UNIVERSIDADE FEDERAL DE JUIZ DE FORA
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA
DOUTORADO EM LINGUÍSTICA

Thales Nascimento Buzan

What can prosodic transfer tell us about L2 prosody acquisition? Data from Brazilian Portuguese Learners of English

JUIZ DE FORA

2026

Thales Nascimento Buzan

What can prosodic transfer tell us about L2 prosody acquisition? Data from Brazilian Portuguese Learners of English

Tese apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do título de Doutor em Linguística.

Orientadora: Profa. Dra. Maria Cristina Lobo Name

Coorientador: Prof. Dr. Juan Manuel Sosa Hernández

JUIZ DE FORA

2026

FICHA CATALOGRÁFICA

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Buzan, Thales Nascimento.
What can prosodic transfer tell us about L2 prosody acquisition? : Data from Brazilian Portuguese Learners of English / Thales Nascimento Buzan. -- 2026.
159 p.

Orientadora: Maria Cristina Lobo Name
Coorientador: Juan Sosa
Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Letras. Programa de Pós-Graduação em Linguística, 2026.

1. Aquisição de L2. 2. Prosódia. 3. Transferência prosódica. 4. Entoação. 5. Percepção da fala. I. Name, Maria Cristina Lobo, orient. II. Sosa, Juan, coorient. III. Título.

FOLHA DE APROVAÇÃO

Thales Nascimento Buzan

What can prosodic transfer tell us about L2 prosody acquisition? Data from Brazilian Portuguese Learners of English

Dissertação apresentada ao Programa de Pós-Graduação em Linguística da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Linguística. Área de concentração: Linguística

Aprovada em 06 de fevereiro de 2026.

BANCA EXAMINADORA

Prof.ª Dr.ª Maria Cristina Lobo Name - Orientadora
Universidade Federal de Juiz de Fora

Prof. Dr. Juan Manuel Sosa Hernández - Co-orientador
Simon Fraser University

Prof.ª Dr.ª Laura Colantoni
University of Toronto

Prof.ª Dr.ª Fatima Hamlaoui
University of Toronto

Prof. Dr. Leônidas José Silva Júnior
Universidade Federal de Pernambuco

Prof.ª Dr.ª Aline Alves Fonseca
Universidade Federal de Juiz de Fora

Juiz de Fora, 07/01/2026.

Documento assinado eletronicamente por **Maria Cristina Lobo Name**, **Professor(a)**, em 06/02/2026, às 14:02, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Aline Alves Fonseca**, **Professor(a)**, em 06/02/2026, às 14:49, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Leônidas José da Silva Jr.**, **Usuário Externo**, em 09/02/2026, às 10:05, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fatima Hamlaoui**, **Usuário Externo**, em 09/02/2026, às 19:35, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Juan Manuel Sosa Hernandez**, **Usuário Externo**, em 11/02/2026, às 08:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Laura Marcela Colantoni**, **Usuário Externo**, em 11/02/2026, às 12:03, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2815051** e o código CRC **229D3D5A**.

Dedico este trabalho aos meus pais que sempre me deram suporte e condições para estudar.

ACKNOWLEDGEMENTS

I would like to start my acknowledgements with a little something about my trajectory up to this point. I entered the master's program at UFJF in 2020, right before the pandemic hit. I had one class in person, and then the whole world shut down. At the time, my research was still incipient and had to be adapted to online data collection and online courses. The research grew fast; I had good results and was fast-tracked to the PhD program in 2021. Doing a PhD humbles you: it makes you realise that no matter how much you study, there will still be a whole lot to learn. Four years later, I have finished the research that I once thought I could never do — and, most importantly, I did so with the help of people I could not have dreamed of meeting. So, my acknowledgements follow this simple but crucial stance: I could never have done it without people — without friends and family, without collaboration and cooperation.

First and foremost, I thank my parents for giving me every possible opportunity to learn and for supporting me along the way.

I thank my friends, most importantly Amber, Alice, Danilo, Daphna, Eric, Fernanda, Gustavo, Ilson, Isabella, Jason, Maja, Larissa, Lucas, Luísa, Max and Ryan (16–2 forever!). Without you, I could never have endured these six years with such ease and happiness.

I thank my advisors, Cristina Name and Juan Sosa. I have never worked so well with anyone as I have worked with Cristina. Her support and guidance, inside and outside the PhD, were crucial in shaping who I am today as a person and as a researcher. I am deeply grateful and honoured to have had the chance to spend all these years under her guidance, from my time as an IC student in 2018 until now. Juan sparked my interest in prosody in every way possible throughout my undergraduate and graduate studies. He has also been a great companion at national and international conferences — it has been a true pleasure and honour to have worked with him.

I thank all the professors I had the honour of meeting in Brazil, namely Aline Fonseca and Flaviane Svartman, for their feedback and help; Ana Paula, for her guidance and friendship throughout our work at our Journal of Linguistics (Revista Gatilho); Gustavo Estivalet, for teaching me how to program my first experiment in jsPsych during the pandemic and for first teaching me statistics; Luciana Lucente, for her valuable feedback and for sharing an excellent Praat script; Márcio Leitão, for his teaching; Mercedes Marcilese, for helping me recruit my first international participants, for her teaching, and for introducing me — back in 2018 — to experiment programming in Psyschope; Tiago Timponi, who offered inestimable support throughout these years; and Ubiratã Kichtofel, for recommending Laura Colantoni as my “sandwich” supervisor and for his excellent teaching on L2 acquisition.

I also thank all the professors I had the honour of meeting in Canada, namely Alexei Kochetov, for letting me use his phonetics lab; Fatima Hamlaoui, who supported me even though she was on sabbatical — I will never forget our first meeting and how much it changed my perception of my own work; Myrto Grigoroglou, for her help and for her excellent teaching on experimental methods; Philip Monahan, who managed to teach me statistics like no one ever could; Jessamyn Schertz, for her kindness and for her excellent teaching of speech perception; and Laura Colantoni, my “sandwich” supervisor and the person who supported me the most during my stay in Toronto and who continues to support me. I am deeply grateful and inspired by your kindness and knowledge.

I thank Janaína Soares and Stefanie Martin for their friendship and for the great partnership we developed in Toronto. I deeply miss our days together.

I thank the examining committee — Laura Colantoni, Fatima Hamlaoui, Leônidas Júnior, and Aline Fonseca — for their valuable feedback.

I thank my graduate program (PPG) for its support throughout these years, under the leadership of Tiago Timponi and with the help of the secretaries Isabel and Isabella.

I thank all the participants who took the time to take part in the many experiments conducted for this research — all of them completely voluntarily. Needless to say, without them none of this would have been possible.

This research was supported by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES, Brazil) [Grant no. 001] and by the *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* (FAPEMIG, Brazil). Without this support, I would never have been able to finish this PhD or to have the experience of being a visiting student at the University of Toronto — which changed my life.

Hopefully, I have shown that research is never done alone, and that I have been accompanied by the best of the best along the way. They believed in me and supported my ambitions, giving me the strength and confidence I needed to finish this work.

Finally, I thank you, the reader, and wish you a pleasant read.

RESUMO

O presente trabalho investiga a aquisição da prosódia em segunda língua (L2), com foco na transferência prosódica do português brasileiro (PB) para o inglês, a partir de dados de produção e percepção de sentenças declarativas de foco amplo e perguntas totais neutras. O estudo tem como objetivos analisar (i) a produção de interrogativas totais em inglês por falantes brasileiros aprendizes de inglês, identificando a ocorrência de influência prosódica da L1; (ii) a percepção de ouvintes nativos de inglês quanto a essas produções, a fim de verificar em que medida a influência prosódica afeta a interpretação da modalidade sentencial; e (iii) os fatores responsáveis por dificuldades perceptuais e por custos de processamento associados ao desencontro entre contorno e função. No PB, perguntas totais neutras são tipicamente realizadas com um contorno ascendente-descendente (L+H*L%) (Moraes, 2008; Castelo e Frota, 2016; Castelo et al., 2018), enquanto no inglês norte-americano o padrão não marcado é ascendente (LH*-H%) (Hedberg, Sosa e Görgülü, 2014). Essas interrogativas foram selecionadas por apresentarem contraste entoacional claro entre as duas línguas, o que as torna particularmente adequadas para investigar transferência prosódica. A motivação do estudo reside tanto na escassez de trabalhos que tratem a influência em seu nível prosódico quanto na ausência de modelos teóricos de aquisição de L2 que incorporem explicitamente a prosódia como componente estruturado da gramática e da competência comunicativa. A pesquisa é composta por um piloto e dois experimentos de produção e percepção. Os dados de produção, obtidos por tarefas de leitura contextualizada automonitorada, revelam que falantes brasileiros (MG, SP e RJ) frequentemente transferem para o inglês o contorno ascendente-descendente típico do PB em perguntas totais neutras. Os experimentos de percepção, realizados com ouvintes nativos de inglês (EUA, Canadá e Inglaterra) e estímulos com filtro low-pass de nativos do português brasileiro e do inglês, demonstram que contornos nucleares L+H*L% são sistematicamente interpretados como declarativos quando presentes em perguntas totais em inglês, resultando em queda significativa na acurácia de identificação de perguntas. Modelos estatísticos de efeitos mistos binomiais indicam que a baixa acurácia (≈ 40%) não está associada à língua materna do falante de forma global, mas a propriedades prosódicas específicas dos estímulos, sobretudo ao contorno nuclear ascendente-descendente ($z = -3.19$, $\beta = -3.07$, $p = 0.0014$). Os resultados motivam a hipótese de que a categorização prosódica depende de sistemas de memória fonológica/acústica que fazem a ponte entre o sinal acústico e a interpretação fonológica. Estudos neurocognitivos sustentam essa visão ao sugerirem que o processamento prosódico se apoia em mecanismos integrativos que articulam a codificação auditiva a rotinas interpretativas mais abstratas, reforçando que a aquisição prosódica em L2 implica construção gradual e reorganização de mapeamentos forma–função. Por fim, os achados têm implicações aplicadas para tecnologias de fala: sistemas de síntese e reconhecimento (TTS/ASR) voltados à naturalidade e confiabilidade comunicativa devem modelar inventários prosódicos sensíveis a língua e variedade. No conjunto, os resultados evidenciam que a transferência prosódica pode comprometer a interpretação da modalidade sentencial em L2 e apontam para a necessidade de integrar a prosódia às teorias de aquisição e ao ensino de línguas.

Palavras-chave: aquisição de L2; prosódia; transferência prosódica; entoação; percepção da fala.

ABSTRACT

This dissertation investigates the acquisition of prosody in a second language (L2), focusing on prosodic transfer from Brazilian Portuguese (BP) to English, based on production and perception data from broad-focus declarative sentences and neutral yes/no questions. The study aims to analyze (i) the production of English yes/no questions by Brazilian learners of English, identifying the occurrence of L1 prosodic influence; (ii) the perception of these productions by native English listeners, in order to assess the extent to which prosodic influence affects sentence-modality interpretation; and (iii) the factors responsible for perceptual difficulties and for processing costs associated with mismatches between contour and function. In BP, neutral yes/no questions are typically realized with a rising–falling contour (L+H*L%) (Moraes, 2008; Castelo and Frota, 2016; Castelo et al., 2018), whereas in North America the unmarked pattern is rising (LH*-H%) (Hedberg, Sosa and Görgülü, 2014). These interrogatives were therefore selected because they present a clear intonational contrast between the two languages, making them particularly suitable for investigating prosodic transfer. The motivation for this study lies both in the scarcity of work addressing influence at the prosodic level and in the absence of L2 acquisition models that explicitly incorporate prosody as a structured component of grammar and communicative competence. The research comprises one pilot study and two production and perception experiments. Production data, obtained through self-paced contextualized reading tasks, show that Brazilian speakers (from MG, SP, and RJ) frequently transfer the BP rising–falling contour to English neutral yes/no questions. The perception experiments, conducted with native English listeners (USA, Canada, and England) using low-pass filtered stimuli from native English and Brazilian speakers, demonstrate that L+H*L% nuclear contours are systematically interpreted as declaratives when occurring in English yes/no questions, resulting in a significant decrease in question-identification accuracy. Binomial mixed-effects statistical models indicate that the low accuracy rate (≈ 40%) is not globally associated with the speaker's native language, but rather with specific prosodic properties of the stimuli, especially the rising–falling nuclear contour ($z = -3.19$, $\beta = -3.07$, $p = 0.0014$). The results motivate the hypothesis that prosodic categorization depends on phonological and/or acoustic memory systems that mediate the link between the acoustic signal and phonological interpretation. Neurocognitive studies support this view by suggesting that prosodic processing relies on integrative mechanisms that bind auditory encoding to more abstract interpretive routines, reinforcing the claim that L2 prosody acquisition involves the gradual construction and reorganization of form–function mappings. Finally, the findings have applied implications for speech technologies: text-to-speech and automatic speech recognition systems (TTS/ASR) aimed at naturalness and communicative reliability should model prosodic inventories that are sensitive to language and variety. Taken together, the results show that prosodic transfer can compromise sentence-modality interpretation in L2 and point to the need to integrate prosody into theories of L2 acquisition and into language teaching.

Keywords: L2 acquisition; prosody; prosodic transfer; intonation; speech perception.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF GRAPHS

LIST OF ABBREVIATIONS

AE	American English
AM	Autosegmental-Metrical (intonation framework)
BE	British English
BP	Brazilian Portuguese
CI	Confidence Interval
EN	English (used in labels such as EN-L1)
F0	Fundamental frequency
GLMM	Generalized Linear Mixed(-Effects) Model
jsPsych	JavaScript library for running behavioral experiments in the browser
L1	First language
L2	Second language
lme4	R package for fitting (generalized) mixed-effects models
MG	Minas Gerais (Brazil)
NC	Nuclear contour
NCType	Nuclear contour type
NE	Native English (used to refer to native-English productions/stimuli)
OR	Odds Ratio
PA	Pitch accent
RF	Rise–fall (contour shape label)
RJ	Rio de Janeiro (Brazil)
RT	Reaction time
SD	Standard deviation
SE	Standard error
SP	São Paulo (Brazil)
ToBI	Tones and Break Indices (annotation system)

US	United States
USA	United States of America
CLI	Cross-Linguistic Influnce

INTRODUCTION

Prosody is a core component of speech through which speakers organize meaning beyond segments. Its properties form a grammatical system and encode linguistic, pragmatic and discourse-related distinctions (Pike, 1945; Bolinger, 1986; Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986; Beckman, 1996; Barbosa, 2019). This matters directly for second language (L2) acquisition because — unlike in L1 development, where prosodic sensitivity emerges early and scaffolds later learning — L2 learners must use prosody for successful interaction from their very first communicative attempts, when their rhythm, intonation and voice quality are immediately evaluated by interlocutors and can facilitate or hinder understanding. This dissertation treats prosody as a structured part of grammar, consisting of abstract phonological categories and systematic form–function mappings. We ground our stance in L2 speech research which shows that prosodic deviations can be especially consequential for perceived naturalness and processing effort, sometimes outweighing segmental inaccuracies in how listeners experience comprehensibility (Munro and Derwing, 1995; Derwing and Munro, 1997, 2005; Hahn, 2004; Kang, 2010).

Despite prosody's importance for communication, most approaches to Second Language Acquisition (SLA) have often prioritized morphosyntax and lexis, leaving phonetics and intonation comparatively under-theorized (Paiva, 2014; VanPatten and Williams, 2015; Colantoni et al., 2015). At the same time, evidence across language pairs suggests that L2 prosody is highly susceptible to cross-linguistic influence, with L1 categories shaping both what learners produce and how they perceive L2 contrasts (Mennen, 2015; Ladd, 1996; Kanaida and Lengeris, 2015; Gut and Pillai, 2014, 2015; Passarella-Reis, 2014; Passarella-Reis et al., 2016; Dias, 2015; among others). In this scenario, this dissertation argues that prosody must be incorporated into models of L2 acquisition not only in production, but also in perception, since prosodic variations may alter the interpretation of an utterance at the level of linguistic categorization.

A central observation motivating this work is that languages may draw on similar prosodic resources while organizing them into different phonological inventories and different mappings between intonational contours and meanings, as formalized in Autosegmental-Metrical approaches to intonation (Pierrehumbert, 1980; Beckman and Elam, 1993). In southeastern Brazilian Portuguese (BP) varieties, neutral yes/no questions are commonly realized with a rising–falling (L+H*L%) nuclear contour pattern, whereas in north American English, neutral yes/no questions are more robustly associated with rising (L*H-H%) nuclei (Moraes, 2008; Castelo and Frota, 2016; Castelo et al., 2018; Hedberg, Sosa and Görgülü, 2014). When BP learners speak English, the learning challenge goes beyond "raising pitch" in questions, they need to acquire a language-specific mapping between nuclear contour shape and interrogative meaning.

Our specific object of investigation is sentence modality identification (question vs. statement) under conditions in which listeners must rely primarily on prosodic structure. To target this, the dissertation adopts a production and perception paradigm based on self-paced recordings of contextualized sentences along with low-pass filtering stimuli in a forced-choice identification task, which reduces access to segmental detail and makes intonational cues comparatively more salient as a basis for categorization. Building on this paradigm, this dissertation evaluates how listeners respond to neutral yes/no questions and broad-focus statements produced by speakers from different L1 backgrounds, and crucially, whether perceptual success is driven by speaker L1 or by nuclear contour type. In doing so, the work addresses a broader theoretical question: when cross-linguistic prosodic influence is observed, what exactly is the locus of the perceptual difficulty — an overall "foreign accent" associated with a speaker group, or a specific mismatch between an unexpected contour and the listener's prosodic inventory?

Guided by this, the dissertation tests the following hypotheses. First, BP learners will show evidence of transfer in the production of English yes/no questions, with rise–fall-like nuclear contours appearing in contexts where English typically favors rises. Second, when segmental cues are reduced, English listeners will not perceive rise–fall contours as part of their prosodic inventory, showing increased misclassification of questions as statements relative to rising contours.

This dissertation is organized as follows. Chapter 1 introduces the research problem and motivates the focus on prosody and prosodic transfer. Chapter 2 reviews influential SLA frameworks and evaluates what they can — and cannot — tell us about prosody, motivating a prosody-specific theoretical approach. Chapter 3 presents the empirical studies in a methodological progression. Finally, we conclude synthesizing theoretical implications, discussing how prosodic information may be stored and accessed, and outlining pedagogical implications and future directions.

1. L1 and L2 Prosody Acquisition: How does it happen?

This chapter establishes a developmental and functional rationale for treating prosody, especially intonation, as central to language learning. Doing so is essential for the dissertation's broader goal of understanding what prosodic transfer can reveal about L2 prosody acquisition, particularly when learners and listeners must rely mainly on intonational structure to infer meaning.

We begin by clarifying what counts as prosody and why it must be treated as a structured component of the grammar, with phonological categories and systematic form–function mappings (Section 1.1). We then trace how prosodic sensitivity emerges before birth and how early exposure privileges rhythm and melodic contours over fine segmental detail (Section 1.2). Building on this, we review how infant-directed speech amplifies prosodic cues — through pitch range, timing, and phrasing — thereby making prosodic structure highly available to the learner during the earliest stages of development (Section 1.3).

The second half of the chapter shifts to L2 acquisition. Unlike infants, L2 learners typically enter a language through immediate communicative pressure, meaning that prosodic patterns are evaluated by interlocutors from the outset and can strongly shape intelligibility, comprehensibility, and social inference (Section 1.4–1.5). We then synthesize evidence that L2 prosody is particularly susceptible to cross-linguistic influence in both production and perception, and that transfer effects can persist even when other areas of proficiency are advanced (Section 1.6). This developmental arc — from early prosodic attunement to adult L2 transfer — sets up the theoretical problem that motivates the next chapter: if prosody is so central to acquisition and communication, how have dominant Second Language Acquisition (SLA) frameworks conceptualized (or failed to conceptualize) its learning?

1.1. What is prosody?

Prosody refers to the suprasegmental properties of speech that structure spoken language beyond individual consonants and vowels. Unlike segmental units, which are realized as discrete and sequential elements, prosodic phenomena extend over larger domains of speech, such as syllables, words, phrases, and entire utterances. They are primarily realized through continuous variation in fundamental frequency (F0), duration, and intensity, which influences voice quality. Crucially, these variations are not merely expressive or paralinguistic; they form an integral part of the grammatical system of natural languages and play a central role in the encoding of linguistic, pragmatic, and discourse-related information.

Early work in phonetics and phonology already recognized the importance of prosody for structuring speech. Pike (1945) emphasized the role of intonation in organizing utterances as coherent communicative units, while Bolinger (1986) argued forcefully that intonation is inherently meaningful rather than ornamental. These insights laid the groundwork for later developments in intonational phonology, particularly within the Autosegmental-Metrical (AM) framework, which established prosody — and intonation in particular — as a structured component of phonological grammar (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986).

Within this framework, prosody is understood as a hierarchically organized system governed by phonological representations and constraints. Beckman (1996) characterizes prosody as the domain responsible for the temporal and melodic organization of speech, encompassing prosodic phrasing, the distribution of prominence, and intonational patterns that signal syntactic structure, information structure, and discourse relations. This organization is commonly modeled through the prosodic hierarchy, which includes levels such as the syllable, foot, prosodic word, phonological phrase, and intonational phrase, each associated with specific phonetic and phonological properties.

Following Barbosa (2019), prosodic structure can be described in terms of a small set of core acoustic parameters and their perceptual correlates. From an acoustic perspective, the primary variables used in the analysis of prosody — intonation in particular — are fundamental frequency (F0), duration, and intensity. F0 corresponds to the rate of vocal fold vibration and constitutes the main physical correlate of melodic variation in speech; it is typically measured in Hertz or semitones and can be systematically manipulated by speakers to convey linguistic and pragmatic distinctions. Duration refers to the temporal extent of linguistic units, mostly measured at the level of syllables and stress groups, while intensity, expressed in decibels, reflects the physical energy of the acoustic signal. These acoustic domains give rise to perceptual or psychophysical correlates, such as perceived pitch, loudness, perceived duration, and voice quality. Importantly, these perceptual dimensions do not map linearly onto their acoustic counterparts: pitch, for instance, is closely related to F0 but is also modulated by intensity and duration, and loudness reflects a logarithmic relationship with acoustic intensity. As a result, prosodic perception depends not on isolated physical cues, but on how multiple acoustic dimensions interact and are categorized by the listener’s perceptual system. This distinction between acoustic parameters and their perceptual interpretation is crucial for understanding how intonational contours are processed, particularly in contexts where listeners are required to rely exclusively on prosodic information.

Intonation has been the focus of extensive theoretical development within the AM approach. In this model, intonational contours are decomposed into discrete phonological units such as pitch accents, phrase accents, and boundary tones, which combine to form nuclear contours (Pierrehumbert, 1980; Beckman and Ayers Elam, 1997). This decomposition allows for a principled distinction between phonological representation and phonetic implementation. While phonetic realizations involve continuous variation in F0 scaling and alignment, phonological categories are discrete and language-specific. This distinction is central for understanding both within-language variation and cross-linguistic differences in intonation.

Gussenhoven (2004) and Ladd (2008) further emphasize that intonation systems are best understood as inventories of contrastive categories, much like segmental phoneme inventories. Languages may share similar phonetic cues — such as rising or falling pitch movements — while differing substantially in how these cues are phonologically categorized and mapped onto meaning. As a result, perceptual sensitivity to pitch movement does not automatically entail correct linguistic interpretation. Rather, interpretation depends on access to the language-specific prosodic categories that constitute a speaker’s or listener’s phonological knowledge.

This notion of a prosodic or intonational inventory is particularly relevant for the present study. Prosodic categories are learned as part of the phonological system of a language and guide both production and perception. From this perspective, prosody is not simply a gradient acoustic phenomenon, but a representational system that mediates how continuous acoustic variation is discretized and interpreted. Gussenhoven (2004) proposes that intonation operates simultaneously at multiple levels: a biological level, reflecting general physiological and perceptual tendencies; a grammatical level, where intonational categories are encoded in the linguistic system; and a pragmatic level, where these categories are interpreted in context. Acquisition of prosody therefore requires not only auditory sensitivity to F0, but also the learning of language-specific mappings between form and function.

Research on Brazilian Portuguese (BP) provides clear evidence for the grammatical status of prosody and for the existence of language-specific intonational patterns. Barbosa (2006, 2019) argues that rhythm and intonation in BP emerge from the interaction between prosodic structure, articulatory dynamics, and discourse organization, resulting in systematic and stable patterns. Studies within the AM tradition have shown that BP exhibits characteristic nuclear contours that differ from those found in languages such as English, particularly in the realization of neutral yes/no questions (Frota et al., 2015; Castelo and Frota, 2016). These differences are not merely phonetic, but involve distinct phonological representations and form–meaning mappings.

Such cross-linguistic differences have important consequences for speech perception and second language acquisition. While listeners are generally highly sensitive to pitch movement, accurate interpretation of intonation depends on whether the encountered contour corresponds to a category in the listener's prosodic inventory. When a contour is absent from that inventory, or when its functional interpretation differs across languages, systematic perceptual misinterpretations may arise. This highlights the need to treat prosody as an integral part of linguistic competence, rather than as a peripheral or purely expressive aspect of speech.

In this dissertation, prosody is therefore conceived as a structured grammatical component, comprising abstract phonological categories that organize and give meaning to acoustic variation. The focus is placed specifically on intonation at the phrasal level, and more precisely on nuclear contours associated with the distinction between broad-focus declaratives and neutral yes/no questions. By examining how Brazilian Portuguese speakers produce English intonation and how native English listeners perceive these productions when segmental information is removed, this study aims to contribute to a deeper understanding of L2 prosodic acquisition, perceptual access, and cross-linguistic influence in second language speech.

1.2. Prosody Before Birth: Prenatal Exposure

A solid body of research indicates that sensitivity to prosody begins before a baby is even born. The human fetus is surrounded by the low-pass filtered sounds of the mother's voice and other environmental speech as they pass through the womb. While the womb muffles fine phonetic detail, the rhythmic and intonational contours of speech carry through. By about the third trimester, the fetal auditory system is sufficiently developed to perceive these prosodic patterns (Narayanan et al., 2022).

As a result, the fetus becomes familiar with the general melody of its native language(s) and even with specific recurrent patterns such as nursery rhymes or frequent words. Classic studies by DeCasper and Spence (1986) showed that newborns prefer listening to their mother's voice over another female voice and even prefer a story or rhyme that was recited by the mother during pregnancy over a new story — implying memory for the prosodic cadence heard in utero. Moreover, newborns just hours or days old have been found to discriminate their native language from an unfamiliar language based on prosody alone. In a landmark study, Mehler et al. (1988) demonstrated that 4-day-old French infants could distinguish French from Russian, presumably by detecting rhythmic differences (French is syllable-timed, Russian more stress-timed). Similarly, newborns of bilingual mothers show equal preference for both languages they heard in the womb (Byers-Heinlein et al., 2010), indicating parallel prenatal learning of two prosodic systems.

Mampe et al. (2009) analyzed the cries of French vs. German newborns. They found that French babies' cries tended to have a rising-falling melody (ending in a higher pitch), whereas German babies more often had a falling melody — mirroring the typical intonation of yes-no questions in French (which often rise) versus German (which often end in a low fall). These differences emerge before the infants have had any postnatal experience with language; they reflect prenatal exposure, as the infants are imitating the general pitch patterns they heard while in the womb (Mampe et al., 2009). In effect, the fetus has begun to "tune" its brain to the ambient language's prosody, and this is observable in neonatal behavior. Such findings powerfully demonstrate that prosody is the first linguistic information acquired: it forms the outer envelope of language that infants latch onto initially. We can thus say that prosody lays the foundation for subsequent language development, giving infants a head-start by the time they are born.

1.3. Infant-Directed Speech (IDS) and Early Prosodic Input

Once born, infants are immersed in a social environment where caregivers naturally scaffold communication. One common phenomenon across cultures is Infant-Directed Speech (IDS) (also known as "motherese" or "parentese") — the special register adults (and older children) use when speaking to babies. IDS is characterized by a suite of exaggerated prosodic features, and these appear to occur in virtually most language communities (Name and Sosa, 2020). Compared to typical adult-directed speech, IDS has higher pitch, wider pitch range with more extreme highs and lows, and greater pitch variability (sing-song intonation). IDS is also slower, with stretched-out vowels and longer pauses, and it often has a distinctive rhythmic quality (e.g. shorter phrases, clearer stress patterns) (Name and Sosa, 2020; Filippa et al., 2025). These features result in an acoustically vivid signal that captures infants' attention. In fact, by as early as 1 month of age, babies show a listening preference for IDS over adult-directed speech — they will turn their heads or suck more to hear the sing-song baby talk, indicating that the exaggerated prosody is highly engaging to them (Fernald, 1985). The functional role of IDS prosody is multifaceted: "prosody in infant-directed speech serves important functions for the infant's attention, regulation, and emotional expression" (Filippa et al., 2025). The warm, exaggerated tone helps modulate infant affect (soothing or exciting them), maintains their interest, and may also communicate linguistic cues in an especially clear way.

Beyond emotional bonding and attention, IDS likely plays a facilitative role in language acquisition. The "prosodic bootstrapping" hypothesis has been discussed over the decades by many scholars (Gleitman and Wanner, 1982; Morgan and Demuth, 1996; Christophe et al.,1997; Christophe et al., 2008). This hypothesis suggests that infants use prosodic patterns in speech as a cue to discover language structure. For example, the exaggerated pauses and intonation of IDS might help infants recognize phrase boundaries or sentence types (questions vs. statements). As early as 5-6 months, infants can already use intonation to distinguish a yes/no question intonation from a declarative intonation in their native language (Frota, Butler and Vigário, 2014). Moreover, IDS often highlights key words (e.g. by pronouncing them louder or on a higher pitch), which may help infants segment words and learn their meanings. There is evidence that the prosodic features of IDS correlate with child language outcomes: mothers who use more exaggerated pitch and stress in IDS tend to have toddlers with larger vocabularies (Ramírez-Esparza et al., 2014). In one study, the prosody of IDS (pitch range etc.) was found to significantly correlate with infants' later vocabulary growth (Han et al., 2024), suggesting that IDS prosody might provide cues that make words more "salient" or learnable for the baby.

At the same time, early prosodic sensitivity is not purely “acoustic” and appears to interact with language-specific experience. For instance, Czeke et al. (2019) report that German-learning infants at 5–6 and 8–9 months did not discriminate Portuguese rising vs. falling contours in IDS-style materials, and the authors explicitly suggest that the extent to which the native language uses morphosyntactic cues to mark questions may shape infants’ ability to use intonation for this function.

Name and Sosa (2020, 2022) have conducted detailed studies on the prosody of IDS in Brazilian Portuguese, shedding light on how caregivers adjust intonation when interacting with preverbal infants. Brazilian Portuguese (BP) is particularly interesting because yes-no questions in adult BP have a characteristic rising-falling intonation rather than a simple final rise. In their study, Name and Sosa (2020) examined the prosody of questions in BP IDS. They recorded natural interactions between Brazilian parents and infants (around 4-11 months old) and analyzed hundreds of caregiver questions. They found that a high proportion of caregivers’ utterances to infants were questions — about one third of all utterances were interrogatives, far higher than in typical adult conversation. Importantly, most of these questions (74%) were “marked” with IDS prosodic features. Caregivers routinely used an exaggerated intonation, higher pitch register, slower tempo, breathier voice quality, or other prosodic modifications when asking questions of their baby. For example, a mother might raise her pitch dramatically at the end of “Você quer papá?” (“Do you want food?”) when addressing her infant, even though in adult BP such yes-no questions would not have such a high rise in Southeast Brazil. Name and Sosa also observed that not all caregivers used the same prosodic cues — some consistently used a very high pitch, others lengthened their vowels etc. — but virtually all employed some form of prosodic exaggeration to mark their utterances as child-directed. Intriguingly, as infants got older, closer to one year, caregivers began posing slightly more genuine (information-seeking) questions and fewer purely rhetorical questions that are less prosodically marked, suggesting that parents tune their prosody and speech acts to the infant’s developing communicative abilities. Overall, this work reinforces that IDS is a prosodically rich input. Caregivers intuitively “intonate” their speech in special ways to engage and teach infants, providing the child with clear exemplars of the melody of the language.

Before turning to L2 prosody, it’s worth noting that prosody’s importance in L1 acquisition is increasingly recognized in educational and clinical contexts. For instance, atypical prosody is often an early sign in developmental disorders like autism or language delay. Encouraging prosodic interaction (through singing, rhythmic nursery rhymes, exaggerated expression) is commonly advised to parents to stimulate infants' language. All this aligns with what has been widely discussed by the prosodic bootstrapping hypothesis: prosody is a key component in how language is first acquired, starting from the womb and then guided by IDS that caregivers adopt with their babies.

1.4. Prosody in L2 Acquisition and Communication

While prosody forms the earliest foundation of linguistic knowledge in the first language, learners face a markedly different landscape when acquiring prosody in a second language. In L1, infants are exposed to prosodic structure long before they rely on speech for intentional communication: they first attune to rhythm, stress and intonation, and only gradually come to map these patterns onto words, syntax and discourse functions. In L2, by contrast, learners typically enter the language through communicative use — they start speaking early and are expected to "get their message across" from the outset. This means that pronunciation, and in particular prosody, becomes a problem much sooner in L2 than in L1: from the very first attempts at communication, learners' rhythm, stress and intonation are evaluated by interlocutors and can immediately facilitate or hinder understanding.

Acquiring a second language's sound system is challenging, and this challenge is not limited to consonants and vowels. L2 learners often struggle even more with prosodic aspects — learning where to place stress in words, how to use intonation appropriately, and how to adopt the rhythm or timing of the new language. Research suggests that prosody may trump segmental pronunciation in influencing listeners' perception and comprehension: an L2 speaker might mispronounce some vowels or consonants yet still be understood, but if their intonation or stress patterns are poorly aligned with the target language, listeners often perceive the speech as unnatural or confusing[1] (Munro and Derwing, 1995). In other words, prosodic deviations are often what make speech sound "foreign" or effortful to process, even when segmental accuracy is relatively high.

[1] The goal is not to erase one's accent (which is absurd), but to enhance intelligibility and enable learners to express themselves closer to the intended meaning.

Furthermore, prosody carries pragmatic and affective meaning that is culturally specific. An L2 learner may unknowingly speak in a tone that native listeners interpret as rude, overly assertive, excessively formal, or uncertain, simply because they have not yet internalized the subtle norms of L2 prosody. Such prosodic mismatch affects social perception: heavier accent or disfluent prosody is associated with negative judgments about competence, credibility or trustworthiness. For example, Dragojevic et al. (2017) show that speakers with stronger foreign accents elicit more negative evaluations because listeners report lower processing fluency — greater difficulty in comprehending their speech — which in turn mediates negative social attribution. Foucart et al. (2020) similarly demonstrate that part of the bias against foreign-accented speakers is driven by the extra processing effort listeners experience when decoding accented or prosodically deviant speech, which reduces perceived fluency and leads to harsher evaluations. In this sense, prosody is not only a linguistic issue but also a social one.

