Prosodic Disambiguation in Japanese

A Comparative Study of Native Speakers and Swedish L2 Learners

Chen Cheng

Supervisor: Dr. Shinichiro Ishihara

Centre for Language and Literature, Lund University

MA in Language and Linguistics, Japanese
SPVR01 Language and Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

June 2026

Abstract

In Japanese, prosodic structure systematically maps onto syntax. Pitch-related features such as downstep serve to signal syntactic boundaries, enabling interpretation of phrase structure in spoken language. While this mapping is consistent and well-established across native speakers, it is not clear if second language (L2) learners are also able to acquire and utilize sentence-level prosodic cues in the same manner.

Previous research on L2 prosody has been for the most part been investigating the acquisition of L2 learners' intonation on lexical level or their perception of prosody for resolving structural ambiguity. In contrast, few studies have been found on how L2 learners produce prosodic phrasing to signal syntactic structure in their spoken language. This thesis addresses that gap by examining whether intermediate and advanced Swedish learners of Japanese can utilize pitch-based features—such as downstep—to differentiate between left and right branching interpretations of ambiguous noun phrases. In addition, it investigates whether their prosodic realization is automatic or adjusted by syntactic awareness.

To this end, speech data was collected from native speakers and learners under two reading tasks: spontaneous and awareness-guided. Results indicate that while in general learners demonstrate native-like pitch patterns (downstep), clear variation within learner group was observed across proficiency levels: advanced learners exhibited more stable, native-like phrasing, but intermediate learners often showed inconsistent pitch pattern. In addition, awareness of different syntactic interpretation resulted in native speakers' prosody pattern to be more salient, but it did not lead to better prosodic realization of L2 learners. These findings highlight the pedagogical value of incorporating sentence-level prosody into L2 instruction—particularly for helping learners bridge the gap between syntactic understanding and prosodic realization, thereby acquiring more native-like intonation and fostering better communication skills.

Keywords: Syntax–prosody mapping, downstep, L2 prosody, awareness, Swedish learners

Acknowledgements

This research could not have been accomplished without the support and encouragement of several individuals.

First and foremost, I would like to show my greatest and sincerest appreciation to my adviser, Dr. Shinichiro Ishihara. As my advisor, he committed himself wholeheartedly to assisting me with this project, checking my progress regularly on a weekly basis. He guided me through every step of this research. Whenever I mailed him questions or doubts regarding my research, he always answered patiently and promptly, so that I could get over each obstacle... Without his ongoing support and insightful guidance, this dissertation would not have been possible.

I am also deeply grateful to all the subjects who volunteered to participate in my research. There were even participants who traveled from Malmö to participate in the experiments, and others who willingly waived the reward offered so that I could recruit more participants. Their interest and commitment were an essential part of the success of this research; without them, this work could never have been accomplished.

Particular gratitude is due to my partner, Marius, who always provides me with emotional support whenever he can. Writing a dissertation like this was such a demanding experience, and when I was tempted to procrastinate, get muddled, or lost confidence when meeting troublesome issues, he was there to criticize me severely when needed, try his best to motivate me, and help me build confidence once again. His tireless encouragement, support, and guidance helped me move forward step by step on a daily basis.

I would also like to thank all of those who helped me publicize my participant recruitment process when I was struggling to get Japanese speakers or Swedish speakers.

I would also like to thank Lund University and the Department of Languages and Linguistics for the financial support that made this research possible.

Finally, I would like to thank myself. Throughout this extremely tough period when I had to juggle between many part-time jobs and my study, I withstood pressure, worked and wrote dissertation simultaneously, overcame perfectionism, and ultimately finished this imperfect but valuable project.

Table of contents

List of figures

List of tables

Abbreviations

LB Left branching

RB Right branching

MaP Major Phrase

MiP Minor Phrase

F0 Fundamental frequency

Romaji Convention

Japanese words are transcribed using the Hepburn system of romanization. Japanese words are shown in italics, and their English translations are in apostrophes. Pitch-accented moras are marked with an apostrophe following the relevant mora in romaji.

Chapter 1 Introduction

Speech often generates structural ambiguity. Nevertheless, listeners are often able to infer the intended meaning. This is supported in part by prosodic cues such as rhythm, stress, and intonation, which mark syntactic structure. These suprasegmental features have been proven to facilitate sentence processing in daily life (Cutler & Isard, 1980; Dahan, 2015).

In Japanese, a pitch-accent language, pitch contours—more specifically the process of downstep—play an important role in encoding syntactic constituents. Downstep is the prosodic lowering of pitch between accented prosodic constituents and is known to mirror the underlying branching structures (Kubozono, 1989; Selkirk & Tateishi, 1991). For instance, in left-branching (LB) constructions like [[na'mano a'yu-no] nio'i] ("smell of raw ayu"), in which all constituents are accented, pitch usually falls steeply and continuously across the word, resulting in a gradual downstep contour. By contrast, right-branching (RB) constructions like [kowa'i [me'-no ya'mai]] ("awful eye disease") exhibit a more gradual falling pitch contour, which is usually broken by pitch reset or metrical boost at the second constituent, indicating the beginning of a new prosodic boundary. This shows that native speakers can use prosodic information, such as pitch movement, to differentiate between LB and RB syntactic structures in their speech. This ability is important in cases of global ambiguity, e.g., modifier–noun phrases with several genitives, where the same structure can permit several syntactic analyses. In such instances, prosodic information serves as a necessary means of syntactic disambiguation. One typical example is tripartite genitive phrases, e.g., [[N1-GEN N2]-GEN N3] (LB) vs. [N1-GEN [N2-GEN N3]] (RB), where one sentence can represent two distinct hierarchical structures, in the case that both interpretations are semantically plausible. In these instances, interpretation can be partially reliant on prosody.

Although the phenomenon of using prosody, such as pitch, to show branching differences and guide intended meanings has been well investigated in native speakers, comparatively less is understood about how Second language (L2) learners use the same prosody strategy to disambiguate structurally ambiguous sentences. Most studies in this area have focused on whether these speakers can rely on prosody to parse structurally ambiguous sentences. In contrast, there is relatively little research conducted on L2 speech prosodic production, particularly whether learners employ pitch contours to disambiguate syntax in speech planning and production.

Among the few existing studies, it has been argued that even proficient L2 learners have problems with acquiring native-like control of prosody (Trofimovich & Baker, 2006). Additional evidence shows that the awareness of syntactic ambiguity on the speakers' part—whether elicited by task instructions or contextual information—can have an impact on their prosodic disambiguation strategies. If speakers recognize possible structural ambiguities, they use more systematic prosodic markers, like boundary pauses and pitch resets, to guide interpretation (Schafer et al., 2000; Snedeker & Trueswell, 2003).

Despite these findings, it is not certain if L2 learners, particularly proficient learners, are able to use prosodic cues in structurally ambiguous contexts, and if their prosody use is influenced by awareness of ambiguity.

To address these issues, the present study examines if intermediate and advanced Swedish learners of Japanese are able to utilize pitch-related features such as downstep and pitch reset—to express syntactic contrasts in structurally ambiguous phrases. The constructions under investigation are tripartite genitive constructions, which allow for left and right branching interpretations. Furthermore, the study examines if the prosodic realizations exhibited by learners differ in relation to their awareness of ambiguity, which is manipulated through task design.

Thus, the main research questions are: 1) Do Swedish learners of Japanese use prosodic cues, particularly pitch-related features such as downstep, in syntactically ambiguous sentences? 2) Does awareness of syntactic ambiguity affect natives and learners' prosodic realization? 3) How does proficiency (intermediate vs. advanced) influence prosodic accuracy and error patterns among learners?

The remaining content of this thesis is organized as follows: Chapter 2 reviews previous research on the prosodic realization of Japanese branching structures, covering relevant theoretical models, prosodic disambiguation, and L2 speech production research. Chapter 3 covers the experimental design, procedure, and materials. Chapter 4 presents the findings. Chapter 5 interprets the results and implications of the findings for theory development and language intonation pedagogy.

Chapter 2 Literature Review

2.1 Japanese Prosody System

To understand the way prosodic information helps syntactic disambiguation in Japanese, one must first identify the major features of the Japanese prosodic system. This section introduces the relevant aspects of Japanese prosody, with a focus on how it is related to syntax and how these relations were described in theories.

2.1.1 Japanese Prosody Structure

Japanese prosody has been examined in several frameworks, which use different terminologies and make different prosodic phrasing assumptions. To prevent terminological confusion, this section concisely surveys three theoretical traditions: the AM/ToBI-based approach, the Minor/Major Phrase framework, and the Syntax–Prosody Mapping hypothesis (Ishihara, 2015).

<table>
<tr><td colspan="2">AM theory / ToBI</td><td colspan="2">Minor/Major Phrase</td><td>Syntax–Prosody Mapping</td></tr>
<tr><td>Pierrehumbert & Beckman</td><td>J_ToBI, X-JToBI</td><td>McCawley, Poser, Kubozono</td><td>Kawahara & Shinya</td><td>Ito & Mester, Selkirk, this chapter</td></tr>
<tr><td rowspan="2">Utterance</td><td rowspan="3">Intonation Phrase</td><td rowspan="2">(not discussed)</td><td>Utterance</td><td>ι_{max}</td></tr>
<tr><td>Intonational Phrase</td><td>PClause or Intonational Phrase (ι)</td></tr>
<tr><td>Intermediate Phrase</td><td>Major Phrase</td><td>Major Phrase</td><td>Phonological Phrase (φ)</td></tr>
<tr><td>Accentual Phrase</td><td>Accentual Phrase</td><td>Minor Phrase</td><td>Minor Phrase</td><td>φ_{min}</td></tr>
</table>

Figure 1 The summary of Japanese prosodic phrasing (taken from Ishihara, 2015, p. 570)

AM theory/ToBI Model

Within this theory, Japanese prosody can be described in terms of two major prosodic constituents: the Accentual Phrase (AP) and the Intonational Phrase (IP) (Pierrehumbert & Beckman, 1988; Venditti, 2005). According to this theory: AP is the domain of lexical pitch accent and has no more than one accent. Its tonal contour can be %L H- H*+L L%. IP, on the other hand, consists of one or more APs and is marked by sentence-final boundary pitch movements like H%, LH%, HL%, and HLH% (Igarashi, 2015). X-JToBI annotation system adheres to this theory, making AP and IP basic units and encoding pitch accent and boundary tone using such label set.

Major/Minor Phrase Terminology

The second tradition, following McCawley (1968), Poser (1984), and Kubozono (1993), employs Minor Phrase (MiP) and Major Phrase (MaP) to define Japanese prosodic domains. This kind of categorization can find its counterpart in previous one, where Minor Phrase approximately corresponds to the AP in ToBI, and Major Phrase can span several Minor Phrases and is sometimes referred to as what AM theory calls IP.

This representation is quite frequent in Japanese descriptive phonology and pitch accent research.

Syntax-Prosody Mapping Framework

More recent work, such as Ito and Mester (2012, 2013) and Selkirk (2009, 2011), promotes a theory called syntax–prosody mapping. In this framework, hierarchical organization of prosodic units are divided into syllable (σ), prosodic word (ω), phonological phrase (φ), and intonational phrase (ι). It replaces the earlier Strict Layer Hypothesis (SLH)—which prohibits recursion and level-skipping—with a less constrained approach. Prosodic domains (φ, ι) are aligned with syntactic constituents according to Match Theory and alignment constraint. The next sections will elaborate more around this family of theories.

2.1.2 Early Syntax–Prosody Mapping Theories

One of the most important topics of study in prosody is the syntax–prosody interface, where the syntactic structure influences prosodic phrasing. Two of the better-known models that have been proposed are the End-based Mapping theory (Selkirk & Tateishi, 1991) and the Branching-based Mapping theory (Kubozono, 1993).

The End-based Mapping theory (Selkirk & Tateishi, 1991) assumes prosodic boundaries to be marked at the edges of syntactic constituents, more specifically maximal projections (XPs).

For instance, in the left-branching construction [[*Na'oya no a'ni no*] *wa'in o*] "Naoya's big brother's win-ACC", XP boundaries fall together closely at the sentence-initial position, and this triggers cumulative downstep across the noun phrases because there are no intermediate φ-boundaries in between.

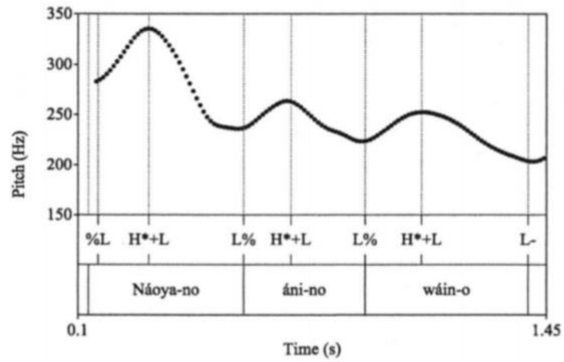

Figure 2 Cumulative downstep: "*Na'oya-no a'ni-no wa'in-o*" (Naoya's big brother's wine-ACC') (taken from Ishihara, 2016, p.1391)

Figure 2 illustrates a typical pitch contour of cumulative downstep in the Tokyo dialect of Japanese. Each accented word (marked by H*+L) shows a progressively lower F0 peak: the peak for *ani-no* is lower than that of *Naoya-no*, and the peak for *wain-o* is even further lowered. This pattern clearly demonstrates how lexical pitch accents induce downstep within an utterance.

As opposed to end-based theories which map prosodic boundaries onto the edges of syntactic constituents (e.g., XPs), branching theory pioneered by Kubozono (1988, 1989, 1993) makes the number and position of syntactic branching the major cue for prosodic structure. It is assumed in this theory that prosodic effects are cumulative: the more syntactic brackets or nested nodes there are at a boundary, the stronger the prosodic effect associated with it will be. In this study, Kubozono contrasted sentences in which the second content word followed an accented word (triggers downstep) or an unaccented word (no downstep). For example, in left-branching structures, where modifiers precede the head noun (e.g., [*Naomi no oneesan no*] *yaringu* 'Naomi's sister's earring'), downstep is typically realized as a staircase-like descent in the fundamental frequency (F0) across successive prosodic units. Each accented phrase triggers a local pitch drop, and the pitch continues to decline steadily throughout the utterance. This pattern reflects a regular chaining of downstep, with no major prosodic boundary interrupting the pitch lowering (Kubozono, 1988).

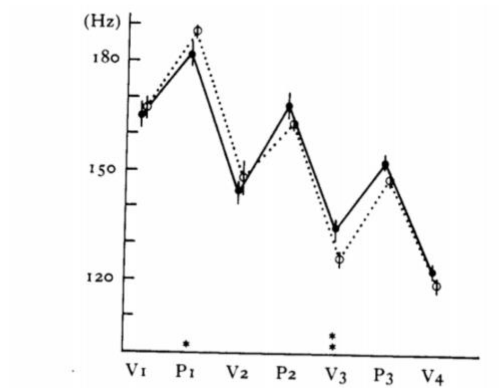

Figure 3 Pitch contour of AAA pattern (dotted line only) for the left-branching structure [*Naomi no oneesan no*] *yaringu* ‘Naomi’s sister’s earring’. (taken from Kubozono,1989, p. 47)

However, the F0 peak of the second word of right branching structure is realized significantly higher than in left branching structures.

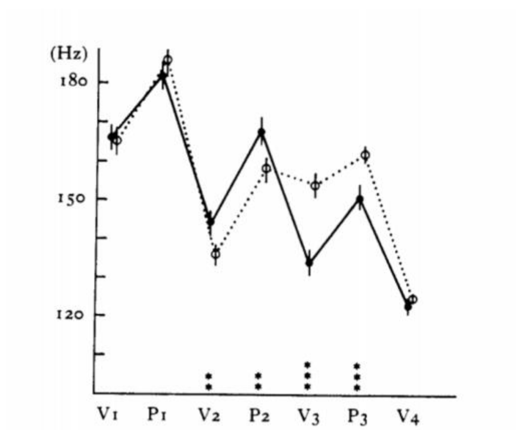

Figure 4 Pitch contour of AAA pattern (solid line only) for the right-branching structure [kowa’i [me’ no ya’mai]] ('terrible eye disease') (taken from Kubozono,1989, p. 45)

As shown in Figure 4, right-branching structures, where modifiers follow the head noun (e.g., *kowai* [*me’* n*o ya’mai*] 'horrible eye disease'), possess a different prosodic contour. Following the initial accented phrase, instead of a constant pitch drop, the pitch of the second component has a tendency to return to a higher range, an effect that can be explained due to the presence of a major prosodic boundary (Beckman & Pierrehumbert, 1988). This boundary effectively breaks the chain of downstep, and thus the second peak (Peak 2) is not significantly lower than the first peak (Peak 1), especially in cases of syntactic break. This difference between left-branching and right-branching structures can be summarized and shown in Figure 5.