Unexpected prosody may therefore lead to misunderstandings, pragmatic misreadings, or negative listener judgments about the speaker's attitude, politeness and competence. Conversely, mastery of L2 prosody can greatly enhance naturalness and communicative effectiveness. Listeners tend to favor speakers whose rhythm and intonation "sound right," even if minor segmental errors remain (Anderson-Hsieh et al., 1992). In sum, prosody is central not only for intelligibility and comprehensibility (Munro and Derwing, 1995), but also for social credibility and overall communicative success in L2 interaction.

Despite this centrality, there is still comparatively little systematization in how L2 prosodic acquisition is treated in mainstream SLA theory. Most major models have either focused on segmental phonology or on morphosyntax and domain-general learning mechanisms, leaving prosody under-theorized or treated as a secondary "accent" issue. At the same time, empirical work has repeatedly shown that prosody is one of the domains most affected by cross-linguistic influence and one of the hardest to fully structure towards target-like patterns in adulthood.

In the next section, we review evidence for cross-linguistic prosodic influence in production and perception, highlighting how L1 prosodic structures constrain L2 processing and how this phenomenon can occur bidirectionally. This provides the groundwork for analysing how major SLA theories have accounted — or failed to account — for prosody, and for motivating the need for dedicated models of L2 prosodic acquisition.

1.5. Challenges in Learning L2 Prosody - Cross-Linguistic Prosodic Influence

One of the main challenges in acquiring second-language prosody is first language cross-linguistic influence (CLI). Experimental research in both production and perception has consistently shown that adult learners tend to carry over L1 prosodic patterns — such as pitch alignment, pitch range, stress placement, phrasing and tune choice — into their L2. At the same time, a growing body of work shows that sustained use of an L2 can reshape aspects of the speaker's native prosody.

In production, a first group of studies documents bidirectional influence between languages. Mennen's (2004) seminal study on Dutch–Greek bilinguals showed that prenuclear rises in both languages were realised with peak alignment patterns that did not fully match monolingual Dutch or monolingual Greek norms: instead, bilinguals produced intermediate alignment values in both languages, clear evidence of L2-induced drift in the L1 as well as L1 transfer into the L2. Similarly, de Leeuw, Mennen and Scobbie (2012) examined late English–German bilinguals and found systematic changes in the intonation of the native language (L1) after long-term immersion in the L2, arguing that L2 experience can lead to phonetic attrition of L1 prosody without loss of overall intelligibility. Chang (2012) went even further by showing that very early stages of L2 learning (American English adults after only one semester of Korean) already produced measurable changes in L1 English segmental and prosodic patterns, pointing to rapid, multifaceted cross-language coupling between L1 and L2 at the phonetic level. Together, these studies demonstrate that prosodic CLI in adulthood is not only L1→L2, but often bidirectional, with L2 experience feeding back into the native prosodic system.

A second large cluster of production studies focuses on L1→L2 CLI in specific learner and contact varieties. Kainada and Lengeris (2015) showed that Greek learners of English produce polar questions with intonational patterns and pitch ranges strongly influenced by Greek, including alignment, pitch span and overall pitch level (rather than adopting English-like contours). Gut and Pillai (2015) found comparable transfer effects in Malay speakers of English: in both information structure and question intonation, these learners rely heavily on L1-based prosodic strategies (e.g.

the distribution of pitch accents and boundary tones), which leads to L2 English patterns that are systematically distinct from native norms but internally coherent within the learner variety. In the BP–English pairing, Passarella-Reis (2014) and Passarella-Reis et al. (2016) show that Brazilian learners frequently produce English yes/no questions with nuclear contours that mirror BP interrogative patterns, and that these BP-like contours are still interpreted as questions by listeners, albeit sometimes with altered pragmatic nuances. More recent work by Buzan et al. (2022) further document intonational influence in the polar questions of Brazilian learners of English, finding robust transfer of BP rising–falling patterns into English and showing that such transfer affects both acoustic realisation and native listeners' processing. Similar contact-driven prosodic transfer has been reported for Romance and other languages — for instance, in Occitan–French (Sichel-Bazin et al., 2012), Spanish–Portuguese contact varieties (Kireva and Gabriel, 2015) and other prosodic contact situations described in *Prosody and Language in Contact* (Delais-Roussarie et al., 2015), where bilingual speakers develop hybrid or restructured intonational systems shaped by both prosodic grammars.

A third set of production studies looks at L2 prosody in typologically closer language pairs and at longer immersion trajectories. Dias (2015) investigated Brazilian learners of Colombian Spanish living in Bogotá and showed partial convergence of L2 Spanish declaratives and yes/no questions to local monolingual patterns, alongside clear traces of BP intonation in both L2 Spanish and L1 BP, thus revealing simultaneous L2 learning and L1 attrition in intonation. Similar findings of increased variability and hybrid contours in bilingual BP–Spanish intonation are reported in later quantitative work (e.g. Silva and Arantes, 2021), which show that bilinguals' L1 and L2 contours are both more variable than monolingual baselines and that learning/attrition levels depend on sentence modality. Extending this contact-within-Romance picture to a long-standing bilingual ecology, Machado's (2024) dissertation on Riverense Spanish (Uruguayan–Brazilian border) shows that bilingual speakers display prosodic systems that are not straightforwardly reducible to either "monolingual Spanish" or "monolingual Portuguese": their intonation exhibits a broader tonal inventory with patterns resembling Portuguese alongside innovative configurations, and the degree of convergence toward the national prestige variety (Montevideo Spanish) is strongly conditioned by generation (with younger speakers aligning more with Montevideo patterns while older speakers retain more distinctive contact patterns). These studies[2] reinforce the idea that even when L1 and L2 are prosodically related (two Romance languages), cross-linguistic influence can still be strong, and that adult learners may stabilise in an interlanguage prosodic system that is neither purely L1 nor purely L2.

Cross-linguistic influence is equally clear in perception studies. Grabe et al. (2003) showed that English, Spanish and Chinese listeners differ in how they categorise English intonational contours, with L1 background shaping sensitivity to tune contrasts and leading to different patterns of confusion between rises and falls. Schmidt and Post and colleagues (2015) demonstrated that listeners' experience with particular L1 contour types (e.g. Australian English high rises) influences how they perceive and categorise rising intonation patterns, again highlighting an L1 filter on prosodic perception. In the BP–English context, Passarella-Reis et al. (2016) and Buzan et al. (2022) showed that both Brazilian and native English listeners can correctly interpret BP-influenced contours on English yes/no questions as interrogatives, but sometimes with delayed reaction times or altered judgments of speaker intention, indicating that CLI in production has processing consequences for listeners. Zhang and Chen (2023) add another perspective from tone–intonation interactions, showing that Mandarin English as Foreign Language (EFL) learners have greater difficulty perceiving English prosodic focus in certain question contexts than in statements, and that their performance reflects L1-based preferences for falling patterns when processing prominence, reinforcing the idea that L1 prosodic categories guide how L2 prosody is heard.

Across these diverse language pairs, modalities and learner populations, the empirical picture appears to be consistent: L2 prosody acquisition in adults is susceptible to cross-linguistic influence. The perceptual categories and production habits established in the L1 constrain how learners process and realise L2 intonation, rhythm and stress, and extended L2 use can in turn reshape L1 prosody. Yet, as we will see in the next section, these robust patterns of cross-linguistic prosodic influence have rarely been fully integrated into general SLA theory. Most major frameworks concentrate on segmental phonology, morphosyntax or domain-general learning mechanisms, leaving prosody either under-specified or marginal. This gap

[2] A parallel line of Brazilian work on L2 prosody (Teixeira; Lima Jr, 2021; Teixeira, 2023; Silva Jr.; Barbosa, 2019; 2020.) adopts a multidimensional approach grounded in multiple acoustic correlates (e.g., rhythmic metrics and other prosodic–acoustic parameters).

underscores the need for models such as the L2 Intonation Learning theory (LILt) (Mennen, 2015), which explicitly place prosodic transfer and development at the centre of L2 speech acquisition.

2. Theoretical Perspectives on Second Language Acquisition

This chapter reviews a set of influential theoretical frameworks in Second Language Acquisition (SLA) in order to evaluate what they can — and cannot — tell us about the acquisition of prosody. The term *framework* is used here as an umbrella label to encompass theories, models, hypotheses, and approaches that have shaped SLA research over the past decades. These accounts differ in their assumptions about what is learned, how learning proceeds, and what kinds of evidence are relevant; collectively, however, they define much of the conceptual space within which SLA has traditionally been theorized (Paiva, 2014; VanPatten and Williams, 2015). They also provide a useful vantage point from which to identify a persistent gap: despite their explanatory scope in domains such as morphosyntax and the lexicon, prosody — particularly intonation — has often remained peripheral or implicit in SLA theory (Colantoni et al., 2015).

To make this discussion coherent and relevant to the present dissertation, the frameworks reviewed here are organized into two broad orientations: domain-general and domain-specific. By domain-general approaches, we refer to accounts that explain language learning primarily through cognitive mechanisms not unique to language, such as associative learning, statistical learning, memory, attention, processing constraints, and interactional feedback. In these approaches, language development is treated as an instance of general learning, and linguistic patterns are expected to emerge from learners' experience with input and use. By domain-specific approaches, we refer to accounts that posit language as being at least partly constrained by an innate, specialized architecture such that acquisition involves the development or reconfiguration of abstract linguistic representations under principled constraints. This distinction is not meant to suggest that all theories fall neatly into one category, but it helps clarify how different traditions conceptualize the nature of linguistic knowledge and the mechanisms that support its acquisition.

Importantly, many SLA frameworks are not self-contained theories of language; rather, they draw on broader theoretical traditions in linguistics, psychology, cognitive science, and the study of complex systems. For instance, behaviorist accounts are historically rooted in learning theory in psychology; generative approaches derive their explanatory goals and formal apparatus from theoretical linguistics; usage-based and connectionist perspectives see language as a byproduct of human experience;

interactionist accounts draw on communication and pragmatics; and dynamic-systems perspectives are grounded in complexity science. Recognizing these broader intellectual lineages is useful here because it helps explain why prosody has been treated unevenly across frameworks: when the object of explanation is defined primarily in terms of morphosyntactic representations or lexical learning, prosody tends to be treated as secondary, as performance, or as something that will "come along" with exposure and use.

The goal of this chapter, therefore, is not to provide an exhaustive history of SLA, but to assess to what extent canonical frameworks provide conceptual room for prosody and how their limitations motivate the need for more targeted models of L2 speech and, ultimately, L2 intonation learning. This critical overview sets up the theoretical rationale for the experimental chapters that follow and provides a basis for returning, in the Discussion, to the role of prosodic representations in both production and perception. We begin by reviewing domain-general frameworks of SLA before turning to domain-specific accounts. Together, these frameworks define most of the theoretical landscape of SLA.

2.1. Domain-General Frameworks

The frameworks reviewed here include behaviorist, usage-based, interactionist, and dynamic-systems approaches.

2.1.1. Structural-Behaviorism Theory on L2 learning

In the mid-20th century, Behaviorism was a dominant paradigm in psychology and influenced early views of language learning. Skinner (1957) famously described language learning as habit formation through stimulus-response reinforcement. In second language learning, this view was manifested in the Audiolingual Method[3] and the Contrastive Analysis Hypothesis (CAH) (Lado, 1957) of the 1950s–60s. The idea was that an L2 is learned by mimicry, memorization, and reinforcement of correct responses. Errors were seen as the result of interference[4] from L1 habits, and drill-

---

[3] Which was "the incorporation of the linguistic principles of the Aural-Oral approach with state-of-art psychological learning theory in the mid-1950s that led to the method that came to be known as Audiolingualism" (Richards and Rodgers, 2001, p. 53).

[4] We often see "interference" as signalling something bad, whereas "influence" as something neutral or good. In any case, some researchers seem to use them interchangeably. It is more common to see "influence" in the recent literature.

based practice (repetition of dialogues, pattern practice) was employed to instill new L2 habits. Pronunciation (including prosody) in audiolingual teaching was certainly not ignored – students were often asked to repeat sentences after a model, paying attention to intonation and rhythm. For example, an English class might drill the intonation of questions by having learners mimic the teacher's rising pitch: "Does he live here?" (rise on *live* and fall at end). Thus, in practice, behaviorist teaching methods recognized that pronunciation and prosody had to be mastered through repetition and corrective feedback.

However, as a theory, behaviorism did not offer much insight beyond "imitate the native speakers." Prosody would be just another behavior to learn via repetition. There were no concept of an internal phonological system or developmental stages — just habits to be formed. By the 1970s, behaviorist accounts of language acquisition had been largely discredited in favor of cognitive and innate approaches. Noam Chomsky's scathing 1959 review of Skinner, along with new evidence of creative language use by learners, led researchers to see language acquisition as more than rote habit formation. For SLA, this meant moving away from pure mimicry models. Error analysis showed learners make systematic errors not explainable solely by L1 habits; interlanguage[5] theory (Selinker, 1972) posited an evolving internal system.

In terms of prosody, behaviorism had no special say beyond predicting that differences between L1 and L2 prosody would cause difficulty (via the contrastive analysis hypothesis). The CAH would claim, for example, that an L2 stress pattern unlike the L1's would be hard to acquire and likely lead to errors — which is often true. But behaviorism gives no mechanism for how a learner moves from "incorrect" to "correct" prosody other than practice and correction. Therefore, it is just a matter of amount of practice for adult learners through intonation drills, but without cognitive awareness or strategy, many still fail to learn and internalize native-like prosody. Indeed, a purely behaviorist model struggles to explain why fossilized accents persist despite massive exposure and perhaps repeated negative feedback. Modern perspectives would point to factors like the learner's attentional focus, identity, or the need for explicit instruction — all which are beyond classic behaviorism.

[5] This is a concept which is adopted by many researchers in SLA. It is understood as the developmental stage of an L2, when the learner uses a linguistic system that is not their native language nor the language being learned (Paiva, 2014).

In summary, behaviorism contributed to the idea of L1 interference and the need for practice, which are relevant to prosody since L1 transfer is real, and practice does help. However, it lacked a richer explanation of developmental processes or constraints, and it treated prosody as just another set of responses to acquire, without acknowledging the cognitive representation of intonation or stress patterns in the learner's mind. As such, behaviorism left prosody largely untheorized except as a "habit" — a gap that later theories needed to fill with more nuanced accounts.

2.1.2. Connectionism and Usage-Based Learning

Connectionist and usage-based theories conceptualize language acquisition as the outcome of general cognitive learning processes such as pattern recognition, analogy, memory, and statistical learning. Rather than assuming an innate, language-specific capacity, these approaches view linguistic knowledge as emerging from learners' cumulative experience with input. As Ellis (2003) explains, connectionism belongs to a broader constructivist tradition that includes functionalist, emergentist, cognitive, and usage-based perspectives, all of which share the assumption that linguistic structure develops through learners' sensitivity to distributional regularities in the input. Under this view, grammatical knowledge is not a fixed system of rules, but a dynamic statistical ensemble that is continuously reshaped as new utterances are processed.

Connectionist models investigate the kinds of representations that can arise when simple associative learning mechanisms are exposed to rich linguistic input (Ellis, 2003). Learning is assumed to be parallel and non-linear, with multiple processes operating simultaneously, in contrast to earlier behaviorist accounts. Although these frameworks differ in their implementation, they share the assumption that both first and second language acquisition rely on domain-general learning mechanisms, even if the extent to which L1 and L2 learning are equivalent remains an open empirical question (Ellis, 2003; Ellis and Wulff, 2015).

Within SLA, connectionist and usage-based approaches have contributed primarily to the study of morphosyntax and lexical development. A substantial body of research has demonstrated how learners can extract regularities from input without explicit instruction, accounting for gradual development, variability, and probabilistic knowledge (Rumelhart and McClelland, 1986; Ellis and Schmidt, 1997; Kempe and

MacWhinney, 1998; Haskell, MacDonald and Seidenberg, 2003; Poerschi, 2004). These approaches have been particularly influential in challenging categorical, rule-based views of acquisition by showing how partial learning and gradient representations naturally emerge from experience.

At the same time, as noted by Ellis (2003) and Paiva (2014), much of this work is focused on L1 acquisition and on a limited set of linguistic domains. Nonetheless, the usage-based emphasis on frequency, salience, and experience offers a general framework within which L2 development can be understood as a process of gradual adaptation shaped by prior linguistic experience.

From a usage-based perspective, prosody can in principle be treated as another set of patterns in the input. An L2 learner exposed to sufficient spoken language could, theoretically, learn that certain pitch movements, rhythmic patterns, or stress distributions are associated with specific communicative functions. For instance, frequent exposure to rising pitch in English yes–no questions could lead learners to statistically associate that contour with interrogative meaning. In this sense, prosodic learning is not fundamentally different from learning syntactic or lexical patterns, as it relies on the same mechanisms of distributional learning and entrenchment.

Connectionist models are particularly well suited to capturing gradience and variability in L2 prosody. Unlike parameter-based accounts, they predict partial learning and context-dependent accuracy, which aligns with observed interlanguage patterns where learners produce target-like prosody in some contexts but not others. They can also accommodate the persistence of L1 influence, as entrenched L1 patterns may compete with emerging L2 patterns, making prosodic remapping gradual and effortful. Adult learners' reduced neural plasticity has been cited as a potential limitation (Sokolik, 1990), although proponents argue that learning mechanisms remain available albeit constrained by strong L1 biases.

Despite their explanatory appeal, usage-based and connectionist approaches offer no explicit account of prosodic representations or categories. Prosody is assumed to be learned implicitly through exposure and use, but the theories do not specify how prosodic patterns are structured in the learner's grammar, how form–function mappings are reorganized across languages, or why certain prosodic features resist change despite extensive input. As a result, while these approaches plausibly explain variability, gradual development, and L1 entrenchment in L2 prosody, they fall short of providing a principled account of prosodic systems. This limitation motivates the need for models that treat intonation as a structured and learnable component of linguistic competence, a gap addressed by prosody-specific approaches discussed in subsequent sections.

2.1.3. Complex Dynamic Systems Theory

Complex Dynamic Systems Theory (CDST), also referred to as Chaos/Complexity Science or Adaptive Complex Systems Theory, encompasses a set of domain-general approaches concerned with describing how patterns emerge and change over time in complex systems (Larsen-Freeman, 1997; de Bot, 2007). Rather than viewing language development as a linear accumulation of rules, CDST conceptualizes it as a nonlinear, emergent, open, and self-organizing process that unfolds across multiple timescales and levels, from individuals to speech communities (Larsen-Freeman and Cameron, 2008). In this view, language is inherently complex because it consists of multiple interdependent subsystems, including phonology, morphology, lexicon, syntax, semantics, and pragmatics.

Within SLA, second language development is understood as a complex adaptive system in which numerous interacting factors shape learning trajectories over time. These include both internal variables, such as age, aptitude, motivation, cognitive style, and learning strategies, and external variables, such as input, interaction, feedback, and properties of the source and target languages (Larsen-Freeman, 1997). Development is therefore expected to be variable and non-monotonic: periods of apparent stagnation may alternate with rapid change, and variability is treated not as noise, but as an inherent property of the system.

One of the main contributions of CDST to SLA is its emphasis on process rather than end state. Instead of focusing on whether learners ultimately converge on a target grammar, CDST highlights how linguistic systems evolve over time through the interaction of multiple subsystems. This perspective has been particularly influential in shifting attention toward longitudinal data and individual developmental trajectories, allowing researchers to capture the dynamic and often unpredictable nature of second language development (Larsen-Freeman, 2015).

CDST also provides a principled framework for understanding why learners differ so markedly in their developmental paths. By treating interlanguage as a dynamic system shaped by multiple interacting constraints, the theory accounts for inter- and intra-learner variability without reducing it to error or deficiency. This has been especially valuable in pronunciation research, where learners frequently show fluctuating patterns of accuracy and fluency that are difficult to explain within linear or stage-based models.

When applied to L2 prosody and pronunciation, a CDST perspective foregrounds the interaction between prosody and other linguistic subsystems. Prosodic development is seen as contingent on factors such as segmental accuracy, fluency, cognitive load, and communicative context. For instance, improvements in segmental pronunciation or grammatical automation may free up attentional resources that allow learners to focus more effectively on intonation and rhythm, while increased fluency may lead to greater stability in prosodic patterns.

Importantly, CDST treats variability in prosody as informative rather than problematic. Inconsistent realization of intonation patterns — sometimes target-like, sometimes non-target-like — is interpreted as the system exploring competing attractor states. Over time, repeated exposure and use may lead the system to self-organize around more stable, and potentially more target-like, prosodic configurations. This perspective aligns well with empirical observations that L2 pronunciation and prosody often develop unevenly, with sudden improvements following extended periods of apparent plateau.

Despite its descriptive strength, CDST does not offer a specific model of how prosodic categories are represented, reorganized, or accessed by learners. Prosody is not ignored, but neither is it singled out for theoretical treatment; it is subsumed within the broader dynamics of the linguistic system. As a result, CDST provides limited predictive power regarding which aspects of prosody are more learnable, which dimensions are more susceptible to change, or how cross-linguistic differences in prosodic inventories are negotiated. In sum, complexity theory offers a valuable lens for understanding variability and developmental trajectories in L2 prosody, but it does not provide a principled account of prosodic systems themselves. This limitation motivates the examination of additional approaches that address prosody and intonation more explicitly.

2.1.4. Interaction, Input and Output Approach in SLA

Michael Long's Interaction Hypothesis, now more commonly referred to as the Interaction Approach, builds on earlier work on input and output in SLA, including Krashen's Input Hypothesis (Krashen, 1982, 1985), Swain's Output Hypothesis (Swain, 1985, 1995, 2005), and Hatch's (1978) emphasis on the role of conversation in language development (Gass and Mackey, 2015; VanPatten and Williams, 2015). Although it is not proposed as a complete theory of SLA, the Interaction Approach has been highly influential and is often described as the dominant interactionist paradigm in the field (Byrnes, 2005). Its explanatory goal is to account for how interactional processes facilitate learning by linking exposure to language with cognitive mechanisms such as attention, working memory, and noticing (Gass and Mackey, 2015).

At the core of the Interaction Approach is the claim that input[6] becomes particularly effective when learners engage in negotiation of meaning during interaction[7]. When communication difficulties arise, interlocutors naturally modify their speech through clarification requests, confirmation checks, repetitions, and reformulations. These interactional adjustments provide learners with negative evidence and enhanced input, which can draw attention to problematic areas and support acquisition (Long, 1996). While the general link between interaction and learning is widely accepted in SLA, it is also acknowledged that interaction alone is unlikely to account for all aspects of second language development (Gass and Mackey, 2015).

One of the main contributions of the Interaction Approach is its explicit integration of environmental and cognitive factors. Rather than treating input as passively received, the theory emphasizes that learning is mediated by learners' selective attention and processing capacity, and that interaction provides conditions under which these resources are optimally engaged (Long, 1996). This has led to a substantial body of empirical work demonstrating that interactional feedback can facilitate development, particularly in vocabulary, morphology, and language-specific syntax.

---

[6] "The language that a learner is exposed to in a communicative context" (Gass and Mackey, 2015, p. 181).

[7] "The conversations that learners participate in" (Gass and Mackey, 2015, p. 183).

The Output Hypothesis complements this view by arguing that language production plays an active role in acquisition. According to Swain (1985, 1995), producing the L2 forces learners to notice gaps between what they want to say and what they can say, thereby priming attention to relevant input. Output can also contribute to automatization, as repeated use of linguistic forms in speaking or writing helps consolidate emerging knowledge. Together, interaction and output highlight the importance of learner engagement and communicative pressure in driving language development.

Despite its focus on spoken interaction, the Interaction Approach does not explicitly address how prosody or phonological representations are acquired. The theory is primarily concerned with lexicon, grammar, and morphosyntax, and prosody is not treated as a specific learning target. Nevertheless, interaction can provide indirect opportunities for prosodic learning. For example, when a learner's intonation leads to misunderstanding, interlocutors may request clarification or reformulate the utterance, implicitly modeling more target-like prosodic patterns. Similarly, difficulties in parsing unexpected intonation may prompt repetition or modification, increasing the salience of prosodic cues.

Interaction also creates contexts in which learners may notice prosodic features. According to the Noticing Hypothesis (Schmidt, 1990), learning requires conscious registration of form–meaning relationships, and interaction can make certain prosodic mismatches noticeable. Likewise, output can draw attention to prosodic gaps when learners realize that their intended meanings are not successfully conveyed. In this sense, interaction and output may contribute to prosodic development by increasing awareness and providing practice opportunities, particularly when prosody affects communicative success.

However, interactionist and output-based approaches offer limited insight into the nature of prosodic representations themselves. They do not predict which prosodic features are more likely to be noticed, which types of feedback are most effective for prosodic learning, or why many learners achieve communicative adequacy while retaining non-target-like prosody. Because interlocutors often accommodate accented speech without explicit correction, interactional pressure alone may be insufficient to drive substantial change in prosody. In sum, while interaction and output highlight the importance of engagement, feedback, and practice in SLA, they do not provide a principled account of how prosodic systems are acquired or reorganized in a second language.

2.2. Domain-Specific Frameworks

We now turn to domain-specific approaches, which treat acquisition as constrained by an innate architecture of linguistic knowledge and review accounts grounded in Krashen's Monitor Model and Universal Grammar. We conclude with a discussion of LILt, a prosody-specific model of L2 intonation acquisition.

2.2.1. Krashen's Monitor Model, input or comprehension hypothesis

In the late 1970s and 1980s, Stephen Krashen's highly influential Monitor Model (Krashen, 1977; 1978) is the first theory to be fully developed for SLA. Krashen proposed that language is primarily acquired through comprehensible input — language that is understandable but slightly above the learner's current knowledge level (*i+1*). Krashen's (1985) theory consists of five main hypotheses: (i) language is acquired subconsciously through exposure rather than conscious learning; (ii) grammatical structures are acquired in a predictable order; (iii) conscious knowledge only monitors output; (iv) meaningful, comprehensible input enables acquisition; and (v) an affective filter can facilitate or hinder input processing. True acquisition occurs when learners understand messages containing elements they are ready to acquire (Krashen, 1982).

Krashen's ideas shifted focus to the importance of understanding messages rather than drilling forms. In this paradigm, anything that improves comprehension of input can aid acquisition. However, Krashen did not explicitly discuss pronunciation or prosody in his main hypotheses. The Input Hypothesis or Comprehension Hypothesis (Krashen, 2004) in principle covers all aspects of language — a learner presumably acquires the sounds, syntax etc. by understanding input — but in practice Krashen's examples and those of his followers were centered on vocabulary and grammar structures. Prosody might be seen as part of the input that gets internalized implicitly. In fact, Krashen (1982, 2013) repeatedly emphasizes that input causes acquisition, not the output, and that speech is a result of this acquisition process, only helping indirectly just because it is part of conversations (it provides input). For instance, a learner who hears a lot of English input should in theory gradually acquire the "feel" for English intonation and stress, as these are part of the language signal being processed. Yet, we often see that even with abundant input, learners do not automatically pick up native-like prosody. Krashen's model provides no mechanism for noticing or correcting pronunciation features since it largely downplays output and feedback.

One could argue that if input is truly comprehensible, the learner is perceiving prosodic cues correctly — and this might help acquisition. For example, hearing expected prosody might help parse sentences (as prosodic bootstrapping suggests), thus aiding comprehension (which then aids acquisition of grammar). But Krashen never articulated this; it's an extrapolation at best. Moreover, his emphasis on comprehensible input did not account for cases where prosody itself needed to be learned even when input was understood. Many adult learners can understand an L2 well (especially in reading) but still speak with poor prosody — their input was comprehensible, yet their output doesn't match native prosody. Krashen's model would predict that as long as they got enough input, their output should eventually align, but that often doesn't happen, indicating some missing piece.

Pronunciation and prosody are areas of performance that often cause anxiety (learners may fear sounding foolish if they exaggerate intonation, for instance). One could speculate that a high affective filter might specifically impede picking up L2 prosody, as learners might not pay attention to the melody of speech when they are stressed. Alas again, Krashen did not delve into such specificities.

Overall, Krashen's framework treated pronunciation as something that would emerge naturally if one simply had enough comprehensible input. This aligns somewhat with the "immersion" philosophy that listening and imitating suffice over time. In reality, for many learners, prosody may require more targeted attention or feedback.

In sum, while the Monitor Model assumes prosody will emerge implicitly through comprehensible input, it offers no mechanism for how prosodic categories are reorganized or why prosody often fossilizes despite massive exposure.

2.2.2. Universal Grammar (UG) and L2 Phonology

Chomsky's generative linguistic theory seeks to characterize the linguistic competence of native speakers and to explain how such competence can be acquired despite the apparent poverty of the input. Deeply rooted in morphosyntax, its primary concern has never been learning as a process, nor second language acquisition specifically, as "its scope does not include a theory of processing, or a theory of learning" (Mitchell and Myles, 2004, p. 92). When applied to SLA, however, generative approaches aim to account for the nature and development of interlanguage competence, understood as a mental and unconscious grammatical system (White, 2015).

Within generativism, comprehension and production are assumed to arise from an abstract linguistic system composed of syntax, phonology, morphology, and semantics, constrained by Universal Grammar (UG). UG is posited as an innate language faculty consisting of universal principles and parameters, which delimit the hypothesis space available to learners. In SLA, the central question concerns the extent to which UG remains accessible in adulthood, with proposals ranging from full access, through partial access mediated by the L1, to no access after a critical period, in which case L2 acquisition would rely exclusively on domain-general cognitive mechanisms (White, 2003, 2015).

Generative approaches have made substantial contributions to SLA by arguing that L2 learners exhibit systematic grammatical knowledge that cannot be fully explained by input frequency alone. Experimental evidence showing sensitivity to abstract constraints supports at least partial access to UG, particularly in syntax. As a result, UG-based research has provided detailed accounts of how learners represent L2 grammars and how interlanguage systems are constrained during development (White, 2015).

Although most UG-oriented SLA research has focused on morphosyntax, phonology has also been incorporated into generative accounts, albeit more marginally. Importantly, evidence suggests that phonological representations in L2 learners are likewise systematic and constrained, rather than random or purely imitative, indicating that abstract grammatical principles may also shape phonological acquisition (Archibald, 1998; Goad and White, 2006).

Within the Principles and Parameters framework, prosody is located at the interface between syntax and phonology, specifically at Phonological Form (PF). Prosodic structure is governed by universal principles — such as stress assignment, prosodic hierarchy, and end-rule application — while cross-linguistic variation arises through parameterization. Languages may differ, for example, in how pitch accents encode focus, how boundary tones map onto clause types, or whether stress is quantity-sensitive (Chomsky, 1995; Ladd, 1996; Truckenbrodt, 1999).

From this perspective, cross-linguistic prosodic influence in L2 acquisition can be understood as parameter transfer. Learners initially project L1 prosodic settings onto the L2, resulting in systematic deviations that reflect an interim grammatical state rather than performance errors. For example, Brazilian Portuguese learners of English may transfer L1 timing or alignment properties, yielding non-target-like nuclear contours. Empirical work on L2 intonation has shown that re-parameterization in prosody is gradual and constrained (e.g., Mennen, 2004; Goad and White, 2006; Colantoni and Steele, 2008), suggesting that prosodic principles remain available even in adult learners.

The Minimalist Program further reinforces the relevance of prosody by foregrounding interface conditions. Under Minimalism, syntactic derivations must be legible at the articulatory–perceptual interface, making the syntax–phonology mapping central to grammatical well-formedness (Chomsky, 1993, 1995; Al-Horais, 2013). Linearization and prosodic phrasing are imposed at PF, and several minimalist proposals explicitly argue that syntactic structure is shaped by phonological requirements (Richards, 2010; Tokizaki, 2006). From this view, acquiring an L2 involves learning language-specific mapping rules between syntax and prosody, rather than prosodic patterns in isolation.

Despite these insights, Universal Grammar does not offer a comprehensive theory of L2 prosody acquisition, nor was it intended to do so. While generative approaches provide a principled way of characterizing prosodic representations and explaining systematic transfer effects, they offer limited explanations for learning mechanisms, developmental trajectories, or the persistence of non-target-like prosody. As Paiva (2014, p. 84) notes, “researchers that adopt the UG assumptions owe the reader a consensus regarding the role of UG in SLA and a theory that explains which are the universal principles or processes in the acquisition of any foreign language”[8]. Thus, although UG provides a valuable representational framework for understanding prosodic structure and cross-linguistic influence, it leaves open critical questions about how prosody is learned and reorganized in a second language —

[8] Original: “Os pesquisadores que adotam os pressupostos da GU ficam devendo a seus leitores um consenso em relação ao papel da GU na ASL e uma teoria que explique quais são os princípios ou processos universais na aquisição de qualquer língua estrangeira.”

questions that motivate the need for prosody-specific models addressed in the following section.

2.2.3. Mennen's L2 Intonation Learning Theory (LILt)

The theoretical models we have presented so far were focused on the learning of domain-general abilities or language-specific abilities but were never concerned directly with speech. When we look at L2 speech learning models[9], we find Flege's SLM (1995) and Best's PAM/PAM-L2 (Best, 1995, Best and Tyler, 2007) which "base their predictions of the relative difficulty or ease of production and perception of non-native speech on comparisons of L1 and the to-be-learned segments" (Mennen, 2015, p. 172). Recognizing the gap in a dedicated theory for L2 prosody, Ineke Mennen proposed the L2 Intonation Learning Theory (LILt) in 2015. Mennen's work is particularly significant because it directly tackles what previous theories largely ignored: a theoretical framework specifically for how second language intonation is acquired in terms of production. LILt is essentially a model that identifies key dimensions along which L1–L2 prosodic differences occur in production and hypotheses about how learners handle these differences.

Mennen (2015) derives generalizations and hypotheses based on an overview of empirical research of L2 intonation production. LILt uses the dimensions of cross-language and cross-varietal variation identified by Ladd (1996), separating the "phonological representation from its phonetic implementation […] and the AM approach" (Mennen, 2015, p. 173).

Therefore, Mennen (2015) points out that languages can differ in their intonation systems on (at least) four dimensions, and thus an L2 learner must adapt on them:

1. Systemic Dimension: This refers to the inventory of prosodic elements in a language and their distribution. For intonation, it means the set of tonal categories or tunes (e.g., what pitch accents and boundary tones exist) and how they combine. For example, English intonation has a category of "falling tone" vs "rising tone" vs "fall-rise", etc., and specific rules on where they occur.

---

[9] SLM and PAM-L2 are primarily concerned with the acquisition of segmental features, and as such, they will not be examined in detail here. Instead, it is pertinent to refer to Mennen (2015), who provides a comprehensive discussion on how LILt extrapolates and formulates hypotheses regarding prosody acquisition, informed by foundational assumptions of L2 speech learning models.

Brazilian Portuguese, by contrast, has a different inventory of nuclear tones, with some overlaps. An English speaker learning BP must learn the BP inventory of intonational patterns, as we shall argue in the Discussion chapter. The systemic dimension is basically phonological: what are the prosodic cues or contrastive patterns of L2 versus L1. LILt hypothesizes that if the L2 has a pattern that is not in the L1's inventory, learners may substitute an L1 pattern (transfer) or have difficulty perceiving the distinction. Evidence for deviations on this dimension come from "L2 learners who fail to produce certain accents that do not form part of the source language inventory" (Mennen, 2015, p. 176).

2. Realisational Dimension: This concerns the phonetic implementation of those prosodic elements. Even if two languages have a similar intonation category (say both have a "yes-no question rise"), they may realize it differently in pitch range or alignment. For instance, a "falling" tone might drop sharply in one language but more gradually in another; a rising accent might peak on the stressed syllable in one language but after it in another. Mennen (2015) gives the example of alignment differences — e.g. Greek vs English, where the timing of pitch peaks differs. The realisational dimension covers these fine details. L2 learners often transfer their native realizations, so a French speaker of English might use a too-narrow pitch range because French's intonation is generally more compressed relative to English's more exuberant pitch excursions; thus, their questions sound monotonic. LILt suggests that mastering L2 intonation requires adjusting these continuous parameters (range, peak alignment, slope, etc.). Mennen (2015) argues that most support for deviations in this dimension is found for intonation, but there are studies concerning alignment (timing) and scaling (height).
3. Semantic Dimension: This dimension concerns the mapping of form to meaning/function. Intonation is meaningful: a particular tune might signal a question, or contrast, or attitude. Languages differ in how they use intonation to signal functions and, for example, in English a rising intonation on a statement can signal uncertainty or that you are asking for confirmation (question-like). In some other languages, that same rising might just indicate continuation or might not be used at all in that context. Mennen's model posits learners must learn which intonational forms map to which pragmatic or grammatical functions in L2. Often learners transfer L1 mappings, which can

lead to pragmatic oddities. An example of this is explored in this study, in which BP yes/no question intonation is commonly marked by a rising-falling intonation in the Southeast whereas Canadian and US English use a rising intonation. LILt emphasizes looking at what the functions of intonation in L2 are and checking if learners acquire those. If a function is L2-specific (like English's use of stress for focus meaning), an L2 learner might not realize to use prosody for that function, unless they notice communicative failure. Deviances in this dimension "may occur in the failure to use intonation to signal certain functions in a language-appropriate way" (Mennen, 2015, p. 176).

4. Frequency Dimension: This refers to how frequently certain prosodic patterns are used in practice in the L2, a dimension not proposed by Ladd (1996). Two languages might have the same intonation patterns and meanings, but differ in preferences. For example, both Japanese and English can lower pitch to signal finality, but perhaps English does it more often. Or both Russian and English can stress words for emphasis, but maybe Russian speakers do it less often or only in certain contexts. A learner might acquire the capability to produce a pattern but not use it at native-like rates. Mennen (2015) includes this dimension to account for usage frequency: even if a learner can do something, do they do it as often as natives? If not, their prosody might still sound off (maybe too stilted or too exaggerated). Conversely, overuse of a pattern can also sound odd (e.g., using a dramatic intonation on every sentence). Mennen (2015) argues that most deviances come from intonational patterns that are more frequent in the learner's L1, where certain patterns occur more frequently than the ones used in their L2.

The value of this classification lies in its ability to facilitate the systematic identification of deviations from native norms, as well as to specify the dimensions along which these deviations most frequently occur. Moreover, it enables structured comparisons among L2 learners across varying proficiency levels, ages of arrival (AOA) and of acquisition, native language backgrounds, speaking styles, and other variables pertinent to the acquisition process.

Mennen's LILt also integrates general SLA concepts like transfer and developmental stages. It acknowledges that L1 transfer is a starting point — learners often impose L1 prosody onto L2 initially on all four dimensions (using L1 tunes, implementation, meaning, and frequency). Over time, with sufficient input and practice, they may adjust each dimension gradually. Also, some dimensions might be “easier” than others: for example, learners might more easily pick up the functional differences if communication breaks down (they learn that to signal politeness they must use a certain intonation, because not doing so caused trouble). However, realisational differences (fine phonetic detail) might persist because they rarely cause outright misunderstanding, only subtle accent. Mennen’s theory therefore could explain why some prosodic features fossilize — if they don’t impede communication, a learner might never notice or bother to change them (e.g. a slightly non-native pitch range).

Since it was first formulated as a research program for L2 intonation (Mennen, 2015), the LILt framework has been empirically tested in several studies with adult learners. Graham and Post (2018), for instance, showed that Japanese and Spanish learners of English differ from native speakers and from each other both in their choice of nuclear contours and in the phonetic realization of pitch accents, with L1 background and proficiency jointly shaping how far learners move along LILt’s systemic and realisational dimensions. Kelly (2022), in a longitudinal case study of Ole Gunnar Solskjaer’s English, found that segmental accuracy improved more rapidly and categorically than the realization of rise–fall pitch accents, suggesting that intonation follows a distinct and slower developmental trajectory and may even undergo phases of hypercorrection, as LILt would predict for suprasegmentals. Zahner-Ritter et al. (2022) examined Mandarin learners of German and showed that learners approximate “non-merger” pitch accent contrasts more successfully than “merger” contrasts, and that proficiency and exposure modulate performance, directly supporting LILt’s assumptions about the role of L1 experience and feature similarity in L2 pitch-accent learning. Building on a series of studies on Czech and German learners, Pešková (2023) reports that L2 Spanish and Italian intonation consists of clusters of L1-like, target-like and hybrid contour patterns that align systematically with LILt’s four dimensions, leading her to propose a Developmental L2 Intonation Hypothesis explicitly grounded in LILt. More recently, work on Uyghur learners of Mandarin and English (Li et al., 2024) has shown clear L1-driven generalisation of final rising boundary tones into both non-native systems and demonstrated that cross-linguistic influence operates simultaneously in the systemic, frequency and realisational dimensions highlighted by LILt.

In conclusion, by distinguishing four dimensions of cross-linguistic variation — systemic, realisational, semantic and frequency — LILt makes it possible to describe L2 intonation not as a vague “foreign accent”, but as a structured interlanguage system, with predictable loci of transfer and development (Mennen, 2015). The empirical studies that have followed (Graham and Post, 2018; Kelly, 2022; Zahner-Ritter et al., 2022; Pešková, 2023; Li et al., 2024) collectively suggest that learners do not simply “fail” to sound native-like, but show patterned deviations that map onto LILt’s taxonomy, and that evolve over time rather than disappearing uniformly. Crucially, the model accommodates both persistence and change: it explains why some L1-based realisational habits can fossilize, while other aspects gradually shift under communicative pressure and increasing proficiency. For the purposes of this dissertation, LILt provides a principled framework to interpret BP–English intonational data: it allows us to ask which dimensions are most affected for our learners, how these effects relate to L1–L2 differences, and why certain prosodic properties may resist change even when segmental accuracy and morphosyntax are already highly advanced. This framework will be revisited in the Discussion chapter to interpret the present production and perception findings in terms of systemic, realisational, semantic, and frequency dimensions.

2.3. Critical Perspective and Conclusion

The theories examined in this chapter were selected because they represent the major intellectual traditions that have shaped SLA research. Together, these frameworks form the canonical foundations of the field and provide the conceptual tools most used to explain language development. Importantly, reviewing these theories reveals a consistent pattern: although each offers valuable insights into grammar, lexicon, or cognitive mechanisms, none provides a principled account of L2 prosody. By analyzing these diverse perspectives, we show how prosody has been historically under-theorized in SLA and why the emergence of models such as Mennen’s LILt is both necessary and timely.

This historical lack of attention is likely due to a combination of factors: pronunciation was long considered a lower-level skill, perhaps less interesting theoretically than syntax; additionally, until recently, measuring prosody or phonological details was technically harder than analyzing grammar, so research gravitated to what was observable in transcripts. The result was that important phenomena of L2 prosody acquisition flew under the radar. Even when we look at the theories on the acquisition of phonology (e.g. Best's and Flege's models), there are no predictions about the acquisition of prosody either. Colantoni et al. (2015) explain that those models were originally conceptualized to account only for segmental phenomena and they speculate two reasons why: researchers may have had difficulties categorizing prosody as a feature that exclusively concerns phonetics and phonology; and the extensive research on segmentals enables data to make predictions compared to L2 prosody. They also point out that even if some researchers have attempted to extend existing models to the study of prosody, when we deal with prosodic features, we must keep in mind that discrete units are not being analysed, that is, those whose substitution may always trigger change in the meaning of word (phonemes). What is at stake here are "problems of non-discrete units that map directly onto meaning" (Colantoni et al., 2015, p. 337).

This consideration is especially important as these units can influence not only the lexical word through lexical stress, but also impact sentence syntax, thereby altering the overall meaning of the sentence. Thus, if a learner must master the prosody-syntax-meaning of the target language, we agree with Colantoni et al. (2015) in considering that extending existing theories may be insufficient and we may have to go beyond phonetics and phonology to study prosodic features with a study of L2 lexicon and syntax acquisition.

In this chapter, we have presented arguments that underscore how important prosody is for language acquisition and use. It is the first aspect of language that infants tune into (before birth), and it remains a key for successful communication in L2. If theories fail to incorporate prosody, we argue that they are incomplete. For instance, a theory of SLA success that ignores pronunciation might celebrate a learner as "highly proficient" based on grammar tests while that learner struggles to be understood in real conversations due to intonation issues.

Mennen's work is a welcome development, offering a prosody-centric theoretical framework. However, the emphasis remains mostly on production and intonation. Other prosodic elements, such as rhythm and stress — which often relate to intonation in language acquisition research — could be explored more thoroughly, along with how intonation is perceived.

In conclusion, the journey of acquiring prosody spans from the womb to adulthood. Going forward, a more integrated theory of SLA should explicitly incorporate phonological and prosodic acquisition alongside grammar and lexicon, treating pronunciation as equally essential to communicative competence.

In the following chapter, we discuss our research on how BP speakers produce English yes/no questions and statements and how they are perceived by native English speakers. We believe this study will offer valuable insights to the field as we work towards a deeper understanding of L2 prosody acquisition.

3. Differences between Brazilian Portuguese and North American English: how can we investigate them?

This chapter presents a research program composed of perception and production experiments that trace the development of empirical inquiry into BP–EN CLI. Across these studies[10] — Buzan et al. (2022), our Experiment 1; Buzan et al. (2026), our Experiment 1.2; and the more recent redesigned study, Experiment 2 — the same overarching question is examined: How do BP speakers produce English intonation, and how do native English listeners perceive these productions when segmental information is removed? The results of this investigation will serve as empirical support for the discussion on the acquisition on L2 prosody.

For production, the experiments used a self-paced reading task methodology and, for perception, we used low-pass filtered stimuli. The cumulative nature of these studies reflects a trajectory common in L2 prosody research: preliminary evidence from a small pilot (Experiment 1) motivates analysis refinement and expansion (Experiment 1.2), which in turn reveals methodological limitations that required a full redesign (Experiment 2). Together, they establish a coherent account of where CLI in this language pair arises, how it affects listener perception, and which nuclear contours are most perceptually disruptive.

First, we present the research program summary, then the major intonational distinctions on broad-focus statements and yes/no neutral questions in BP and Canadian English. We then review methodological frameworks for isolating prosodic information in perception (Section 3.4) and for eliciting controlled yet naturalistic productions (Section 3.5). Finally, Sections 3.6–3.8detail the three experiments that implement this research program.

3.1. The Research Program: Rationale and Experimental Roadmap

The three experiments presented in this chapter form a cumulative research program designed to isolate the specific conditions under which CLI from Brazilian Portuguese affects the perception of English yes/no questions. Although they differ in scope and methodological refinement, all three studies address the same central hypothesis: BP speakers may carry over their L1 rising–falling nuclear contour into

[10] This research was approved by the Ethics Committee of the Federal University of Juiz de Fora under CAAE: 50628521.3.0000.5147 and of University of Toronto under RIS Protocol Number: 45390.

English, and this contour — absent from the English intonational system — may hinder native English listeners' interpretation of interrogativity when segmental information is removed.

Experiment 1 (Buzan et al. 2022) served as the initial proof-of-concept. Conducted under pandemic constraints, it combined remote production data with a low-pass filtered perception task. Despite limitations in recording quality, it established two crucial findings: (i) BP learners do produce English questions with the characteristic BP rising–falling (RF) contour, and (ii) English listeners systematically misinterpret RF-ending questions as statements. This pilot study revealed the phenomenon and provided the empirical motivation for further investigation.

Experiment 1.2 (Buzan et al. 2026) expanded the participant pool, introduced detailed ToBI-based contour annotation, and improved analytical rigor through mixed-effects modeling. It demonstrated that apparent perceptual disadvantages for BP stimuli stemmed not from speaker L1, but from specific nuclear contour shapes. The study also showed that item-level acoustic variability — uncontrolled in the original recordings — could mask or exaggerate perceptual effects. These findings underscored the need for cleaner stimuli and more systematic control over production conditions.

Experiment 2, the redesigned study, addressed these issues by collecting high-quality recordings under controlled conditions and implementing stricter filtering and exclusion criteria. This version provided the most rigorous test of the hypothesis: by minimizing item-level variability and ensuring consistent prosodic forms, it isolated nuclear contour shape as the decisive predictor of perceptual outcomes. Rising–falling contours again produced strong misclassification effects, while rising contours produced by BP speakers patterned perceptually with those of native English speakers.

Together, these three experiments chart a methodological progression — from initial observation, to refinement, to controlled verification — that enables a principled assessment of where CLI arises and how it affects listener perception.

3.2. Autosegmental-Metrical Theory and ToBI: Framework and Limitations

The present study adopts the Autosegmental-Metrical (AM) model of intonation as its theoretical framework for representing prosodic structure. Originally developed in Pierrehumbert's (1980) seminal work and further elaborated in Beckman and Pierrehumbert (1986) and Ladd (2008), the AM model treats intonation as a structured component of phonological grammar composed of discrete tonal units. Within this framework, intonational contours are analyzed as sequences of high (H) and low (L) tonal primitives that combine into pitch accents, phrase accents, and boundary tones. These tonal units are represented on an autonomous tonal tier and associate with metrically prominent syllables and prosodic boundaries. Crucially, AM distinguishes between phonological representation and phonetic implementation: while surface F0 contours are continuous and gradient, they are interpreted as realizations of discrete tonal categories. This distinction is particularly relevant for second language research, where the central question often concerns whether learners have acquired the underlying prosodic categories of the target language or merely approximate their phonetic realization.

The AM framework is especially suited for cross-linguistic comparison because it provides a common representational vocabulary for describing intonational systems across languages. It allows researchers to speak in principled terms about tonal inventories, distributional constraints, and form–meaning mappings. In the context of the present dissertation, AM provides the theoretical basis for arguing that Brazilian Portuguese and North American English differ not only in phonetic pitch movement but in the structure of their nuclear contour inventories and in how these contours map onto sentence modality.

Despite its wide adoption and explanatory power, the AM framework is not without limitations. One of the most persistent critiques concerns the tension between categorical representation and phonetic gradience. Although AM posits discrete tonal categories, empirical evidence shows that intonation exhibits substantial gradient variation in alignment, scaling, and pitch range. Ladd (2008) acknowledges that the boundary between categorical and gradient phenomena is not always straightforward, and Arvaniti (2011, 2016) has argued that some aspects of intonation may resist strict discretization. Continuous variation in pitch height or alignment can carry systematic pragmatic meaning, raising questions about whether all relevant distinctions are adequately captured by a limited inventory of H and L tones. In this sense, AM may impose categorical structure onto phenomena that are, at least in part, gradient in nature. The tonal labels employed in AM analyses should therefore be understood as analytical abstractions rather than direct acoustic realities.

Another theoretical challenge concerns the determination of phonological contrast. Establishing whether two contours constitute distinct categories or phonetic variants often requires functional and distributional evidence that is not always straightforward to obtain. Prieto (2014) and Ladd (2008) discuss the difficulty of defining clear criteria for phonological contrast in intonation, especially in languages where contour differences correlate with subtle pragmatic distinctions rather than categorical grammatical contrasts. This issue becomes even more complex in L2 research, where learner productions may fall between L1 and L2 norms. It is not always self-evident whether a learner's contour represents a mis-specified phonological category, a phonetic implementation difference, or an emerging interlanguage system. These considerations underscore that AM provides a representational framework but does not eliminate the interpretive dimension of prosodic analysis.

To operationalize AM assumptions in empirical data, the present study employs ToBI (Tones and Break Indices) conventions. ToBI was developed as a standardized annotation system for encoding tonal categories and prosodic boundaries in speech corpora. It provides a practical method for labeling pitch accents, phrase accents, boundary tones, and degrees of prosodic juncture. It is important to emphasize that ToBI is not itself a theory of intonation but rather a transcription system grounded in AM principles. Its strength lies in enabling systematic annotation, cross-study comparability, and statistical modeling of tonal categories. In the present research, ToBI-based annotation allows nuclear contours to be classified in a consistent manner and incorporated as predictors in mixed-effects models.

However, ToBI annotation also presents methodological limitations. Inter-annotator reliability studies have shown that agreement is generally high for broad contour distinctions such as rising versus falling patterns but decreases for more fine-grained tonal labels. Studies such as Syrdal and McGory (2000) and Yoon et al. (2004) demonstrate that even trained annotators may disagree on alignment distinctions or complex bitonal configurations. This variability reflects both acoustic ambiguity and theoretical interpretation. Furthermore, ToBI systems are language-specific and sometimes differ in their label inventories, which can complicate cross-linguistic comparison. The proliferation of language-specific ToBI variants has raised concerns about whether differences in labeling reflect genuine phonological contrasts or analytical conventions. Additionally, ToBI primarily encodes categorical tonal events and does not systematically capture gradient phonetic parameters such as pitch range scaling or subtle alignment differences, which may nonetheless be perceptually meaningful.

These limitations are acknowledged in the present dissertation. The use of AM and ToBI does not imply that intonation is purely discrete, nor that tonal labels exhaust the richness of prosodic phenomena. Rather, these frameworks are adopted as analytic tools that allow systematic comparison of contour types across languages and speakers. The focus of the present study is on nuclear contour shape and its perceptual consequences under controlled conditions, specifically low-pass filtering, which isolates pitch movement from lexical content. Given this objective, the categorical distinction between rising, falling, and rise–fall contours provides a theoretically motivated and empirically tractable basis for analysis. At the same time, the interpretation of results remains sensitive to the possibility that gradient phonetic detail may also contribute to perception.

In sum, Autosegmental-Metrical theory and ToBI annotation provide a coherent and widely accepted framework for representing intonational structure, enabling principled cross-linguistic comparison and quantitative modeling. Their limitations, particularly concerning gradience, contrast determination, and annotation variability, are recognized and taken into account in the interpretation of findings. The present study therefore employs these tools with theoretical caution, acknowledging that they offer structured representations of prosodic categories rather than exhaustive models of intonational reality.

3.3. Broad-Focus Declarative Intonation in North American English and Brazilian Portuguese

Broad-focus declarative sentences in North American English typically have a falling intonation pattern. In Autosegmental-Metrical (AM) terms this is often transcribed with a high nuclear pitch accent followed by low phrase and boundary tones (e.g. H*L-L%). In other words, the stressed syllable of the sentence’s final prominent word is assigned a high tone, and the pitch then falls to a low level by the end of the utterance (Beckman and Elam, 1993; Pierrehumbert, 1980; Bartels, 1999). Acoustically, this corresponds to a high F0 peak on the nuclear syllable followed by a decline in pitch throughout any post-tonic material and the final pitch usually reaches the speaker's low register at the boundary (L%).

Key acoustic characteristics of these declarative contours include the timing of the F0 peak (peak alignment) and the pitch range of the fall. In North American English (including Canadian), the nuclear H* often has an earlier peak alignment — the F0 peak typically occurs near the middle or earlier part of the stressed vowel (or just after its onset) in broad-focus statements (Patience et al., 2018). By contrast, question intonation tends to show later peak alignment. A recent study of Western Canadian English showed that in statements the F0 peak aligns around the end of the tonic syllable (on average within a few milliseconds of the vowel's end) (Kim and Arnhold, 2024). The overall pitch range in a neutral declarative is usually mid-to-high at the nucleus, downstepping from any preceding accents, and then drops to low. This can give Canadian declaratives a perceptibly "flat" or level ending as the voice trails off. For example, an utterance like "Marge has a new computer." would be produced with a high pitch on "-puter" followed by a steep fall to low at the very end (Patience et al., 2018). Such contours correspond to a ToBI transcription of H*L-L%.

Kim and Arnhold (2024) explored the way focus is expressed in terms of acoustic correlates and/or pitch accents in Canadian English would differ from Mainstream American English (MAE). Despite minor regional prosodic differences, Standard Canadian English intonation closely resembles General American patterns for statements. That is, an unmarked broad-focus statement in Toronto, Vancouver, or Winnipeg will typically use a falling nuclear tone. However, one phonological subtlety noted in recent work is that the inventory of pitch accents in Canadian English may differ slightly. The authors show that Western Canadian speakers displayed a strong preference for H* accents in broad focus, using very few rising pitch accents like L+H* in that context. In fact, the contrast between H* and L+H* (a bitonal "low+high" accent often associated with emphatic or contrastive focus in American English) may be neutralized in Canadian speech (Kim and Arnhold, 2024). This suggests that Canadian English intonation might have a simplified pitch accent system (at least in some regions), relying more on duration and phrasing to mark focus. Overall, though, a neutral declarative intonation in Canada is realized as a high accent followed by a low tail, i.e. a falling contour.

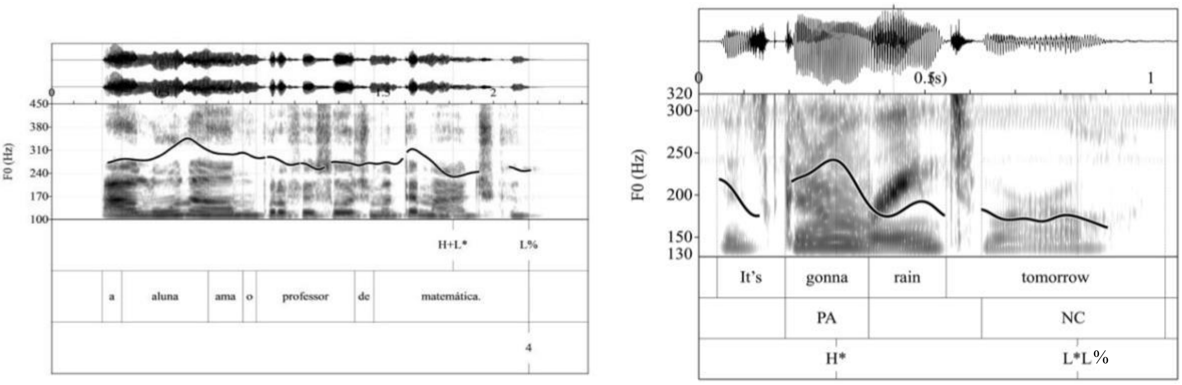
It should be noted that Canadian speakers, like those in many other English varieties, can also produce declaratives with non-falling intonation for pragmatic reasons. One well-known feature is the High Rising Terminal (HRT) or "uptalk," where a statement ends in a rising pitch rather than a fall. In ToBI notation this might be transcribed as L*H-H% (or similar), effectively applying a yes/no question-like tune to a declarative sentence. This intonation does not signal that the sentence is a grammatical question, but it adds an interpersonal or discourse meaning. In Canadian English, uptalk has been observed as early as the 1990s and is common especially in casual speech and the rising declarative intonation typically starts around the final accented syllable with a low pitch and then glides up to a high boundary tone (Sando, 2009). For example, a speaker might say *"We left around five o'clock."* with a rise at the end, conveying that they are soliciting confirmation or keeping the listener engaged, rather than simply stating a fact.

Broad-focus declarative statements in Southeast Brazil (Rio de Janeiro, São Paulo, Minas Gerais and Espírito Santo) are characterized by a falling intonation contour. The nuclear pitch accent is typically a high-to-low falling tone (transcribed as H+L* or H*), followed by a low boundary tone L%. In ToBI notation this can appear as H* L-L% — a high accented syllable with a low phrase and boundary tone (Frota et al., 2015; Castelo and Frota, 2016). Acoustically, the pitch peak usually occurs at or just before the stressed syllable, then falls through the remainder of the word to a low level at the end of the intonational phrase. This gives statements a final low pitch and the final stressed syllable (or any post-stressed syllable) is realized at a low F0, signalling closure.

Figure 1, we bring examples of this type of sentence in both BP, taken from Castelo and Frota (2016) of a speaker from the state Minas Gerais, and American English[11], taken from a female American participant of our study, side by side.

[11] Although some of the examples cited in this chapter come from broader North American English or US English data, our main empirical focus is on Canadian English, which patterns vary similarly to General American in terms of broad-focus statements and yes/no questions.

Figure 1: Broad focus declaratives in BP and EN



Source: Left: BP example taken from Castelo and Frota (2016, p.102). Right: stimulus recorded by a native Canadian English participant of this study

3.4. Neutral Yes/No Question Intonation in North American English and Brazilian Portuguese

Yes–no questions in Canadian English generally exhibit a rising intonation pattern, much like other varieties of English. The prototypical yes/no question tune in AM/ToBI terms is often a low-rise contour, transcribed as H-H% — a high phrasal tone followed by a high boundary tone — or L*H-H% — here the nuclear pitch accent is low (L*), and the phrase accent and boundary tone are high, producing a final rise (Hedberg, Sosa and Görgülü, 2014). The effect is that the pitch stays low or mid throughout the questioned proposition until the last stressed syllable, then rises sharply on or after that syllable into a high boundary. Empirical corpus studies of North American English confirm that this L H*-H% "low rise" is the unmarked yes/no question contour, occurring far more frequently than any other pattern (Hedberg, Sosa and Görgülü, 2014). For instance, a yes/no question like "Did you <u>eat</u>?" would typically have a low flat pitch on "eat" followed by a rise on the final boundary (Hedberg, Sosa and Görgülü, 2014). Canadian speakers use this same default tune for genuine information-seeking questions.

Acoustically, the yes/no question contour is characterized by a final F0 rise that often reaches a higher pitch level than the speaker's neutral baseline. The magnitude of the rise can vary, but questions often involve an expanded pitch range at the end. Measurements have shown that questions yield a larger pitch excursion than statements: e.g. the F0 change from the low point to the high peak at the end is significantly greater in yes/no questions than in declaratives (Patience et al., 2018). In production, Canadian English questions may start at a slightly higher overall pitch level as well, sometimes described as a higher “register” or elevated baseline for questions (though this effect is subtle and speaker-dependent).

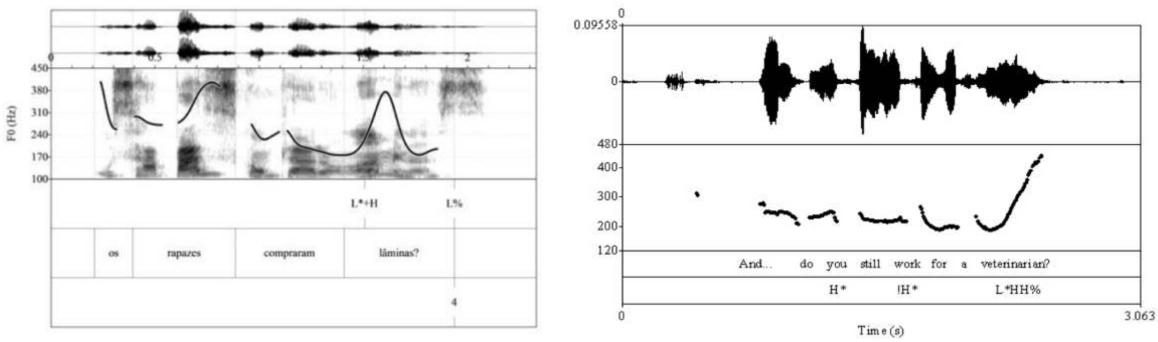
Importantly, the timing of the F0 peak is later in questions. In statements, nuclei peak early, and in yes/no questions we often see a delayed peak accent (sometimes labeled as L*+H or L+H*). The F0 peak occurs after the accented vowel, often in the following syllable or just at the phrase edge (Patience et al., 2018). For example, in *“Are you coming?”* the word “coming” might carry a low accent on [com-], and the actual F0 apex might not occur until into the [-ing] or even on a following hesitation. Patience et al. (2018) show that in Canadian data, F0 peaks in yes/no questions were found on average over 100ms after the end of the nuclear vowel, whereas in statements the peak was roughly synchronized with the end of the vowel. This alignment difference corresponds to a phonological distinction: questions tend to have a L+H pitch accent (rising pitch accent), versus the H* accent of statements. The resulting contour is a low plateau on the nuclear syllable followed by a rise, which listeners perceive as the questioning intonation.

In ToBI notation, several variants of rising question tunes have been noted in Canadian English. Aside from the common L*H-H% (low rise), speakers may sometimes use a low plateau followed by a mid or high rise (transcribed as L*L-H% in some analyses) or a high starting rise (H*L-H% or H*H-H%) for certain pragmatic effects (Shokeir, 2008). The phonological labels can differ, but essentially all indicate a rising boundary tone (H%) that makes the utterance sound like a question. The choice of a higher vs. lower rise may correspond to subtle pragmatic nuances (see below). Canadian English does not routinely use a final falling tone for yes/no questions — a high boundary is the norm. In fact, a yes/no question pronounced with a falling contour (H*L-L%) is very marked and typically conveys a special meaning (e.g. disbelief or challenge) rather than a neutral question (Hedberg, Sosa and Görgülü, 2014). Hedberg, Sosa and Görgülü (2014) found that falling yes/no questions comprised only a small fraction of their North American corpus, and when used, these falling questions were “non-genuine” — not sincere requests for unknown information. In Canadian English, we can observe this in echo questions or incredulous repetitions. For example, if speaker A says, “I aced the exam,” Speaker B might ask with a falling tone, “You aced the exam?”, to express surprise or scepticism. The falling pitch (H*L-L%) signals that B isn’t genuinely asking for confirmation but rather expressing incredulity. By contrast, a rising yes/no question like “You aced the exam?...” (with final rise) would be a genuine inquiry confirming the information.

Neutral Y/N questions in Brazilian Portuguese (BP) typically end with a rising-falling contour in the Southeast Brazil (also known in the literature as the Central varieties) (Moraes, 2008; Castelo and Frota, 2016; Castelo et al., 2018). In AM terms this is often transcribed by some authors (Moraes, 2008; Moraes and Colamarco, 2007) with a rising bitonal pitch accent within the stressed syllable and then a simple boundary fall (e.g. L+H*L%) and by other authors (Frota, Cruz et al., 2015) with the low tone as the nucleus of the rising pitch accent (e.g. L*+HL%). Cruz et al. (2022) report that more recently there has been an uncompromising annotation (L+H) for cases under discussion (Frota and Moraes, 2016). We have adopted the high tone as the accented one as we believe it best fitted the data we gathered in our production collection.

In Figure 2, side by side, we see an example taken from Castelo and Frota (2016) of a neutral Y/N question produced by a BP speaker[12] from the state Minas Gerais (MG), Southeast Brazil, and an example taken from Hedberg et al. (2014) of the same type of question being produced by a native English speaker from the US. In BP, questions rely on intonation only, and there is no linear syntagmatic distinction between a question and a statement. However, in EN there is also inversion or an auxiliary verb.

Figure 2: Production of a neutral Y/N question in BP and EN



Source: Left in BP, right in EN (Castelo and Frota, 2016, p. 108; Hedberg, Sosa and Görgülü, 2014, p. 13)

[12] Translation: Did the guys buy blades?

These cross-linguistic differences between BP and Canadian English intonation motivate the central hypothesis tested in this chapter: that BP speakers will sometimes transfer the BP rising–falling contour to English yes/no questions, and that this contour will be perceptually disruptive for native English listeners when segmental information is removed.

3.5. Methodology of L2 Prosody Perception

Investigating whether prosody alone conveys sentence modality requires a methodology capable of eliminating lexical, syntactic, and semantic cues while preserving pitch movement. Low-pass filtering has become the standard technique for this purpose (Munro 1995; Grabe et al. 2003; Radu et al. 2018). In this method, all frequencies above a speaker-specific threshold are removed, rendering the stimulus unintelligible while maintaining its melodic contour. This allows researchers to test whether listeners can rely solely on intonation to distinguish questions from statements.

Grabe et al. (2003) showed that English, Spanish, and Mandarin listeners accurately distinguished rising from falling contours under low-pass filtering, suggesting that pitch direction is a robust cross-linguistic cue. Radu et al. (2018) found that L1-Spanish L2-English speakers also reliably identified English question contours, indicating strong perceptual access to L2 intonation patterns even without segmental cues. Passarella dos Reis et al. (2016) extended this approach to BP–EN, demonstrating that BP learners had difficulty distinguishing English rising contours from their own rising–falling patterns, often interpreting both as declaratives.

The key insight for the present section is that low-pass filtering allows researchers to test whether a specific contour shape — such as BP's RF — obscures interrogativity for native English listeners. Because the technique forces listeners to rely solely on pitch, perceptual confusions can be directly attributed to intonational structure rather than lexical bias or syntax.

Across all experiments in this research program, stimuli were low-pass filtered. Every stimulus was played twice, separated by 500 ms of silence, and listeners classified the stimulus as either a "Question" or a "Statement".

Across studies, low-pass filtering provided a controlled environment to test how English listeners perceive BP-like prosodic contours, particularly the rising-falling (RF) nuclear pattern as a question or statement, when no segmental information is present.

3.6. Methodology of L2 Prosody Production

A central challenge in L2 prosody research is devising an elicitation method that is natural enough to reflect learners' habitual prosodic patterns while controlled enough to allow systematic acoustic and ToBI-based analysis. Context-guided self-paced reading tasks have become standard for eliciting sentence-type intonation (Pierrehumbert 1980; Beckman and Elam 1997; Frota et al. 2015; Patience et al. 2018). This methodology was adopted across all experiments in this research program and three principles support its use.

First, contextualized prompts minimize pragmatic variability. Without context, learners may impose emotional or discourse-driven intonational patterns, producing contours that deviate from canonical sentence modality. Contexts constrain speakers to neutral interpretations, producing consistent broad-focus statements and neutral yes/no questions.

Second, scripted reading tasks ensure segmental and syntactic uniformity across repetitions and across speakers. Since the analytical target is the nuclear contour rather than segmental accuracy, controlling the lexical material prevents unwanted variability in prosodic structure. Each target sentence was therefore preceded by a brief contextual prompt and produced three consecutive times, yielding comparable nuclear contours suitable for annotation.