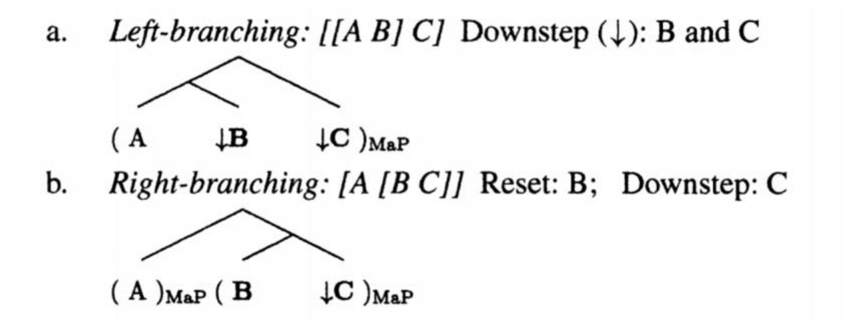

Figure 5 A comparison of the prosodic outcomes of left-branching and right-branching syntactic structures in Japanese taken from (Ishihara, 2016, p. 1391).

2.1.3 Recursive Prosodic Phrasing in Japanese and Re-evaluating Downstep Domains in Japanese

Whereas above-mentioned traditional models foresee binary oppositions—either cumulative downstep or total reset—some spoken speech tends to show cases that cannot be easily accommodated within a rigidly layered analysis. Ito and Mester (2012) challenged the early traditional layered hierarchy and developed a novel theory of Japanese prosodic structure that is called the recursive prosodic phrasing theory. Rather than presupposing distinct categories for each prosodic level, the recursive theory invokes a single phonological phrase category (φ) that is embeddable recursively, in the same way that syntactic constituents are.

This model violates the Strict Layer Hypothesis (Selkirk, 1984). Instead, recursive phrasing proposes several φ-projections within one sentence, with minimal and maximal φ determined by structural prominence rather than categorical type. In this way, domains such as MiP and MaP are redefined not as distinct categories but as relational projections of the same prosodic unit.

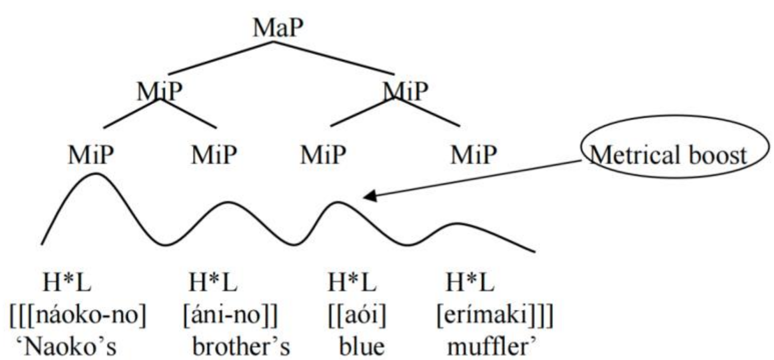

Figure 6 Recursive φ in Genitive Constructions (taken from Ito & Mester, 2012, p.9)

Figure 6 illustrates how recursive prosodic phrasing accounts for pitch patterns in genitive structures. Here, the string "Naoko's brother's blue muffler" (Japanese: náoko-no áni-no aói erímaki) consists of four lexically accented items (each marked H*L). In a traditional flat MaP–MiP view, we should hear successive downstepping through the chain, each F0 peak lower than the previous one.

However, as is evident from the Figure 6, the pitch peak on aói is unexpectedly high. According to Ito and Mester (2012), [*náoko-no áni-no*] forms one unit, and [*aói erímaki*] is another, and both are housed within a higher-level MaP. This recursive embedding resulted in metrical boost—a F0 rise that counters the predicted pattern of downstep. Such an effect is impossible to be explained within traditional non-recursive, flat prosodic models, but it arises naturally from the recursive structure.

Production experiments also reveal that pitch register is not fully reset at syntactic boundaries or focus positions. For example, Ishihara (2016), in a controlled re-examination of downstep in Japanese, demonstrates that although the rise of F0 typically happens at syntactic edges such as those in right-branching constructions, this reset of pitch is often incomplete. That is, the F0 peak for genitive constructions such as [N1-GEN [N2-GEN N3]], [N2-GEN N3] is regarded as a new prosodic domain. However, pitch on N2 does not fully return to the initial pitch range established at the beginning of the utterance but rises only partially.

Rather than ruling out pitch reset altogether, Ishihara (2016) indicate that gradient realizations of pitch resetting fit the production data more strongly (Ishihara, 2016, pp.1421-1423). Based on this, pitch reset is not all or nothing, and its presence or absence cannot be the sole test for MaP boundary determination. To account for these results, Ishihara makes use of the recursive prosodic phrasing model of Ito and Mester (2012, 2013). This model facilitates across-MaP downstep, whereby the onset pitch of a new MaP (e.g., at N2) is only partially blocked by the register of the prior MaP, so that there is incomplete resetting of the pitch. As such, what might appear as a lack of reset actually may be a reflection of the within-MaP and across-MaP downstep interaction.

This study thus builds on the recursive model to see how Swedish L2 learners of Japanese, like native speakers, use pitch cues to disambiguate genitive structures (N1-GEN N2-GEN N3). More focus is placed on whether L2 learners exhibit gradient downstep and reset patterns, an indication sensitivity to hierarchical prosodic phrasing.

2.2 Prosodic Disambiguation of Syntax in Native Japanese Speakers

2.2.1 Prosodic Disambiguation in Native Speech Production

Following the theoretical basis of syntax–prosody mapping, recent empirical research has offered strong evidence that native speakers employ prosodic markers—such as pitch reset, downstep, boundary pauses, and phrase-final lengthening—to disambiguate syntactic structures during real-time comprehension and production. (e.g., Snedeker & Trueswell, 2003; Schafer et al., 2000). These prosodic strategies seem to be implicit and systematic when marking syntactic information.

First of all, evidence from Japanese production studies has showed that native speakers can spontaneously use prosodic features to capture underlying syntactic structure. For example, Venditti and Yamashita (1994) acoustically examined the utterance *Mari ga yonda* "Mary read it" both in simplex sentences and complex ones. Acoustic analysis showed that when the clause was within a complex structure, it exhibited significant pitch lowering and final lengthening compared to when it was in isolation. This indicates that even when lexical items do not change, speakers adjust prosodic realization to reflect syntactic embedding.

In addition, Azuma (1997) further tested a syntactically ambiguous sentence *Nara-de (#) taoreta yooji-o hakonda*, which can mean "carried a child who fell in Nara" or "carried a child who had fallen in Nara". The two readings were linked to different prosodic realizations: the former involved a pause and pitch reset at the boundary (#), whereas the latter did not. In follow-up perception tests with varied pitch contour (F0) and pause length, F0 was the predominant cue for syntactic interpretation.

Wolff et al. (2008) also presented empirical evidence that clear prosodic boundaries in production contribute to the marking of non-canonical word orders in Japanese. Specifically, object–subject–verb scrambled sentences were better understood when speakers placed a prosodic break following the fronted object. This finding indicates that Japanese speakers make sentence structure information clearer by using prosody to signal syntactic deviations from fundamental word order in speech. Also, contrastive intonation, marked by the L+H* pitch accent, has been found to contribute to referential ambiguity resolution through marking of discourse contrastive sets (Dahan et al., 2002).

However, the use of prosodic cues is not totally unified among all native speakers. Research in Japanese has also shown that speakers have individual differences in how they use prosody to mark syntactic structures. For example, Hirose (2006) found that in language production experiments, different speakers rely on prosody in different ways: some rely more on fundamental frequency changes, while others rely on pauses or changes in duration of certain parts in speech.

2.2.2 Perceptual Evidence for Prosodic Disambiguation

Some perceptual studies in Japanese have also been carried out to verify the alignment between prosodic cues and syntactic structure. Uyeno et al. (1980, cited in Venditti et al., 2014, p. 300) found that in Japanese, intermediate phrase boundaries help to clarify the scope of adverbial modifiers. For example, in the sentence "*kjo'nen a'nda e'rimaki-ga nusuma'reta*" (The scarf I knitted last year was stolen), *kjo'nen* (last year) can have two different syntactic scopes: If *kjo'nen* (last year) modifies the entire sentence, an inter-phrase boundary will appear after *kjo'nen* (last year) and a pitch reset will occur on the word *a'nda* (knit), marking a larger syntactic unit. If "last year" modifies only the verb "knit", no pitch reset will occur, but a downstep will appear, indicating that this part is processed as a complete prosodic unit.

Uyeno (1980) also experimented with the influence of pitch contours on the interpretation of structurally ambiguous relative clause attachment sentences. Using tape versions with systematically modified pitch patterns of sentences such as *Kyō koronda otona ga waratta* "The adult who fell today laughed" or "The adult laughed today", he played these to 34 native speakers of Japanese. The outcome was that if pitch went to an early high (e.g., on Kyō) and fell step by step in the course of the sentence, structure was heard as left-branching (i.e., [*Kyō koronda otona*] subject). However, if pitch was kept high in the later part of the sentence (e.g., on "*otona*"), subjects heard the sentence with a center-embedded clause (i.e., the main clause *otona ga waratta*). Uyeno's results straightforwardly showed that Japanese listeners are using the global pitch contour, specifically the slope and position of F0 peaks, to select grammatical structure in instances of ambiguity.

In summary, experiments on native Japanese speakers have shown that prosodic cues are employed in both production and perception of ambiguous syntactic structure. Speakers spontaneously adjust pitch, duration, and phrasing to indicate syntactic embedding and ambiguity, while listeners also use pitch contours, especially pitch reset and downstep to guide structural interpretation.

2.2.3 Effects of Ambiguity Awareness on Prosodic Disambiguation

While many studies have shown that native speakers use prosodic markers to indicate syntactic structure and disambiguate, there has been increasing investigation exploring whether such prosodic disambiguation is automatic or only triggered when the speaker is conscious of structural ambiguity. In other words, does prosodic marking reflect automatic parsing biases, or is it a controlled strategic procedure?

Several experiments in other languages have investigated how speakers indicate syntactic boundaries with the help of prosodic markers, especially in cases of uncertainty. Work by Schafer et al. (2000) and Allbritton et al. (1996) provides strong support that speakers employ more systematic and salient prosodic marking when they are aware of sentence-level ambiguity—a strategy referred to as strategic prosodic marking.

In a study by Schafer et al. (2000), researchers designed a cooperative game task in which participants used structurally ambiguous sentences to achieve specific communicative ends. Even when context was enough to disambiguate the syntax, speakers automatically marked syntactic boundaries through prosodic means, such as inserting pauses, changing patterns of intonation, or lengthening particular syllables. Especially with early closure sentences (e.g., "When that moves, the square…"), speakers typically added a prosodic break after "moves", whereas for late closure sentences (e.g., "When that moves the square, it…"), the break was after "square". Such differences manifested both in word durations and pitch contours, showing that disambiguation cues relying on prosody are being utilized even when it is possible to be achieved through context alone.

On the other hand, Allbritton et al. (1996) found that in more traditional reading-aloud tests, speakers who are uninformed of sentences'ambiguity rarely produced sufficient prosodic cues to assist listeners in disambiguating. Even where the sentences were ambiguously formed, prosodic realization was neutral and without the durational or pitch contrasts needed to disambiguate.

Some scholars also support that speakers only use prosody to disambiguate when they are aware. For example, Snedeker and Trueswell (2003) hypothesized that speakers would only use active prosodic cues to disambiguate in the case that they perceive the existence of referential ambiguity under the presented context. Specifically, when two spoons and two cats are present in the context, and one cat has a spoon, the target sentence such as "Tap the cat with the spoon" will create ambiguity because this sentence has two interpretations. The speakers make syntactic pauses at this point by using prosodic boundaries (such as placing a pause after "cat"). However, if there is only one cat in the context (i.e., there is no ambiguity), speakers do not normally give redundant prosodic cues.

Contrary to this, Kraljic and Brennan (2005) demonstrated that in highly interactive conversations, even if the context itself is clear and unambiguous, speakers use prosodic indicators to disambiguate and this is not based on the specific needs of the listener. This means that in real interactive contexts, prosodic disambiguation is not only driven by "ambiguity awareness", but also by the interlocutor's goal to convey a message. Some researchers have suggested that this can be related to the speaker's own personality, e.g., cooperativeness. Schober and Brennan (2003) pointed out that when the speaker is a child, the speaker will more likely show cooperation, so he will use prosodic devices more frequently and more explicitly to disambiguate syntax.

These findings might imply that prosodic marking of ambiguity is not determined by syntactic structure in isolation, but also by pragmatic factors such as communicative intent and speaker's awareness state, and communicative intent. When speakers themselves are aware of the potential for ambiguity or are explicitly required to disambiguate, their own prosodic strategy becomes more systematic and consistent, thus facilitating listener interpretation.

Thus, in experimental design that aims at simulating real-life language communication, manipulation of the variable of ambiguity awareness might be one of the essential factors that can affect the prosodic change.

2.3 Disambiguation Strategies Using prosody in L2

Compared to native Japanese speakers, L2 learners might not be that capable in producing prosodic cues indicating syntactic structure. To see how Swedish learners of Japanese arrive at or depart from native-like prosody, it is first necessary to summarize the theoretical models accounting for L2 phonological and prosodic acquisition.

2.3.1 Theoretical Approaches to L2 Intonation Acquisition

Prosody acquisition can be seen within the category of intonation acquisition. One of the earliest accounts of L2 Intonation acquisition is The Markedness Differential Hypothesis (MDH). According to this hypothesis, marked features that are less common cross languages, are more difficult to acquire. The Ontogeny Phylogeny Model (OPM) develops this idea further by asserting that prosodic features with marked values are highly susceptible to transfer effects and might not be completely acquired unless they are typologically unmarked or accessed through input. (Major, 2001). Experimental evidence in favor of Eckman's MDH comes from Rasier and Hiligsmann (2007), who conducted a comparison of L2 Dutch and L2 French prosody accent strategies. They provide evidence that typological differences of the L1 strongly affect the acquisition of L2 prosody, particularly for accent placement. Additionally, they discovered a relationship between pause errors and accent errors and demonstrated that prosodic errors grow over time.

Also, Flege's (1995) Speech Learning Model (SLM) suggests that, the acquisition of L2 sounds that are distinct from L1 categories is challenging because learners tend to categorize them as the same as their L1 categories. According to SLM, perceptually distinct L1 sounds from the target L2 sounds facilitate it to form a new L2 sound category and are therefore easier to acquire—

at least perceptually, if not always in production. If L1 and L2 sounds are similar, then they fall into one category, and it becomes more difficult to learn. SLM also states that the L2 pronunciation will develop incrementally in terms of increased exposure to the target language.

Earlier models indeed have significantly contributed our understanding of L2 segmental perception and phonemic categorization, they mention little in the way of suprasegmental domains. Unlike single sounds, intonation deals with dynamic patterns across extended parts of speech, governed by both phonological structure and discourse-level function. In response to this shortage, Mennen (2015) introduced the L2 Intonation Learning Theory (LILt), a multi-dimensional theory that accounts for why it is not possible for second language learners to produce native-like intonation.

Based on Mennen (2015), such problems arise because the learner's L1 and L2 differ in four domains: (1) the systemic dimension, which refers to the distinctive intonation patterns and their distribution in the language, (2) the realizational dimension, which deals with how these patterns are actually realized in speech (e.g., pitch, duration, and intensity, etc), (3) the semantic dimension, which deals with how intonation is used to express meaning and function, and (4) the frequency dimension, which analyzes how often specific intonation patterns are used in spontaneous speech. As LILt implies, when learners' L1 and L2 are quite dissimilar in any of these areas, they will tend to produce unnatural or inappropriate intonation. Mennen (2015) also argues that these problems are not simply a question of learning new intonation patterns but also of unlearning habits of L1 perception. For example, a student can have difficulty with using a rising intonation on a question if his or her L1 uses a specific pitch pattern to indicate the same function. Although the model is consistent with current theories such as the Speech Learning Model and the Perceptual Assimilation Model in proposing that perception is central, LILt further suggests that knowing both how intonation is created and how it is used to convey meaning is important in order to predict where learners will have difficulty. The model is now a "working theory", i.e., it is still open to revision with more research, but it already gives us a basis for exploring and addressing L2 intonation problems.

While the above theoretical models serve well to offer frameworks for L2 intonation acquisition from a phonological and cognitive perspective, it is also necessary to consider the development of learners' prosodic systems over time. Interlanguage development perspectives serve to be helpful in this matter. Following language acquisition, intonation patterns of L2 learners come closer to the target language step by step but systematic overgeneralization is common (Archibald, 1997; Hiligsmann & Rasier, 2002). For example, Grosser and Wieden (1989) found that learners overuse the most prominent accentuation patterns of the target language in early development, and this might be transferred subsequently to contrastive accentuation, which is a case of cross-linguistic influence as well as trouble in modulating intonation patterns to discourse context.

2.3.2 Significant Factors Influencing L2 Phonology Acquisition

In L2 phonology acquisition, many factors affect the learner's mastery of the target language's phonological features. These factors mainly include the age of acquisition (AoA), L1 transfer, syntactic complexity, and the ability to use prosodic features.

AoA

Many studies have explored AoA on L2 phonological acquisition, but there are different views in the academic community. Some researchers support the Critical Period Hypothesis (CPH), believing that the younger the age, the more conducive it is to achieve a native-like pronunciation level. For example, Tahta, Wood, and Lowenthal (1981a, as cited in Long, 1990) found that there is almost no accent transfer when learning a L2 before the age of 6, while accent transfer is almost inevitable when learning L2 after the age of 12. Krashen and Scarcella (1979, as cited in Long, 1990) also pointed out that although older learners perform faster in the initial stage of language learning, younger learners can eventually show higher levels of ultimate attainment in phonological acquisition.