Third, this method aligns with a large tradition in prosody research. Grice et al. (1997) used context-guided reading to analyze German and Italian intonation; Frota et al. (2015) applied similar techniques in Romance languages; and Kim and Arnhold (2024) adopted scripted contexts to study focus marking in Canadian English. These precedents validate the use of controlled reading tasks for examining phonological categories such as pitch accents and nuclear contours in both L1 and L2 systems.

In this dissertation, production data were analyzed in Praat. All recordings were segmented, annotated for pitch-accent type and nuclear contour, and acoustically inspected. Nuclear contours were labeled using ToBI conventions, distinguishing rising (R), falling (F), rising–falling (RF), and sustained (S) patterns, with tonal alignment verified relative to the stressed syllable. Acoustic measurements of minimum and maximum F0 in the nuclear region were extracted to compute pitch span (F0 range), supporting quantitative analyses of contour shape.

Although specific recording conditions varied across experiments — remote recordings during the pandemic in Experiments 1 and 1.2, and controlled laboratory recordings in Experiment 2 — the core methodology remained consistent. Experiment 2 implemented improved hardware, controlled recording environments, and standardized procedures to reduce item-level variability and yield higher-quality stimuli for perception testing.

Across all studies, the production methodology provided a controlled yet naturalistic basis for evaluating how BP speakers produce English sentence-type intonation and how these productions influence subsequent perception.

3.7. Experiment 1 – Buzan et al. (2022)

The first study we present, Buzan et al. (2022), served as a pilot investigation into the presence of cross-linguistic intonational influence among BP learners of English and its perceptual consequences for both EN native listeners and BP learners. We hypothesized that prosodic transfer would occur in BP speakers in yes/no questions and that this transfer would hinder native English perception.

Although modest in scale and conducted under pandemic constraints, it introduced the methodological foundations used throughout the dissertation: context-guided production data, multiple repetitions of each item, and the use of low-pass filtered stimuli for the perceptual tasks. It also revealed the central pattern that motivated all subsequent work — the RF contour characteristic of BP yes/no questions appeared in EN productions by BP speakers, and this contour systematically misled native EN listeners into interpreting those utterances as statements.

Experiment 1 was the first of three stages of our investigation, and it provided crucial early evidence that prosodic transfer in this language pair is not only common, but also functionally significant.

3.7.1. Production participants

The production included twenty-three speakers: nineteen BP speakers from Southeast Brazil (Minas Gerais, Rio de Janeiro, São Paulo) and four native English speakers (two American, two British), all aged between twenty and twenty-five and balanced by gender. Seven BP speakers were excluded because they did not follow recording instructions, often due to background noise, hesitations, or failure to record all items. The final dataset therefore contained productions from twelve BP speakers (seven women and five men) and four native English speakers, sixteen in total. Based on two experienced English teachers' evaluation on their English productions, the 12 BP Speakers were divided into advanced (5), intermediate (2) and beginner (5) levels. Table 1 shows the participants summary.

Table 1: Production participants

Group	N	Gender Breakdown	Origin
BP Speakers	12	7 women, 5 men	Southeast Brazil
Native EN Speakers	4	2 women, 2 men	2 from USA, 2 from Leicester (UK) 1 man and 1 woman from each country

Source: Own elaboration.

3.7.2. Production Task and Methodological Rationale

Participants recorded twelve sentences consisting of six statements and six yes/no questions (Appendix 1). Each sentence was preceded by a brief contextual prompt intended to minimize pragmatic variation. Each sentence was recorded three times consecutively without pausing or restarting the recording, resulting in thirty-six tokens per participant.

Due to COVID-19 restrictions, the recordings were made by participants in their homes using personal mobile devices. Instructions were delivered in PDF format along with the contexts and target sentences.

3.7.3. Acoustic and Phonological Analysis

The goal of the production analysis was to identify instances in which BP speakers used the RF contour typical of southeastern BP neutral questions when uttering English Y/N questions.

3.7.4. Production Results

The analysis revealed clear and recurrent instances of L1 prosodic transfer. Across all BP participants, approximately one third of all question tokens showed the RF pattern. Declaratives showed no change, as expected.

Table 2 summarizes the distribution of RF realizations for three representative BP speakers in items Q9–Q12. These speakers exemplify the degree of variability observed in the dataset. Participants E and G displayed the highest number of RF productions, each producing multiple repetitions of Q10 and Q12 with the BP-like configuration.

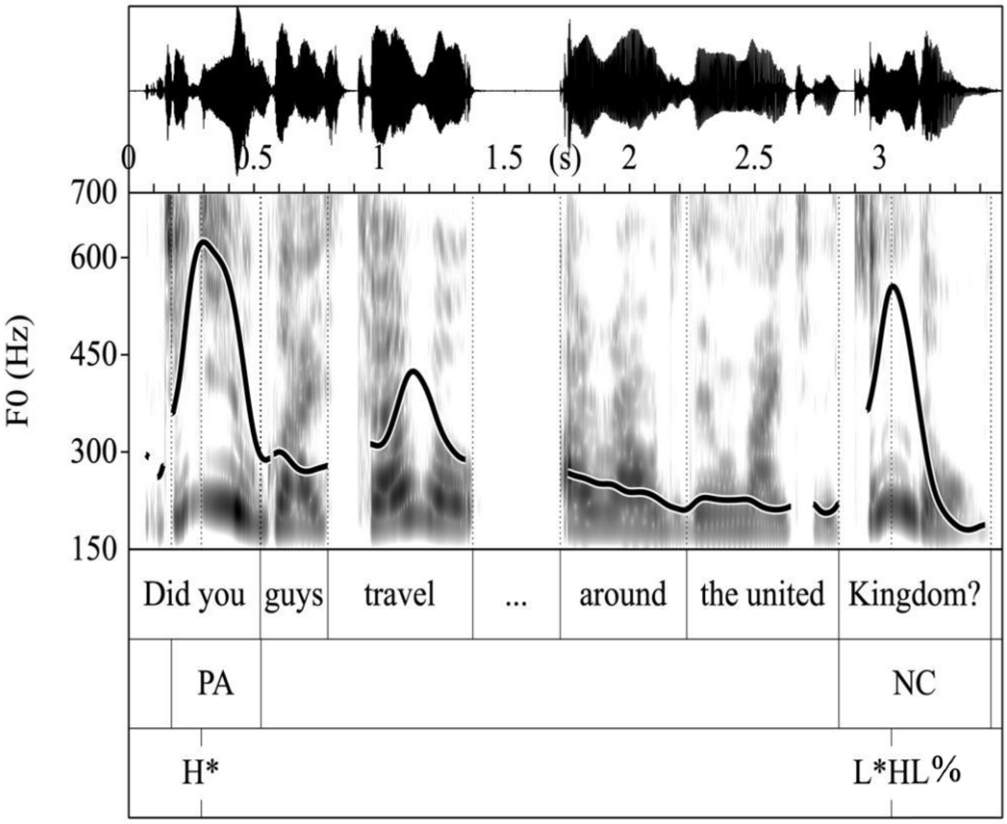
Table 2: Distribution of rise-fall nuclear contours across participants and question stimuli

PARTICIPANTS	QUESTION 9	QUESTION 10	QUESTION 11	QUESTION 12	TOTAL PER PARTICIPANT
PARTICIPANT E	-	1^{st}, 2^{nd} and 3^{rd} recordings	3^{rd} recording	3^{rd} recording	5
PARTICIPANT G	3^{rd} recording	1^{st}, 2^{nd} and 3^{rd} recordings	-	2^{nd} recording	5
PARTICIPANT F	-	1^{st} and 2^{nd} recordings	-	-	2

Source: Own elaboration.

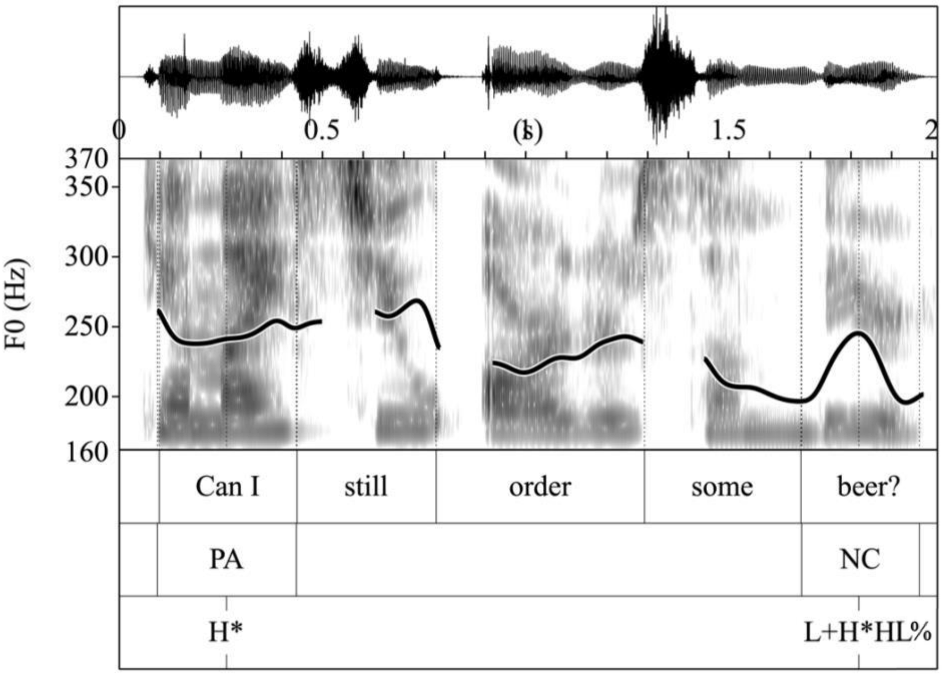
We chose two contours that best illustrate this phenomenon. Figure 3 presents an example of Q10 (“Did you guys travel around the United Kingdom?”) produced by Participant D, where the nuclear contour traces the characteristic BP rise on the nuclear syllable followed by a marked fall. Figure 4 shows Q11 (“Can I still order some beer?”) produced by Participant E, which displays a similar RF trajectory.

Figure 3: Question 10 – participant D, beginner EN-L2 speaker



Source: Own elaboration.

Figure 4: Question 11 – participant E, beginner EN-L2 speaker



Source: Own elaboration.

Given that EN neutral Y/N questions rely on a rising boundary tone (typically L*H-H%), the BP rise-fall contour represents a substantial structural deviation likely to affect listener interpretation.

3.7.5. Pilot Perception Task with English Listeners

The results of the acoustic analysis revealed that approximately 33% of the total questions in the collected sample produced by native BP speakers exhibit rise-fall intonation patterns. We aim to assess to what degree this prosodic influence is perceived by testing the perception of both native English and native BP listeners.

3.7.6. Perception Method

We conducted perception tasks with native EN (L1-EN group) and native BP listeners (L1-BP group). Regarding the L1-EN group, we included two versions[13] of the same forced-choice task: a version conducted via Google Forms, which was replaced for a version programmed in jsPsych and done through the Cognition[14] platform. Both versions are exemplified in section 3.7.8.

Twenty-seven stimuli were selected from the BP and EN recordings (APPENDIX 1). As for the selection criteria, we had 17 questions, of which: 2 are from AE, 3 are from BE, and 13 are from BP speakers; and 10 statements, of which: 2 are from AE, 4 are from BE and 4 are from BP (APPENDIX 8). At this stage of the research, we were not accounting for the different varieties and for the annotation of the stimuli (see Table 6 and 7 in section 3.8.3, Experiment 1.2, for the ToBI annotation of the stimuli).

All items were low-pass filtered in Praat using the pass-hann-band filter. Speaker-specific frequency cutoffs were applied (approximately 300–600 Hz for male voices and 600–800 Hz for female voices) to ensure that lexical information was removed while preserving the prosodic contour. Participants heard each stimulus twice and classified it as either a “Question” or a “Statement”.

---

[13] The Google Forms version was done while the author was concluding a coding course. Once that was done, the second version was coded using jsPych.
[14] https://www.cognition.run

3.7.7. Perception Participants

We tested fifteen L1-EN listeners, nine in the offline version and six in the online version. The L1-BP listener group tested later, and they were given only the Cognition link and had fifteen BP speakers, with varied self-assessed L2 proficiency levels. The aim was to examine whether BP learners themselves perceive English rising contours and BP RF contours differently when segmental and lexical information are removed. Recruitment was done through social media.

Table 3 presents a systematization of the L1-EN participants information. Participants age ranged between 18 to 45 years old.

Table 3: L1-EN participants description

Participants	Country of birth	L1	L2
Participant 1	US	English	Spanish, French
Participant 2	US/California	English	-
Participant 3	US/California	English	-
Participant 4	Spain/US	English	Spanish
Participant 5	US	English	-
Participant 6	US/New Deli	English	-
Participant 7	UK	English	Turkish, Spanish and Italian
Participant 8	US	English	-
Participant 9	US/Kentucky	English	-
Participant 10	England/Northamptonshire	English	Portuguese
Participant 11	US/Texas	English	-
Participant 12	Argentina	English	Portuguese
Participant 13	Canada/Vancouver	English	-
Participant 14	Canada	English	-
Participant 15	US/Utah	English	Spanish and Portuguese

Source: Own elaboration.

We see that most participants are from the US (N=10), but we also have participants from Canada (N=2), the UK (N=2) and Argentina (N=1). We kept all participants that reported having English as their L1 regardless of them speaking

Portuguese or not being born in an English-speaking country due to the extreme difficulty of recruiting[15] participants during the pandemic.

For the L1-BP group, seen on Table 4, we collected place of residence, age range and their self-assessed L2 listening proficiency through a Likert scale, since they were listening to English. Participants age ranged between 18 to 55 years old.

Table 4: L1-BP participants description

Participants	Where do you live?	Proficiency
Participant A	Paraíba	Low
Participant B	Uberlândia-MG	Very High
Participant C	Juiz de Fora	High
Participant D	Juiz de Fora	Average
Participant E	Juiz de Fora	Very High
Participant F	Juiz de Fora	Very High
Participant G	São Paulo-SP	Average
Participant H	Juiz de Fora	Average
Participant I	Teresópolis-RJ	Very High
Participant J	Juiz de Fora	Very Low
Participant K	Juiz de Fora	Very High
Participant L	Goiânia	High
Participant M	Governador Valadares-MG	Average
Participant N	Juiz de Fora	Very High
Participant O	Itaperuna-RJ	Low

Source: Own elaboration.

All participants claimed to have some English knowledge, and most participants are from Southeast Brazil (N=13).

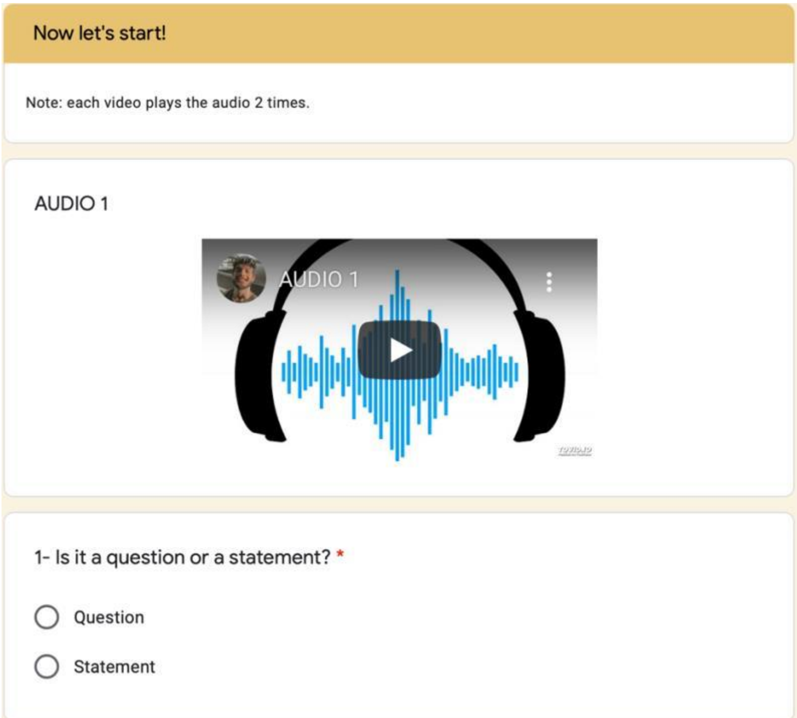
[15] Research conducted in Brazil is prohibited from offering compensation for participation according to our national Ethics Committee. Therefore, every participation was completely voluntary across all tasks in this dissertation.

3.7.8. Procedure

Regarding the Google Forms test, the participant read the instructions and provided personal information, such as country of origin, whether they spoke another language besides English, and age range before beginning to listen to the stimuli.

The test involved clicking on a video displaying a fixed image and the audio of a specific stimulus. This stimulus repeated twice with a one-second interval. Participants then judged whether the audio was a question or a statement by selecting the appropriate option below the video. The order of the words "statement" and "question" varied in each audio, both in the question and in the alternatives, so that the order of presentation would not bias the participants' choice. Figure 5 below shows a trial of this test.

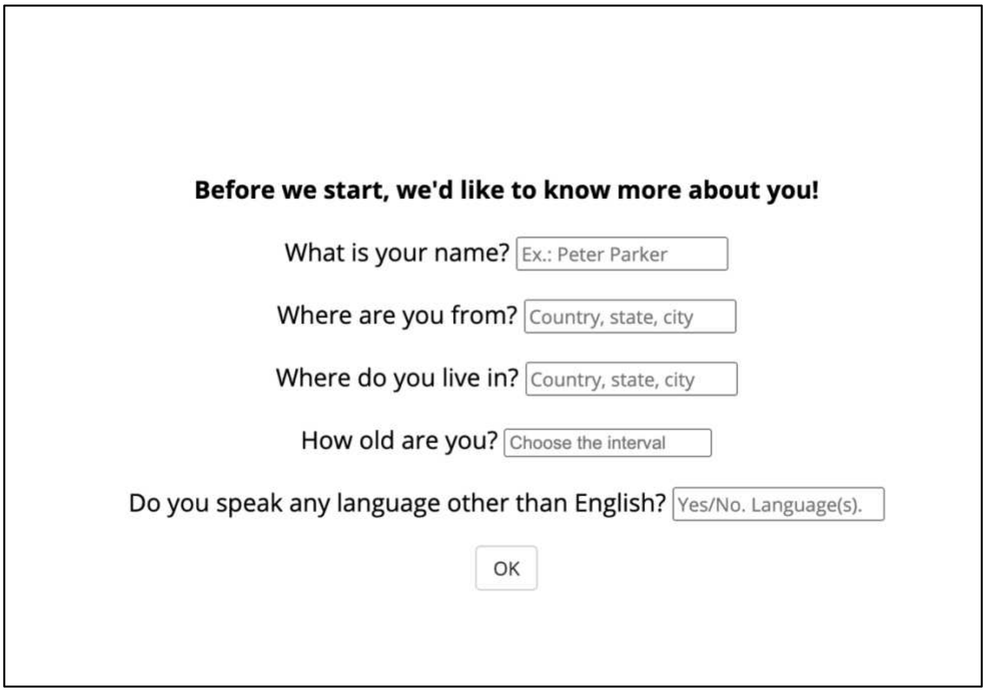
Figure 5: Example of a Google Forms trial



Source: Own elaboration.

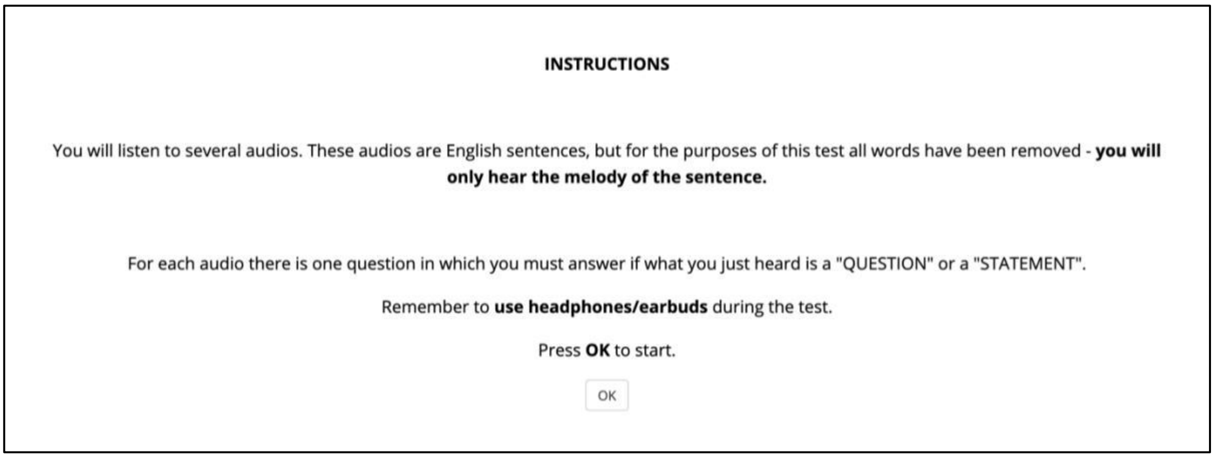
We refined this forced-choice task by programming it with the jsPsych library (De Leeuw, 2015) and applying it through the Cognition platform. There was no training in either version of this task (1 and 1.2), which we addressed in Experiment 2. The following figures show the windows that the L1-EN listeners saw on their screens. The first screen presents an informed consent form in English; the second contains some sociocultural questions; and the third contains basic information and instructions for taking the test, such as explaining the type of choice the participant will make and the importance of using headphones to perform the activity. Immediately after the third screen, the test begins. We measured accuracy and reaction times. Figures 6 and 7 are examples of the second and third screens, respectively.

Figure 6: Initial screen with sociocultural questions for native English speakers



Source: Own elaboration.

Figure 7: Instructions for native English speakers

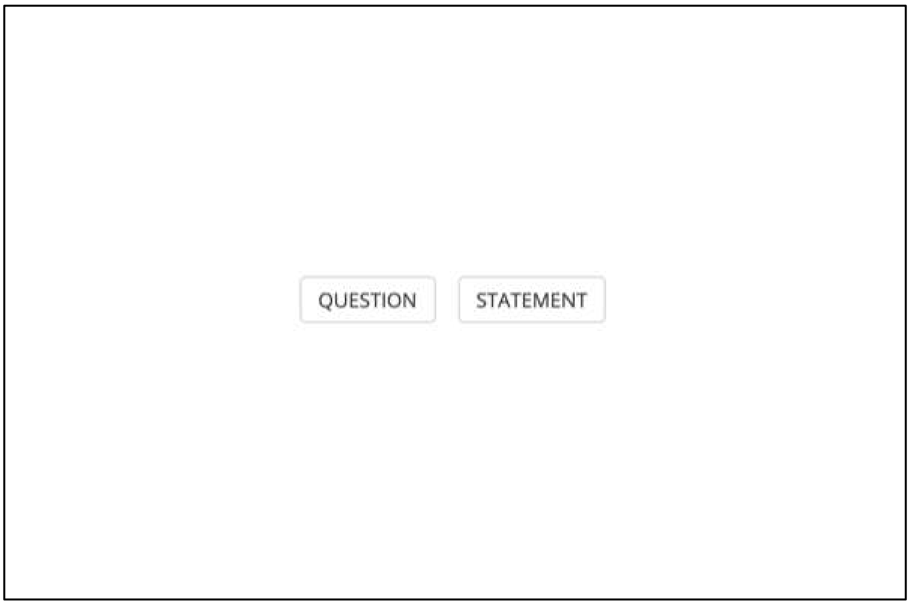


Source: Own elaboration.

The test begins after the instruction screen. Participants are presented with 27 screens, one for each audio, in a pseudo-randomized order. This type of randomization was chosen to ensure that there would not be many audios of the same nature (question/statement) being presented consecutively.

Each audio plays twice, with a one-second interval between repetitions. On the same screen, the "Question" and "Statement" buttons are displayed (Figure 8), and they must be selected by clicking with the mouse or tapping the cell phone screen immediately after the end of the second repetition. There is a 500ms fixation point between stimuli. Below is an example of the auditory stimuli screen.

Figure 8: Decision screen



Source: Own elaboration.

The sociocultural information, Likert scale and instructions screens for the L1-BP group are in the Appendix 2. We present the results of both versions of the task (for the L1-EN group) as we understand them to be essentially the same and part of a pilot experiment.

3.7.9. Results

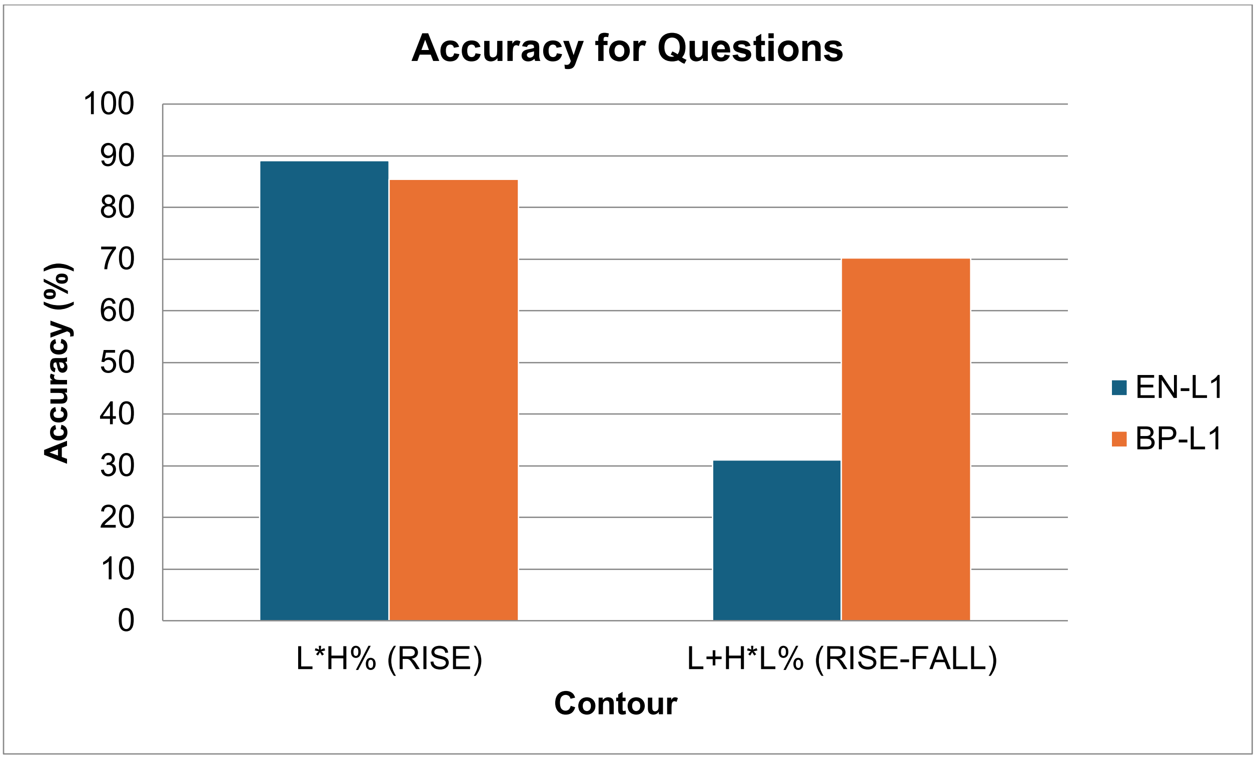
The L1-EN group showed a consistent pattern. Stimuli produced with a rising contour characteristic of English questions were identified as “Question” at high rates.

In contrast, stimuli that contained the RF contour were overwhelmingly misidentified as “Statement”.

The L1-BP group recognized English rising-final questions at high rates, indicating awareness of the English pattern. They also recognized BP-like RF contours as questions, which is expected given their familiarity with this contour in their L1. However, mean reaction time results showed that RF-contour questions led to longer processing times than English rising questions. Due to the scarcity of data, we could not run a statistical analysis.

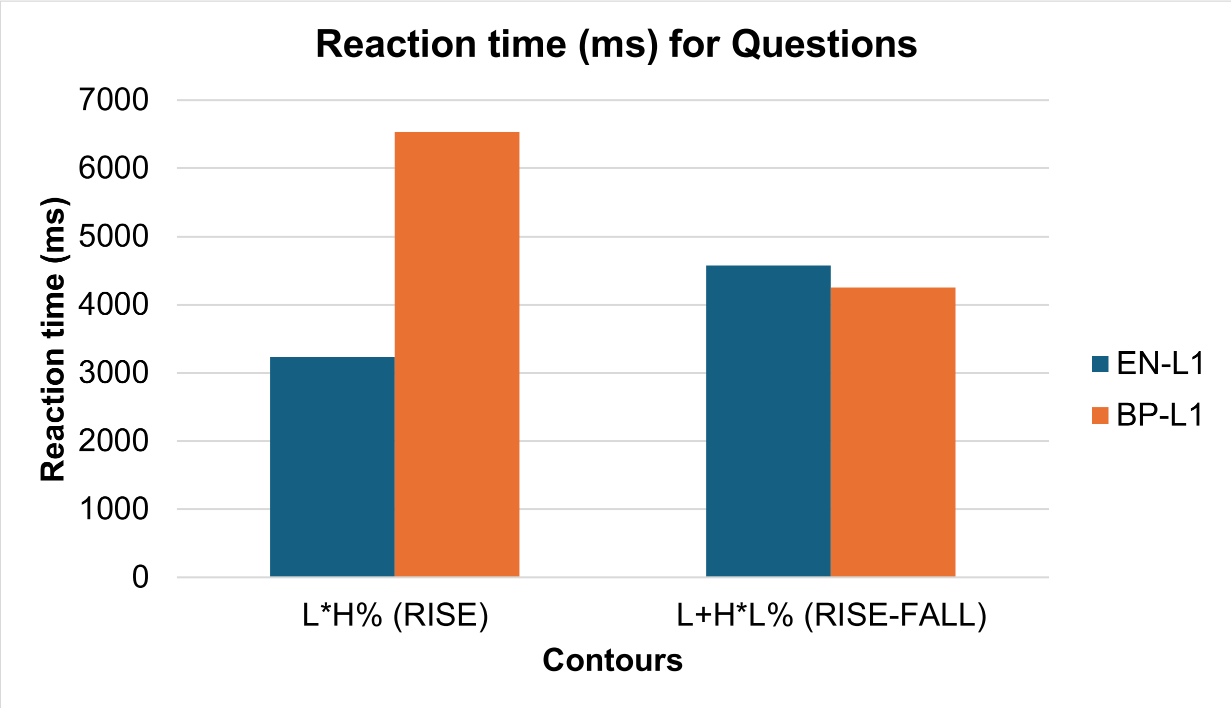
The accuracy and mean RT are displayed in figures 9 and 10 for both groups.

Figure 9: Mean accuracy for questions for both groups



Source: Own elaboration.

Figure 10: Mean reaction time for questions for both groups



Source: Own elaboration.

The perceptual results revealed that the RF contour is not simply different from English patterns but actively misleading for English listeners. They tended to interpret RF-ending questions as statements, demonstrating that prosodic transfer can hinder communicative clarity especially when segmental cues are removed[16], a result we do not see on the L1-BP group. The fact that we have a mismatch in accuracy for both groups suggests that we managed to capture the access to the implicit prosodic knowledge of both groups.

Reaction time results are congruent with this interpretation. The L1-EN group had longer response times to answer RF-contour questions than rising questions, suggesting additional processing effort or uncertainty triggered by the unfamiliar contour shape. The L1-BP group findings suggest that although BP listeners ultimately interpret both contour types as questions, their processing is not symmetrical. The RF contour, while familiar, may still introduce perceptual “noise,” possibly due to ambiguity about whether the contour corresponds to a neutral or expressive reading.

[16] Naturally, when they are not removed, they should be able to hear the inversion.

3.7.10. Limitations of Experiment 1

Experiment 1 was foundational but constrained. The primary limitation was the variability in recording conditions, as all productions were collected remotely using personal smartphones. Differences in microphone quality, background noise, speaking distance, and environment introduced some acoustic variability that could not be fully controlled in analysis. The sample size was also relatively small, both in the production and perception components, limiting the possibility of robust statistical modeling. Additionally, certain stimuli appeared more prone to prosodic influence than others, notably Q10, suggesting that sentence-specific prosodic properties influenced the likelihood of transfer. Without being able to run a mixed-effects modeling structure, it was not yet possible to distinguish reliably between effects attributed to speaker L1, sentence type, or idiosyncratic properties of particular items. Also, the RT results must be taken with a grain of salt due to the unbalanced number of participants, 9 for the EN-L1 group and 15 for the BP-L1 group.

These limitations motivated a new data collection in Experiment 1.2 and a more systematic experimental design implemented in Experiment 2, including controlled recording environments, expanded participant pools, and the incorporation of random-effects structures in all statistical analyses.

3.7.11. Summary of Experiment 1

Experiment 1 provided the first empirical evidence that BP learners may transfer the BP rising–falling (RF) contour to English yes/no questions, and that this transferred contour can disrupt interrogative interpretation for native English listeners. English listeners consistently misclassified RF-ending questions as statements, whereas rising-final questions were correctly identified at high rates. BP listeners, in turn, recognized both English rising and BP RF contours as questions, though RF questions elicited longer reaction times, suggesting an additional processing cost even for familiar contours.

As a pilot study conducted under pandemic constraints, Experiment 1 also exposed key methodological challenges that shaped the subsequent stages of the research program. Remote recordings produced notable acoustic variability, limiting the degree to which item-level differences, sentence type, and speaker L1 could be disentangled analytically. The small sample further restricted the use of mixed-effects modeling, and several stimuli appeared disproportionately prone to transfer, raising the possibility that perceptual effects might be tied to stimulus-level prosodic properties rather than to broader L1 categories.

Taken together, these findings established three insights that guided the redesign of the next experiments:

1) RF transfer is both frequent and perceptually consequential;
2) perceptual difficulty may arise from the nuclear contour itself rather than from speaker L1; and
3) a more controlled recording environment and larger participant pool are required to isolate the locus of this effect.

Experiment 1 therefore laid the conceptual and methodological groundwork for Experiment 1.2, which refined the analytical procedures and expanded sample size, and for Experiment 2, which implemented a fully controlled production–perception pipeline to test whether contour shape alone predicts perceptual success.

3.8. Experiment 1.2 – Buzan et al. (2026)

Experiment 1.2, presented in Buzan et al. (2026), constitutes the expansion of native English listener participant pool of Experiment 1. Building on the preliminary findings from Experiment 1, this experiment used a new and larger participant sample, and a significantly more robust annotation and analysis pipeline. Although the stimuli used in this experiment were also drawn from remote recordings collected during pandemic conditions, the study sought to overcome the limitations of the pilot by separating the English varieties (American and British) in the analysis and by conducting a detailed phonological and acoustic analysis of the stimuli before their use in perception testing.

The central goal of Experiment 1.2 was to examine how North American listeners, mainly Canadian English listeners, our only EN-L1 group[17], perceive low-pass filtered Y/N questions and statements produced by speakers of three L1 backgrounds: Brazilian Portuguese (BP), American English (AE), and British English (BE). This allowed the study to assess whether perceptual difficulty arises from cross-linguistic influence associated with L1, BP intonation or whether it results from more

[17] We did not test L1BP listeners on Experiments 1.2 and 2 due to time constraints. We further discuss this topic in the future directions section.

general factors, such as the shape of the nuclear contour or item-specific acoustic variation. The distinction of AE and BE stimuli also made sense from an ecological validity[18] standpoint.

Experiment 1.2 preserved the essential low-pass perception logic described in Section 3.4. In contrast to the pilot study, however, we incorporated a carefully controlled annotation of nuclear contour types and an expanded mixed-effects modeling framework, allowing for more precise identification of the perceptual predictors that account for listener performance.

3.8.1. Perception participants – Experiment 1.2

Twenty-three L1-English listeners participated in the perception experiment. All participants were living in Toronto at the time of testing and ranged from twenty-five to thirty-five years of age. One participant was excluded for being a native speaker of Spanish rather than English, and one was excluded due to low accuracy scores < 0.4, which suggested possible inattention to the task. The final sample consisted of twenty L1-English listeners and they are displayed in Table 5.

[18] Ecological validity concerns how well experimental materials reflect the linguistic environments listeners naturally experience. Because Canadians routinely encounter multiple English varieties, including both North American and British English through daily interaction and media exposure, incorporating AE and BE stimuli makes the task more representative of their real-world perceptual input.

Table 5: Participants' description

PARTICIPANT	ID	COUNTRY/CITY OF BIRTH	L2S SPOKEN
P1	109	Canada, Ontario, Toronto	German
P2	144	Canada, Ontario, Hamilton	None
P3	147	Canada, Ontario, Toronto	None
P4	12	USA, Texas, Midland	None
P5	171	Toronto, Canada	None
P6	118	Canada, Ontario, Toronto	Spanish, French, Japanese
P7	126	Canada, Ontario, Toronto	Japanese, Mandarin Chinese, Korean
P8	141	Burlington, Ontario, Canada	None
P9	149	Hamilton	None
P10	154	Canada, Ontario, Hamilton	None
P11	160	Toronto ON, Canada	None
P12	157	Hamilton, ON, Canada	Hindi, Marathi
P13	159	Canada, Ontario, Toronto	Japanese
P14	199	Toronto	Taiwanese Mandarin
P15	23	USA, Utah, Provo	Spanish, Portuguese (EU)
P16	96	USA, North Hampshire	None
P17	7	Canada	None
P18	31	Canada, Vancouver	None
P19	152	Canada, Ontario, Hamilton	None
P20	98	USA, Oregon	Spanish

Source: Own elaboration.

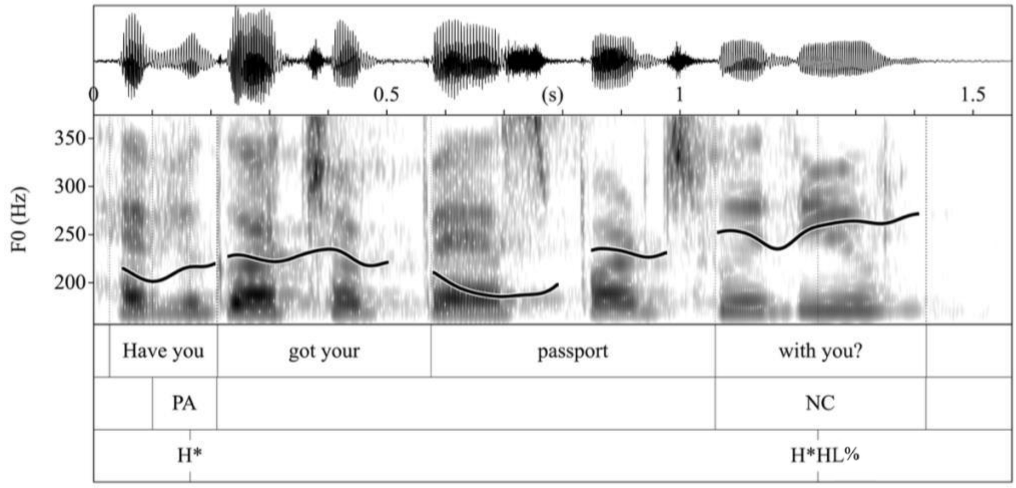
3.8.2. Stimuli

The stimuli consisted of twenty-seven utterances selected from the production recordings described in the Buzan et al. (2022), our Experiment 1. No distractor stimuli were included and there was no training.

3.8.3. Annotation of Nuclear Contours

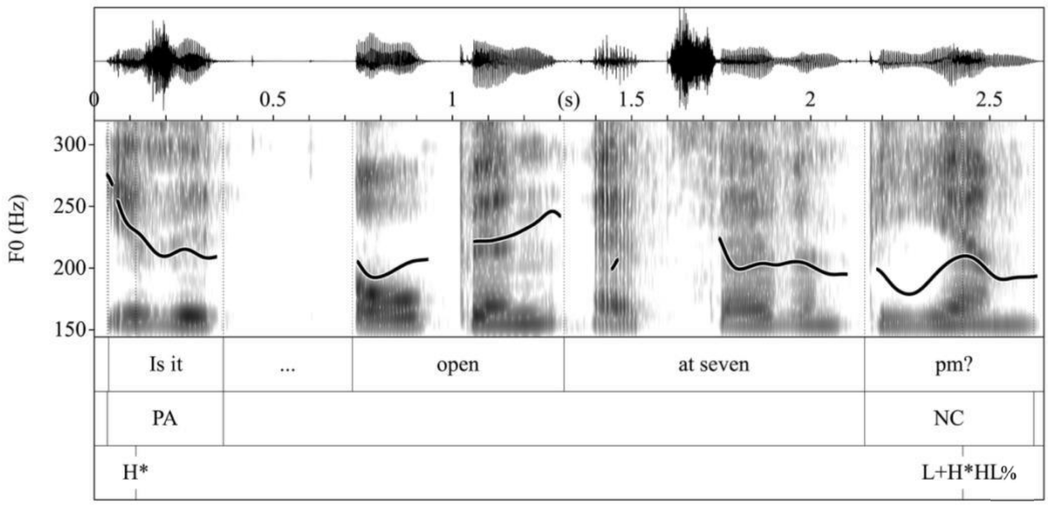
A detailed ToBI-based annotation was conducted to classify the nuclear contours of all stimuli. The author of this dissertation and an experienced professor in annotation independently inspected the pitch trajectory from the nucleus to the boundary: downward movements were labelled as falling (H*L%, H*+L% for questions; L*L%, H+L*L%, H*+L% for statements), upward movements as rising (L*H%), and plateau-like trajectories as sustained (L+H*HL%, L*HL%, H*HL%). Figures 11 and 12 show two sustained contours.

Figure 11: Question 16 – participant F, beginner EN-L2 speaker



Source: Own elaboration.

Figure 12: Question 23 – participant E, beginner EN-L2 speaker



Source: Own elaboration.

Finally, tonal alignment relative to the stressed syllable was also checked. This procedure ensured consistent annotation across speakers and sentence types. Table 6 summarizes the types of nuclear contour[19] by speaker L1 for questions and Table 7 for statements:

Table 6: Question's nuclear contour types and their speaker's L1

L1	NC	Type	Count
AE	L*H%	Rise	2
BE	H*+L%	Fall	1
BE	H*L%	Fall	1
BE	L*H%	Rise	1
BP	L*H%	Rise	5
BP	L+H*HL%	Sustained	3
BP	L+H*L%	Rise-fall	2
BP	H*HL%	Sustained	1
BP	L*HL%	Sustained	1

Source: Own elaboration.

Table 7: Statement's nuclear contour types and their speaker's L1

L1	NC	Type	Count
AE	H+L*L%	Fall	2
BE	H+L*L%	Fall	3
BE	H*L%	Fall	1
BP	H+L*L%	Fall	3
BP	L*L%	Fall	1

Source: Own elaboration.

Looking at questions, as shown in Table 6, AE speakers produced only rising nuclear contours (L*H%). BE speakers produced both rising (L*H%) and falling contours (H*L%, H*+L%)[20]. BP speakers showed the greatest variability, producing

[19] All initial pitch accents are H*.
[20] In H*L%, we have the high target on the stressed syllable independently of the low boundary tone, whereas in H*+L%, we have the low target part of the pitch accent being associated with the accent, yielding the start of the fall inside the nuclear accent.

rising-falling (L+H*L%)[21], rising (L*H%), and sustained contours (L+H*HL%, L*HL%, H*HL%). For statements, only falling contours were found, with two realizations: L*L% and H+L*L%.

The distribution of nuclear contours in statements in Table 7 shows a clear predominance of falling contours across all three L1 groups. AE statements were consistently realized with H+L*L% falls. BE also favored falling contours, particularly H+L*L%, but with some variation, as a smaller number of H*L% contours were also produced. BP showed a concentration of H+L*L% falls, complemented by a smaller number of L*L% falls. Overall, this confirms that in the dataset, falling nuclear contours were the choice for statements, while cross-variety differences emerged in the relative use of H+L*L% *versus* L*L% and H*L%. As a reminder, Figures 1 and 2 illustrate examples of a rising–falling nuclear contour, typical of southeastern BP and indicative of prosodic transfer into English productions.

3.8.4. Acoustic Measurements

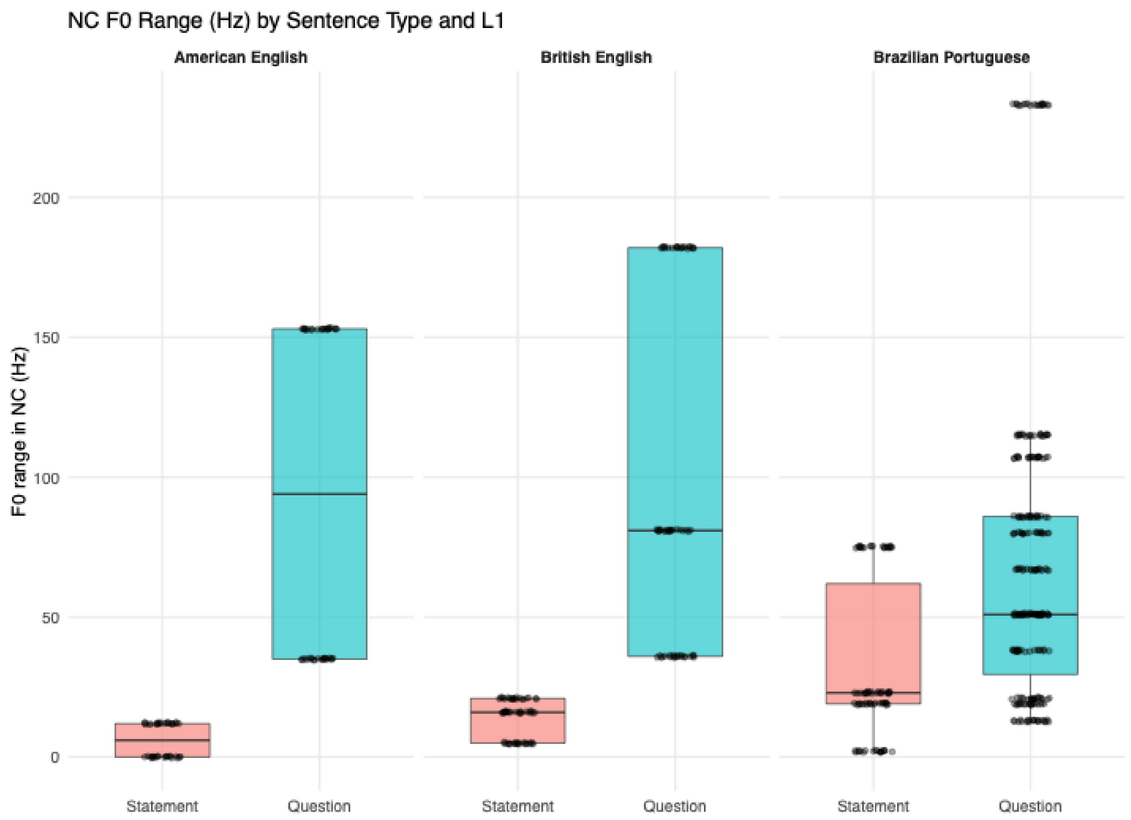
In addition to phonological annotation, the nuclear contour region of each stimulus was inspected acoustically. This choice is motivated by both theoretical and methodological considerations. In Autosegmental-Metrical models of intonation, the nuclear configuration is widely treated as the most informative locus for encoding and recovering sentence modality (e.g., statement vs yes/no question), because it contains the final tonal specification and the phrase-edge tones that most directly contribute to the global contour and to listeners' categorical interpretations. Prenucelar regions were not measured since comparatively they are more variable across speakers and styles and tend to contribute less consistently to modality distinctions, particularly when the task requires a forced categorical judgment.

Minimum and maximum F0 values were extracted with a Praat script (Appendix 3) for the nuclear region, allowing calculation of the F0 range. This measurement provided a quantitative indicator of pitch span, which could then be tested as a

[21] The "rise-fall" patterns in BE and BP are not equivalent. BE "rise-fall" contours are typically accent-driven (often analyzed as a falling nuclear pitch accent such as H+L* followed by L%), while the BP rise-fall observed here (L+H*L%) comprises a rising pitch accent (L+H*) followed by a low boundary tone (L%), meaning that the fall is largely realized as an edge movement toward the phrase boundary rather than being encoded as a falling pitch accent.

predictor of perceptual accuracy. Graph 1 of F0 ranges for statements and questions across the three L1 groups appears below.

Graph 1: NC F0 Range by Sentence Type and L1



Source: Own elaboration.

AE and BE display a clear categorical split: statements display a very small pitch range, while questions are marked by wide pitch spans. In contrast, BP shows a smaller difference in pitch range than the other two languages, with considerable overlap between sentence types. This acoustic overlap suggests that pitch span in nuclear contours may be a less reliable cue to sentence modality in BP, potentially contributing to lower perceptual accuracy.

3.8.5. Perception Task Procedure

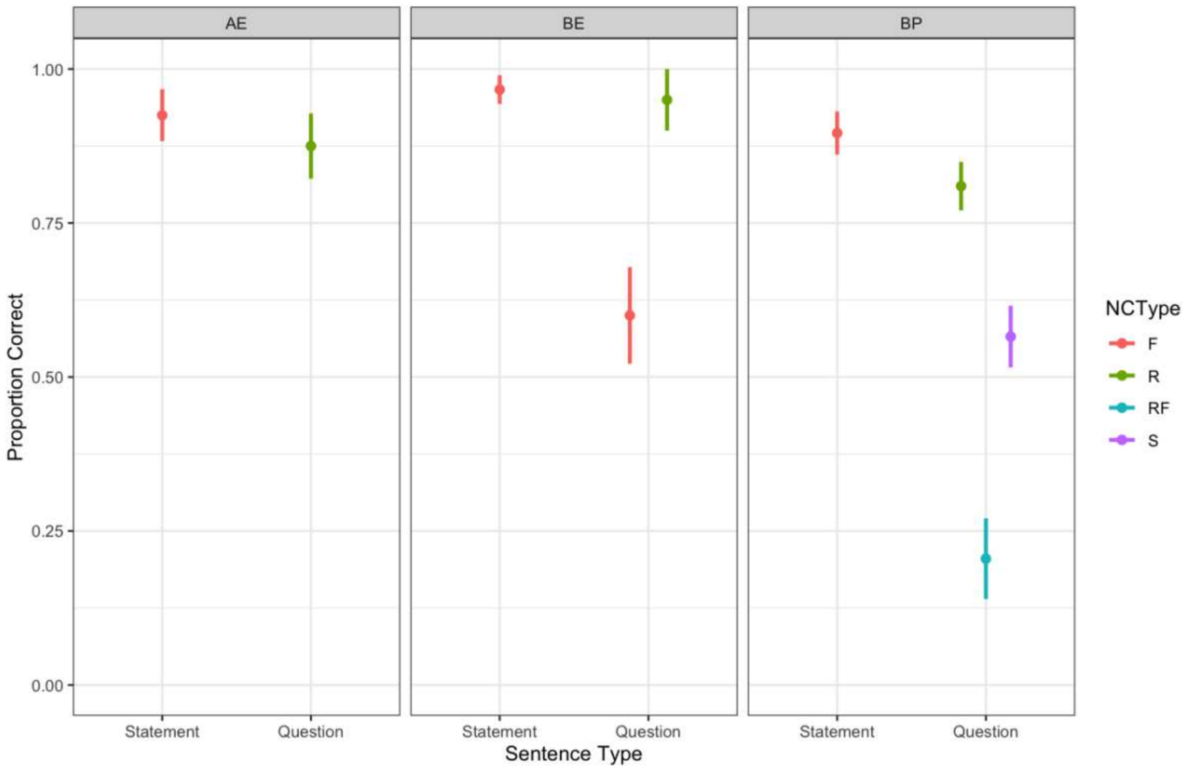
Procedure was the same as in section 3.5.

3.8.6. Results

This section presents the results for the perception task. This study aimed to examine the effects of speaker L1 (American English, British English and Brazilian Portuguese), sentence type (question and statement), and nuclear contour (NC) on Canadian English listeners' accuracy in identifying statements and questions. The initial pitch accent type was excluded due to lack of variance, since it was predominantly H*. This imbalance limits the interpretability and reliability of fixed-effect estimates for less frequent categories.

We predicted (1) lower accuracy for questions ending in a rising-falling contour, which are characteristic of beginner BP-L1 speakers from the Southeast Brazil, than in the ones ending with a rise; (2) high accuracy score for statements. Graph 2 shows the mean accuracy for both questions and statements by nuclear contour type and speaker L1, averaged across participants.

Graph 2: Mean accuracy by nuclear contour type and speaker L1



Source: Own elaboration.

Statements were identified with consistently high accuracy across L1s. In contrast, question accuracy was more variable and depended on contour category: in the BP condition, RF questions showed the lowest accuracy, whereas R questions patterned comparatively high, with S questions intermediate.

Because accuracy is binary and responses are repeated across both participants and stimuli, we modeled accuracy using generalized linear mixed-effects models using the lme4 package in R (R Core Team, 2021). The models include crossed random intercepts for Participant and stimulus to capture substantial item-to-item variability; models without an item random effect were frequently singular and overstated fixed-effect certainty.

Table 8: Effect of Nuclear Contour type

Fixed effects	β (estimate)	OR	SE	z-value	p-value
(Intercept)	2.88	17.76	1.07	2.70	0.007
L1BE	1.12	3.06	1.44	0.78	0.438
L1BP	0.08	1.08	1.14	0.07	0.946
Sentence Type (Question)	-3.56	0.03	1.64	-2.17	0.030
NCTypeR	2.96	19.28	1.76	1.68	0.093
NCTypeRF	-1.19	0.30	2.18	-0.55	0.584
NCTypeS	1.38	3.98	1.92	0.72	0.473

Source: Own elaboration.

Table 9: Interaction of L1 and Sentence Type

Fixed effects	β (estimate)	OR	SE	z-value	p-value
(Intercept, AE Statements)	2.87	17.69	1.61	1.78	0.075
L1 BE	1.31	3.70	2.20	0.60	0.552
L1 BP	0.33	1.39	2.02	0.16	0.871
SentenceType (Questions)	-0.52	0.59	2.25	-0.23	0.816
L1 BE x Questions	-2.37	0.09	2.97	-0.80	0.426
L1 BP x Questions	-1.60	0.20	2.62	-0.61	0.541

Source: Own elaboration.

In the main item-controlled model shown in Table 8, the effect of sentence type remains reliable: questions are significantly less likely to be identified correctly than statements (Estimate = −3.56, SE = 1.64, z = −2.17, p = .030; OR = 0.03, 95% CI

[0.00, 0.71]). Once item and participant variability are included[22], L1 does not contribute robust differences in accuracy (L1BE $p = .438$; L1BP $p = .946$), suggesting that the lower BP accuracy observed descriptively is not captured as a stable main effect independent of item difficulty. Regarding contour type, there is only weak evidence for an advantage of R contours relative to the baseline contour category (NCTypeR Estimate = 2.96, SE = 1.76, $z = 1.68$, $p = .093$; OR = 19.28, 95% CI [0.61, 607.11]), while RF and S do not show reliable differences from baseline (ps ≥ .47). The wide confidence intervals for contour effects reflect substantial uncertainty once stimulus variability is appropriately modeled.

Table 9 shows Canadian listeners with a high baseline accuracy for AE statements, with the intercept indicating a strong tendency toward correct identification ($\beta = 2.87$, $z = 1.78$, $p = .075$). However, once stimulus- and participant-level variability are controlled for, there was no reliable evidence that accuracy differed by stimulus L1: neither BE stimuli ($\beta = 1.31$, $z = 0.60$, $p = .552$) nor BP stimuli ($\beta = 0.33$, $z = 0.16$, $p = .871$) significantly changed accuracy relative to AE. Likewise, sentence modality did not yield a reliable main effect in this model, with questions not differing significantly from statements overall ($\beta = -0.52$, $z = -0.23$, $p = .816$). Finally, the L1 × sentence type interaction terms were not significant (BE × Questions: $\beta = -2.37$, $z = -0.80$, $p = .426$; BP × Questions: $\beta = -1.60$, $z = -0.61$, $p = .541$), indicating that once item difficulty is accounted for, the dataset does not support a robust interaction whereby questions behave differently across L1s.

When we look at the odds ratios, the model still suggests a high baseline likelihood of correct identification for AE statements (OR = 17.69), but the uncertainty around this estimate is large (95% CI [0.75, 415.79]). Importantly, none of the odds-ratio contrasts provide reliable evidence for systematic L1 or sentence-type differences: BE stimuli are numerically higher than AE (OR = 3.70, 95% CI [0.05, 275.86]) and BP stimuli are also slightly higher than AE (OR = 1.39, 95% CI [0.03, 72.71]), but both effects are highly uncertain and non-significant. Similarly, questions are numerically less accurate than statements (OR = 0.59, 95% CI [0.01, 48.48]) and the interaction terms trend in a negative direction (BE × Questions OR = 0.09, 95% CI [0.00, 31.91]; BP × Questions OR = 0.20, 95% CI [0.00, 34.36]), yet these estimates

[22] We must keep in mind that lower accuracy in questions could be related to variability in the stimuli; the stimuli for statements are more similar than those for questions.

again come with wide confidence intervals. Overall, these results indicate that after controlling for stimulus-level variability, which is substantial in this dataset, the model does not provide precise evidence for an L1-by-sentence-type accuracy asymmetry, even though descriptive patterns in Graph 2 suggest lower performance on questions, particularly for BP stimuli with RF nuclear contour.

3.8.7. Interpretation

Experiment 1.2 shows that Canadian English listeners' accuracy in distinguishing questions from statements in low-pass filtered speech seems to be shaped primarily by sentence type and by the distribution of nuclear contour types across the stimulus set, rather than by broad L1 categories. Descriptively, questions were more often misclassified than statements, and the error pattern was strongly directional: across conditions, listeners tended to interpret question stimuli as statements more often than the reverse. This asymmetry was most pronounced for BP stimuli, where a substantial number of BP questions were heard as statements, while BP statements were largely identified correctly.

Graph 2 helps locate where these Question→Statement confusions arise. Statements were identified with consistently high accuracy across L1s, whereas question accuracy depended on contour category. In the BP condition in particular, RF questions yielded the lowest accuracy, while R questions patterned comparatively high and S questions were intermediate.

When item-level variability is modeled explicitly, this descriptive BP disadvantage is not recovered as a robust fixed effect of L1, and the L1 × sentence type interaction is not reliably supported. The mixed-effects analyses show substantial stimulus-to-stimulus variability, indicating that a considerable portion of performance differences reflects the specific items sampled in each condition. In other words, Experiment 1.2 suggests that the perceptual challenge is tied to particular stimulus configurations — especially those associated with low-performing question contours — rather than to BP as a global L1 category.

Overall, Experiment 1.2 refines the interpretation of Experiment 1 by showing that the locus of perceptual difficulty was concentrated in the question domain and was closely linked to contour distributions and item difficulty. This directly motivated the redesign implemented in Experiment 2, which aimed to improve control over stimulus composition and recording conditions and to reduce item-driven variability so that contour-based effects could be isolated more reliably.

3.8.8. Limitations of Experiment 1.2

Experiment 1.2 was limited by stimulus-level variability. Models that did not include an item random effect were frequently singular, and once stimulus intercepts were included, the large item-to-item variance substantially reduced the precision of fixed-effect estimates. As a result, although the descriptive results consistently point to a concentration of errors in questions—and particularly in low-performing BP question contours—mixed-effects comparisons of BP questions against AE and BE questions were not estimated with high precision.

Finally, the inclusion of multiple English varieties increased ecological validity but also introduced additional prosodic variability that was not central to the main research question. Together, these limitations motivated the controlled stimulus redesign and recording procedures adopted in Experiment 2, which were intended to achieve more balanced contour distributions and stronger control over item-level variability.

3.8.9. Summary of Experiment 1.2

Experiment 1.2 extended the findings of Experiment 1 incorporating systematic contour annotation, and adopting mixed-effects modeling with crossed random intercepts for participants and stimuli. Descriptively, listeners showed high accuracy for statements across stimulus L1s. The accuracy summary indicated that question performance varied by nuclear contour category, with BP RF questions exhibiting the lowest accuracy, R questions showing comparatively higher accuracy, and S questions intermediate.

At the same time, stimulus-level variability played a substantial role in performance. Once item-level variability was explicitly modeled, L1 effects and L1 × sentence type differences were not robust, indicating that the apparent BP disadvantage in questions is closely tied to the particular stimulus set and to contour distributions rather than to a categorical L1 effect. These constraints — uneven contour distributions across L1s, partial confounding between contour and speaker group, and strong item-driven variance — motivated the fully controlled stimulus design and recording procedures implemented in Experiment 2 to isolate contour-based effects under improved methodological conditions.

3.9. Experiment 2 – Redesigned Production and Perception Study

Experiment 2 represents the culmination of the methodological and theoretical developments initiated in Experiments 1 and 1.2. The first two studies showed that perceptual difficulty is driven by nuclear contour shape, especially the RF pattern characteristic of BP yes/no questions. Experiment 1.2 showed that this difficulty was amplified by item-level acoustic variability, especially in stimuli produced under uncontrolled recording conditions during the pandemic. These observations made clear that a more rigorous and controlled methodological approach was required.

Experiment 2 was therefore designed anew, beginning with new production stimuli carefully recorded under controlled circumstances. The goal was to minimize acoustic variability, ensure clear and consistent recording of intended contours, and provide a more stable basis for subsequent perception testing.

The perception component of Experiment 2 adopted the same low-pass filtering logic, but with improved filtering procedures, more consistent acoustics, and systematic stimulus exclusion based on auditory and acoustic criteria. This experiment therefore provides the strongest empirical test of how specific nuclear contour shapes — rather than broad L1 categories or item artifacts — influence English listeners' identification of sentence modality.

3.9.1. Experiment 2 Production Participants

The production component included eleven native speakers of BP and nine native speakers of North American English (NE).

All NE speakers either participated in person at the Phonetics Laboratory at the University of Toronto or, when necessary, via zoom under quiet and controlled conditions. Recordings done via zoom did not affect the quality of the F0. When participants (N=7) were tested in the laboratory, stimuli were recorded in a soundproof booth[23] and a Samson Q2U microphone.

---

[23] We thank Professor Alexei Kochetov for letting us use the phonetics lab at UofT.

The study included two participants, one from Malaysia (P7) and another from Iran (P3), both of whom were not born in English-speaking countries. Nonetheless, they were classified within the NE group, as they relocated to Canada prior to the age of ten and attended schools where English was the medium of instruction.

BP speakers were all born in the Southeast Brazil — Rio de Janeiro (RJ) or Minas Gerais (MG) — and reported moderate daily usage of English as an L2, ranging from five to thirty percent of daily language use. All participants were tested in person at the NEALP lab at the Federal University of Juiz de Fora under quiet and controlled conditions and were recorded using a SteelSeries 9 headset, ensuring a similar high-quality input. Tables summarizing participant characteristics appear below. Note that the percentage of English use and average proficiency in Table 13 was self-assessed by the participants.

Table 10: Native English participants

Participant	Sex	Age	Other L1	City of birth	State of birth	Country of birth	L2	% EN use
P1	F	22	French	Los Angeles	California	USA	No	85%
P2	M	21	Spanish	Orlando	Florida	USA	No	85%
P3	M	19	Farsi	Tehran	Tehran county	Iran	French	90%
P4	M	23	No	Fairfax	Virginia	USA	Japanese	90%
P5	F	21	No	New York	New York	USA	Spanish/ German	100%
P6	F	30	No	Vancouver	British Columbia	Canada	German	100%
P7	F	24	No	Kuala Lumpur	Federal Territory of Kuala Lumpur	Malaysia	Mandarin/ Cantonese /Malay	50%
P8	F	19	Russian/ French	Toronto	Ontario	Canada	No	75%
P9	F	20	No	Ottawa	Ontario	Canada	French	95%

Source: Own elaboration.

Table 11: Native Brazilian Portuguese participants

Participant	Sex	Age	L2	City of birth	State of birth	% EN use	Average proficiency
P1	M	28	English	Juiz de Fora	MG	20%	Advanced
P2	M	23	English/Japanese	Valença	RJ	15%	Advanced
P3	F	23	English/Spanish	Volta Redonda	RJ	25%	Intermediate
P4	F	23	English/Spanish	Petrópolis	RJ	10%	Advanced
P5	F	23	English/French	Itaperuna	RJ	30%	Intermediate
P6	F	26	English	Juiz de Fora	MG	5%	Intermediate
P7	F	23	English	Bom Jesus de Itabopoana	RJ	30%	Intermediate
P8	F	25	English	Barbacena	MG	30%	Advanced
P9	F	24	English/Spanish	Juiz de Fora	MG	30%	Advanced
P10	F	24	English/French	Juiz de Fora	MG	10%	Intermediate
P11	M	42	English/Spanish	Juiz de Fora	MG	30%	Advanced

Source: Own elaboration.

3.9.2. Stimuli and Procedure

As in the previous experiments, the target tokens included broad-focus statements and neutral yes/no questions, each introduced by a brief contextual prompt. For the test trials, twelve sentences were constructed — six questions and six statements — in quasi-mirrored pairs (for example, “This is my computer” and “Is this your computer?”). Twenty-four distractor sentences were also created, comprising exclamations, narrow-focus statements, and yes/no questions (refer to Appendix 4).

The set of sentences was divided into two recording lists, each presented in a different pseudo-randomized order to minimise potential order effects. Participants utilized a MacBook Air M1 to complete the task using PowerPoint, structured such that the initial slide displayed only the context, followed by a second slide with both the contextual prompt and the target sentence (see Appendix 5). Instructions directed participants to silently read the context, proceed to the next slide, and then articulate the target sentence three consecutive times — a self-paced task. Following three practice trials, participants commenced the test recordings. A break was indicated by a grey slide, at which point participants were required to complete a linguistic background questionnaire (Appendix 6). Upon completion, participants resumed recordings until the conclusion of the task.

3.9.3. Phonological and Acoustic Annotation

Contours were categorized as rising (R) — H*H%, H*H-H%, L*H%, L*H-H%, L+H*H% — falling (F) — H*L%, H*L-L%, H*L%, H*L-L%, H+L*L%, L*H%, L*L% — rising–falling (RF) — L+H*L% — or sustained (S) — H*!H%, L*!H%, H*!H%, L*!H% — depending on the direction and alignment of the pitch movement. Initial pitch accents were categorized as rising (R) — H*, L+H* — and falling (F) — H*L, L*, 93% of PAs were R.

Tables 14 and 15 show the nuclear contour distribution across the stimuli and Graph 3 shows the distribution of NC types. If we group the NCs by type, we have 89.5% L1-BP questions with a R contour and 97.9% L1-BP statements with a F contour; 91.5% L1-NE questions with a R contour and 89.6% L1-NE statements with F contour.

Table 12: Distribution of BP NC contours per sentence type

L1	Sentence Type	NC Type	NC	Tokens	Percent	Grouped percentage
BP	Question	F	H*L%	1	2.08	89.5
BP	Question	F	H*L-L%	3	6.25	
BP	Question	R	H*H%	5	10.4	
BP	Question	R	H*H-H%	1	2.08	
BP	Question	R	L*H%	13	27.1	
BP	Question	R	L*H-H%	19	39.6	
BP	Question	R	L+H*H%	1	2.08	
BP	Question	RF	L+H*L%	2	4.17	10.5
BP	Question	S	H*!H%	2	4.17	
BP	Question	S	L*!H%	1	2.08	
BP	Statement	F	H*L%	5	10.4	97.9
BP	Statement	F	H*L-L%	3	6.25	
BP	Statement	F	H+L*L%	19	39.6	
BP	Statement	F	L*H%	1	2.08	
BP	Statement	F	L*L%	19	39.6	
BP	Statement	S	H*!H%	1	2.08	2.1

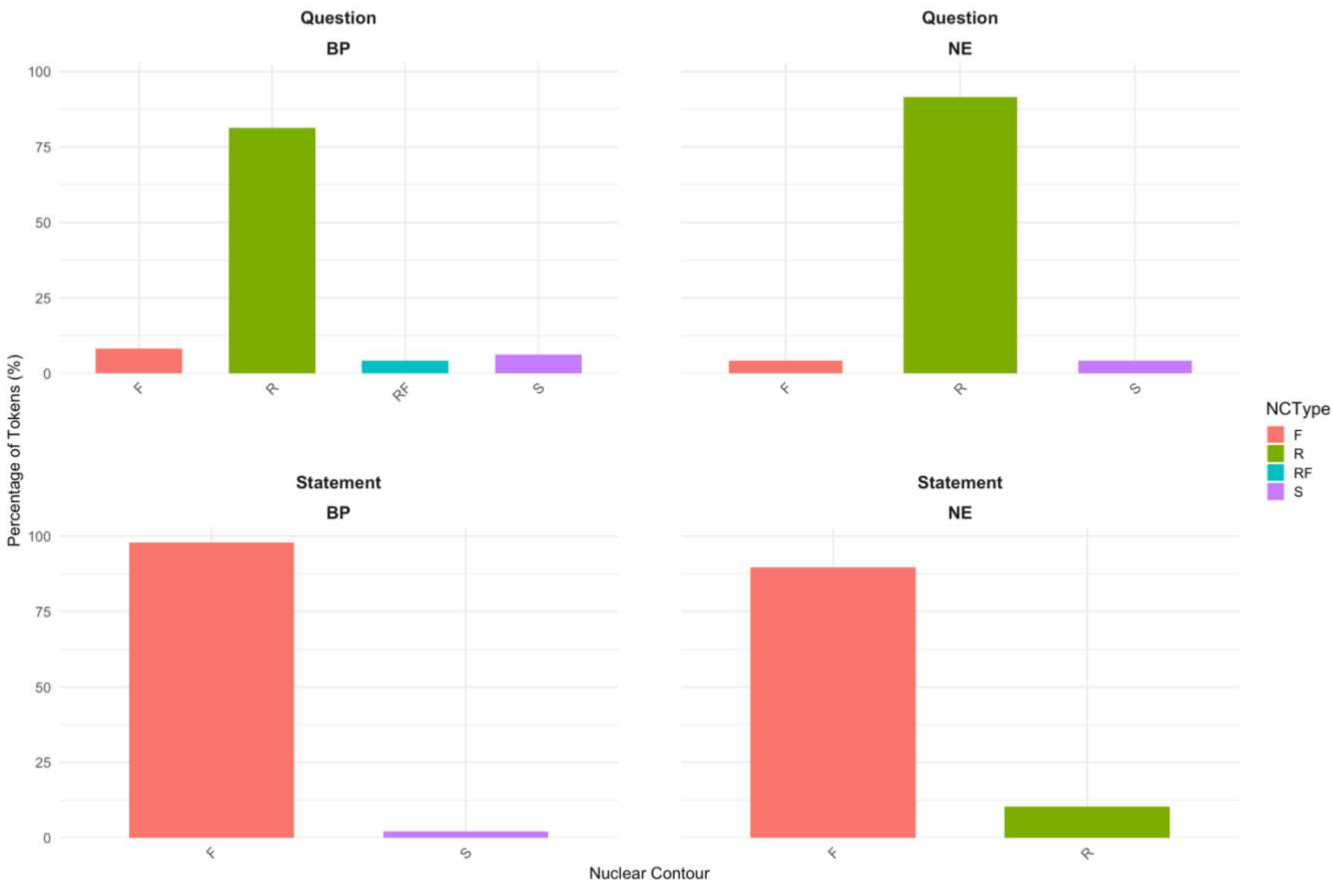
Source: Own elaboration.