However, other scholars have reservations about the idea that "the younger the better". For example, Hatch (1983) and McLaughlin (1984, as cited in Long, 1990) believe that existing research data do not clearly support this view. In addition, Snow (1983, as cited in Long, 1990) even called the idea that "children have more advantages in speech acquisition" a "fantasy" and a "myth", indicating that this view is not generally recognized.

Specific experiments also reveal that the "rate advantage" of older learners in speech acquisition is usually short-lived. For example, studies by Olson and Samuels (1973), Ervin-Tripp (1974), Ekstrand (1976, 1978), Grinder, Ottomo, and Toyota (1962), and Snow and Hoefnagel-Hohle (1978, as cited in Long, 1990) have shown that in the short term, older learners can adapt faster to L2 phonological features under limited language exposure time due to their stronger cognitive skills and test-taking skills.

L1 Transfer & Markedness

Many studies, such as what have reviewed above, have also explored the influence of L1 prosodic transfer on L2 speech acquisition, covering prosodic features such as intonation, stress, rhythm, and pauses. For example, Archibald (1998) noted that Polish and Hungarian learners of English tend to stick to L1 prosodic patterns, over-stressing nouns or stressing sentence-final positions. This agrees with Backman (1979) and Broselow (1988), who illustrated that Spanish learners of English maintained a syllable-timed rhythm, failing to adapt to English stress-timed characteristics.

However, not all studies conclude the same way in terms of L1 prosodic transfer. While Jilka (2000) and Van den Doel (2006) addressed the negative impact of L1 transfer on prosodic processing and, more particularly, on stress distribution and intonation contours, Mennen (2006) further distinguished between two forms of prosodic transfer: phonological interference and phonetic interference. According to Mennen (2006), phonological interference mainly stems from the difference between the tonal inventory in L1 and L2. For example, a learner who uses falling tone to indicate the finality of a sentence in his native language may wrongly apply this intonation pattern to all sentences in L2. Phonetic interference is realized through the various phonetic realization of the identical phonological tones in L1 and L2. For instance, stress is conveyed by pitch change in L1, but in L2 it may be conveyed by loudness or length. If the learners don't take on such more subtle differences, intonation errors will result.

2.3.3 Types of L2 Intonation Errors

Current studies have found out that L2 learners have a hard time acquiring native-like intonation patterns, particularly when prosodic markers are used to indicate syntactic boundaries or to clarify structures. Early studies have shown that L2 learners generally overuse pitch accents in the early stages of language learning. Grosser (1993) and Wieden (1993) found in longitudinal studies that German learners tend to add accents to almost every word when learning English. This phenomenon is not limited to English learners, but has also been found in other language groups. For example, similar prosodic errors have been observed in Polish and Hungarian speakers learning English (Archibald, 1997), French speakers learning Dutch (Hiligsmann & Rasier, 2002; Rasier, 2003), Spanish learners learning English (Backman, 1979), and Dutch learners learning English (Jenner, 1976; Willems, 1982).

Mennen (2007) provided a helpful definition of two sources of non-native intonation: phonological errors, which include inappropriate choice of intonational categories (e.g., kind of pitch accent or where), and phonetic errors, which include incorrect realisation of otherwise accurate targets for intonation, e.g., misplaced peaks of F0 or limited pitch range. It is most appropriate for research that asks how L2 speakers use prosody to resolve syntactic ambiguity since it proposes that even advanced learners possessing the relevant prosodic categories will not always be able to produce disambiguating contours due to marginal deviations. Interestingly, Mennen (2007) summarized a number of empirical studies demonstrating that phonetic implementation errors are long-lasting after even highly proficient learners, particularly in situations where precise alignment of prosodic boundaries with syntactic structure is necessary.

Some scholars tested L2 Japanese learners' pitch acquisition, as Japanese pitch accent is a feature of the Japanese language that distinguishes words by accenting particular morae in most Japanese dialects. Shport (2008) demonstrated that L2 Japanese learners do not acquire pitch accent patterns inductively. She required her test takers to distinguish between minimal triplets such as gekka, kohaku, and tanka where only kohaku had the unaccented pattern consistent with the target. It proved challenging for the learners to distinguish between such patterns, an indication that they naturally perceive pitch accent as salient unless explicitly taught.

2.3.4 Specific Prosodic Challenges for Swedish Learners of Japanese

While there has been increasing research in L2 prosody acquisition, studies specifically targeting Swedish learners of Japanese, and more specifically their command of pitch boundaries in spoken production, are noticeably absent.

Among the few studies that systematically tackle this issue is Nagano-Madsen (2015), which examined the acquisition of a number of prosodic parameters in Japanese by Swedish learners, including pitch accent, intonational phrasing, and information structure. In this research, 15 intermediate to advanced learners of Swedish provided recorded speech samples that were examined in PRAAT software. Eleven prosodic features were coded for systematic occurrence, with each being awarded a score when realized consistently. The findings indicated that intonational phrase boundaries, specifically initial F0 reset, were some of the earliest-appearing features, whereas pitch accent and focus marking continued to be challenging—even at advanced levels.

One of the primary findings was that learners did not produce the steep F0 fall characteristic of Japanese pitch accent. Instead, they used a consistent F0 rise or shallow declination across the word, resulting in a flat pattern that was neither native-like Japanese nor within Swedish norms. This suggests that the learners were not simply transferring their L1 prosody, but building a new, hybrid intonational system—an interlanguage prosody (Nagano-Madsen, 2015). Further, the research found an initial "reset phase" in which learners started sentences on a flat pitch, and a high-pitch reset at the start of main syntactic units, out of which increasingly more native-like IPs evolved over time.

In a follow-up study, Nagano-Madsen (2018) expanded this study by comparing the perception and production of L2 prosody in both Japanese and Mandarin for Swedish learners. This study confirmed that Swedish learners preferred F0 rise to F0 fall, a feature which was evident in both languages. Learners also had trouble perceiving pitch fall distinctions, which led to difficulties in distinguishing between Japanese questions (sentence-final F0 rise) and statements (F0 fall). In phrasing, learners used "upstep" patterns—where pitch increases across successive phrases—contrary to the "downstep" found in native Japanese speech. This reversal of pitch boundary approach also indicates that there is a target language and Swedish L1 prosodic tendency mismatch.

These researches provide important information on how pitch control can be particularly difficult for Swedish learners. While Japanese and Swedish are both often classified as pitch accent languages, they are typologically distinct in their prosodic systems. In Swedish, pitch accents (Accent 1 and Accent 2) are primarily associated with the timing of F0 movements and have minimal lexical function (Bailey, 1990). Japanese pitch accent, by contrast, involves fixed positions of pitch drop that are crucial to both lexical identity and prosodic phrasing (Kawahara, 2015). This structural difference might predict that while Swedish learners have experience with pitch accent, they may not have the prosody awareness or perceptual sensitivity required to handle pitch boundaries in Japanese.

Nagano-Madsen (2018) has also shown that the Swedish learners tend to produce greater F0 on the topic compared to the focus, the reverse of the native F0 pattern. This lack of alignment reflects a more fundamental difficulty in learning how prosodic cues combine with syntax and pragmatics in Japanese.

2.4. Summary of Literature and Research Gap

There is extensive evidence to prove that prosody plays a critical role in resolving syntactically ambiguous sentences. Prosodic characteristics such as pitch contour, pause, and phrase-final lengthening can be useful cues for syntactic structure in native speakers. Experiments such as those of Schafer et al. (2000) and Kraljic & Brennan (2005) have shown that when speakers pay attention to marking ambiguity, they utilize more perceptible prosodic boundaries to aid the understanding of listeners—a process generally referred to as strategic prosodic marking. Allbritton et al. (1996), however, found that naive speakers who did not explicitly mark syntactic ambiguity were inclined to create neutral prosody, insufficient for disambiguation. These results indicate that prosodic realization is highly task condition- and speaker awareness-sensitive.

Furthermore, there is also growing interest in how these variables interact with L2 prosody. Much of this research, however, has been focused on perception, or segmental rather than suprasegmental features.

Apart from that, there are also other clear gaps which exist in the limited literature of L2 prosody acquisition. To begin with, previous studies have conducted perception tests and investigated how L2 learners utilize prosodic patterns to disambiguate ambiguous sentences. However, there have been few studies using syntax-prosody mapping view and examining how L2 learners utilize prosodic cues in encoding syntactic information when they produce speech. There is also little information about how awareness of syntactic ambiguity works in syntax-prosody mapping. Finding out about this can help know if it is possible for L2 learners' prosody to become as automatic and communicative as it is in the speech of native speakers.

Third, there are not many systematic investigations of prosodic marking of syntactic information in Swedish speakers learning Japanese. Since Japanese and Swedish are pitch-accent languages, but they are different in how pitch is used to show linguistic information. This prosodic typological difference makes Swedish learners of Japanese an ideal population for studying how the L1 prosodic system influences L2 acquisition. Nagano-Madsen (2015, 2018) is one of the few in-depth studies of the issue. The research exhibited that while L2 learners were able to acquire larger prosodic constituents like intonational phrase boundaries relatively early, represented by early F0 reset, the accurate production of pitch accent of lexicon and focus-related F0 movements remained as an issue even at advanced levels. Swedish L2 learners tended to substitute L1-like steep F0 drops with flat or rising contours, producing a prosodic system that did not belong to either L1 or L2 patterns. This was considered an interlanguage prosody process that was shaped by both transfer and internal reorganization.

Overall, most of the earlier studies on L2 prosody have dealt with either phonetics or lexical-level pitch accents. In contrast, relatively less investigation has been done to the question of how L2 learners use prosody at the sentence level to indicate syntactic structure and distinguish meaning. This gap is especially wide in languages like Japanese, where prosodic boundaries often align with syntactic boundaries. However, it remains unclear whether L2 learners can use these cues automatically and systematically, as native speakers do.

Chapter 3 Methodology

3.1 Introduction

In order to examine the way Swedish L2 learners of Japanese map syntactic structure onto prosody, this research conducted a controlled speech production experiment between native speakers of Japanese and Swedish learners. The experiment had three broad research questions: 1) Do Swedish learners of Japanese use pitch-related prosodic cues such as downstep, in structurally ambiguous sentences with left- and right-branching reading? 2) Does awareness of syntactic ambiguity affect learners' prosodic realization of branching contrasts? 3) How do learners in different awareness conditions differ in their prosodic accuracy and error patterns? The following sections explain participant recruitment protocols, stimulus, and analytical procedures.

3.2 Participants

14 participants were initially recruited for this research. Two of them, one Japanese native speaker and one Swedish learner, were allocated to pilot testing, which was used to confirm the experimental design, task clarity, and questionnaire usability. These two participants' data were excluded from the analysis because of background noise in their recordings and program errors happening during the experiment which negatively affected the participants' pitch. Later, the data of the 10 subjects were calculated, 4 being native speakers of Japanese (3 females, 1 male) and 6 being Swedish learners of Japanese (3 females, 3 males). The recruitment occurred primarily through networking or online. Most Swedish participants were recruited through the Japanese Studies program in Lund University. The Japanese participants were recruited from the local Japanese community in Skåne, including exchange students and long-term residents, using social media sites such as Facebook groups. All the Japanese subjects were living in Skåne at the time of the experiment. The Swedish subjects were students or recent graduates of Japanese Studies studies or persons with at least three years' independent study. No official placement test was given, but the subjects reported their own Japanese ability in listening, speaking, reading, and writing through a language background questionnaire (Appendix B). According to their self-evaluation and educational background, all the participants fell within the intermediate to advanced range (B1–C1). As for phonetic training, five of the Swedish learners had not received any formal Japanese phonetic training, and one had received some. Most depended on informal exposure through reading out loud, imitation, or talking with natives. Their weekly contact with Japanese

varied from 1 hour to 6 hours of self-study, coursework, and media. In order to keep anonymity and also to label each participant's data, all the participants were assigned a code based on nationality (J/S), gender (F/M), and number. For instance, "JF1" indicates Japanese Female 1 and "SM2" indicates Swedish Male 2.

Table 1 Overview of Swedish participants

Participant ID	Language Proficiency (Self-reported)	Age Range	Phonetics Training	Current Weekly Exposure to Japanese (Study/Media)	Years of Learning Japanese	Holding of Japanese Language Certificate
SF1	B1–B2	18–30	No	<1 hour	5	No
SF2	C1–C2	18–30	No	3–6 hours	9	Yes
SF3	B1–B2	30–40	No	1–3 hours	2	Yes
SM2	B1–B2	18–30	No	1–3 hours	3	No
SM4	B1–B2	18–30	No	3–6 hours	2	No
SM5	C1-C2	18–30	Yes	3–6 hours	4	No

Table 2. Overview of Native Japanese Participants

Participant ID	Region (Prefecture, Japan)	Area	Age Range	Can Speak Standard Japanese?
JF1	Hyōgo	Kansai	18–30	Yes
JF2	Chiba	Kantō	40–50	Yes
JF4	Tokyo	Kantō	18–30	Yes
JM2	Kanagawa	Kantō	40-60	Yes

Participation in the experiment was voluntary, and each participant was compensated with a cinema ticket, along with a small souvenir brought from China by the researcher as a token of appreciation. Before the experiment began, all participants signed a written informed consent form in accordance with ethical guidelines.

The experiment took place offline in a quiet classroom at SOL (Centre for Languages and Literature), Lund University, using a high-quality external microphone to ensure clear audio recordings.

3.3 Materials

3.3.1 Linguistic stimuli

The experimental stimuli consisted of five syntactically ambiguous Japanese sentences, each of the form *N1-GEN N2-GEN N3*, embedded in the frame sentence *Kore wa ~ desu* (“This is ~”). These tripartite genitive structures were formulated to allow for two syntactic interpretations: left branching ([[N1-no N2]-no N3]) or right branching ([N1-no [N2-no N3]]).

All of the items were inspired by Hirose and Mazuka’s (2024) stimuli structures which included a color noun (N1), a modifying noun (N2), and a head noun (N3). For instance, in *ore’nji-no me’ron-no ra’ito* ("orange melon light"), the adjective noun *ore’nji* ("orange") may either bind to *meron* ("melon"), resulting in the interpretation "a light with a melon which is orange in color" (left-branching: [[*ore’nji-no meron*]-*no raito*]), or bind to raito ("light"), resulting in an interpretation of "an orange light with a melon (which can be any color)" (right-branching: [*ore’nji-no* [*me’ron-no ra’ito*]]).

Thus, these two analyses, according to whether N1 modifies N2 or N3, are mapped onto contrasting syntactic structures: right-branching and left-branching structures.

All five of the ambiguous sentences adopted this format, with common terms (e.g., household items, animals, fruit) for semantic plausibility and familiarity. All words in the stimuli have pitch accent, thus enabling the downstep prosodic pattern in Japanese.

Table 3. Overview of stimuli sentences

No.	Stimulus Sentence	Left-Branching Interpretation ([[N1-GEN N2]-GEN N3])	Right-Branching Interpretation ([N1-GEN [N2-GEN N3]])
1	*ore’nji-no me’ron-no ra’ito*	a light with a melon pattern that is orange in color	an orange light that has a melon pattern on it
2	*kimi’dori-no ya’mori-no zu’bon*	pants with a light green gecko pattern	light green pants with a gecko pattern

3	*gō'rudo-no re'mon-no ka'mera*	a camera with a golden lemon pattern	a gold-colored camera with a lemon pattern on it
4	*gurī'n-no o'ruka-no do'resu*	a dress with a green orca pattern	a green dress with an orca pattern on it
5	*mura'saki-no go'rira-no me'gane*	glasses with a purple gorilla pattern	purple glasses with a gorilla pattern on it

Semantic naturality and interpretability of the two readings were secured by pilot testing with one native Japanese and consultation with a Japanese teacher. Efforts were paid to minimize lexical or semantic bias towards either of interpretation, so that prosodic realization would reflect spontaneous structural processing. Clear pitch extraction was possible through the use of lexical items that lack voiceless obstruents in prosodic focus positions (e.g., /s/, /k/, /t/) where feasible.

Along with the five ambiguous ones, ten filler sentences were constructed—five with unambiguous left-branching and five with unambiguous right-branching.

Table 4. Overview of filler sentences

No.	Filler Sentences	Translation
1	*Na'oya no a'ni no ne'ko*	Naoya's older brother's cat
2	*Ma'riko no jī'nan no me'gane*	Mariko's second son's glasses
3	*Yū'ji no a'ni no zu'bon*	Yūji's older brother's pants
4	*Aniu'e no bu'ka no ru'kku*	My brother's subordinate's bag
5	*Na'omi no ane no ka'mera*	Naomi's older sister's camera
6	*Yū'ji no mi'dori no zu'bon*	Yūji's green pants
7	*Na'na no a'ka no mō'fu*	Nana's red blanket
8	*A'ni no a'o no nō'to*	My older brother's blue notebook
9	*Ri'na no ku'ro no bū'tsu*	Rina's black boots
10	*Mi'na no mura'saki no ne'kutai*	Mina's purple necktie

One interpretation was considered one item, so there were 20 intems constructed in total. All sentences were syntactically identical and matched for length (around 15–20 moras).

3.3.2 Visual Context for Interpretation

In Task 1, the target sentences were all presented with a single visual display that was constructed to lead either a left branching or right branching interpretation. For example, for the sentence *Kore wa gurī'n no o'ruka no do'resu desu* "This is the green orca dress", a picture of a black dress on which there is a green orca would provide visual evidence for the LB reading, in which the color adjectives modify the animal (i.e., "a dress with a green orca pattern"). Control items, being unambiguous, were also paired with literal visual representations to provide naturalness and stylistic consistency.

Figure 7 Illustration of "*Kore wa gurī'n no o'ruka no do'resu desu*" with LB interpretation

In Task 2, the visual display was manipulated to contrast two possible meanings explicitly. The two images were displayed vertically for the subjects—for instance, green pants with a gecko and a green gecko on black pants—along with a structurally ambiguous sentence such as *Kore wa kimi'dori no ya'mori no zu'bon desu*. An arrow was used to mark the target referent, and this is meant to cue the participants to highlight the pointed object and speak the corresponding interpretation of the ambiguous structure out loud. The contrastive presentation was meant to elicit more conscious syntactic disambiguation and have the subjects produce prosody in accordance with one of the two structures.

Figure 8 Example of Visual Contrast Used in Task 2 to Elicit Syntactic Disambiguation

In order to reduce the possibility of perceptual distraction or visual salience effects, all the images were designed in an identical minimalist line-draw style with the same design program (Canva). Color was restricted to 2–3 colors in each image, and size, visual complexity, and composition were identical for all the objects. This visual consistency meant that any variation in prosodic realization would be attributable to syntactic processing and not to confounded visual difference.

All visual stimuli were checked by native Japanese speakers to ensure that each picture reliably biased its target interpretation without causing semantic ambiguity.

3.3 Preparation

Overall, 20 items and their corresponding visual context were presented in Task 1: 10 ambiguous items (5 intended to favor left-branching readings and 5 right-branching), and 10 unambiguous filler sentences (5 with left-branching and 5 with right-branching). In Task 2, only the previous 10 ambiguous items are used.

In order to be able to assess the intra-subject pitch variability, every item was repeated four times in each task, thus totaling 80 trials for Task 1 and 40 sentences for Task 2. Presentation order of the stimuli list was pseudo-randomized, so that no identical item appeared in direct sequence.

The tasks were carried out in a silent study or group room at the Centre for Languages and Literature (Språk- och litteraturcentrum) at Lund University. Written informed consent (Appendix A) had been obtained from all participants prior to the experiment, recognizing their voluntary participation and anonymization of their data according to ethical guidelines.

The experiment was conducted using a PsychoPy script. Both task started with a welcome screen followed by the trials involving displaying the sentences' and finished with a thank-you message and a brief break to prevent fatigue.

Stimuli were displayed on a MacBook Air, and audio data were recorded using a Zoom H5 handheld recorder attached to a Shure WH20XLR head-mounted microphone. Recordings were made at 44.1 kHz and 16 bits of resolution, giving us studio-quality recordings with hardly any detectable ambient noise.

Two pilot subjects, who are from Sweden and Japan, were invited for testing prior to running the actual data collection. The pilot session was for the purpose of testing the readability of instructions, the audibility of stimuli, and the quality of recording equipment and so on.

3.4 Ethical Considerations

All participants were given a simple and clear information sheet detailing the purpose of the study and the tasks involved (such as voice recording). They were informed that participation was voluntary and how data would be utilized anonymously for the research purpose. Written consent was signed by all participants prior to data collection. All audio recordings were randomly assigned speaker IDs and stored safely on password-protected devices. No personally identifiable data was linked to the audio files or the results of the analysis.

Participants were also informed that they were completely free to take breaks, ask questions, or withdraw from the study at any point without consequences. Participants were not financially compensated, but a movie ticket sponsored by Lund University and a thank-you gift from China were provided by the researcher for their effort and time. Since the experiment was carried out offline, scheduling was based on each participant's convenience. They were also free to change or cancel the date based on their sudden schedule change. The researcher was very careful to treat all the participants equally and with respect. The researcher was careful to treat all the participants equally and with respect, and strove to create a friendly and relaxed atmosphere for all participants.

3.5 Experiment Procedure

Pre-task

Prior to the experiment, participants received an information sheet describing the purposes, procedures, and uses of the data for the study. It was given in Japanese to the native speakers and in English to Swedish learners. Written informed consent was also given by all participants. They were informed of their right to withdraw or halt at any time.

In order to familiarize the participants with the process, a warm-up exercise was performed wherein they read three to ten practice sentences. This was done so as to familiarize them with the interface, and facilitate comprehension of the task.

Task 1: Unconscious Reading

In the first task, participants were instructed to read 80 sentences presented sequentially on a computer screen aloud . The sentences comprised of the following types:

5 syntactically ambiguous sentences with examples of LB-interpretation

5 ambiguous sentences with RB-interpretation visuals,

5 different LB filler sentences

5 unambiguous RB filler sentences.

Every sentence was accompanied by a picture aiming at biasing the target reading implicitly without prompting. Images and sentences appeared one at a time in PsychoPy, and individuals pressed a key to advance at their own pace.

Since this study examines pitch patterns at the sentence level, measures were taken to minimize confounding variables and enhance phonetic precision. In order to limit exposure to uncontrolled lexical pitch errors, all stimuli presented to Swedish subjects were furigana-annotated and IPA-based pitch accent marked as shown in the figure 9. These were intended to guide subjects to pronounce the words correctly without disrupting syntactic processing.

Figure 9 Sample sentence with Furigana and IPA-Based pitch accent marking

These supports were clearly specified as pronunciation support only and were not intended to affect syntactic interpretation. The subjects were reminded by means of written instructions and orally prior to the task—that the final goal was natural spontaneous prosody production.

Task 2: Disambiguated (Conscious) Reading

The second experiment tested the effect of explicit awareness of syntactic ambiguity on prosody. A subset of the same ambiguous items (5 LB, 5 RB) as the previous task was presented with visual contrasts. These image pairs were designed to make the ambiguous structure of each sentence readily apparent and assist participants in attending to the intended meaning. Figure 9 presents an example of such a visual distinction.

They were told to read each sentence in a manner that made the meaning clear, as signaled by the visual prompt. They were told to clarify the sentence through intonation, but they were also told to be spontaneous and read as naturally as possible.

Figure 10 Example of visual contrast used in task 2 to encourage prosodic disambiguation

Every target sentence was read four times in a randomized order and it was ensured that no identical items were displayed consecutively, yielding 40 read sentences per subject during conscious reading.

3.6 Data Processing

The collected speech data went through a multi-stage processing routine to ensure rigorous acoustic analysis of prosodic features. It entailed automatic scripting, followed by careful manual data cleaning and spanned several weeks.

Step 1: Audio segmentation

All recordings were initially imported into Adobe Audition, where every utterance was segmented manually. Segmented files were then labeled systematically to identify speaker ID and item ID to allow for accurate alignment of audio files and experimental stimuli.

Step 2: Praat word boundary labeling

Using Praat scripts written for this purpose, each sentence was divided into five approximate prosodic units: (1) topic phrase *kore wa*, (2) first genitive phrase (N1-GEN), (3) second genitive phrase (N2-GEN), (4) head noun (N3), and (5) sentence-final copula *desu*. The boundaries were hand-adjusted so that actual speech onset/offset was closely approached.

Step 3: Labeling of pitch points (min1, max, min2)

F0 (Fundamental frequency) targets were marked on the three target noun phrases (N1, N2, N3) with a semi-automated Praat script.[1] Three measurement points were marked for each noun:

$F0_{min1}$: first pitch valley that is immediately preceding the pitch rise (e.g., beginning of the noun),

$F0_{max}$: the peak of the pitch (typically over the first vowel of the noun),

$F0_{min2}$: the last pitch drop following the peak, usually aligning with the end of the noun.

For consistency, pitch tracks were visually checked and spurious values due to technical error were removed.

Step 4: Extraction of F0

From smoothed pitch tracks, Praat scripts pulled out F0 values at the marked points for the three target words. These were written into formatted tables for later processing.

Step 5: Speaker/item mapping and tabulation of data

All the acoustic data retrieved were compiled in Excel tables, together with the data of each speaker, item information, condition (LB/RB), and task number. Each sentence occurrence was labeled with its respective measurement points, in order to compare speakers and conditions crosswise.

3.7 Pitch Pattern Diagnosis and Acoustic Analysis

The contours were diagnostic measures of the following prosodic predictions:

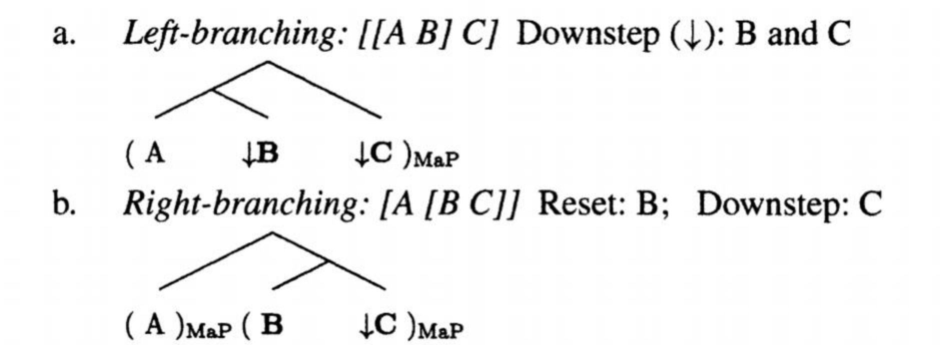

Figure. 11 A comparison of prosodic realization in both branching structures (taken from Ishihara, 2016, p.1396).

Left-branching structures (e.g., [[N1-GEN N2-GEN] N3]) are to be regarded as one Major Phrase, with cumulative downstep across constituents ($N1_{max} > N2_{max} > N3_{max}$).

[1] Segmentation and acoustic annotation were done manually by the author, then checked partially by one Japanese linguistics expert for precision and consistency.Word boundaries were lablled according to common acoustic cues like formant transitions, waveform discontinuities, and F0 contour movement.

According to early prosody analyses in downstep (e.g., Selkirk & Tateishi, 1991; Kubozono, 1993), right-branching structures (e.g., [N1-GEN $_{MaP}$ [N2-GEN N3] $_{MaP}$]) are routinely treated as two separate Major Phrases (MaP), usually triggering a pitch reset at N2. The reset routinely creates a local F0 peak ($N2_{max}$) that is slightly lower than $N1_{max}$.

To demonstrate these theoretical findings, Figure 12 and Figure 13 show some example pitch contours produced by a Japanese native speaker (JF4) from Tokyo in Task 1. These figures were created by Praat's picture function and illustrate the prosodic differences between left and rightbranching structures.

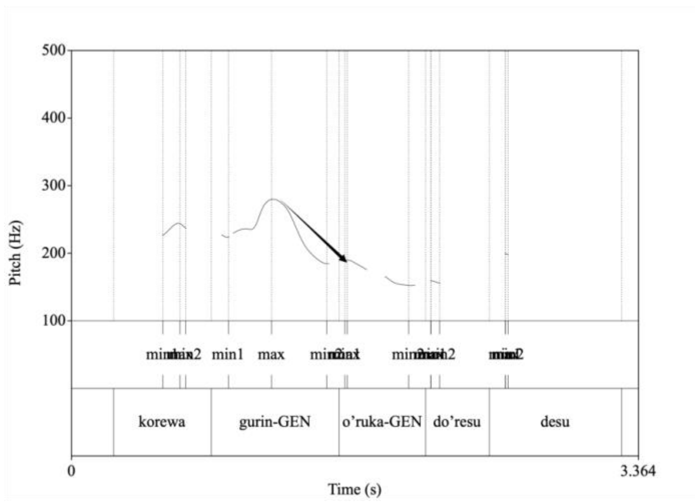

Figure 11 A representative pitch contour of a left-branching reading of the same sentence, produced by speaker JF4 (LB Item 04, Repetition 4 in Task 1). In this interpretation, "guri'n no" modifies do'resu ("dress"), forming the structure [green [orca's dress]].

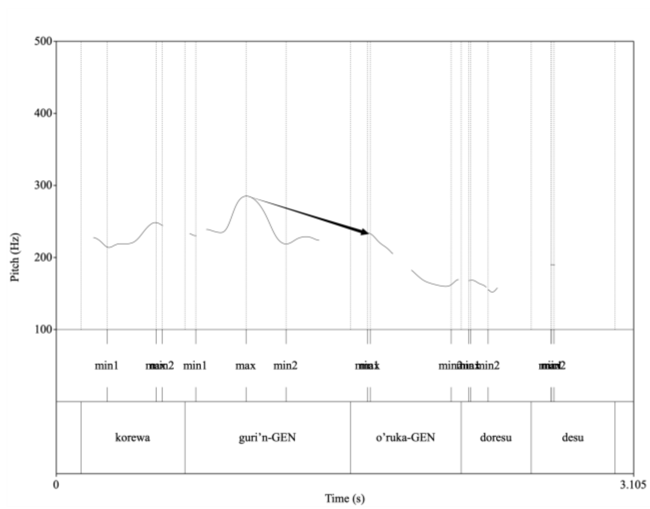

Figure 13 A representative pitch contour of a right-branching reading of the sentence *kore wa gurin no oruka no doresu desu* ("This is the dress with a green orca") (LB Item 04, Repetition 4 in Task 1). In this structure, "*guri'n no*" ("green") modifies o'ruka ("orca"), forming the constituent [[green orca]'s dress].

As Figure 12 and 13 show, the right-branching condition has a higher pitch at N2 ($N2_{max}$) compared to that of its left-branching counterpart. Relative pitch difference between $N1_{max}$ and $N2_{max}$ is smaller for the right-branching condition, meaning that $N2_{max}$ is nearer to matching $N1_{max}$ height.

In order to examine how prosodic realization indicates underlying syntactic structure, this study targeted F0 as the main acoustic cue. Three F0 measurement points ($F0_{min1,}$ $F0_{max}$ and $F0_{min2}$) were sampled for all three target nouns (N1, N2, N3) of every sentence.

This yielded nine measurement points per sentence (3 nouns × 3 points), enabling fine-grained tracking of the contour over the trajectory of the tripartite genitive phrase.

In order to investigate how participants prosodically realize structurally ambiguous sentences in line with left-branching and right-branching structures, and to be able to compare the F0 values across speakers with different pitch ranges, all the F0 values are linearly normalized, which is a method being used in Ishihara (2016). The highest mean peak F0 of each speaker (which is set to be the $F0_{max}$ of N1) was set at 1.0, and the lowest mean valley, which is typically the second lowest point in the N3 was given the position 0.0. This was computed from the following equation:

$$F0_{norm} = \frac{\mathrm{F0-R2}}{R1-R2}$$

Where: R1 is the speaker-specific mean of all $N1_{max}$'s F0 values, and R2 is the speaker-specific mean of all $N3_{min2}$' F0 values.

What is worth mentioning is that, although in the formula, the highest reference point is set to be $N1_{max}$, in a few voice samples of L2 learners, the F0 at $N2_{max}$ was higher than at $N1_{max}$, resulting normalized values above 1, their individual patterns can still be well-maintained in the normalized data. Thus, this method maintains each speaker's pitch pattern and make different speakers' pitch data to be compared.

To visualize overall trends in pitch contour, all normalized F0 values were also grouped together by syntactic condition (LB or RB) and measurement point (e.g., $N1_{max}$, $N2_{min1}$). The mean normalized F0 was calculated within each group across all native speakers. This allowed average pitch trajectories to be generated for each condition. The data was then analyzed using linear mixed-effects models in R. The aim was to examine whether statistically significant differences existed in F0 realization patterns between structural conditions and participant groups.

Chapter 4 Results

4.1 Prosody of Ambiguous Syntactic Structures Among Native Speakers in Unconscious Reading (Task 1)

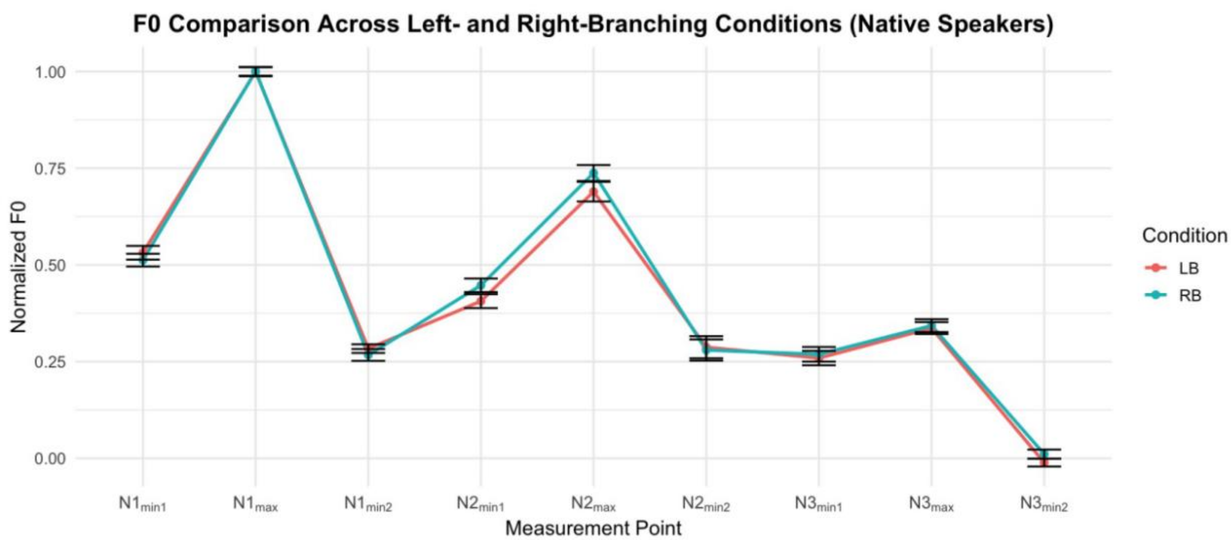

Figure 14 F0 Comparison Across Left- and Right-Branching Conditions (Native Speakers)

Figure 14 presents the resulting F0 contour lines for native speakers, separately plotted for left- and right-branching conditions. Each line is the grand average of the normalized values of all native speakers at each of the nine measurement points. In the LB condition, a pitch declination is apparent: $F0_{max}$ values gradually decrease from N1 to N2 to N3. From the graph, this pattern is in line with the predictions of the cumulative downstep model, which follow from the assumption that in left-branching structures, the three constituents form one overall major prosodic phrase, and thus result in uninterrupted pitch lowering across constituents (Selkirk & Tateishi, 1991).