Table 13: Distribution of NE NC contours per sentence type

L1	Sentence Type	NC Type	NC	Tokens	Percent	Grouped percentage
NE	Question	F	L*L%	2	4.26	4.2
NE	Question	R	H*H%	1	2.13	91.6
NE	Question	R	L*H%	28	59.6	
NE	Question	R	L*H-H%	14	29.8	
NE	Question	S	L*!H%	2	4.26	4.2
NE	Statement	F	H*L%	9	18.8	89.6
NE	Statement	F	H*L-L%	6	12.5	
NE	Statement	F	L*L%	27	56.2	
NE	Statement	F	L*L-L%	1	2.08	
NE	Statement	R	L*H%	4	8.33	8.33
NE	Statement	R	L*H-H%	1	2.08	2.08

Source: Own elaboration.

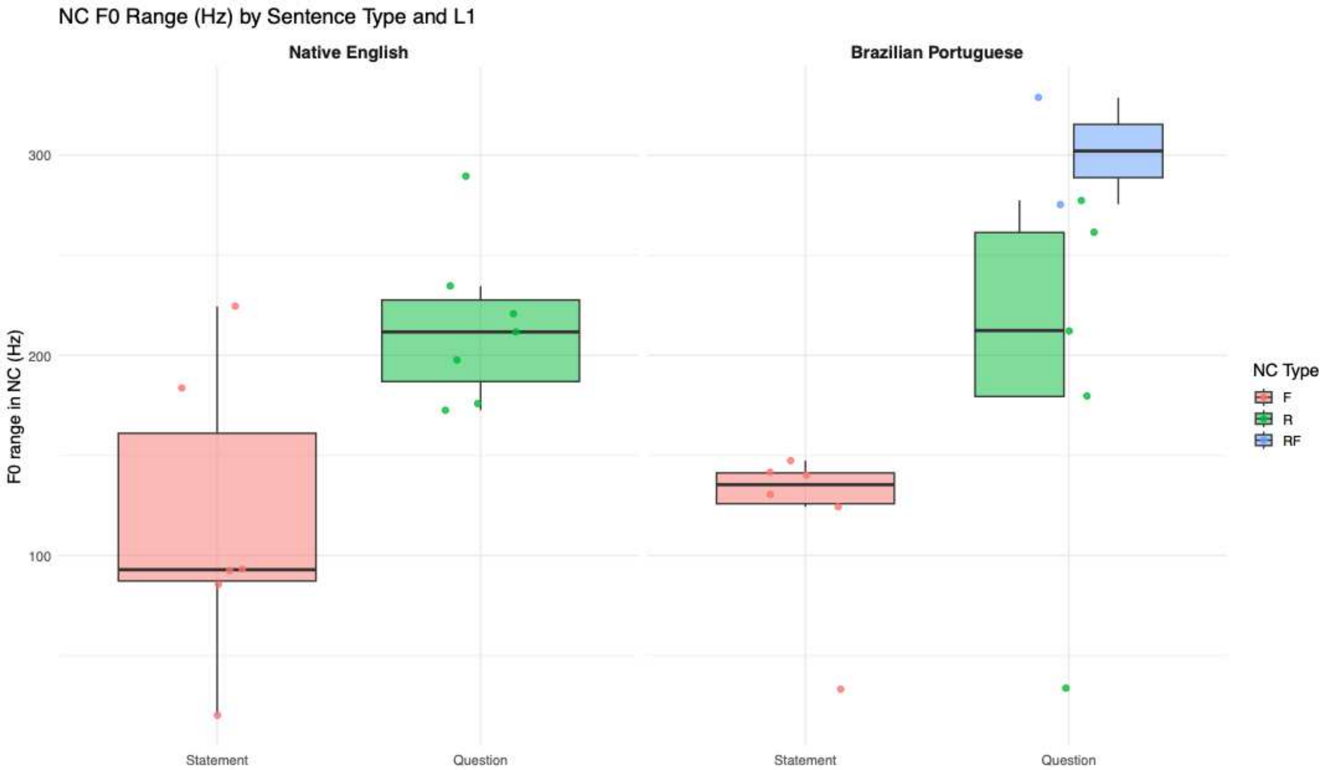
Graph 3: Distribution (%) of NC types per sentence type and L1



Source: Own elaboration.

Graph 4 shows the F0 range by sentence type and L1. We extracted the F0 range through the Contour Clustering software (Constantijn, 2023; Constantijn, 2025).

Graph 4: Nuclear Contour range by sentence type and L1



Source: Own elaboration.

North American English speakers consistently produced rising contours for questions and falling contours for declaratives. Pitch spans in NE questions were substantially larger than those in statements, and the timing of the nuclear rise aligned with expectations for North American Y/N interrogatives.

Brazilian Portuguese speakers displayed a larger repertoire of contour shapes for questions. Rising contours were common, as were falling contours, but RF contours and sustained plateaus were also attested on a smaller scale. Statements were generally falling but showed somewhat greater variation in pitch span.

Acoustically, NE speakers exhibited a categorical split between statements and questions, while BP speakers showed overlap in nuclear contour F0 range. This overlap suggests that contour direction and alignment, rather than pitch span alone, may be the primary perceptual cue distinguishing sentence types in BP.

These production patterns provided the basis for stimulus selection in the perception experiment.

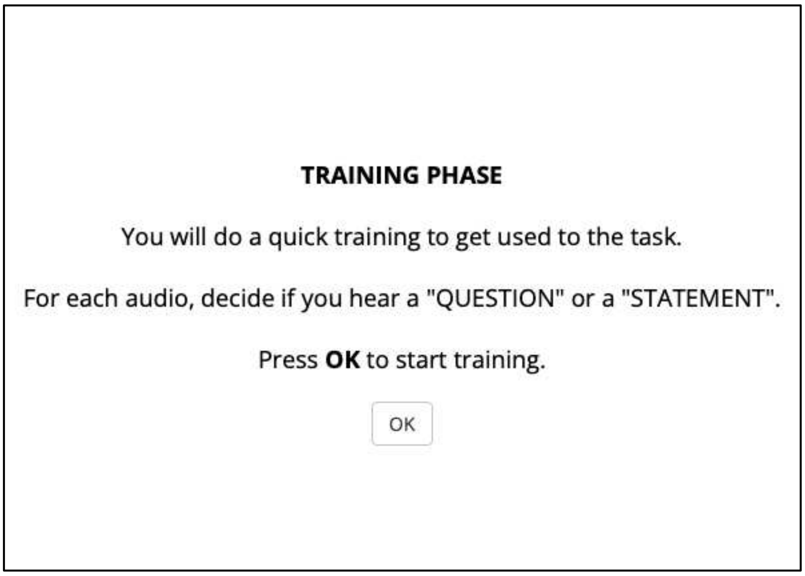
3.9.4. Experiment 2 - Perception study

Experiment 2 was programmed on the same platform as Experiment 1 and followed this timeline: consent form, personal information screen, training phase, test phase and end screen. We made improvements to the coding (APPENDIX 7), collected more demographical data (country of birth and place of residence — country, state/province and city) from participants, replaced the age range for exact age numbers, added a training phase and randomized all stimuli (opposed to pseudo-randomizing in Experiment 1). Figure 15 shows the improved personal information screen and Figure 16 the new training phase screen.

Figure 13: Personal information screen

Source: Own elaboration.

Figure 14: Training phase screen



Source: Own elaboration.

The training phase was composed of three stimuli, two questions and one statement, pass-filtered and produced by native English speakers.

3.9.5. Participants

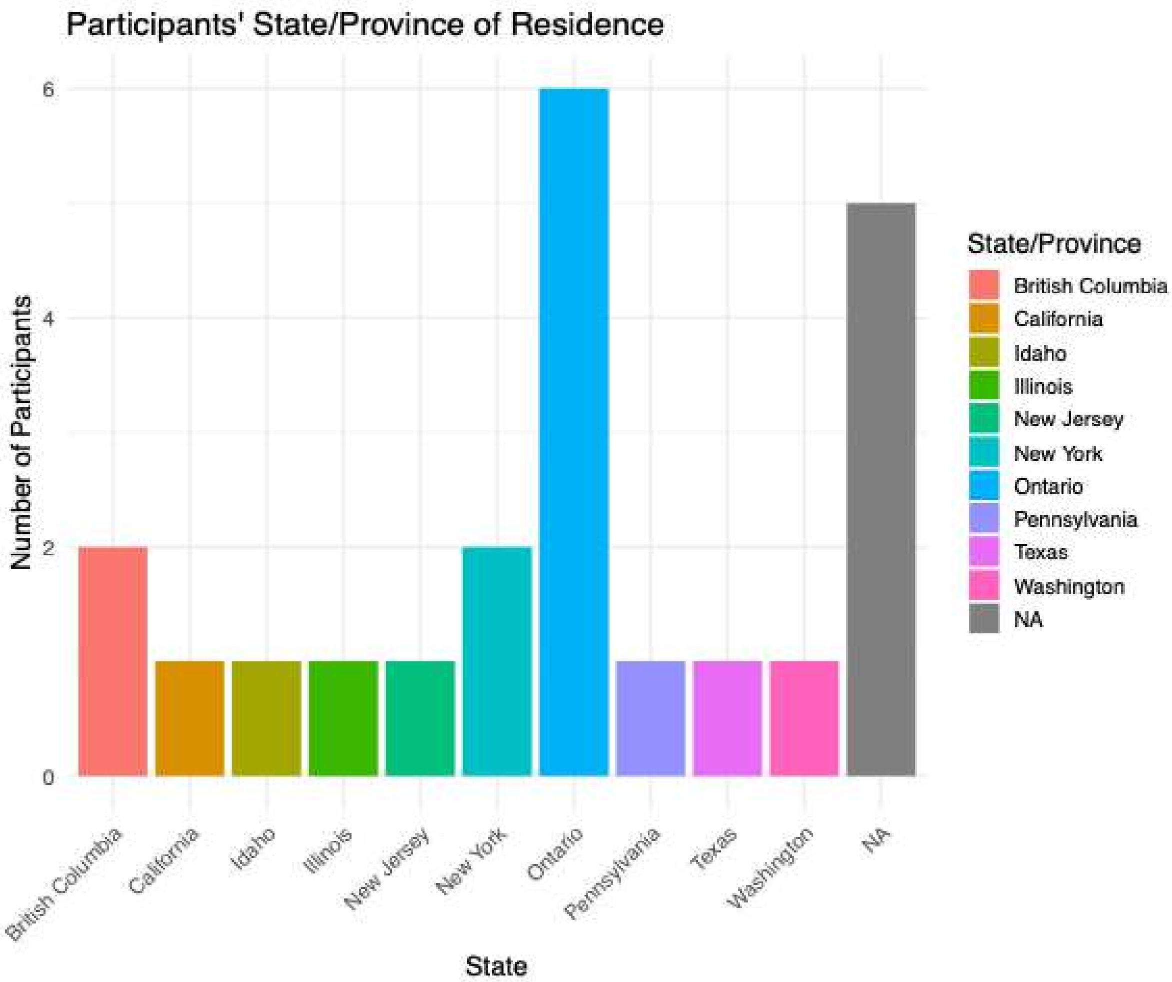
Twenty-five L1-EN listeners participated in the perception experiment. Three participants were excluded, two participants spoke Portuguese as their L2 and one was not from an English-speaking country. Therefore, we had twenty-two participants, fifteen from the US (answers varied between US, United States, USA and America) and 8 from Canada, with a mean age of 27 years. Table 16 shows the general participant's description. In Graph 5, we show participants' state/province of residence — NA data is from participants who preferred not to respond to this question.

Table 14: Participants' description

Participant	Country of birth	Country of residence	State/Province	City	Age	L2
1	Canada	Canada	Ontario	Toronto	21	Spanish
2	US	United States	Texas	Houston	22	Urdu
3	United States	France	-	Paris	23	French, Spanish
4	United States	Canada	British Columbia	Vancouver	36	
5	US	Germany	-	-	36	German
6	United States	United States	Idaho	Kuna	55	Spanish
7	US	Belgium	-	-	25	French, Swahili
8	United States	Canada	Ontario	Toronto	18	
9	United States	United States	New York	Plainview	19	Korean
10	America	United States	New York	New York City	19	French
11	USA	Hungary	-	-	36	Hungarian, Turkish
12	Canada	Canada	Ontario	Scarborough	20	French
13	United States	United States	New Jersey	Closter	18	Korean
14	United States	United States	Illinois	Chicago	26	Arabic
15	Canada	Canada	Ontario	Burlington	55	
16	United States	United States	California	San Francisco	22	Italian
17	Canada	Beirut	-	Lebanon	23	
18	Canada	Canada	Ontario	Toronto	30	
19	Canada	Canada	Ontario	Toronto	25	German
20	Canada	Canada	British Columbia	Vancouver	22	Korean
21	United States	United States	Pennsylvania	Philadelphia	23	French
22	United States	United States	Washington	Seattle	20	French

Source: Own elaboration.

Graph 5: Participants' state/province of residence



Source: Own elaboration.

Table 16 and graph 5 allow for a clear visualization of the diverse participant pool, which comprises of eight US states and 2 opposite Canadian provinces and eleven L2s — which we argue it represents the great bilingual community of what we consider to be a native English speaker.

3.9.6. Stimuli and Filtering

We sought to select stimuli that represented the data we had collected in production. For the training phase, we selected two questions with L*H% and one statement with L*L% as their nuclear contours. The following tables represent the 28 test stimuli, 14 L1BP and 14L1EN, with 7 questions and 7 statements from each L1 group. We did not include stimuli with sustained NCs and there were no distractors.

Tables 17 and 18 show the distribution of stimuli by the L1BP group and tables 19 and 20 show it by the L1EN group.

Table 15: NC contour for questions by L1BP

NC Contour	Number of stimuli
L+H*L%	2
L*H-H%	2
L*H%	2
H*H%	1

Table 16: NC contour for statements by L1BP

NC Contour	Number of stimuli
H+L*L%	2
L*L%	2
H*L-L%	1
H*L%	2

Table 17: NC contour for questions by L1EN

NC Contour	Number of stimuli
L*H%	4
L*H-H%	3

Table 18: NC contour for statements BY L1EN

NC Contour	Number of stimuli
H+L*L%	1
L*L%	5
H*L%	1

Sources: Own elaboration.

All stimuli selected from the production set underwent low-pass filtering in Praat, using a filtering script[24]. Each stimulus played twice, with 100ms of silence before the audio started playing and 500ms of silence between each repetition.

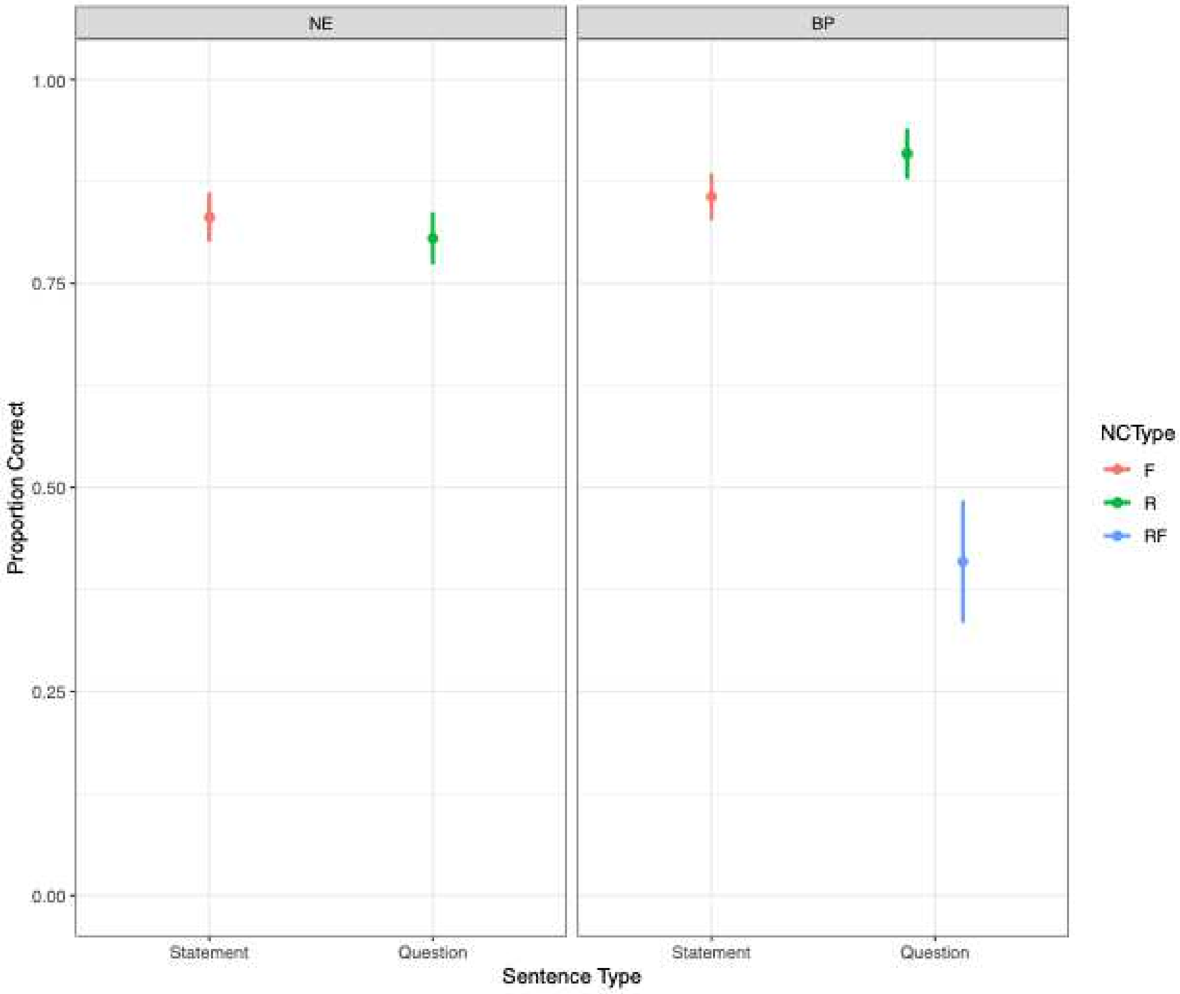
[24] We thank Professor Luciana Lucente for sharing the script PURR2004 1.0 by Petra Wagner (2004), available for download in https://github.com/pabarbosa/prosodia-experimental/commit/bb65ca9172b486c66741eed6f2f5c0e042c9049c.

3.9.7. Results: Accuracy

Upon inspection of the results, one BP question with a rising NC underperformed (<30% accuracy) compared to the other questions (>65% accuracy). When we listened back to problematic stimulus, we noticed that there was a sharp noise in the middle of the sentence, which could have misguided participants. Therefore, we decided to remove it from the analysis, leaving the BP question stimuli with 6 items.

Graph 6 shows the accuracy summary by L1 of the stimuli, Native English (NE) and Brazilian Portuguese (BP) and NC, fall (F), rise (R) and rise-fall (RF).

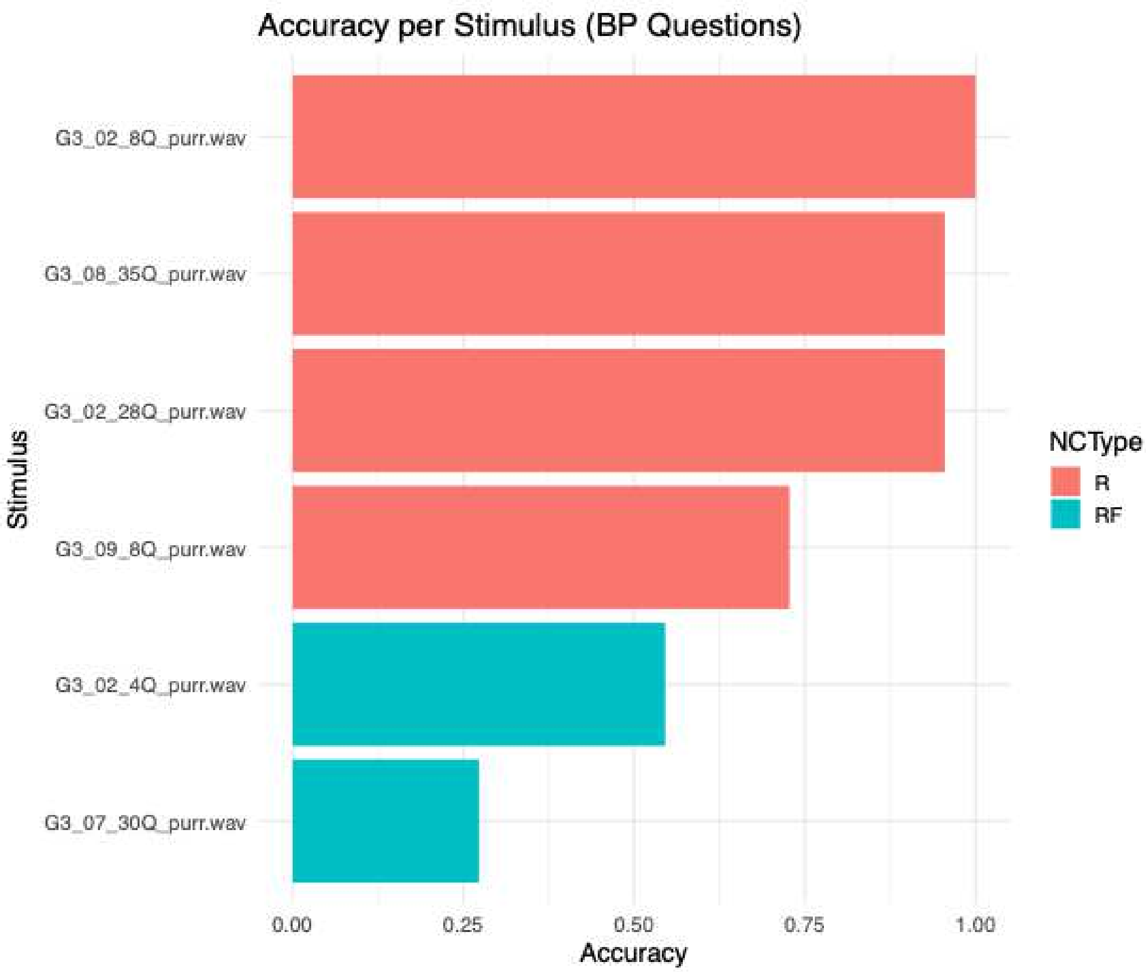
Graph 6: Accuracy summary of sentence type by NC and L1



Source: Own elaboration.

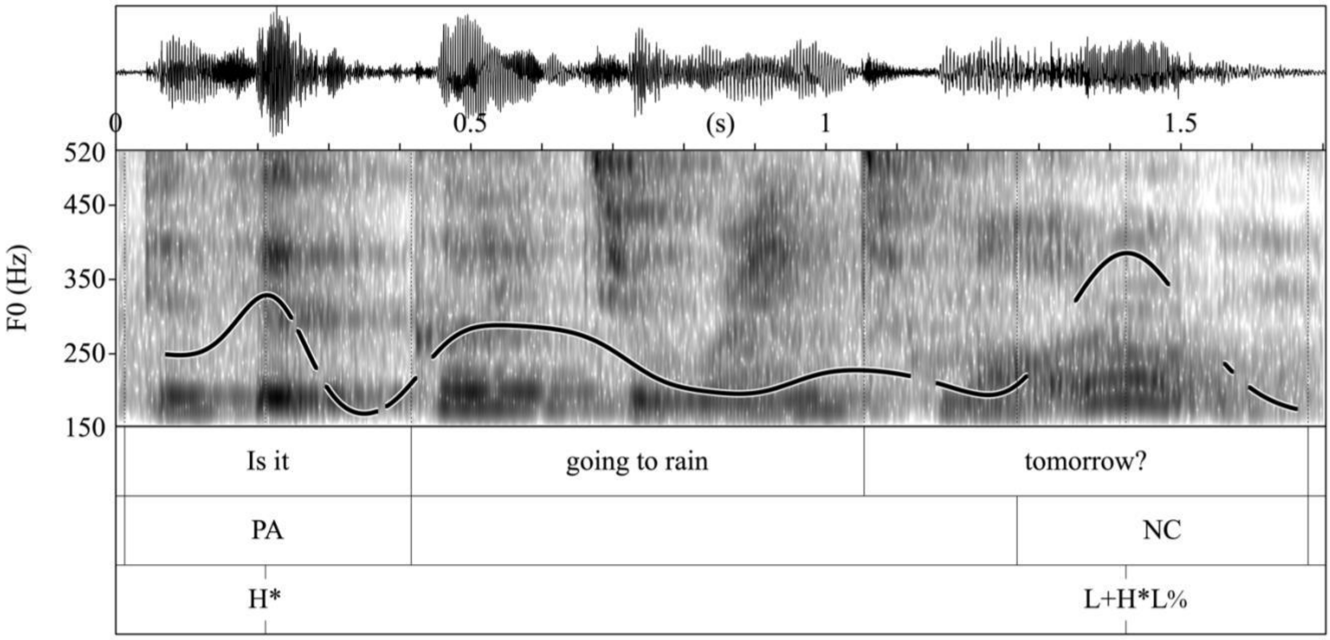
Accuracy was very high (>75%) for F and R nuclear contours and, as expected, RF had very poor accuracy (<50%). Graph 7 shows accuracy of BP questions by NC type, which displays the RF underperforming when compared to the R questions. We have one stimulus just above 50% and the other just over 25% accuracy — both are shown in Figures 17 and 18.

Graph 7: Accuracy of BP questions by NC type



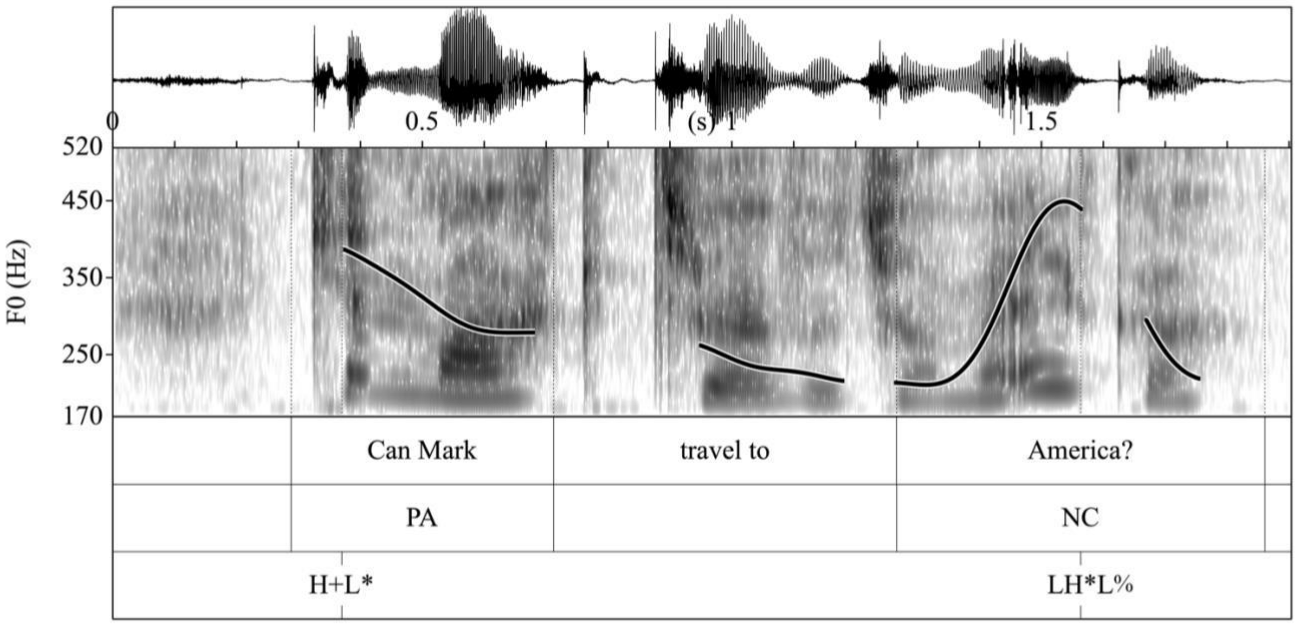
Source: Own elaboration.

Figure 15: Stimulus G3_07_30Q



Source: Own elaboration.

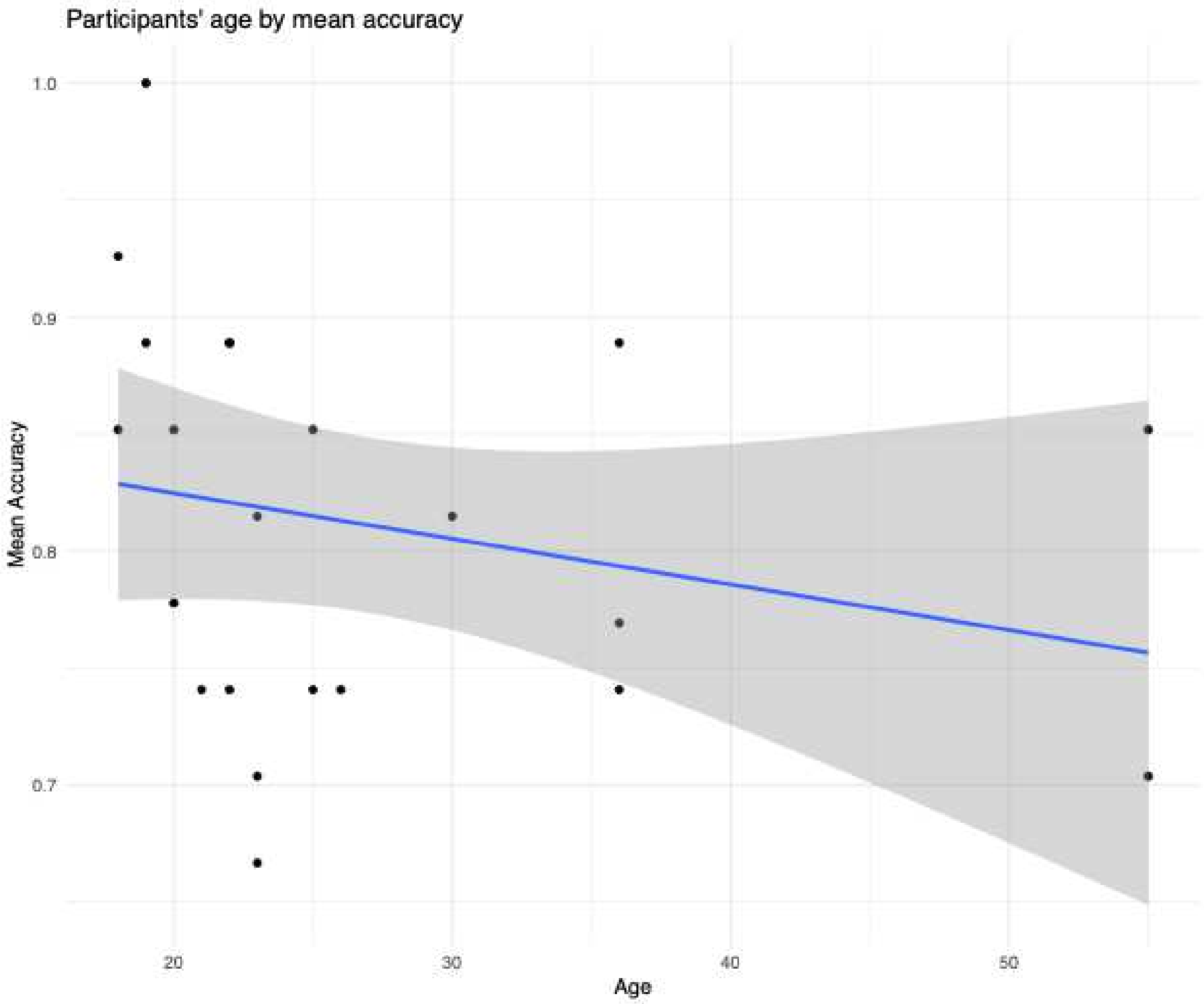
Figure 16: Stimulus G3_02_4Q



Source: Own elaboration.

We also plotted, in Graph 8, age by mean accuracy. Interestingly, we can see a decrease in accuracy with age, although all participants had very high accuracy (>65%).

Graph 8: Participants' age by mean accuracy



Source: Own elaboration.

3.9.8. Results: Statistical Analysis

A mixed-effects model[25] was fitted, including random intercepts for both participants and stimuli, providing the most statistically appropriate structure for the full dataset. The inclusion of stimulus-level random effects significantly improved model fit, indicating that perceptual performance varies substantially across specific items.