Conversely, the RB condition exhibits a slightly shallower pitch drop on N2. That is, the $F0_{max}$ at N2 appears higher than for the LB condition.

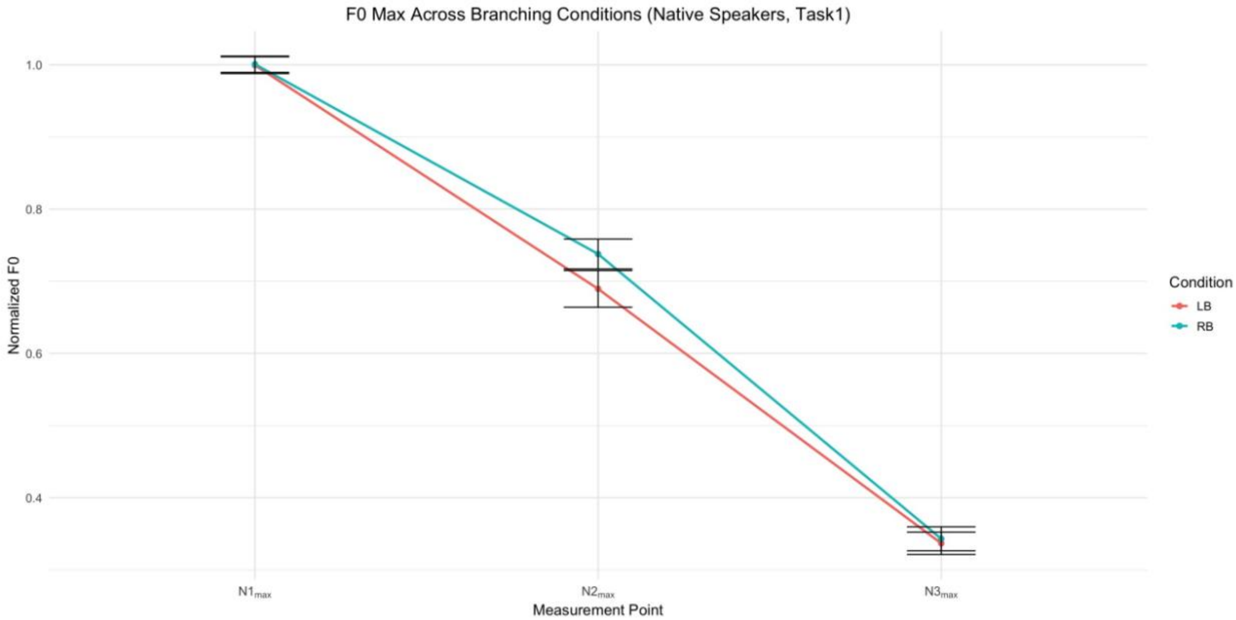

Figure 15 F0 at peak (max) positions within syntactic constituents (N1, N2, N3) in left- and right-branching conditions[2]

Fig. 15 depicts a line graph of $F0_{max}$ of all nouns (N1, N2 and N3), showing only the overall downstep pattern without the local peaks and valleys associated with each noun.

In the LB condition, a downstep pattern is observed systematically ($N1_{max} > N2_{max} > N3_{max}$), while the RB condition illustrates a slight pitch rise at N2.

Because sample size of Japanese speakers is small (N=4), in order to ascertain if the pitch height at N2 ($F0_{max}$) is different for the two interpretations of the sentences (LB & RB), a Wilcoxon rank-sum test was carried out on the F0 values normalized at $N2_{max}$ in both conditions for native speakers in Task 1.

Although the result of the test does not permit us to conclude that the median Normalized F0 at N2 max is different for the two conditions (test stat. $U = 2783.0$, $p = 0.155$) this could, in part, be due to the small sample size.

[2] In chapter 4, error bars represent ±1 standard error of the mean F0, calculated across speakers for each measurement point

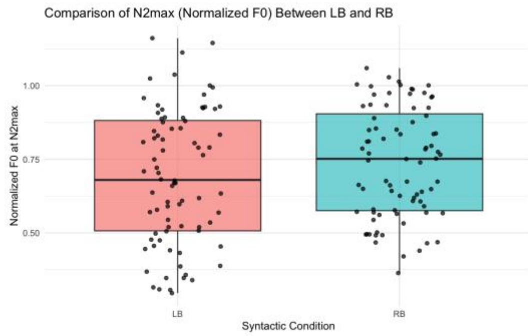

Figure 16 Comparison of $N2_{max}$ across LB and RB conditions among 4 Japanese speakers

A boxplot (Figure 16) displays the distribution of the normalized F0 values for both conditions at $N2_{max}$. According to the plot, RB has a slightly higher median pitch than LB, but because at a significance level of 5%, this difference is not considered statistically significant.

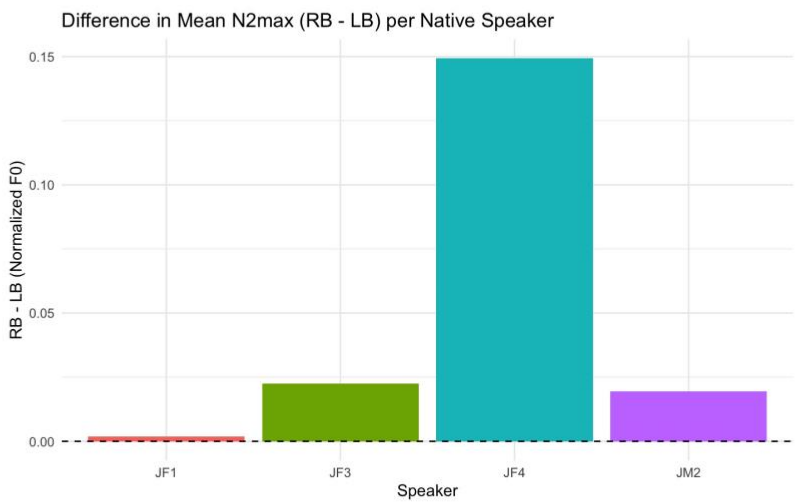

Figure 17 Comparison of $N2_{max}$ across LB and RB conditions for individual Japanese speakers

According to figure 17, the normalized F0 differences at $N2_{max}$ for the Right-Branching and Left-Branching conditions for individual native speakers are different. The values represent the average difference per speaker, calculated by subtracting LB from RB.

Of the four native speakers: JF4 (a female from Tokyo) has a clear pitch rise when in the RB condition, with a difference of more than 0.15. This shows that this speaker realized a rather strong pitch accent at N2 in RB. JF3 and JM2 show minor but clear increase in RB over LB. JF1 shows almost no distinction between the two states. These findings suggest that there may be an overall trend for larger $N2_{max}$ in RB, but the extent of the difference is quite variable across speakers. Only one speaker (JF4) has a strong RB > LB effect.

For speaker JF4, a Wilcoxon signed-rank test was then conducted between pitch height at $N2_{max}$ in the RB and LB conditions. The result showed a statistically significant difference ($V = 182$, $p = 0.0027$), i.e., that JF4 indeed had a tendency to realize higher pitch peaks at N2 in the RB condition compared to the LB condition. This reveals a clear prosodic difference in line with the syntactic interpretation, at least for this particular speaker.

4.2 Prosody Realization Among Swedish L2 learners in Task 1

To examine whether Swedish L2 learners are sensitive to using different pitch patterns in syntactically ambiguous constructions when they read out sentences without knowing their ambiguity, normalized F0 contour across three content words (N1, N2, N3) under LB and RB conditions were compared.

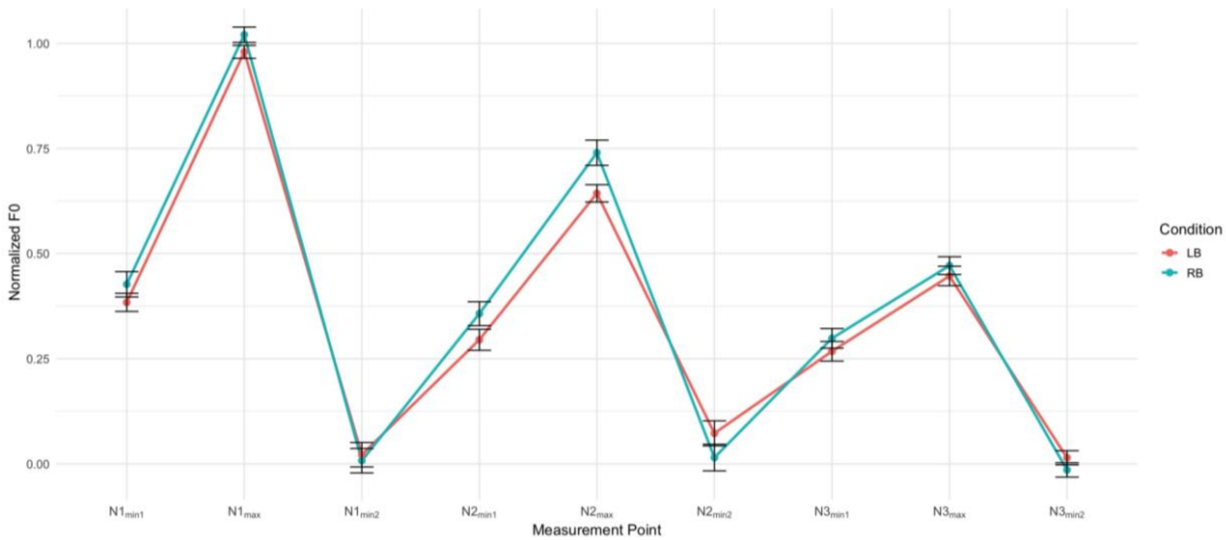

Figure 18 Mean normalized F0 contours across nine pitch measurement points (min1 –max–min2) for learners under left-branching (LB) and right-branching (RB) conditions in Task 1

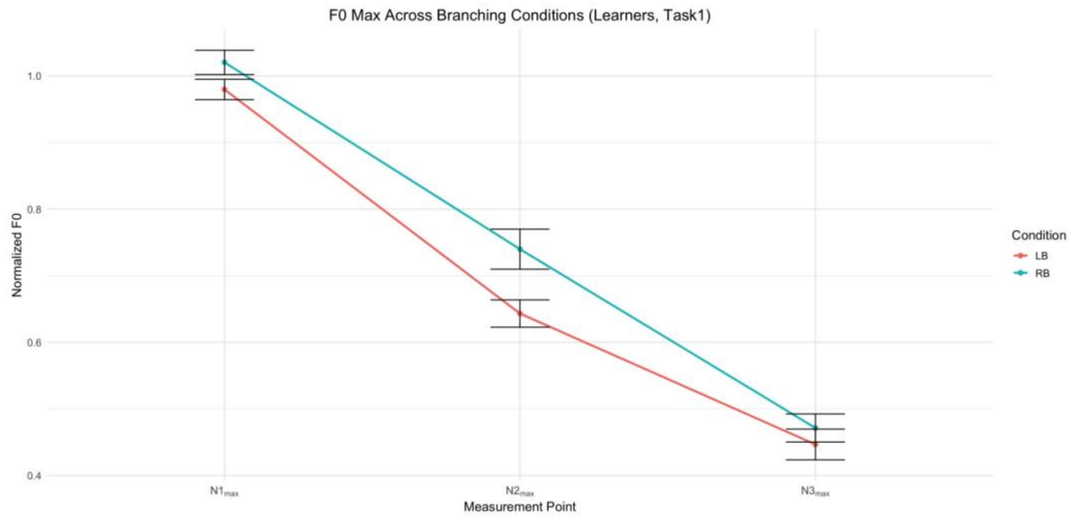

Figure 19 Mean normalized F0 values at maximum pitch positions (N1max, N2max, N3max) for learners under left-branching (LB) and right-branching (RB) conditions in Task 1.

Figure 18 shows Swedish L2 learners' average F0 contours at nine measurement points of pitch (min1–max–min2) on three content words (N1, N2, N3) for the LB and RB structures of Task 1 (unconscious reading). The shape of the overall contour shows a downstep pattern, which is characteristic of maximum pitch of each content words drop gradually. There are clear pitch peaks at $N1_{max}$, $N2_{max}$, and $N3_{max}$. The trajectories for LB and RB conditions are also virtually identical at most points.

At $N2_{max}$, where previous studies indicate possible disambiguation cues by pitch prominence, RB is slightly higher in F0 than LB, though the difference appears small.

A linear mixed-effects model was then used to test whether the F0 peaks at N2 of learner group varied between left- and right-branching structures in Task 1. The model showed a significant main effect of condition (Estimate = 0.097, $p < .001$), i.e., that L2 learners had a higher pitch at $N2_{max}$ with the right-branching interpretation compared to the left-branching one. This suggests that even in the absence of overt awareness of ambiguity, spontaneously, learners exhibit some sensitivity to the overall structure of sentences.

4.3 Prosody Comparison Between Native Speakers and L2 Learners in Unconscious Reading

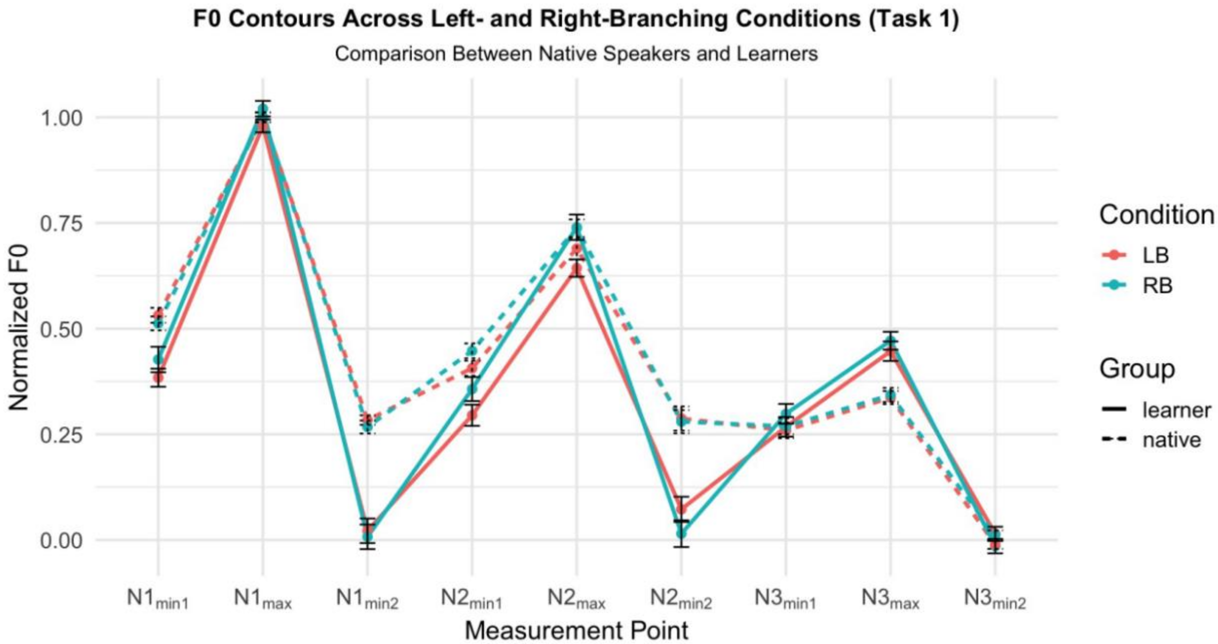

Figure 20 Pitch contours of both groups in task 1 under two conditions

Figure 20 illustrates the average F0 contours generated by native speakers and L2 learners on nine measurement points in left-branching and right-branching syntactic structures. The pitch values were normalized for each speaker as a way of controlling for between-speaker variation in pitch range. For both groups (native and learner), F0 values at each measure point were averaged and their standard errors calculated over all tokens. These have been graphed as lines with error bars, with red for LB condition and green for RB condition. Solid and dashed lines were used to differentiate the two groups.

According to the graph, both groups exhibit rising and falling contours with local peaks at $N1_{max}$ and $N2_{max}$. Yet, the L2 learners exhibit generally lower F0 values at various low points (e.g., $N1_{min2}$, $N2_{min2}$). Another seeming difference exists in $N3_{max}$, where learners generate higher pitch values than natives. These visual trends encourage additional examination by way of mixed-effects modeling.

The nine measurement points' pitch values were separately compared for the L2 learners and the native speakers under the left-branching condition and right branching condition with linear mixed-effects model. The model included group, measurement point, and their interaction as fixed effects and speaker as a random intercept.

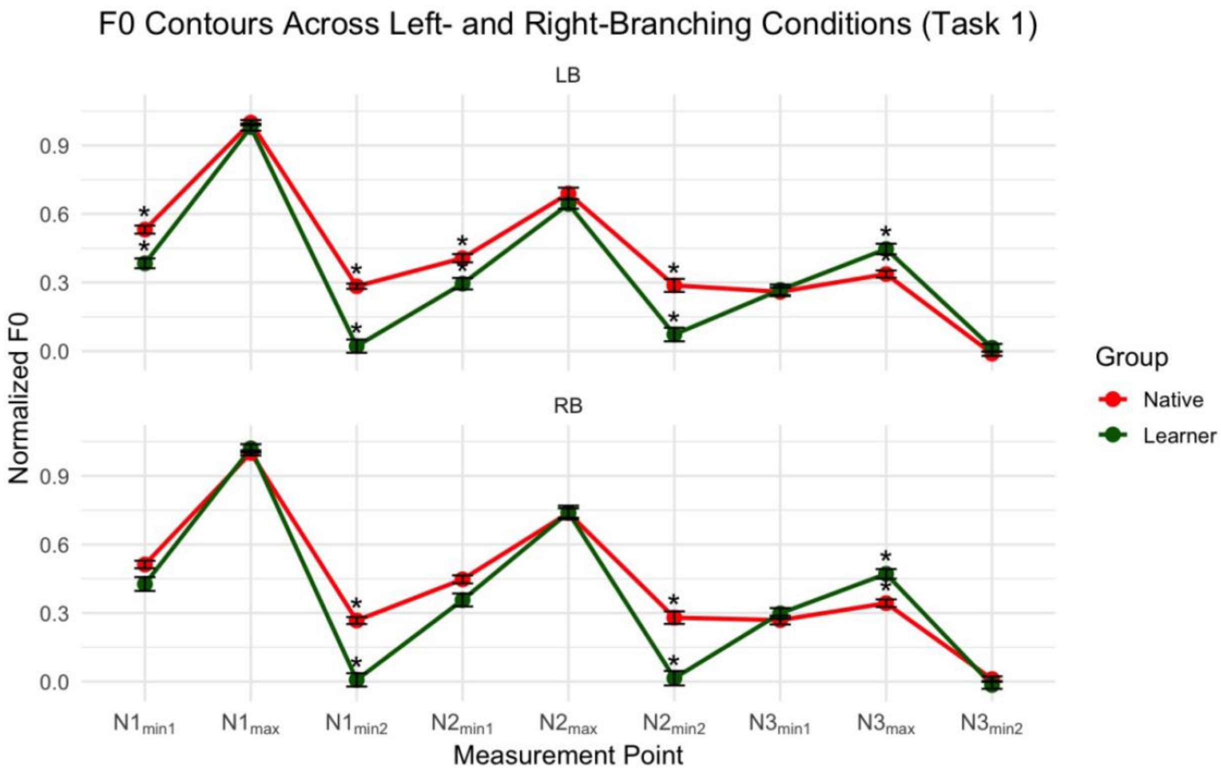

Figure 21 F0 contours of two groups across two conditions in task 1

The results revealed some significant differences. To begin with, significant between-group differences were found at $N1_{min1}$ (LB), $N1_{min2}$ (both LB and RB), $N2_{min1}$ (LB), and $N2_{min2}$ (both LB and RB). These positions normally occur in prosodic valleys of the sentence, and the findings indicate that the pitch of learners drops more steeply than that of native speakers. In contrast, at the sentence-final peak point ($N3_{max}$), learners had significantly greater F0 values than native speakers for both LB and RB conditions. However, no differences were found at the mid-sentence high pitch peak points ($N1_{max}$, $N2_{max}$) or at the other low points ($N3_{min1}$, $N3_{min2}$), where pitch values for the two groups were closer to each other. These results indicate that the most significant differences between native speakers and learners are in the first pitch drops and final pitch peak.

4.4 Awareness Effects on Prosody Realization (Task 1 vs. Task2)

Apart from the spontaneous and unconscious reading task (Task 1), a follow-up task was also done to see if speakers were manipulating their prosody to render structural interpretation more marked when they were consciously attending to syntactic ambiguity. In this task, participants were presented with two contrastive intepretation of the same sentence.

In order to study how task awareness influences prosodic realization in native Japanese speakers and Swedish L2 learners, pitch contours between Task 1 (unconscious reading) and Task 2 (conscious reading) were compared. Two syntactic conditions, LB and RB, were analyzed separately in order to observe awareness-related differences. Normalized F0 values for nine measurement points were compared in each condition. Particular focus was on $N2_{max}$, a prosodically prominent location, to see if task-induced awareness resulted in more salient cues like pitch reset.

First, to illustrate each group's pitch, a overall pitch contour plot was generated using R studio to visually show the differences in prosodic patterns between the two groups across the two tasks.

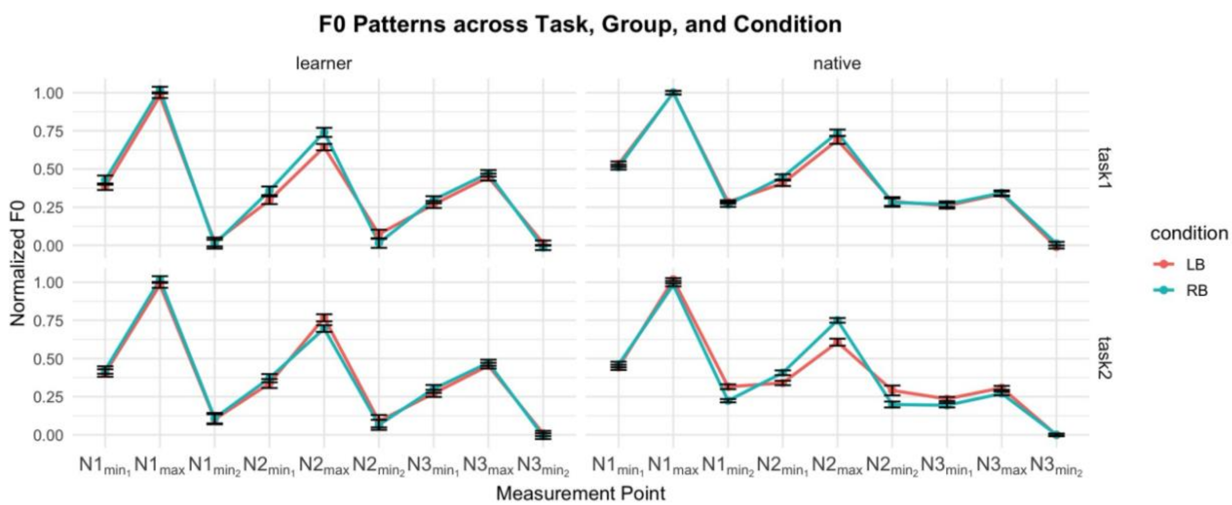

Figure 22 Mean normalized F0 contours (min_1–max–min_2) for N1–N3 across LB and RB conditions in two tasks (task 1 = unconscious reading, task 2 = conscious reading), separately plotted for learners and native speakers. Error bars represent ±1 SE

To investigate if ambiguity awareness influenced pitch realization, linear mixed-effects models were constructed for each group (native speakers vs. L2 learners), each syntactic condition (left-branching vs. right-branching), and the pitch of an important measurement point that is supposed to differ in both conditions: $N2_{max}$.

A linear mixed-effects analysis revealed a significant task effect for native speakers at $N2_{max}$ for the left-branching condition (Estimate = –0.082, $p < .001$), with significantly lower pitch in reading in Task 2 (with structural awareness) than Task 1 (spontaneous reading). This indicates that native speakers are able to enhance prosodic downstep for syntactic disambiguation when they are aware of underlying ambiguities.

However, unlike in the left-branching condition, a linear mixed-effects analysis did not reveal a significant task effect for native speakers at $N2_{max}$ for the right-branching condition (Estimate = 0.012, $p = .564$). This suggests that native speakers did not change pitch at this point based on their understanding of structure, which may show that right-branching structures are less marked phonologically.

In terms of learner group's data, the linear mixed-effects model revealed a significant task effect at $N2_{max}$ for L2 learners in the left-branching condition (Estimate = 0.124, $p < .001$). Contrary to native speakers, who lowered their pitch in Task 2, learners utilized significantly higher F0 when they were aware of the ambiguous structure. In the right-branching condition, the linear mixed-effects model showed no task effect at $N2_{max}$ for the L2 learners (Estimate = –0.044, $p = .205$). While there was some decrease in pitch in Task 2, the difference was not significant.

4.5 Individual Variation in L2 Prosody Realization Among L2 Learners

Although the previous sections have discussed the general prosodic trends of L2 learners in unconscious (Task 1) and conscious (Task 2) reading tasks, group-level research of this kind can potentially mask some L2 learners' individual differences. To develop a clearer understanding of the manner in which L2 learners differ in their utilization of pitch cues, therefore, this section discusses individual learner patterns, with special interest in deviation from or approximation to native-like prosody.

Among all of the L2 learners, SM5 and SF2 both reported themselves as advanced learners, having better skills in listening, speaking, writing and reading. Also, they reported to have extensive exposure to Japanese and high usage in daily life. They were thus categorized as the advanced group. The other four learners (SM2, SM4, SF1, and SF3), on the other hand, were categorized as intermediate, either by virtue of their Japanese Language Proficiency Test (JLPT) N3 certificate or their completed undergraduate studies in Japanese at Lund University.

To see their pitch pattern, first, error rates were computed separately for both groups. For the line graphs, the percentage of violations of prosodic boundaries in all left-branching structures in both tasks for each learner was calculated. The two kinds of errors that were taken into consideration were: (1) $N2_{max} > N1_{max}$ and (2) $N3_{max} > N2_{max}$.

The Y-axis of the graph represents the percentage of sentences, under the LB condition, where a learner made the prosodic error $N2_{max} > N1_{max}$—i.e., the pitch peak of the second noun was higher than that of the first noun. This is a violation of the expected downstep pattern of Japanese, in which pitch peaks are to successively lower (N1 > N2 > N3).

In the graph, for example, a 0.35 score for a learner indicates that in 35% of the sentences they had produced under the LB condition, N2's pitch peak was higher than that of N1. This implies that more than one-third of their productions differed from the expected prosodic marking of syntactic structure in Japanese.

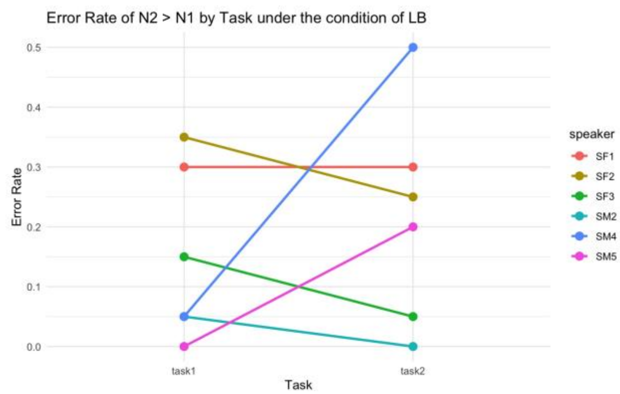

Figure 23 Error rate of $N2_{max} > N1_{max}$ by task under the condition of LB (SF2 and SM5 are advanced learners, and others are intermediate learners)

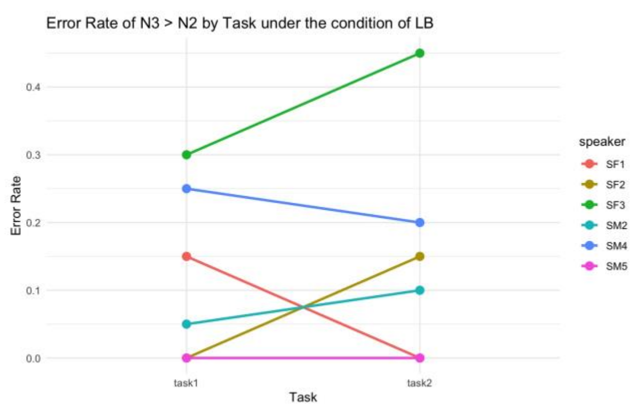

Figure 24 Error rate of $N3_{max} > N2_{max}$ by task under the condition of LB (SF2 and SM5 are advanced learners, and others are intermediate learners)

Figure 21 displays proportion of $N2_{max} > N1_{max}$ errors in the LB condition, and Figure 22 displays another type of error, i.e., where $N3_{max} > N2_{max}$ in the same condition. In both error patterns (N2 > N1 and N3 > N2), learners also exhibited great variability in prosodic performance between tasks. SM4, for instance, under LB condition, shows a sudden rise in error rate in Task 2, to 0.5. SF2 and SF1 show fairly steady and small error rates, and SM2 and SF3 show improvement with moderately lowered error rates. In contrast, SM5 moves from error-free performance in Task 1 (0 error) to a rather low error rate of 0.2 in Task 2. In Task 2, SF3's error rate rises slightly to 0.45 from 0.3, whereas that of SF1 falls significantly, with errors going down from 0.15 to 0. SF2, SM2, and SM4 show minimal fluctuation across tasks. SM5 retains an error rate of 0 for both tasks.

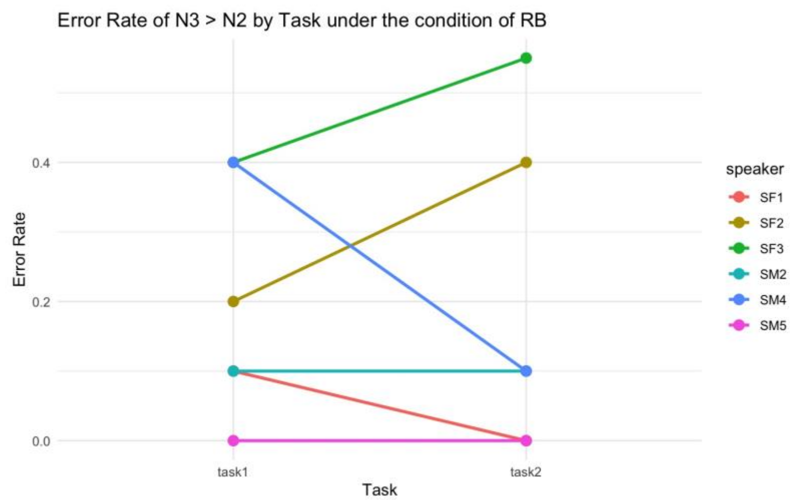

Figure 25 Error rate of $N3_{max} > N2_{max}$ by task under the condition of RB (SF2 and SM5 are advanced learners, and others are intermediate learners)

Figure 25 plots how frequently the learners produced unexpected pitch peaks—a rise from $N2_{max}$ to $N3_{max}$ in the RB condition.

Among the subjects, SM5, a self-rated advanced learner, who claimed to have received systematic intonation training, attained a strikingly native-like prosodic structure with almost no downstep errors in either error type.

To further illustrate individual pitch trajectories and compare learners' prosodic patterns with those of native speakers, the following line graphs present normalized F0 contours across measurement points and task conditions.