[25] Accuracy ~ StimL1 * Sentencetype + (1 | Participant) + (1 | stimulus)

Table 19: Fixed effects of the generalized linear mixed-effects model for overall accuracy

Parameter	Estimate (β)	OR	SE	z-value	p-value
Intercept (NE – Statement)	1.80	6.03	0.39	4.57	<0.001
StimL1BP	0.16	1.17	0.54	0.29	0.78
SentencetypeQuestion	-0.24	0.78	0.53	-0.46	0.65
StimL1BP × SentencetypeQuestion	-0.33	0.72	0.78	-0.42	0.67

Source: Own elaboration.

Once this variability is accounted for, none of the fixed effects reach significance. There is no evidence that BP stimuli are harder than English ones (StimL1BP), nor that questions are globally more difficult than statements after controlling for stimulus-level variation. The non-significant interaction further indicates that BP questions are not disproportionately harder than English questions.

This model reveals that perceptual difficulty is not explained by language (L1) or sentence type in general, but by specific stimuli. This aligns with the finding that the RF contour is responsible for poor accuracy — an effect that disappears when stimuli are pooled under broader categories.

This indicates that English listeners did not find BP stimuli intrinsically more difficult than NE stimuli. The variability observed in raw accuracy was entirely attributable to specific items and, as subsequent analyses showed, to specific nuclear contours.

When nuclear contour type (NCType) is included, the model's explanatory power increases substantially. Table 22 shows the model output[26].

[26] Accuracy ~ StimL1 + Sentencetype + NCType + (1 | Participant)

Table 20: Fixed effects of the mixed-effects model including nuclear contour type.

Parameter	Estimate (β)	OR	SE	z-value	p-value
Intercept (NE – Statement – F)	1.51	4.52	0.20	7.38	<0.001
StimL1BP	0.46	1.58	0.25	1.85	0.06
SentencetypeQuestion	-2.34	0.10	0.38	-6.16	<0.001
NCTypeR	2.40	11.02	0.40	5.97	<0.001

Source: Own elaboration.

We see that two strong predictors emerge:

- Sentence type: Questions are far less accurate than statements (OR = 0.10). This reflects the intrinsic perceptual difficulty of identifying interrogatives from low-pass filtered signals.
- Nuclear contour type: NCType is the strongest effect in the entire dataset. R contours are identified with 11 times higher odds of correct classification than the baseline (F). This may indicate that the contour is driving perceptual accuracy. Stimulus L1 shows only a marginal effect (p = .06), suggesting no reliable disadvantage for BP compared to English.

Therefore, NCType is the primary predictor of perceptual success, confirming that the BP RF contour — not BP stimuli in general — produces systematic misclassification. To isolate the source of perceptual difficulty in BP stimuli, a final model[27] was fitted using only BP-produced questions.

[27] Accuracy ~ NCType + (1|Participant) + (1|stimulus)

Table 21: Fixed effects of the mixed-effects model for BP questions only

Parameter	Estimate (β)	OR	SE	z-value	p-value
Intercept (R)	2.66	14.31	0.68	3.89	<0.001
NCTypeRF	-3.07	0.05	0.96	-3.19	0.0014

Source: Own elaboration.

This model isolates the core perceptual problem by focusing exclusively on BP questions. The contrast coded here compares RF questions against R questions. The intercept reflects performance on R contours only with very high predicted probability (≈ 93% accuracy) and discrimination (OR = 14.3).

On the other hand, the RF penalty is extremely strong (β = -3.07) with low predicted probability (accuracy ≈ 40%) and only 5% of the odds of R being correctly identified (OR = 0.05).

3.9.9. Conclusion

Experiment 2 strongly suggests that nuclear contour shape, rather than speaker L1, is the determining factor in English listeners' perception of sentence modality under low-pass filtering. The distinction between BP and AE English seen in Experiment 1.2 has disappeared, likely pointing out that methodological issues were driving this result rather than perceptual cues. Based Experiment 2 results, we can say that BP speakers who use an English-like rising contour produce stimuli that are perceptually similar enough to those produced by NE speakers.

The persistent perceptual penalty observed in Experiment 1 and 1.2 for BP questions is thus attributable to the presence of the rising–falling contour. This contour is not part of the English inventory for neutral Y/N questions, and its use by BP speakers induces misclassification of interrogatives as statements. Importantly, when BP speakers use rising contours, the perceptual penalty vanishes.

Experiment 2 also illustrates the importance of high-quality stimuli for prosodic perception research. Item-level acoustic anomalies, such as weak harmonic structure after filtering or unexpected mid-sentence noise, significantly distort perceptual outcomes. Once these anomalies are accounted for, the structure of the results becomes clearer and more theoretically interpretable.

3.9.10. Summary of Experiment 2

Experiment 2 provided the most controlled test of prosodic transfer in this dissertation. By eliminating recording artifacts and modeling item-level variability, it showed that perceptual difficulty arises exclusively from specific nuclear contour shapes — particularly the rising–falling pattern — and not from speaker L1. These findings consolidate the evidence from Experiments 1 and 1.2, establishing nuclear contour type as the primary locus of cross-linguistic influence in BP–EN intonational transfer.

3.10. General Conclusions

The three experiments reported in this chapter form a coherent research program on prosodic transfer in BP–EN yes/no questions and statements. Taken together, they show that the main source of perceptual difficulty for native English listeners is not the fact that the speaker is Brazilian, nor the mere presence of a question, but the specific nuclear contour shape used by BP speakers when producing English interrogatives. In particular, the rising–falling (RF) contour characteristic of neutral yes/no questions in southeastern Brazilian Portuguese emerges as the primary locus of cross-linguistic influence in this language pair.

Experiment 1 established the phenomenon and its perceptual consequences under highly constrained conditions. BP learners frequently transferred the RF contour into English questions, and native English listeners systematically misclassified these RF questions as statements when segmental information was removed by low-pass filtering. BP listeners, by contrast, recognized both English rising and BP RF contours as questions, although RF questions elicited longer reaction times, suggesting additional processing demands even in the L1. At the same time, the small sample size and the uncontrolled nature of the remote recordings limited the extent to which item-level differences, sentence type, and speaker L1 could be disentangled.

Experiment 1.2 refined this picture by expanding the listener sample, introducing systematic ToBI annotation, and incorporating mixed-effects modeling. With these methodological improvements, the apparent disadvantage of BP stimuli found in the pilot study no longer held once item-level variability was taken into account. Speaker L1 and sentence type did not reliably predict accuracy. Instead, nuclear contour type emerged as the main predictor of perceptual performance. Rising contours were identified with very high accuracy, falling contours produced intermediate results, and RF contours were associated with a marked drop in accuracy. These findings indicated that the perceptual problem was tied to prosodic form rather than to broad L1 categories. At the same time, uneven distributions of contour types and residual recording variability still prevented a fully controlled assessment of RF effects.

Experiment 2 provided the most rigorous test of these emerging hypotheses. By collecting new production data from BP and North American English speakers under controlled conditions in Toronto and Juiz de Fora, and by carefully selecting and filtering stimuli, this study minimized acoustic artifacts and allowed for a cleaner mapping between contour shape and perceptual outcome. The production results showed that both BP and NE speakers overwhelmingly used rising contours for questions and falling contours for statements, but that BP speakers also produced a small but robust set of RF questions, mirroring the BP pattern described in the literature. In perception, once random variation due to specific items was included in the models, neither L1 nor sentence type alone explained listeners' performance. When nuclear contour type was added, however, its effect was strong and systematic. Rising contours yielded very high predicted accuracy, falling contours were somewhat less accurate, and RF questions showed a substantial penalty, with predicted accuracy dropping to around chance. A final model restricted to BP questions confirmed that the contrast between rising and RF contours fully captured the observed difficulty: when BP speakers used English-like rising contours, their productions were perceived as questions at rates comparable to those of native English speakers; when they used RF contours, interrogativity was often lost.

From a theoretical perspective, these findings support a view of prosodic transfer as contour-specific rather than L1-global. The problem for English listeners is not that the speaker is Brazilian, nor that BP and English differ abstractly in "question intonation", but that BP offers learners an additional nuclear pattern that is not part of the English inventory for neutral yes/no questions. When that pattern is carried over into English, it maps poorly onto the native repertoire of interrogative tunes and is therefore misinterpreted as a statement. At the same time, the results show that BP speakers are capable of producing English-like rising contours which are fully intelligible to English listeners under low-pass conditions.

Methodologically, the chapter also illustrates the importance of linking production and perception in L2 prosody research. Low-pass filtering proved to be a powerful tool for isolating the contribution of intonation to modality judgments, while mixed-effects modeling with random intercepts for participants and stimuli made it possible to separate contour-level effects from item noise. The progression from remote, pandemic-era recordings in Experiment 1 to controlled laboratory data in Experiment 2 shows how incremental improvements in recording quality, annotation, and statistical modeling sharpen theoretical conclusions.

Finally, the three experiments converge on a unified account of where perceptual ambiguity arises in BP–EN prosodic interactions. In this language pair, the critical locus of cross-linguistic influence lies in the nuclear contour of yes/no questions, particularly in the contrast between rising and rising–falling tunes. English listeners have little difficulty with BP productions that use English-like rising contours, but they struggle when confronted with RF questions, which they systematically misclassify as statements under low-pass filtering.

In the following section, we discuss how this research relates to the literature and the future directions.

Closing Thoughts and Future Directions

This dissertation set out to test an underexplored idea in L2 prosody acquisition: if intonation is part of linguistic knowledge, then cross-linguistic influence should be observable not only in production, but also in how listeners perceive and categorize prosodic patterns when segmental information is unavailable.

Across three experiments, the results converge on a conclusion: perceptual difficulty for English listeners was not primarily driven by speaker L1 (BP vs. English) but by the shape of the nuclear contour — specifically, by the presence of the Brazilian Portuguese rise–fall pattern in yes/no questions. Crucially, this perceptual penalty was already visible in the earliest pilot stage (Experiment 1), indicating that the core problem was detectable even before methodological refinements increased experimental control.

In Chapter 2, we reviewed domain-general and domain-specific approaches to SLA, and we highlighted a persistent asymmetry: while these frameworks have generated rich predictions for segmental learning, they have rarely treated prosody as a central object of acquisition. Domain-general perspectives can readily accommodate the notion that learners become sensitive to distributional properties of the input; however, these approaches often under-specify what is represented when learners "learn prosody." In contrast, domain-specific accounts — particularly those grounded in phonological representation — have clearer commitments about inventories, categories, and mappings between form and function, but have historically emphasized segments and syllable structure rather than intonational grammar.

Our empirical results make this gap difficult to ignore. If perception were driven mainly by bottom-up acoustics, then the rise–fall pattern should remain interpretable as interrogative as long as its acoustic salience is preserved. Yet across experiments, English listeners systematically treated BP rise–fall yes/no questions as less interrogative under low-pass conditions, while rising contours yielded high accuracy — regardless of whether the speaker was BP or English. This indicates that listeners were not simply tracking raw pitch movements, they were evaluating an acoustic signal through the lens of category-level prosodic knowledge, i.e., whether a contour is representative as a question-like object in their prosodic/phonological inventory.

Methodologically, low-pass filtering was essential because it reduced segmental cues while preserving global prosodic structure (especially F0 movement).

Under these conditions, native English listeners performed well when the contour aligned with rising interrogativity, and they failed when the contour patterned as rise–fall. In Experiment 2, the statistical models made this particularly clear: once nuclear contour type was included, the explanatory power of the model increased substantially, and speaker L1 ceased to be a reliable predictor.

This pattern supports an interpretation in which listeners have fine-grained access to F0 information, however perception is constrained by phonological expectations about the prosodic inventory. The rise–fall question contour in BP appears to be treated as unfamiliar or unexpected to the interrogative set for native English listeners, leading to a robust interrogativity loss. The same logic also explains why the perceptual penalty vanishes when BP speakers produce rising question contours: when the prosodic object matches the listener's expectations for an English yes/no question, the signal is readily categorized as interrogative. In this sense, the results are not only about acoustic discriminability; they are about the categorization of prosodic form-function pairings.

Mennen's L2 Intonation Learning Theory (LILt) was developed to explain why L2 learners often struggle with intonation even when segments are intelligible. LILt is particularly useful here because it decomposes cross-language differences into dimensions that generate concrete hypotheses about where deviation is likely to arise: the systemic dimension (inventory differences), the realisational dimension (phonetic implementation), the semantic dimension (form–function mapping), and the frequency dimension (distributional dominance of patterns). In Chapter 2, we already argued that LILt provides a prosody-specific framework that earlier SLA theories lack; the present findings allow us to connect those dimensions to a coherent empirical line of thought.

First, our production-driven motivation aligns most directly with the systemic and semantic dimensions. Brazilian Portuguese and North American English differ in how neutral yes/no questions are typically implemented. When BP EN-L2 speakers produce English yes/no questions with BP-like rise–fall contours, they are plausibly drawing on an L1 prosodic inventory and an L1 mapping between interrogativity and nuclear contour shape. Under LILt, this is exactly the kind of context in which transfer is expected: the learner recruits an available L1 option to express an L2 function, producing a contour that is well-formed in BP but not target-like in English.

Second, our perception findings are congruent with LILt's implications by showing that these dimensions are not only relevant to learners' output, but also to listeners' interpretation. The persistent rise–fall penalty suggests that for English listeners, the rise–fall pattern does not behave like an interrogative category. This supports a strong version of the claim that prosodic learning involves building (or revising) a prosodic inventory and the mappings that license it.

Third, the frequency dimension provides a plausible route for explaining why the rise–fall contour is so difficult for English listeners in the first place. If a contour is rare or non-dominant in the listener's experience as an interrogative, then under limited information conditions (low-pass filtering) it is reasonable that the system defaults to the most frequent, most robust interrogativity cue in the listener's grammar — yielding systematic misclassification. The fact that the same contour can still be interpreted in other conditions (e.g., with richer segmental context, discourse, and pragmatic information) is compatible with the idea that frequency and contextual redundancy jointly determine how strongly the system commits to a category.

While LILt offers a principled way to theorize production and the loci of transfer, our perception results raise a deeper question that LILt only indirectly addresses: if an L2 learner must acquire the target prosodic inventory, how is that prosodic knowledge stored, stabilized, and accessed online?

One promising path is to treat prosodic categorization as supported by memory systems that bridge the acoustic signal and phonological interpretation. If we adopt Baddeley's (2009) conception of memory, which distinguishes systems primarily by duration and treats memory as a set of structures and processes operating over those structures, storage is seen as temporary and constitutes one component within a broader cognitive architecture. To reach the notion of *phonological memory*, we draw on the phonological loop introduced in Baddeley and Hitch's (1974) multicomponent model of working memory. Working memory is conceived as a limited-capacity system that simultaneously stores and processes information during complex cognition. It comprises (i) a central executive responsible for attentional control (selection, manipulation, shifting, and interfacing with long-term memory) and (ii) two temporary storage systems: the visuospatial sketchpad (visual/spatial information) and the phonological loop, which temporarily maintains phonological information for a few seconds via a storage buffer plus a rehearsal/refreshing mechanism. Baddeley (2009) further proposes that the phonological loop supports language acquisition (L1 and L2), particularly vocabulary learning, and accounts for a wide range of verbal-memory phenomena.

Within this framework, we define phonological memory — often discussed under the label phonological short-term memory (Mota, 2015) — as the ability to hold phonological material for roughly two seconds and to generate phonological representations (e.g., phonetic categories) that can later be consolidated into long-term memory (Kogan, 2022). However, what counts as "phonological material" remains underspecified, and recent operationalizations have focused overwhelmingly on segmental content (phonemes, syllables, non-speech tones) and on short-term storage, with comparatively little work addressing long-term representations. This creates a gap for prosody: although prosody is grounded in suprasegmental relations across syllabic units and can be characterized through physical correlates (F0, duration, intensity) and perceptual correlates (pitch, loudness, perceived duration, voice quality) (Barbosa, 2019), its relationship to phonological memory has rarely been examined. We can hypothesize that such prosodic information is stored — at least in part — in a short- and long-term phonological/prosodic memory system, and that these representations must be accessible both for native speakers in speech tasks and for L2 learners as they acquire the target language's prosodic inventory.

The working memory literature suggests that prosodic discrimination, especially when segmental information is unfamiliar or reduced, draws on both storage and processing resources. In a sentence-level prosodic discrimination task conducted in a language unknown to the children (French), Stepanov, Kodrič and Stateva (2020) report that children's performance correlates with working memory measures (forward/backward digit span and non-word repetition), and argue that both the storage and processing components of working memory are implicated in prosodic discrimination, rather than the task being purely perceptual. More broadly, aptitude research has also highlighted working memory and phonological coding as central to language learning capacity, with non-word repetition commonly treated as an index of the relevant memory resources that support the encoding and manipulation of novel sound patterns (Chan, Skehan and Gong, 2011). Together, these findings motivate the hypothesis that successful L2 prosodic learning requires not only sensitivity to F0, but also memory-supported mechanisms that stabilize and provide access to prosodic representations during online categorization. Kogan (2022) further distinguishes between phonological short-term memory and acoustic memory, arguing that the latter may store pre-categorical acoustic detail that is especially relevant when listeners must encode and compare fine-grained auditory patterns. In the context of intonation, this distinction is attractive: nuclear contours are dynamic patterns over time and perceiving them under degraded input plausibly depends on the fidelity with which pitch trajectories are temporarily retained and compared.

This dissertation did not directly measure working memory, acoustic memory, or neural responses, so any strong mechanistic claim would be premature. Still, the pattern of results motivates a promising research agenda: if prosodic categories constrain perception, then learners must develop memory-supported mappings between stored contour representations and communicative functions. This leads to the hypothesis of a prosodic memory component — either as an extension of phonological short-term memory or as an interaction between pre-categorical acoustic storage and phonological categorization — that would be specifically relevant for acquiring intonational inventories.

Neurocognitive work reinforces the plausibility of a specialized or systematically organized substrate for prosodic processing. In a broad review of the brain basis of language, Friederici (2011) notes that sentence-level prosody recruits a right-hemisphere temporo-frontal network. Lesion and neuroimaging evidence likewise suggests that prosody is not localized to a single "prosody module" but instead relies on distributed systems whose contribution varies with the modality (linguistic vs. affective), the level of representation (e.g., lexical/phonemic vs. sentential), and the acoustic cues emphasized by the task (Baum and Pell, 1999; Wildgruber et al., 2006; Belyk and Brown, 2014).

From the standpoint of this dissertation, the key point is conceptual: if prosody is grounded in neurocognitive systems that encode temporally extended auditory patterns and link them to linguistic interpretation (Friederici, 2011), then L2 prosody learning cannot be reduced to learning "pitch height" or "melody" as surface features. Instead, it plausibly involves developing stable neural/representational routines that support (i) encoding pitch trajectories over time, (ii) categorizing them relative to a language-specific inventory, and (iii) mapping them to communicative functions under uncertainty. Converging neuroimaging and brain-stimulation evidence further supports this view by arguing that prosody perception can proceed via dual routes — dorsal and ventral pathways — within right-hemisphere fronto-temporo-parietal networks (Sammler et al., 2015), consistent with an architecture integrating auditory analysis with higher-level interpretive routines (Wildgruber et al., 2006; Friederici, 2011).

This study also aligns with and expand on Derwing and Munro's (1995) work on the perceptual salience of prosody in L2 speech. While earlier studies have shown that listeners can reliably distinguish between rising and falling intonational contours when segmental cues are removed (Grabe et al. 2003; Radu et al. 2018; Passarella dos Reis et al. 2016; Buzan et al. 2022), the present results suggest that this capacity may be compromised when the pitch contours do not align with L1 expectations. Such mismatches appear to challenge listeners' interpretative strategies, revealing that prosodic transfer can diminish comprehensibility when segmental information is absent. This reinforces the view advanced by Derwing and Munro (1995) that prosodic deviations — especially at the nuclear contour level — can have a disproportionate impact on listener comprehension, often outweighing the role of segmental accuracy. In particular, our data echo their observation that certain types of intonational divergence led to reduced intelligibility, not due to accent per se, but because they increase listener effort and processing difficulty (Derwing and Munro 1995; Munro 1995).

In this sense, instructionally, the findings motivate the argument of that prosodic instruction should be more prominently featured in English teaching for L1-BP speakers. First, learners need to be made aware that English yes/no questions are typically realized with rising nuclear contours in North America and that BP-like rise–fall contours in South America can change how an utterance is categorized. Second, learners should be trained to perceive and produce these contrasts under variable conditions, since real-world listening often resembles "degraded" perception (noise, distance, low quality audio, divided attention). Work on L2 prosody pedagogy argues that prosody instruction can be effective, particularly when it is systematic, includes perception–production links, and targets functions that matter for comprehensibility rather than only aesthetic "native-likeness".

In concrete terms, the current results suggest that teaching should prioritize (i) nuclear contour inventories for high-functional-load sentence types (polar questions vs. statements), (ii) form–function mappings (what contours *do* in interaction), and (iii) feedback that helps learners reweight cues away from L1 defaults. Given the robustness of the rise–fall penalty in perception, purely implicit exposure is unlikely to guarantee change for many learners; explicit comparison between BP and English contours and guided practice should be treated as essential rather than supplementary.

If the rise–fall contour is systematically “misread” by English listeners as non-interrogative under reduced cues, then L2 learners who transfer BP question contours into English are at risk of producing speech that is segmentally intelligible yet pragmatically unsuccessful — precisely the kind of mismatch that can undermine comprehensibility and interactional success. (Munro and Derwing, 1995; Derwing and Munro, 1997; Derwing and Munro, 2005). This aligns with broader pronunciation research emphasizing that suprasegmentals can disproportionately affect perceived naturalness and communicative success (Hahn, 2004; Kang, 2010).

Instructionally, the findings motivate explicit prosody teaching in at least two senses. First, learners need awareness that intonation patterns systematically affect how utterances are interpreted (Derwing and Munro, 2005; Chun and Levis, 2021). Second, learners should be trained to perceive and produce these contrasts under variable conditions, since real-world listening frequently occurs under adverse circumstances (noise, transmission limitations, and divided attention), which reshape cue weighting and speech recognition demands (Mattys et al., 2012).

Future work can extend the present findings along three complementary axes. First, cross-varietal BP research is needed. Because BP intonation is not uniform (Castelo et al., 2018; Castelo and Frota, 2016), studies should test varieties with predominantly rising yes/no question contours with speakers from Northeast Brazil (João Pessoa, Aracaju and Salvador) to determine whether the perceptual penalty is contour-dependent across BP varieties in the same way. This would refine the claim that the issue is not BP per se, but specific nuclear contours associated with regional systems.

Second, listener experience could be brought into the design more directly. Testing English L1 listeners with high proficiency in BP (or with extensive exposure to BP) can reveal whether increased familiarity attenuates the rise–fall penalty, and whether the perceptual system can effectively add (or re-map) a rise–fall contour as interrogative category through experience. In the same vein, BP heritage speakers in English-dominant contexts are a theoretically powerful group: they allow us to test whether early exposure to BP intonation changes the contour inventory and its mappings even when English becomes dominant, providing a bridge between L1 acquisition, heritage phonology, and L2 perception.

Third, the cognitive–neural speculation raised above could become an empirical program. Adding measures of working memory, acoustic memory, and (where feasible) neurophysiological responses would allow future studies to test whether individuals with greater auditory/working memory resources are more resilient to contour "mismatch" under low-pass conditions, and whether training reshapes the encoding and categorization of nuclear contours over time.

In addition, future work should treat response times as a potentially informative window into prosodic processing. RTs can reveal processing cost even when accuracy is near ceiling (e.g., slower categorization for "mismatching" contours), but they are inherently composite measures, reflecting perceptual encoding and decision processes. In the pilot experiment, RTs were collected precisely with this motivation, but time constraints and software/platform limitations prevented implementing reliable RT logging (and latency control) in the subsequent experiments. Still, incorporating RTs in future designs remains a valuable path to pursue, especially if combined with accuracy and individual differences.

Finally, we can look at the rapidly improving speech technology. If human listeners treat certain contour types as outside their interrogative inventory, then AI voice pattern recognition and synthesis systems that aim for naturalness and communicative reliability must encode not only acoustic fidelity but also language- and variety-specific prosodic inventories. The present results therefore motivate evaluation benchmarks for Text to Speech and Automatic Speech Recognition (TTS/ASR) that include intonational well-formedness and function recognition across varieties, rather than focusing exclusively on segmental accuracy.

This dissertation makes three core contributions. Empirically, it provides converging evidence — across a pilot, an expanded replication, and a controlled redesign — that perceptual difficulty in BP–English prosodic transfer is driven primarily by nuclear contour type, with a particularly strong penalty for rise–fall yes/no question contours under low-pass conditions. Theoretically, it strengthens the case that prosody must be explicitly incorporated into models of L2 speech acquisition in both production and perception: the relevant generalizations are not adequately captured by approaches that treat prosody as secondary or purely acoustic. By linking the findings to LILt's dimensions, the dissertation also shows how a prosody-specific framework can generate interpretable predictions about where L2 deviation will arise and why it matters for communication.

Methodologically, the work illustrates how careful stimulus control, annotation, exclusion criteria, and modeling decisions sharpen theoretical interpretation —

demonstrating that prosodic transfer cannot be responsibly reduced to speaker L1 labels when intonational categories cut across speakers and varieties.

Finally, in terms of resources and broader impact, the data and analyses are intended to contribute to ongoing efforts to document and model Portuguese prosody, including the Interactive Atlas of the Prosody of Portuguese, by adding updated evidence on BP nuclear contours and their perceptual consequences for non-native listeners.

REFERENCES

AL-HORAIS, N. The Minimalist Program and its new insight to the concept of Universal Grammar. **Journal of Universal Language**, v. 14, n. 2, p. 79–112, 2013.

ANDERSON-HSIEH, J.; JOHNSON, R.; KOEHLER, K. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. **Language Learning**, v. 42, n. 4, p. 529–555, 1992.

ARCHIBALD, J. A formal model of learning L2 prosodic phonology. **Second Language Research**, v. 10, n. 3, p. 215–240, 1994.

______. Second language phonology, phonetics, and typology. **Studies in Second Language Acquisition**, v. 20, n. 2, p. 189–211, 1998. DOI: 10.1017/S0272263198002046

ARVANITI, A. The representation of intonation. In: HARDCASTLE, William J.; LAVER, John; GIBBON, Fiona (ed.). **The handbook of phonetic sciences**. 2. ed. Oxford: Wiley-Blackwell, 2010. p. 757–780.

______. Analytical challenges in the study of intonation: categorization and tonal variation. *Laboratory Phonology*, Berlin, v. 7, n. 1, p. 1–31, 2016.

BARBOSA, P. A. **Incursões em torno do ritmo da fala**. Campinas: Pontes, 2006.

______. **Prosody and language contact**. Cambridge: Cambridge University Press, 2019.

______. **Prosódia**. 1. ed. São Paulo: Parábola, 2019. (Linguística para o ensino superior, 2).

BARTELS, C. **The intonation of English statements and questions: a compositional interpretation**. New York: Garland, 1999.

BAUM, S. R.; PELL, M. D. The neural bases of prosody: insights from lesion studies and neuroimaging. **Aphasiology**, v. 13, n. 8, p. 581–608, 1999. DOI: 10.1080/026870399401957.

BECKMAN, M. E.; PIERREHUMBERT, J. B. Intonational structure in Japanese and English. **Phonology Yearbook**, v. 3, p. 255–309, 1986.

BECKMAN, M. E. The parsing of prosody. **Language and Cognitive Processes**, v. 11, n. 1–2, p. 17–67, 1996.

BECKMAN, M. E.; AYERS ELAM, G. **Guidelines for ToBI labelling (version 3.0)**. Columbus: Ohio State University, 1997.

BELYK, M.; BROWN, S. Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. **Social Cognitive and Affective Neuroscience**, v. 9, n. 9, p. 1395–1403, 2014.

BERTONCINI, J. Perception of speech sounds in early infancy. **Current Psychology of Cognition**, v. 13, n. 5, p. 515–536, 1993.

BEST, C. T. A direct realist perspective on cross-language speech perception. In: STRANGE, W., ed. **Speech perception and linguistic experience**. Timonium: York Press, 1995. p. 171–204.

BEST, C. T.; TYLER, M. D. Nonnative and second-language speech perception: commonalities and complementarities. In: BOHN, O.-S.; MUNRO, M. J., eds. **Language experience in second language speech learning**. Amsterdam: John Benjamins, 2007. p. 13–34.

BOLINGER, D. **Intonation and its parts: melody in spoken English**. Stanford: Stanford University Press, 1986.

BOŠKOVIĆ, Ž. Principles and parameters theory and Minimalism. In: DEN DIKKEN, M., ed. **The Cambridge handbook of generative syntax**. Cambridge: Cambridge University Press, 2013. p. 233–321.

BÜRING, D. **Intonation and meaning**. Oxford: Oxford University Press, 2016.

BUZAN, T.; NAME, C.; SOSA, J. Intonational interference in English-L2 Brazilian speakers: production and perception. In: [S.l.: s.n.], 2022. p. 828–831. Available at: https://www.isca-speech.org/archive/speechprosody_2022/buzan22_speechprosody.html. Accessed on: Oct. 4, 2023.

BUZAN, T. N.; NAME, C.; SOSA, J.; COLANTONI, L. Native English speakers' perception of questions and statements produced by L1 Brazilian Portuguese and L1 English speakers. 2026. Manuscrito submetido para publicação.

BYBEE, J. **Language, usage and cognition**. Cambridge: Cambridge University Press, 2011.

BYERS-HEINLEIN, K.; BURNS, T. C.; WERKER, J. F. The roots of bilingualism in newborns. **Psychological Science**, v. 21, n. 3, p. 343–348, 2010. DOI: 10.1177/0956797609360758.

BYRNES, H. Review of task-based language learning and teaching. **The Modern Language Journal**, v. 89, p. 297–298, 2005.

CASTELO, J.; FROTA, S. The intonational phonology of Brazilian Portuguese. In: FROTA, S.; PRIETO, P., eds. **Intonation in Romance**. Oxford: Oxford University Press, 2016. p. 118–151.

CASTELO, J.; FONSECA, A.; COLLISCHONN, G.; HENRIQUE, P.; FROTA, S. The perception of yes-no questions across varieties of Brazilian Portuguese. **Filologia e Linguística Portuguesa**, v. 20, n. esp., p. 11–25, 2018.

CHANG, C. B. Rapid and multifaceted effects of second-language learning on first-language speech production. **Journal of Phonetics**, v. 40, n. 2, p. 249–268, 2012. DOI: 10.1016/j.wocn.2011.10.007.

CHOMSKY, N. A minimalist program for linguistic theory. In: HALE, K.; KEYSER, S. J., eds. **The view from Building 20: essays in honor of Sylvain Bromberger**. Cambridge, MA: MIT Press, 1993. p. 1–52.

______. **The minimalist program**. Cambridge, MA: MIT Press, 1995.

CHRISTOPHE, A.; NESPOR, M.; GUASTI, M. T.; VAN OOYEN, B. Prosodic structure and syntactic acquisition: the case of the head–direction parameter. **Developmental Science**, v. 1, n. 1, p. 1–14, 1997. DOI: 10.1111/1467-7687.00001.

CHRISTOPHE, A.; MILLOTTE, S.; BERNAL, S.; LIDZ, J. Bootstrapping lexical and syntactic acquisition. **Language and Speech**, v. 51, n. 1–2, p. 61–75, 2008. DOI: 10.1177/00238309080510010501.

CHUN, D. M.; LEVIS, J. M. Prosody in second language teaching: methodologies and effectiveness. In: GUSSENHOVEN, C.; CHEN, A., eds. **The Oxford Handbook of Language Prosody**. Oxford: Oxford University Press, 2020. Edição online: 10 fev. 2021. DOI: 10.1093/oxfordhb/9780198832232.013.41.

COLANTONI, L.; STEELE, J.; ESCUDERO NEYRA, P. R. **Second language speech: theory and practice**. Cambridge: Cambridge University Press, 2015.

COOPER, R. P.; ASLIN, R. N. Preference for infant-directed speech in the first month after birth. **Child Development**, v. 61, n. 5, p. 1584–1595, 1990. DOI: 10.2307/1130766.

CRUZ, M.; CRESPO-SENDRA, V.; CASTELO, J.; FROTA, S. Asking questions across Portuguese varieties. In: CRUZ, Marisa; FROTA, Sónia (org.). **Prosodic variation (with)in languages: intonation, phrasing and segments**. Sheffield: Equinox Publishing, p. 36–70, 2022.

NARAYANAN, D. Z.; TAKAHASHI, D. Y.; KELLY, L. M.; HLAVATY, S. I.; HUANG, J.; GHAZANFAR, A. A. Prenatal development of neonatal vocalizations. **eLife**, v. 11, e78485, 2022. DOI: 10.7554/eLife.78485.

CZEKE, N.; ZAHNER, K.; RIMPLER, J.; BRAUN, B.; FROTA, S. German infants do not discriminate Portuguese rising vs. falling contours. **Pôster** apresentado em: Workshop on Infant Language Development (WILD), out. 2019. DOI: 10.15496/publikation-27450.

DE BOT, K. Dynamic systems theory and SLA. **International Journal of Applied Linguistics**, v. 17, n. 3, p. 421–443, 2007.

DE LEEUW, E.; MENNEN, I.; SCOBBIE, J. M. Singing a different tune in your native language: first language attrition of prosody. **International Journal of Bilingualism**, v. 16, n. 1, p. 101–116, 2012.

DECASPER, A. J.; FIFER, W. P. Of human bonding: newborns prefer their mothers' voices. **Science**, v. 208, n. 4448, p. 1174–1176, 1980. DOI: 10.1126/science.7375928.

DECASPER, A. J.; SPENCE, M. J. Prenatal maternal speech influences newborns' perception of speech sounds. **Infant Behavior and Development**, v. 9, n. 2, p. 133–150, 1986. DOI: 10.1016/0163-6383(86)90025-1.

DELAIS-ROUSSARIE, E.; AVANZI, M.; HERMENT, S. (org.). **Prosody and language in contact**. Berlin: Springer, 2015.

DERWING, T. M.; MUNRO, M. J. Accent, intelligibility, and comprehensibility: evidence from four L1s. **Studies in Second Language Acquisition**, v. 19, n. 1, p. 1–16, 1997. DOI: 10.1017/S0272263197001010.

______. Second language accent and pronunciation teaching: a research-based approach. **TESOL Quarterly**, v. 39, n. 3, p. 379–397, 2005. DOI: 10.2307/3588486.

DIAS, E. C. O. **Declarativas e interrogativas totais no espanhol L1 e L2 falado em Bogotá: uma contribuição para estudos prosódicos**. 2015. 273 f. Tese (Doutorado em Linguística) – Universidade Federal de Santa Catarina, Florianópolis, 2015.