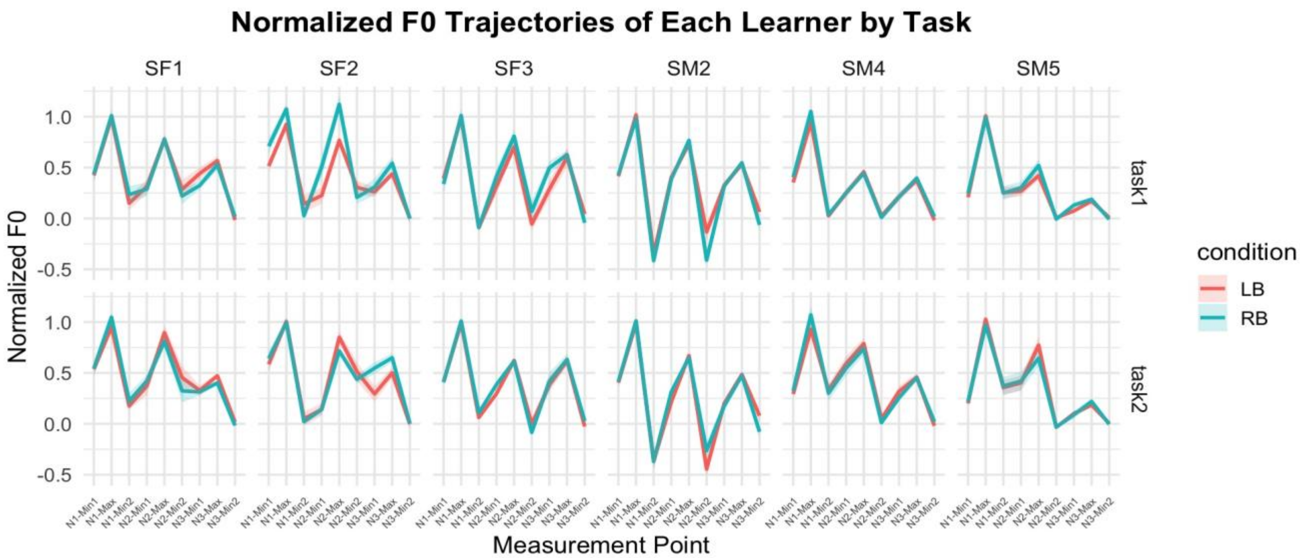

Figure 26 Pitch contour of each learner by task (SF2 and SM5 are advanced learners, and others are intermediate learners)

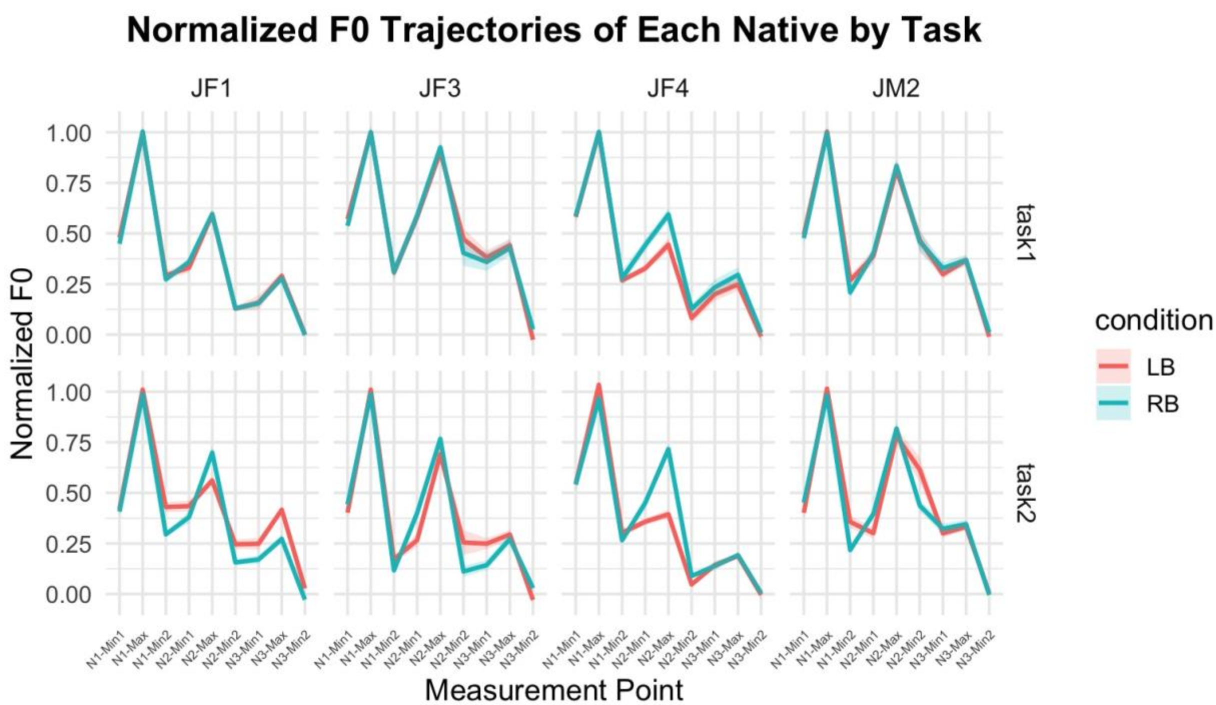

Figure 27 Pitch contour of each native by task (among all participants, JF4 is from Tokyo)

The pitch contour trajectories in Figures 24 and 25 indicate clear differences within and between groups in the use of prosody across tasks.

Among the learners, SM2—an intermediate participant—showed very salient F0 changes, sudden slope-down and slope-up changes in pitch. The close overlap of the LB and RB lines, however, shows that there was no adjustment to syntactic boundaries and systematic differences between task 1 and task 2. SF3, who is another intermediate learner, showed downstep pattern correctly in task 1, specifically with $N2_{max}$ at RB condition higher than its counterpart in LB condition. However, in task 2, there was no expected downstep pattern between N2 and N3, and both conditions showed highly similar pitch contours. As for SM4, another intermediate participant, although his pitch contours of LB and RB are almost the same at first glance, but actually he elevated his pitch at $N2_{max}$ than its counterpart in task 1.

SF1 showed a reversed trend for both tasks. The LB condition in the task 2 yielded higher $N2_{max}$ values than the RB condition. This reversal of the natural order of pitch cues may reflect that there is no clear strategy to correctly use pitch to mark syntactic boundary in RB condition.

As for advanced learners, SF2 demonstrated a productive pattern of syntactic-prosody alignment: under Task 1, her $N2_{max}$ pitch under the RB condition rose appreciably—close to the height of $N1_{max}$—that approximates native pitch reset behavior very closely and reflects growing sensitivity to syntactic structure through prosody. SM5, the other advanced learner, showed a stable prosodic pattern in Task 1. That is, the RB condition also observed a clear rise in pitch at $N2_{max}$. However, this pattern was not continued in Task 2. Specifically, SM5 showed an unexpected rise in pitch in the LB condition, where the $N2_{max}$ pitch was as high as or even higher than that of the RB counterpart. A possibility is that careful attention to sentence structure differences in Task 2 induced error or inappropriate use of pitch, particularly in a speaker who is already sensitized to pitch pattern changes.

In native speakers, JF4 was the one who used pitch reset consistently in all tasks. In both RB readings, she pushed $N2_{max}$ much higher than the LB counterpart, revealing her sensitivity to sentence boundaries. Given that JF4 is the sole Tokyo participant, this is in line with previous research that standard Tokyo Japanese speakers more frequently exhibit pitch reset effects.

The other native speakers (such as JF1, JM2, JF3) revealed less distinct patterns in Task 1, with LB and RB lines being closely comparable. However, in Task 2, almost all native speakers showed more pronounced RB–LB divergence, with $N2_{max}$ rising in the RB condition, suggesting more use of prosodic disambiguation strategies under syntactic sensitivity.

The results show that the manipulation of awareness (Task 2) helped the native speakers in prosodic disambiguation but created additional variability for the L2 learners. Native-like prosodic responses to structural ambiguity were found only in a subgroup of advanced learners.

Chapter 5 Discussion

This chapter presents the results with reference to the three research questions and prior research on Japanese prosody and L2 prosodic acquisition.

The main aim of the current research was to study how and if Swedish learners of Japanese realize Japanese structurally ambiguous noun phrases prosodically compared to native Japanese speakers and whether or not their prosody is influenced by ambiguity awareness.

Despite the fact that earlier research has either concentrated on word-level pitch errors or compared L2 production with idealized native contours, the present study sought to focus on sentence-level prosody, in addition to investigating variability within L2 speakers themselves.

The experiment had three research questions: 1) Do Swedish learners of Japanese use prosodic cues, particularly pitch-related features such as downstep, in syntactically ambiguous sentences? 2) Does awareness of syntactic ambiguity affect natives and learners' prosodic realization? 3) How does proficiency (intermediate vs. advanced) influence prosodic accuracy and error patterns among learners?

5.1 RQ1: Prosodic Cue Use in Structurally Ambiguous Sentences Between Two Groups

This study first investigated how native speakers and L2 learners use pitch to mark sentence structure. It used simple noun phrases (e.g., N1-GEN-N2-GEN-N3) to create branching ambiguity that would correspond to left- and right-branching meanings. All pitch values at each measure point were converted to normalized data so as to eliminate speaker differences, as explained in Chapter 4.

The results showed that intermediate and advanced Swedish L2 learners showed a general trend of downstep from N1 to N2 to N3, which is near-native-like prosodic phrasing. This might indicate that Swedish learners of intermediate and advanced Japanese proficiency level have developed an implicit awareness of the prosody pattern of Japanese noun phrases, particularly the expected downstep pattern across constituents. However, there are also noticeable variations within learner's group. Intermediate learners had more variable pitch patterns, higher error rates with pitch's rises and falls not necessarily matching sentence structure. Advanced learners produced more even patterns, showing they had a stronger grasp of sentence rhythm.

Another finding of this study is that under the right-branching condition in task1, some native speakers did reset their pitch at $N2_{max}$ and some did not show difference between two branchings. In Task 1, for instance, native speaker JF4 produced a distinct pitch rise at $N2_{max}$ in the RB condition, in contrast with the more drastic downstep contour in the LB control. The rise can be taken to mark the occurrence of a prosodic boundary, presumably a Major Phrase boundary, between N1 and the constituent [N2–N3]. This boundary is attested by the measured pitch contour and aligns with earlier findings that MP boundaries evoke pitch resetting at its left edge (Selkirk & Tateishi, 1991). As for the reason why only JF4 showed this pitch reset pattern in $N2_{max}$ in right branching structure under task 1, That is probably due to the fact that JF4 was the only participant in the native group who was originally from Tokyo. Previous experimental studies on Japanese prosody (Selkirk & Tateishi, 1986, 1991; Kubozono 1993), including those reporting consistent pitch reset or downstep, have primarily used standard Tokyo Japanese data to analyze. It is therefore possible that dialectal variation influenced the realization of prosodic boundaries in the current study. Although all participants reported having the ability to speak standard Japanese, native speakers from non-Tokyo regions may have internalized regional prosodic features that subtly interfere with the canonical Tokyo Japanese pitch patterns. This may explain why in task1 the pitch reset pattern observed in JF4 was more significant than in other native speakers. This experiment only recruited one participant that is from Tokyo, so this must be a limitation. To understand if dialectical background is the reason that caused other Japanese participants' pitch reset in right branching structure to be insignificant, future research must include a larger population of native speakers from Tokyo and also areas outside Tokyo. More strict comparison could be able to separate the influence of regional effect on pitch reset.

Interestingly, some Swedish L2 learners also exhibited a similar pitch pattern to natives in Task 1. Although the task was reading out stimuli sentences with visual depiction that would guide participants towards a specific interpretation, learners in general also had a significant pitch rise at $N2_{max}$ under the right-branching condition, which is similar to native-like prosodic phrasing. This implies that although L2 speakers are not overtly aware of structural ambiguity, they may still produce prosodic boundary cues naturally like native speakers. Downstep is not a typical boundary cue for Swedish prosody, so the phenomenon observed is less likely to be attributable to L1 transfer. However, it should be noted, that the current sample of learners was relatively small and skewed toward higher-level speakers. As advanced learners in this study produced exceedingly low error rates and more native-like pitch patterns, it is possible that their performance disproportionately affected the group-level results. This means that the prosodic behavior found here could be representative of a more advanced subsection's capabilities instead of that of the general population of L2 learners. Additional research with a larger, proficiency-balanced sample is necessary to determine the generalizability of these findings and to contribute further to the understanding of how prosodic acquisition unfolds at various points in L2 development.

5.2 RQ2: Effects of Syntactic Awareness on Prosody Realization

In Task 2, where branching ambiguity was made extra obvious, participants were expected to make the prosody pattern salient to reflect the intended meaning and thus marking the syntactic boundary more clearly.

After observing native speakers' individual pitch contour graphs between task 1 and task 2, it is not hard to find that native speakers indeed showed a more perceptible prosodic difference between left-branching and right-branching sentences. Notably, the F0 value of $N2_{max}$ was considerably lower in the left-branching condition, making the downstep pattern more apparent. Most speakers intentionally elevated the pitch at $N2_{max}$ in the RB condition in task 2 to mark the beginning of a new prosodic unit. In task 1, which is to elicit natural and automatic reading, most native speakers showed no difference in their pitch usage between LB and RB conditions. This aligns well with one finding in Allbritton et al. (1996): speakers usually do not supply enough prosodic cues on standard reading tasks when they do not know the ambiguous structure and they will use more salient cues when they are aware of the ambiguity.

The finding implies that native speakers utilized pitch cues more deliberately, such as enhancing pitch height at $N2_{max}$ with the aim of signaling syntactic structure when contrast between meanings was made salient.

On the other hand, L2 learners as a whole showed a different direction of trend between task 1 and task 2 overall. Unexpectedly, in the RB condition of Task 2, the learners demonstrated a lower pitch at $N2_{max}$ compared to its LB counterpart, reversing the pattern observed for native speakers. Thus, even though Task 2 showed participants the contrast between the two meanings, L2 learners did not manage to show more differences as natives in the RB and LB conditions. Notably, for advanced learner, such as SF2 and SM5, the difference between the two types of branching was clearer in Task 1 compared to Task 2. When they were given the instruction of natural reading (Task 1), both learners had a clear rise in pitch at $N2_{max}$ in the RB condition compared to the left-branching one, which is similar to native speakers. However, this difference in pitch became weaker or even lost under Task 2 when they were aware of ambiguity in sentences.

This variability can partly be accounted for by the design of task. Some participants acknowledged that they could not easily guess out the purpose of the design and thus could not successfully produce pitch differences based on two potential interpretations. They claimed that they believed that they have conveyed the intended interpretation in the experiment. The learners' data thus indicate that, although they noted the ambiguity and syntactic distinctions, their prosody patterns were obviously individualized and not systematic.

5.3 RQ3: Individual Variability in Prosody Accuracy and Error Patterns

Apart from making inter-group comparisons, looking into individual differences within the Swedish learner group also showed some trends, since prosody usage was seen to have individual differences among native speakers, one would expect the intra-group differences to be even greater in the L2 group. Intermediate learners often produced more irregular pitch patterns and made mistakes even in associating speech rhythm with sentence structure. For instance, in both LB and RB readings, N3 is always located at the lower end of one prosodic domain, thus no pitch peak is expected in the $N3_{max}$. However, after calculating error rates, it is clear that there are still intermediate learners who produce high pitch at $N3_{max}$ in their speech, which showed that they had not yet acquired the use of sentence-level prosody and that learners' knowledge of prosodic boundaries was not stable at this level. Also, for a few intermediate learners, such as SM2, the pitch range in their speech is quite big such as very low pitch valleys and overly high peaks. This could reflect some learners being extra cautious of their pronunciation accuracy or their inclination of marking each noun phrase separately without paying attention to producing sufficient smooth speech.

Advanced learners' data, in contrast to that of intermediate learners, exhibited more consistent and more accurate prosodic performance, with limited errors in downstep or phrasing patterns in both tasks. Both advanced learners (SM5 and SF2) came very close to that of the native group, using sentence-boundary downstep consistently and making a clear distinction between left and right branching structures. Also, their pitch range between valleys and peaks are highly close to that of native speakers. Notably, one advanced learner reported that they had training lessons on Japanese intonation, including private lessons focused on prosody and regular practice alone using shadowing and imitation. Although only one participant (SM5: an advanced learner) reported having special training in intonation, they also indicated that native speaker imitation is their regular practice. That involves recording their own voice, playing it back to check, and shadowing native speakers' speech. This provides some evidence for previous claims that prosodic characteristics, although subtle and challenging to acquire, are not unlearnable and unacquirable through guided input and feedback (Mennen, 2015). Therefore, pedagogical methods that involves explicit practice, especially through intonation-focused reading or shadowing, are likely to help learners internalize pitch-based disambiguation strategies and improve sentence-level fluency in the target language. This study is limited by the low small size, but if there are many participants, with a range of Japanese language levels, maybe this tendency to have native-like prosody could be established more strongly.

Future intonation teaching models might therefore take into account the inclusion of prosody-based practice earlier in L2 teaching, particularly for learners whose L1 lacks similar phonological systems. This practice might facilitate earlier boundary cue sensitivity and greater accuracy in learners' syntax-to-prosody mapping.

5.4 Limitations

This study has a number of limitations worthy of note. In terms of methodology, the sample size of participants is quite small, comprising only four native Japanese speakers and six Swedish Japanese learners. The limited number of subjects limited the study to be small-scale statistics analysis, which in turn constrained the generalizability of the results obtained and enhances the possibility of reported effects being due to individual difference as opposed to systematic tendencies. Recruitment process was very hard to find qualified Swedish participants whose Japanese proficiency has reached intermediate or advance. Also, due to the fact that audio recording sessions were designed to take place offline to make sure audio quality, there were not many learners and natives that could attend offline. Thus, sample size is quite limited.

Another limitation is that although most L2 learners have graduated from Japanese major, their Japanese proficiency is self-reported, except for two who have obtained JLPT certificates. Future research could conduct Japanese language test for more rigorous evaluation.

Additionally, the population of recruited participant is biased toward intermediate-level respondents, in part because of difficulties in recruiting advanced subjects, potentially constraining the validity of comparisons across proficiency levels. On top of that, although the majority of words in the stimuli list were common in daily life and thus being familiar to the L2 participants, pitch accent information of each word was overtly indicated to aid their pronunciation. However, this control may not entirely account for the actual way learners produce pitch accents individually because some of them might have wrong pitch patterns without relying on the pitch marking. In hindsight, instead of taking lexical pitch accent as a controlled variable, including it as part of the study—such as examining if intermediate and advanced learners can use pitch correctly both in word-level and sentence-level prosody might be more meaningful.

The task 2 design has limitation as well. In task 2, the subjects were presented with two interpretations of the same sentence. Although this is effective in bringing out conscious prosodic decisions, this design could inadvertently introduce narrow focus on specific constituents. According to Ishihara (2016), focus can independently raise F0. It is then difficult to isolate whether specific pitch patterns in the taks 2 are due to syntactic structure or focus-induced emphasis. However, when considering other kinds of design like interaction-based or dialogue tasks, which are liable to cause shifting focus and more complicated information structure, the current contrastive reading task probably reduces uncontrolled focus effects. Furthermore, subjects read alone, and they were required to still maintain natural reading. Thus, though focus-related interference can not be ignored, the task design might be a compromise.

Moreover, to test downstep pattern, the stimuli were created structurally simple, comprising of only a few brief noun phrases. This method captured pitch-related patterns more, but may also have made the results less generalizable. That is to say that, the simple structures, while being useful for checking prosodic patterns, may not represent complex sentence patterns learners are exposed to in spontaneous speech, where prosodic phrasing is often integrated with more profound meaning and discourse context.

In addition, the research was carried out solely in production, and there was no perception data on examining how L2 learners perceive prosodic cues and parse ambiguous structures. This means that one cannot tell whether the learners were able to perceive prosodic differences, even though they were not able to produce them consistently. A perception follow-up study would be able to show whether the learners are prosodically aware but not implementing it in speech.

Finally, from a theoretical standpoint, the current thesis did not make direct cross-linguistic comparison between Japanese language and Swedish language, which also possesses lexical pitch accent but differs in prosodic structure and use. Investigating possible transfer or interference effects could have revealed more about the course of L2 prosodic development. This thesis adopted Nagano-Madsen's (2015) interlanguage prosody perspective, interpreting learners' speech as part of an emergent interlanguage system rather than as a direct result of L1 transfer. However, integrating both interlanguage and L1 transfer perspectives may have yielded a more comprehensive understanding of how learners acquire prosody patterns.

Chapter 6 Conclusion

This research has investigated the use of pitch-related prosodic cues—downstep—by Swedish learners of Japanese in resolving syntactic ambiguity.

Three general research questions guided the study:

1) Do Swedish learners of Japanese use prosodic cues, particularly pitch-related features such as downstep, in syntactically ambiguous sentences?

2) Does awareness of syntactic ambiguity affect natives and learners' prosodic realization?

3) How does proficiency (intermediate vs. advanced) influence prosodic accuracy and error patterns among learners?

In order to answer these questions, speech data were gathered from Japanese native speakers and Swedish learners of Japanese in two task conditions: unconscious reading and ambiguity-aware reading. Test stimuli involved noun phrases that were open to two interpretations, left-branching or right-branching. Test items were carefully screened for lexical familiarity and pitch accent patterns to minimize confounding variables unrelated to the downstep phenomenon in Japanese.

The results indicate that native speakers, especially speakers from Tokyo, employed pitch changes in natural reading to indicate whether the sentence is interpreted as left branching or right branching, particularly when they were conscious of it.

Intermediate and advanced Swedish learners of Japanese, in general, can exhibit downstep pattern in syntactically ambiguous noun phrases. However, Swedish learners' prosodic realizations changed significantly across proficiency levels. Intermediate learners had a tendency for inconsistent prosodic phrasing and boundary marking, suggesting struggle in interfacing syntax and prosody. Advanced learners, on the other hand, demonstrated more native-like control of pitch, particularly for right-branching interpretations, and manifested clearer downstep realization with fewer phrase-level errors, especially one advanced learner had intonation training and demonstrated better prosody performance.

When they are conscious of syntactic ambiguity, native speakers of Japanese tend to make their pitch elevation at the beginning of new prosodic domain more salient to signify a right-branching meaning (e.g., [N1-GEN [N2-GEN N3]]).

On the other hand, most Swedish learners could not produce this increased pitch despite awareness of syntactic ambiguity. Even the higher-level learners produced greater N2 pitch in left-branching sentences in the awareness condition compared to the unconscious condition, which means that awareness of syntactics was not always capable of enhancing the right prosodic marking, but rather led them to apply their default intonation strategies incorrectly.

In conclusion, these findings have methodological implications for future research. The results suggest that syntactic awareness, as elicited by the experimental task, does not necessarily improve prosodic realization and can, in some cases, produce overcorrection or unnatural pitch alteration. This highlights the need for task design precautions in L2 prosody studies so that task-induced awareness will not intrude upon the prosodic patterns under investigation. From a pedagogical perspective, this research highlights the need for explicit instruction in prosody for L2 learners of Japanese, a language where sentence-level pitch plays a role in communication. For this, pedagogical strategies like pitch imitation, intonation shadowing, and boundary pitch training can help learners grasp the mapping of syntax and prosody. This will strengthen their ability to have more native-like speech.

References

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 714.

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics*, *35*(1), 167-182.

Archibald, J. (1998). *Second language phonology.* John Benjamins Publishing Company.

Azuma, J. (1997). Nihongo no tōgo kyōkai ni okeru F0 to mōra-chō no furumai ni tsuite [F0 and mora timing at the syntactic boundaries of Japanese language]. In O. B. Kenkyūkai (Ed.), *Bunpō to Onsei* [Grammar and Speech] (pp. 21–44). Kurosio Shuppan.

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

Bailey, L. M. (1990). *A feature-based analysis of Swedish pitch accent and intonation* [Unpublished doctoral dissertation]. University of Delaware.

Beckman, M. E., & Pierrehumbert, J. (1988). *Japanese tone structure*. MIT Press.

Broselow, E. (1988). Prosodic phonology and the acquisition of a second language. In *Linguistic Theory in Second Language Acquisition* (pp. 295-308). Springer Netherlands.

Cutler, A., & Isard, S. D. (1980). The production of prosody.

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.

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

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, *92*(1), 233-277.

Grosser, W., & Wieden, W. (1989). Principles of word-stress assignment in Austrian learners of English. *Englisch als Zweitsprache*, *89*, 102.

Grosser, W. (1993). Aspects of intonational L2 acquisition. In B. Kettemann & W. Wieden (Eds.), *Current Issues in European Second Language Acquisition Research* (pp. 81–94). Gunter Narr Verlag.

Hirose, K., Abe, Y., & Minematsu, N. (2006, May). Detection of fillers using prosodic features in spontaneous speech recognition of Japanese. In *Proceedings of Speech Prosody 2006* (pp. 187-190). International Speech Communication Association.

Hirose, Y., & Mazuka, R. (2024). Developmental changes in the interpretation of an ambiguous structure and an ambiguous prosodic cue in Japanese. In M. Koizumi (Ed.), *Interaction between linguistic and nonlinguistic factors* (pp. 255-274). De Gruyter Mouton.

Igarashi, Y. (2015). 13 Intonation. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 525–568). De Gruyter Mouton.

Ishihara, S., & Kubozono, H. (2015). 14 Syntax-phonology interface. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 569-618). De Gruyter Mouton.

Ishihara, S. (2016). Japanese downstep revisited. *Natural Language & Linguistic Theory*, *34*(4), 1389-1443.

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.

Itô, J., & Mester, A. (1993). Japanese phonology: Constraint domains and structure preservation. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 817-838). Blackwell.

Ito, J., & Mester, A. (2012). Recursive prosodic phrasing in Japanese. *Prosody matters: Essays in honor of Elisabeth Selkirk*, 280-303.

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

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

Jilka, M. (2000). *The contribution of intonation to the perception of foreign accent* [Doctoral dissertation]. University of Stuttgart.

Kawahara, S. (2015). The phonology of Japanese accent. In H. Kubozono (Ed.), *The handbook of Japanese phonetics and phonology* (pp. 445–492). De Gruyter Mouton.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?. *Cognitive Psychology*, *50*(2), 194-231.

Kubozono, H. (1989). Syntactic and rhythmic effects on downstep in Japanese. *Phonology*, *6*(1), 39-67.

Kubozono, H. (1993). *The organization of Japanese prosody*. Kurosio Publishers.

Long, M. H. (1990). The least a second language acquisition theory needs to explain. *TESOL Quarterly*, *24*(4), 649-666.

Major, R. C. (2001). *Foreign accent: The ontogeny and phylogeny of second language phonology*. Routledge.

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). Springer Berlin Heidelberg.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. *Trends in Linguistics Studies and Monographs*, *186*, 53.

Myrberg, S., & Riad, T. (2015). The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics*, *38*(2), 115-147.

Nagano-Madsen, Y. (2015). Acquisition process of L2 Japanese intonation by Swedish learners-Interlanguage or prosodic transfer?. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)* (Paper No. 0820, pp. 1-4). The University of Glasgow.

Nagano-Madsen, Y. (2015). Prosodic phrasing unique to the acquisition of L2 intonation-an analysis of L2 Japanese intonation by L1 Swedish learners. In *Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015)* (pp. 820-823). International Speech Communication Association.

Nagano-Madsen, Y. (2018). Perception and production of L2 prosody by Swedish learners–summary and application to the teaching using Japanese and Chinese data. *FONETIK 2018*, *45*, 608-612.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology: With a new foreword* (Vol. 28). Walter de Gruyter.

Nespor, M., & Vogel, I. (2007). *1986, Prosodic phonology*. Foris.

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

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*, 169-182.

Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: The role of the partner. In *Handbook of Discourse Processes* (pp. 128-169). Routledge.

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

Selkirk, E. (2009). On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure. *Journal of the Linguistic Society of Japan*, 136, 1–39.

Selkirk, E. (2011). The syntax‑phonology interface. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 435‑484). Blackwell.

Selkirk, E., & Tateishi, K. (1991). Syntax and downstep in Japanese. In I*nterdisciplinary Approaches to Language: Essays in honor of S.-Y. Kuroda* (pp. 519-543). Springer Netherlands.

Shport, I. (2008). Acquisition of rhythm: Evidence from spontaneous L2 speech. *The Journal of the Acoustical Society of America*, *123*(5_Supplement), 3328-3328.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103-130.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition*, 28(1), 1-30.

Uyeno, T. (1980). Comprehension of relative clause construction and pitch contours in Japanese. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics*, *14*, 225-236.

Van den Doel, R. (2006). How friendly are the natives? An evaluation of native-speaker judgements of foreign-accented British and American English [Doctoral dissertation]. Utrecht University.

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. *Prosodic Typology: The Phonology of Intonation and Phrasing*, 172-200.

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). Oxford University Press.

Venditti, J. J., & Yamashita, H. (1994). Prosodic information and processing of temporarily ambiguous constructions in Japanese. In *Proceedings of ICSLP 1994* (pp. 1147-1150). International Speech Communication Association.

Venditti, J. J., Jun, S. A., & Beckman, M. E. (2014). Prosodic cues to syntactic and other linguistic structures in Japanese, Korean, and English. In *Signal to Syntax* (pp. 287-311). Psychology Press.

Wieden, W. (1993). Aspects of acquisitional stages. In B. Kettemann & W. Wieden (Eds.), *Current Issues in European Second Language Acquisition Research* (pp. 125–135). Gunter Narr Verlag.

Wolff, S., Schlesewsky, M., Hirotani, M., & Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language*, *107*(2), 133-157.

Appendix A Informed consent for Swedish and Japanese participants.

I hereby give my consent to participate in a study in linguistics at Lund University in Sweden.

This study is about the analysis of Japanese speech produced by natives and Swedish learners of Japanese. For this purpose, I will be presented stimulus sentences in Japanese and read them aloud for recording. Also, I will fill in a questionnaire asking about my basic information or language learning experiences and language proficiency. The recording data and the answers to the questionnaire will be used only for research purposes. All of the data obtained from this study will be presented in a way that the participants cannot be identified through the data.

By my signature below, I certify that:
• I understand that my participation is voluntary and I may withdraw at any time without giving any explanation.
• I have received assurance that the data will be dealt with completely anonymously.
• I understand that the results of the study will be presented at scientific conferences and in journal articles without revealing the identity of the participants.
• I received sufficient information prior to the experiment. On completion, I will have the opportunity to ask questions.
• I will receive compensation after carrying through the whole experiment.

This consent form is signed in two copies, one for me and one for project documentation.

Place and date: Lund University; ____________________________
Signature: ________________________________

For more information, you may contact Chen <ch1886ch-s@student.lu.se>
Lund University Master's programme in language and linguistics (Japanese specialisation)

Appendix B Questionnaire for Swedish participants

Confidentiality Notice:
All information collected in this questionnaire will be kept strictly confidential and used solely for academic research purposes. Your personal data will be anonymized and will not be shared with third parties. By proceeding, you consent to participate in this study.

Name:

Age:

Gender: ☐ Male ☐ Female ☐ Other ☐ Prefer not to say

City in Sweden you are from:

Self-Reported Japanese Proficiency Level:
(Please check the option that best describes your current Japanese ability. Only check 1 option for each skill)

Listening

☐ I can understand only a few basic words or greetings.

☐ I can understand simple conversations when spoken slowly and clearly.

☐ I can follow daily conversations and short videos with familiar topics.

☐ I can understand TV shows, movies, and lectures on both concrete and abstract topics.

☐ I can fully understand fast, complex speech, even on unfamiliar topics.

Speaking

☐ I can say a few basic words or phrases.

☐ I can introduce myself and talk about familiar topics with help.

☐ I can handle short conversations on everyday topics.

☐ I can express my opinions clearly in discussions and presentations.

☐ I can speak fluently and spontaneously in academic/professional settings.

Reading

☐ I can read hiragana and katakana.

☐ I can understand simple sentences and basic vocabulary.

☐ I can read manga, short news articles, and daily materials.

☐ I can understand academic texts and newspapers.

☐ I can critically read complex and abstract texts (e.g., literature, research papers).

Writing

☐ I can write basic words and fixed expressions.

☐ I can write short sentences about daily life.

☐ I can write emails and short essays with some mistakes.

☐ I can write formal texts and reports in Japanese.

☐ I can write complex texts with high accuracy and proper style.

Language Learning Background

1. What native languages do you speak?

(Please check all that apply) Note: By "native languages", I mean the languages you were exposed to at home during early childhood, including those spoken by your parents or guardians.

☐ Swedish

☐ English

☐ Other: ___________________________

2. At what age did you start learning Japanese? If you had breaks during your study, please specify the periods.
Example: 2007–2010, 2012–present

3. How often do you use Japanese outside of your university studies?
(e.g., reading manga/novels, watching anime/movies, speaking with friends/partners, using at work, listening to music, etc.)

Activity	**Daily**	**Several times a week**	**Once a week**	**Rarel y**	**Neve r**
Reading manga or novels	☐	☐	☐	☐	☐
Watching anime, dramas, or movies	☐	☐	☐	☐	☐
Talking with Japanese friends/partners	☐	☐	☐	☐	☐
Using Japanese at work/school	☐	☐	☐	☐	☐

Listening to Japanese music/podcasts	☐	☐	☐	☐	☐
Writing in Japanese (e.g., SNS, diaries)	☐	☐	☐	☐	☐

4. How often do you use Japanese with native speakers?(e.g., only in class, with friends, partner, or family members)
(Please tick the frequency for each context)

Context	Daily	Several times a week	Once a week	Rarely	Never
In language classes	☐	☐	☐	☐	☐
With Japanese-speaking friends	☐	☐	☐	☐	☐
With Japanese-speaking partner/family member	☐	☐	☐	☐	☐
Through online conversations/language exchange	☐	☐	☐	☐	☐
In part-time job / volunteer / professional settings	☐	☐	☐	☐	☐

5. On average, how often do you study Japanese per week?

(Including classes, self-study, watching Japanese materials with the aim of learning, etc.)

☐ Less than 1 hour per week

☐ 1–3 hours per week

☐ 3–6 hours per week

☐ 6–10 hours per week

☐ More than 10 hours per week

6. How do you usually practice your Japanese pronunciation?

(You can select multiple options.)

☐ I practice alone by reading aloud.

☐ I mimic native speakers from anime, dramas, or podcasts.

☐ I practice with native speakers during conversation.

☐ I use pronunciation-focused apps or tools.

☐ I have received correction or feedback from teachers.

☐ Other (please specify): ____________________________

7. Have you ever received structured training focused specifically on Japanese pronunciation?

(e.g., pronunciation-focused university courses, private lessons dedicated to pronunciation, pronunciation workshops, etc.)

☐ Yes

☐ No

If yes, please briefly describe the type of training you received:

8. Do you have any Japanese language certificates? (e.g., JLPT, Kanken, etc.) ☐ Yes ☐ No
If yes, please provide the following information:

Test	When Taken	Score / Level

Thank you very much for your participation!

Appendix C Japanese participant information form

日本実験参加者情報収集フォーム

Confidentiality Statement / 機密保持声明

This information is collected solely for the purpose of this research study. All personal data will be kept strictly confidential and will only be used for research purposes. Your privacy is our priority.

この情報は本研究の目的のみのために収集されます。すべての個人データは厳重に機密として保持され、研究目的にのみ使用されます。あなたのプライバシーは研究の優先事項です。

Basic Information / 基本情報

- **Name / 氏名: ____**
- **Participant Number / 参加者番号: ________________________________**
- **Gender / 性別: □ Male / 男性 □ Female / 女性 □ Other / その他 □ Prefer not to say / 回答しない**
- **Age / 年齢: ____________ years old / 歳**

Language Background / 言語背景

- **Languages spoken / 話せる言語: [Multiple selections allowed / 複数選択可]**
- **□ Japanese / 日本語**
- **□ English / 英語**
- **□ Other / その他: ____________________________________**

Can you speak standard Tokyo Japanese? / 標準東京語を話せますか?

- **□ Yes / はい**
- **□ No / いいえ**
- **□ Partially / 一部できる**

Regional Background / 地域的背景

- **Place of origin / 出身地:**
- **Prefecture / 都道府県: ___________________________________**
- **Participant's Signature / 参加者の署名: ______________________________ Date / 日付: _________________**
- **Researcher's Signature / 研究者の署名: ______________________________ Date / 日付: _________________**