DRAGOJEVIC, M.; GILES, H.; BECK, A. C.; TATUM, N. T. The fluency principle: Why foreign accent strength negatively biases language attitudes. **Communication Monographs**, v. 84, n. 3, p. 385–405, 2017. DOI: 10.1080/03637751.2017.1322213.

EIMAS, P. D.; SIQUELAND, E. R.; JUSCZYK, P. W.; VIGORITO, J. Speech perception in infants. **Science**, v. 171, n. 3968, p. 303–306, 1971. DOI: 10.1126/science.171.3968.303.

ELLIS, N. C. Constructions, chunking, and connectionism. **Language Learning**, v. 53, p. 1–64, 2003.

ELLIS, N. C.; WULFF, S. Usage-based approaches to SLA. In: VANPATTEN, Bill; WILLIAMS, Jessica (org.). **Theories in second language acquisition: an introduction**. 2. ed. New York; London: Routledge, 2015. p. 63–82.

ELLIS, N. C.; SCHMIDT, R. Morphology and longer distance dependencies: laboratory research illuminating the A in SLA. **Studies in Second Language Acquisition**, v. 19, n. 2, p. 145–171, 1997. Disponível em: http://www.jstor.org/stable/44488681. Acesso em: 3 jan. 2026.

FERNALD, A. Four-month-old infants prefer motherese. **Infant Behavior and Development**, v. 8, p. 181–195, 1985.

______. Intonation and communicative intent in mothers' speech to infants: is the melody the message? **Child Development**, v. 60, n. 6, p. 1497–1510, 1989. DOI: 10.2307/1130938.

FILIPPA, M.; TISSOT, H.; MANCINELLI, T.; et al. Maternal and paternal infant-directed speech is modulated by child age. **Scientific Reports**, v. 15, art. 13624, 2025. DOI: 10.1038/s41598-025-98047-3.

FLEGE, J. E. Second-language speech learning. In: STRANGE, W. (ed.). **Speech perception and linguistic experience**. Timonium: York Press, 1995. p. 233–277.

FOUCART, A.; COSTA, A.; MORIS-FERNANDEZ, L.; HARTSUIKER, R. Foreignness or processing fluency? On understanding the negative bias toward foreign-accented speakers. **Language Learning**, v. 70, n. 4, p. 974–1016, 2020. DOI: 10.1111/lang.12413.

FRIEDERICI, A. D. The brain basis of language processing: from structure to function. **Physiological Reviews**, v. 91, n. 4, p. 1357–1392, 2011. DOI: 10.1152/physrev.00006.2011.

FROTA, S. (coord.). **Interactive Atlas of the Prosody of Portuguese Project (InAPoP)**. Web platform. University of Lisbon, 2012–2015.

FROTA, S.; CRUZ, M.; SVARTMAN, F.; COLLISCHONN, G.; FONSECA, A.; SERRA, C.; OLIVEIRA, P.; VIGÁRIO, M. Intonational variation in Portuguese: European and Brazilian varieties. **Phonology**, v. 32, n. 2, p. 231–274, 2015.

FROTA, S.; MORAES, J. Intonation in European and Brazilian Portuguese. In: WETZELS, Willem Leo; COSTA, João; MENUZZI, Sérgio (org.). **The handbook of Portuguese linguistics**. Hoboken, NJ: John Wiley & Sons, 2016. p. 141–166. DOI: 10.1002/9781118791844.ch9.

FROTA, S.; BUTLER, J.; VIGÁRIO, M. Infants' perception of intonation: Is it a statement or a question? **Infancy**, v. 19, n. 2, p. 194–213, 2014. DOI: 10.1111/infa.12037

FROTA, S.; PRIETO, P. (ed.). **Intonation in Romance**. Oxford: Oxford University Press, 2015.

GASS, S. M.; MACKEY, A. Input, interaction, and output in second language acquisition. In: VANPATTEN, B.; WILLIAMS, J. (eds.). **Theories in Second Language Acquisition: An Introduction**. New York: Routledge, 2015. p. [s.p.].

GLEITMAN, L. R.; WANNER, E. Language acquisition: the state of the art. In: WANNER, E.; GLEITMAN, L. R. (eds.). **Language Acquisition: The State of the Art**. Cambridge: Cambridge University Press, 1982. p. 3–48.

GOAD, H.; WHITE, L. Ultimate attainment in interlanguage grammars: a prosodic approach. **Second Language Research**, v. 22, p. 243–268, 2006. DOI: 10.1191/0267658306sr268oa

GRABE, E.; KOCHANSKI, G.; COLEMAN, J. Quantitative modelling of intonational variation. In: SOLÉ, M. J.; RECASENS, D.; ROMERO, J. (eds.). **Proceedings of the 15th International Congress of Phonetic Sciences**. Barcelona: Universitat Autònoma de Barcelona, 2003. p. 2477–2480.

GRAHAM, C.; POST, B. Second language acquisition of intonation: the case of English learners of Japanese and Spanish. **Journal of Phonetics**, v. 66, p. 1–17, 2018. https://doi.org/10.1016/j.wocn.2017.09.005.

GUSSENHOVEN, C. **The Phonology of Tone and Intonation**. Cambridge: Cambridge University Press, 2004.

GUT, U.; PILLAI, S. Prosodic marking of information structure by Malaysian speakers of English. In: DELAIS-ROUSSARIE, E.; AVANZI, M.; HERMENT, S. (org.). **Prosody and Language in Contact**. Berlin: Springer, 2015. p. 51–71.

HAHN, L. D. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. **TESOL Quarterly**, v. 38, n. 2, p. 201–223, 2004. DOI: 10.2307/3588378.

HAN, M.; DE JONG, N. H.; KAGER, R. Relating the prosody of infant-directed speech to children's vocabulary size. **Journal of Child Language**, v. 51, n. 1, p. 217–233, 2024. DOI: 10.1017/S0305000923000041.

HANSEN EDWARDS, J. G.; ZAMPINI, M. L. (eds.). **Phonology and Second Language Acquisition**. Amsterdam: John Benjamins, 2008.

HARDISON, D. M. Generalization of computer-assisted prosody training: quantitative and qualitative findings. **Language Learning & Technology**, v. 8, n. 1, p. 34–52, 2004. DOI: 10.64152/10125/25228.

HASKELL, T. R.; MACDONALD, M. C.; SEIDENBERG, M. S. Language learning and innateness. **Cognitive Psychology**, v. 47, p. 119–163, 2003.

HATCH, E. M. Discourse analysis and second language acquisition. In: HATCH, E. M. (ed.). **Second Language Acquisition: A Book of Readings**. Rowley: Newbury House, 1978.

HAYES, B. Metrical stress theory: principles and case studies. Chicago: **University of Chicago Press**, 1995

HEDBERG, N.; SOSA, J.; GÖRGÜLÜ, E. The meaning of intonation in yes/no questions in American English: a corpus study. **Corpus Linguistics and Linguistic Theory**, v. 13, n. 2, p. 321–368, 2014.

JUN, Sun-Ah (ed.). **Prosodic typology: the phonology of intonation and phrasing**. Oxford: Oxford University Press, 2005.

KAINADA, E.; LENGERIS, A. Native language influence on the production of second language prosody: the case of Greek learners of English. **Journal of the International Phonetic Association**, v. 45, n. 3, p. 269–287, 2015.

KALAND, C. Contour clustering: a field-data-driven approach for documenting and analysing prototypical f0 contours. **Journal of the International Phonetic Association**, v. 53, n. 1, p. 159–188, 2023. DOI: 10.1017/S0025100321000049.

______. **Contour Clustering** (Version 2025-03) [Computer software]. 2025. Disponível em: https://constantijnkaland.github.io/contourclustering/. Acesso em: 3 jan. 2026.

KANG, O. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. **System**, v. 38, n. 2, p. 301–315, 2010. DOI: 10.1016/j.system.2010.01.005.

KELLY, N. Change across time in L2 intonation vs. segments: a longitudinal study of the English of Ole Gunnar Solskjaer. **Languages**, v. 7, n. 3, p. 210, 2022.

KEMPE, V.; MACWHINNEY, B. The acquisition of case marking by adult learners of Russian and German. **Studies in Second Language Acquisition**, v. 20, n. 4, p. 543–587, 1998. DOI: 10.1017/S0272263198004045.

KIM, J.; ARNHOLD, A. Phonetic and phonological aspects of prosodic focus marking in Canadian English. **Laboratory Phonology: Journal of the Association for Laboratory Phonology**, v. 15, n. 1, p. 1–38, 2024. DOI: 10.16995/labphon.9316.

KIREVA, E.; GABRIEL, C. Rhythmic properties of a contact variety: comparing read and semi-spontaneous speech in Argentinean Porteño Spanish. In: DELAIS-ROUSSARIE, E.; AVANZI, M.; HERMENT, S. (Org.). **Prosody and Language in Contact**. Berlin: Springer, 2015. p. 149–168.

KOGAN, V. The role of acoustic memory and phonological short-term memory in non-native perception: which memory matters more and when? In: KKESE, E. (ed.). **Speech Perception and Production in L2**. Newcastle upon Tyne: Cambridge Scholars Publishing, 2022. p. 236–266.

KRASHEN, S. D. Some issues related to the Monitor Model. In: BROWN, H. D.; YORIO, C.; CRYMES, R. (Eds.). **On TESOL'77 – Teaching and Learning English as a Second Language: Trends in Research and Practice**. Washington: TESOL, 1997. p. [s.p.].

______. The Monitor Model for second language acquisition. In: GINGRAS, R. C. (Ed.). **Second Language Acquisition & Foreign Language Teaching**. Washington: Center for Applied Linguistics, 1978. p. [s.p.].

______. **Principles and Practice in Second Language Acquisition**. Oxford: Pergamon Press, 1982.

______. **The Input Hypothesis: Issues and Implications**. London: Longman, 1985.

______. Applying the Comprehension Hypothesis: Some Suggestions. Paper presented at: **13th International Symposium and Book Fair on English Teaching (ETA-ROC)**, Taipei, 13 Nov. 2004. Available at: https://www.sdkrashen.com/content/articles/eta_paper.pdf. Accessed on: 2 Jan. 2026.

______. **Second Language Acquisition: Theory, Applications, and Some Conjectures**. Cambridge: Cambridge University Press, 2013.

KUHL, P. K. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. **Perception & Psychophysics**, v. 50, n. 2, p. 93–107, 1991. DOI: 10.3758/BF03212211.

______. Infants’ perception and representation of speech: development of a new theory. In: DE BOYSSON-BARDIES, B.; DE SCHONEN, S.; JUSCZYK, P.; McNEILAGE, P.; MORTON, J. (eds.). **Developmental Neurocognition: Speech and Face Processing in the First Year of Life**. [S.l.]: Springer, 1992. p. 259–274.

______. Early language acquisition: phonetic and word learning, neural substrates, and a theoretical model. In: GAZZANIGA, M. S. (ed.). **The Cognitive Neurosciences III**. Cambridge, MA: MIT Press, 2004. p. 837–854.

LADD, D. Robert. **Intonational Phonology**. 2nd ed. Cambridge: Cambridge University Press, 2008.

LADO, R. **Linguistics Across Cultures**. Ann Arbor: University of Michigan Press, 1957.

LARSEN-FREEMAN, D. Chaos/complexity science and second language acquisition. **Applied Linguistics**, v. 18, n. 2, p. 141–165, 1997.

______. Complexity theory. In: VANPATTEN, B.; WILLIAMS, J. (eds.). **Theories in Second Language Acquisition: An Introduction**. New York: Routledge, 2015. p. [s.p.].

LARSEN-FREEMAN, D.; CAMERON, L. **Complex Systems and Applied Linguistics**. Oxford: Oxford University Press, 2008.

LECANUET, J. P. Fetal sensory competencies. **European Journal of Obstetrics and Gynecology and Reproductive Biology**, v. 75, n. 1, p. 81–85, 1997. DOI: 10.1016/S0301-2115(97)00162-4.

LI, T.; FENG, H.; JIA, Y.; MA, R.; QI, Y. Cross-linguistic influence on intonation acquisition: a study on the production of L2 Mandarin and L3 English intonations by Uyghur speakers. In: **SPEECH PROSODY 2024**, Leiden. **Proceedings of Speech Prosody 2024**. [S.l.]: ISCA, 2024. p. [s.p.].

LONG, M. H. **The role of the linguistic environment in second language acquisition.** In: RITCHIE, W. C.; BHATIA, T. K. (Eds.). **Handbook of Second Language Acquisition.** San Diego: Academic Press, 1996. p. 413–468.

MACHADO, V. S. **Spanish prosody on the Uruguayan-Brazilian border: cross-generational and social variation.** 2024. 321 f. Tese (Doutorado em Linguística Hispânica) – University of Toronto, Toronto, 2024.

MAMPE, B.; FRIEDERICI, A. D.; CHRISTOPHE, A.; WERMKE, K. **Newborns' cry melody is shaped by their native language.** Current Biology, v. 19, n. 23, p. 1994–1997, 2009.

MATTYS, S. L.; DAVIS, M. H.; BRADLOW, A. R.; SCOTT, S. K. **Speech recognition in adverse conditions: A review.** Language and Cognitive Processes, v. 27, n. 7–8, p. 953–978, 2012. DOI: 10.1080/01690965.2012.705006.

MEHLER, J.; BERTONCINI, J.; BARRIERE, M.; JASSIK-GERSCHENFELD, D. **Infant recognition of mother's voice.** Perception, v. 7, n. 5, p. 491–497, 1978. DOI: 10.1068/p070491.

MENNEN, I. **Bi-directional interference in the intonation of Dutch speakers of Greek.** Journal of Phonetics, v. 32, n. 4, p. 543–563, 2004.

______. **Beyond segments: Towards an L2 intonation learning theory.** In: DELAIS-ROUSSARIE, E.; AVANZI, M.; HERMENT, S. (Eds.). **Prosody and Language in Contact.** Berlin: Springer, 2015. p. 171–188.

MENNEN, I.; DE LEEUW, E. **Beyond segments: Prosody in SLA.** Studies in Second Language Acquisition, v. 36, n. 2, p. 183–194, 2014. DOI: 10.1017/S0272263114000138.

MENNEN, I.; REUBOLD, U.; ENDES, K.; MAYR, R. **Plasticity of native intonation in the L1 of English migrants to Austria.** Frontiers in Psychology, v. 13, 2022.

MITCHELL, R.; MYLES, F. (Eds.). **Second Language Learning Theories.** 2nd ed. London: Hodder Arnold, 2004.

MOON, C.; COOPER, R. P.; FIFER, W. P. **Two-day-olds prefer their native language.** Infant Behavior and Development, v. 16, n. 4, p. 495–500, 1993.

MOON, C.; LAGERCRANTZ, H.; KUHL, P. K. **Language experienced in utero affects vowel perception after birth: A two-country study.** Acta Paediatrica, v. 102, n. 2, p. 156–160, 2013.

MORAES, J. **The pitch accents in Brazilian Portuguese: Analysis by synthesis.** In: **Proceedings of Speech Prosody 2008.** p. 389–397, 2008.

MORAES, J.; COLAMARCO, M. **Você está pedindo ou perguntando? Uma análise entonacional de pedidos e perguntas no português do Brasil.** Revista de Estudos da Linguagem, v. 15, n. 2, p. 113–126, 2007. Available at:

http://www.periodicos.letras.ufmg.br/index.php/relin/article/view/2451/2405.
Accessed on: Jan. 3, 2026.

MORGAN, J. L.; DEMUTH, K. (eds.). **Signal to syntax: bootstrapping from speech to grammar in early acquisition**. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

MORSE, P. A. The discrimination of speech and nonspeech stimuli in early infancy. **Journal of Experimental Child Psychology**, v. 13, n. 3, p. 477–492, 1972. https://doi.org/10.1016/0022-0965(72)90076-6

MOTA, M. B. Sistemas de memória e processamento da linguagem: um breve panorama. **Revista Linguística**, Rio de Janeiro, v. 11, n. 1, p. 205–215, jun. 2015. https://doi.org/10.17074/2238-975X.2015v11n1p205

MUNRO, M. J.; DERWING, T. M. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. **Language Learning**, v. 45, n. 1, p. 73–97, 1995.

______. The functional load principle in ESL pronunciation instruction: An exploratory study. **System**, v. 34, n. 4, p. 520–531, 2006. https://doi.org/10.1016/j.system.2006.09.004

NAME, C.; SOSA, J. M. The prosody of questions in Brazilian Portuguese infant-directed speech. In: **Proceedings of the 10th International Conference on Speech Prosody**. 2020. p. 334–341.

NAME, C.; SOSA, J. M. Infant-directed speech in Brazilian Portuguese: The prosody of questions. **DELTA: Documentação de Estudos em Linguística Teórica e Aplicada**, v. 38, n. 3, 2022.

NAME, C. Language perception development. In: BABATSOULI, E. (ed.). **An anthology of bilingual child phonology**. [S.l.]: Springer, 2022. p. 15–42. https://doi.org/10.1007/978-3-031-13545-3_2

NARAYANAN, D. Z.; TAKAHASHI, D. Y.; KELLY, L. M.; et al. Prenatal development of neonatal vocalizations. **eLife**, v. 11, e78485, 2022.

PAIVA, V. L. **Aquisição de segunda língua**. São Paulo: Parábola Editorial, 2014.

PAPOUSEK, M.; BORNSTEIN, M. H.; NUZZO, C.; PAPOUSEK, H.; SYMMES, D. Infant responses to prototypical melodic contours in parental speech. **Infant Behavior and Development**, v. 13, n. 4, p. 539–545, 1990. https://doi.org/10.1016/0163-6383(90)90022-Z

PASSARELLA DOS REIS, L.; GONÇALVES, A. R.; SILVEIRA, R. Perception of intonational patterns and speaker's intentionality in English yes-no questions produced by Brazilians. **Revista de Estudos da Linguagem**, v. 24, n. 1, p. 65, 2016.

PASSARELLA DOS REIS, L. Descendo ou rising? Padrões entoacionais de perguntas totais da língua inglesa e da língua portuguesa por falantes brasileiros. **Revista X**, v. 1, n. 0.2014, 2014. Disponível em: https://revistas.ufpr.br/revistax/article/view/34266. Acesso em: 20 jul. 2021.

PATIENCE, M. MARASCO, O.; COLANTONI, L.; KLASSEN, G.; RADU, M.; TARAROVA, O. Initial pitch cues in English sentence types. In: **SPEECH PROSODY 2018**, 2018, Poznań, Poland. *Proceedings…* [S.l.]: ISCA, 2018. p. 463–467. DOI: 10.21437/SpeechProsody.2018-94

PEŠKOVÁ, A. **L2 Spanish and Italian intonation: accounting for the different patterns displayed by L1 Czech and German learners**. Berlin: Language Science Press, 2023.

PIERREHUMBERT, J. B. **The phonology and phonetics of English intonation**. 1980. Thesis (PhD in Linguistics) — Massachusetts Institute of Technology, Cambridge, 1980.

PIKE, K. L. **The intonation of American English**. Ann Arbor: University of Michigan Press, 1945.

POERSCHI, J. M. Simulações conexionistas: a inteligência artificial moderna. **Linguagem em (Dis)curso**, v. 4, n. 2, p. 441–458, jan./jun. 2004.

PRIETO, P. The intonational phonology of Catalan. In: FROTA, Sónia; PRIETO, Pilar (ed.). **Intonation in Romance**. Oxford: Oxford University Press, 2015. p. 43–80.

RAMÍREZ-ESPARZA, N.; GARCÍA-SIERRA, A.; KUHL, P. K. Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. **Developmental Science**, v. 17, n. 6, p. 880–891, 2014. DOI: 10.1111/desc.12172.

RICHARDS, J. C.; RODGERS, T. S. **Approaches and methods in language teaching**. 2. ed. Cambridge: Cambridge University Press, 2001.

RICHARDS, N. **Uttering Trees**. Cambridge, MA: MIT Press, 2010.

RADU, M.; et al. The perception and interpretation of sentence types by L1 Spanish–L2 English speakers. **Linguistic Approaches to Bilingualism**, v. 8, n. 5, p. 611–636, 2018.

ROSENBERG, A.; HIRSCHBERG, J.; NAKATANI, C. Prosodic and lexical effects on prominence perception. **Speech Communication**, v. 52, n. 3, p. 343–353, 2010. https://doi.org/10.1016/j.specom.2009.12.004

SAMMLER, D.; GROBRAS, M.-H.; ANWANDER, A.; BESTELMEYER, P. E. G.; BELIN, P. Dorsal and ventral pathways for prosody. **Current Biology**, v. 25, n. 23, p. 3079–3085, 2015. DOI: 10.1016/j.cub.2015.10.009.

SCHMIDT, R. The role of consciousness in second language learning. **Applied Linguistics**, v. 11, p. 129–158, 1990.

SCHMIDT, E.; POST, B. The development of prosodic features and their contribution to rhythm production in simultaneous bilinguals. **Language and Speech**, v. 58, pt. 1, p. 24–47, 2015.

SCHUMANN, J. H. **The pidginization process: a model for second language acquisition**. Rowley, MA: Newbury House, 1978.

SELINKER, L. Interlanguage. **International Review of Applied Linguistics in Language Teaching (IRAL)**, v. 10, p. 209–231, 1972.

SHOKEIR, V. Evidence for the stable use of uptalk in South Ontario English. **University of Pennsylvania Working Papers in Linguistics,** v. 14, n. 2, p. 16–24, 2008.

SICHEL-BAZIN, Rafèu; BUTHKE, Carolin; MEISENBURG, Trudel. Language contact and prosodic interference: nuclear configurations in Occitan and French “statements of the obvious”. In: **SPEECH PROSODY**, 6., 2012, Shanghai. Shanghai: Tongji University Press, 2012. p. 414–417. DOI: 10.21437/SpeechProsody.2012-105.

SILVA, C.; ARANTES, P. Quantitative analysis of fundamental frequency in Spanish (L2) and Brazilian Portuguese (L1): evidence of learning and language attrition. **Journal of Speech Sciences**, Campinas, v. 10, n. 00, p. e021003, 2021. DOI: 10.20396/joss.v10i00.15779.

SKINNER, B. F. **Verbal behavior**. New York: Appleton-Century-Crofts, 1957

SOKOLIK, M. E. Learning without rules: PDP and resolution of the adult language learning paradox. **TESOL Quarterly**, v. 24, n. 4, p. 685–696, 1990.

SPENCER, K. A.; DECASPER, A. J. Prenatal experience with low-frequency maternal-voice sounds influences neonatal perception of maternal voice. **Infant Behavior and Development**, v. 10, n. 2, p. 133–142, 1987. https://doi.org/10.1016/0163-6383(87)90017-8

STEPANOV, A.; KODRIČ, K. B.; STATEVA, P. The role of working memory in children’s ability for prosodic discrimination. **PLOS ONE**, v. 15, n. 3, e0229857, 2020.

SWAIN, M. Communicative competence: some roles of comprehensible input and comprehensible output in its development. In: GASS, S.; MADDEN, C. (Eds.). **Input in Second Language Acquisition**. Rowley, MA: Newbury, 1985. p. 235–253.

______. The Output Hypothesis: just speaking and writing aren’t enough. **Canadian Modern Language Review**, v. 50, p. 158–164, 1993.

SWAIN, M. Three functions of output in second language learning. In: COOK, G.; SEIDLHOFER, B. (Eds.). **Principle and Practice in Applied Linguistics**. Oxford: Oxford University Press, 1995. p. 125–144.

______. Focus on form through conscious reflection. In: DOUGHTY, C.; WILLIAMS, J. (Eds.). **Focus on Form in Classroom Second Language Acquisition**. Cambridge: Cambridge University Press, 1998. p. 64–81.

______. The Output Hypothesis: theory and research. In: HINKEL, E. (Ed.). **Handbook of Research in Second Language Learning and Teaching**. Mahwah, NJ: Lawrence Erlbaum, 2005. p. 471–483.

SILVA JR., L. J.; BARBOSA, P. A. Speech rhythm of English as L2: an investigation of prosodic variables on the production of Brazilian Portuguese speakers. **Journal of Speech Sciences**, v. 8, n. 2, p. 37-57, 2019. DOI: 10.20396/joss.v8i2.14996.

______. Speech rhythm of English as L2: the influence of duration and F0 on foreign accent investigation. In: CONGRESSO BRASILEIRO DE PROSÓDIA, 1., 2020. **Anais do Congresso Brasileiro de Prosódia**, v. 1, p. 59-62, 2020.

SYRDAL, A. K.; MCGORY, J. Inter-transcriber reliability of ToBI prosodic labeling. In: **Proc. 6th International Conference on Spoken Language Processing** (ICSLP 2000), vol. 3, 235-238. 2000. p. 235–238. doi: 10.21437/ICSLP.2000-521

TEIXEIRA, L. A. S.; LIMA JR., R. M. Análise do desenvolvimento do ritmo do inglês-L2 por brasileiros por meio de três métricas rítmicas. **Revista X**, v. 16, n. 5, p. 1258-1292, 2021. DOI: 10.5380/rvx.v16i5.81413.

TEIXEIRA, L. A. S. A multidimensional analysis of English-L2 rhythm development. **Ilha do Desterro**, v. 76, n. 3, p. 223-250, 2023. DOI: 10.5007/2175-8026.2023.e94742.

TOKIZAKI, Hisao. **Linearizing structure with silence: a minimalist theory of syntax-phonology interface.** 2006. Tese (Doutorado em Linguística) — University of Tsukuba, Tsukuba, 2006.

TRUCKENBRODT, Hubert. On the relation between syntactic phrases and phonological phrases. **Linguistic Inquiry**, v. 30, n. 2, p. 219–255, 1999

VANPATTEN, B.; WILLIAMS, J. (Eds.). **Theories in Second Language Acquisition: An Introduction**. 2. ed. New York: Routledge, 2015.

WERKER, J. F.; TEES, R. C. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. **Infant Behavior and Development**, v. 7, n. 1, p. 49–63, 1984. https://doi.org/10.1016/S0163-6383(84)80022-3

WHITE, L. **Second Language Acquisition and Universal Grammar**. Cambridge: Cambridge University Press, 2003.

______. Linguistic theory, Universal Grammar, and second language acquisition. In: VANPATTEN, B.; WILLIAMS, J. (Eds.). **Theories in Second Language Acquisition: An Introduction**. New York: Routledge, 2015. [pages not provided].

WILDGRUBER, D.; ACKERMANN, H.; KREIFELTS, B.; ETHOFER, T. Cerebral processing of linguistic and emotional prosody: fMRI studies. **Progress in Brain Research**, v. 156, p. 249–268, 2006. DOI: 10.1016/S0079-6123(06)56013-3.

WITKO, A.; GHIMENTON, A. Early vocabulary acquisition in infants: new insights from naturalistic data. **Journal of Child Language**, v. 46, n. 3, p. 477–502, 2019. https://doi.org/10.1017/S0305000918000571

YOON, T.; CHAVARRIA, S.; COLE, J. Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI. In: **INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING, 8.**, 2004, Jeju Island. Jeju Island: ICSLP, 2004. p. 2729–2732.

ZAHNER-RITTER, K.; ZHAO, T.; EINFELDT, M.; BRAUN, B. How experience with tone in the native language affects the L2 acquisition of pitch accents. **Frontiers in Psychology**, v. 13, 903879, 2022.

ZHANG, X.; CHEN, Y. Chinese EFL learners' perception of English prosodic focus. In: **INTERSPEECH 2023**. [S.l.]: ISCA, 2023. p. 92–96. Available at: https://www.isca-archive.org/interspeech_2023/zhang23da_interspeech.html. Accessed on: January 3, 2026.

APPENDIX 1: Experiments 1 and 1.2 recording stimuli

Phrase 1:

You've been to the supermarket to do some shopping for your mother. She's asked you to buy tomatoes, apples, and oranges. When you arrive home, you tell her:

⇒ The supermarket was out of tomatoes.

Phrase 2:

You work at a fruit shop at a street fair. A client seems to be very interested in your apples. You ask him:

⇒ Would you like to buy them?

Phrase 3:

You work at a company. Your boss asks you if your coworker has already arrived and you respond:

⇒ I don't believe he's arrived yet.

Phrase 4:

You are out on a walk when your mom calls you. She complains that she can't hear you very well. You go somewhere else and ask her:

⇒ Can you hear me now?

Phrase 5:

You are a tour guide and, before you and your group leave the hotel for your daily sightseeing, you ask them:

⇒ Have you got your passports with you?

Phrase 6:

You and your friends are planning to hang out at a nice restaurant. As you want to arrive early, you ask them:

⇒ What time does the restaurant open? Is it open at 7pm?

Phrase 7:

You and your friend are in a candy shop and they ask you if you're going to buy a chocolate bar that's on the shelf. You reply:

⇒ I'd definitely buy this chocolate if it was cheaper.

Phrase 8:

You and your friends were on an outdoor walk yesterday and today they want to go again. You decide not to go because your best friend isn't feeling very well. You tell your friends:

⇒ Her feet really hurt from yesterday.

Phrase 9:

You are scheduling a meeting with your coworkers. One of them suggests that one week from now would be the best option and you say:

⇒ This time next week I'll be travelling abroad.

Phrase 10:

Your friends just got back from a trip to London. As soon as you meet you ask them:

⇒ Did you guys travel around the United Kingdom?

Phrase 11:

You are chatting with your friend about a complicated situation happening at the building that they live in. After she'd told you about a series of absurd things that happened you say:

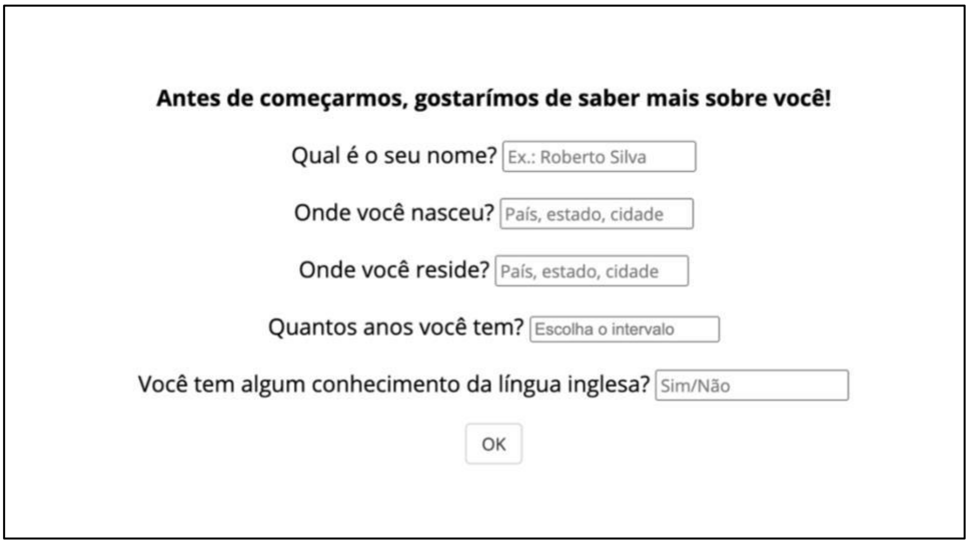
$\Rightarrow$ If it keeps going like this, you'll have to move out pretty soon.

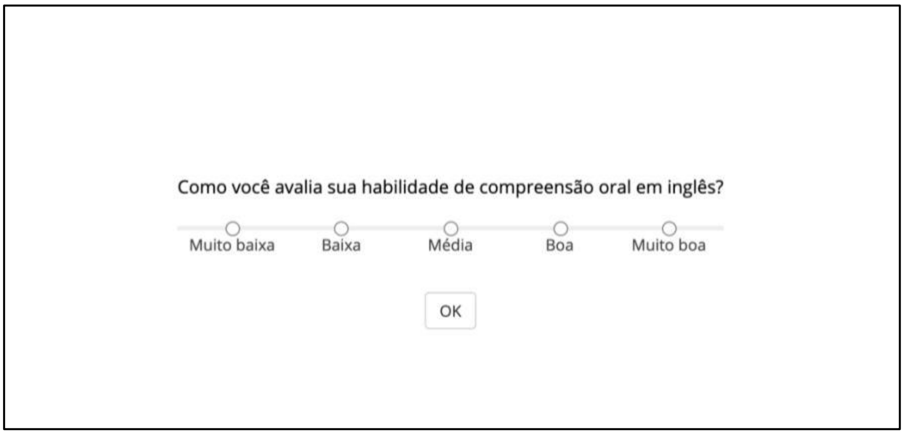
Phrase 12:

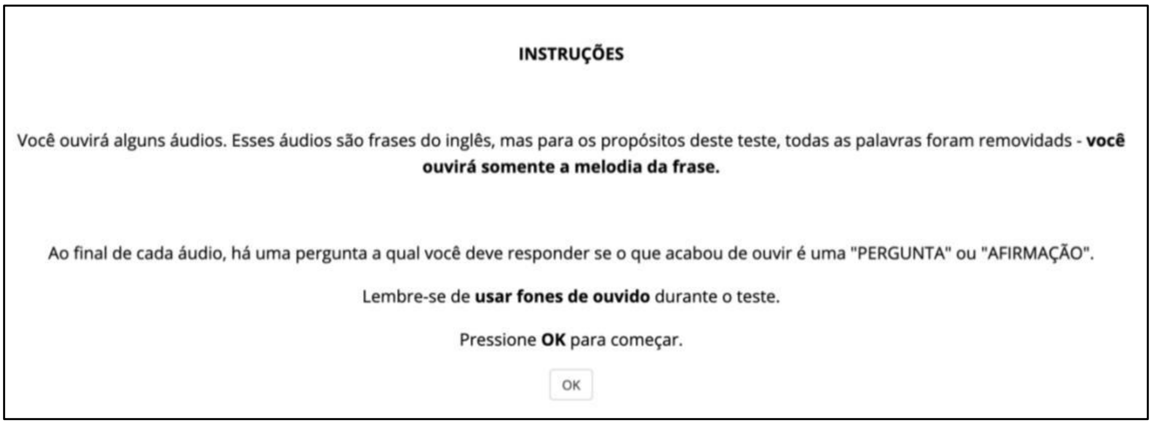
You're heading home when you decide to have some beer at a nearby pub. You realize the pub is almost closing. You ask the bartender:

$\Rightarrow$ Can I still order some beer?

APPENDIX 2: Screens for the L1-BP group in Experiment 1







APPENDIX 3: Praat script for extracting min and max F0 values

```
##get measurements.praat
##original script created by Christan Kroos, Rikke Bundgaard-Nielsen, Michael Tyler
##modified by Mark Antoniou
## C 2010 MARCS Auditory Laboratories

#
===============================================================
# 1. Because of a bug in Praat 'Show intensity' cannot be scripted.
# Open an editor on an abritrary sound and make sure that 'Show intensity' is ticked
before you run this script
# 2. The comma-separated file is a .txt file. Import this into Excel and use the text
import wizard to change columns to text format where necessary.
#
===============================================================

#create list of filenames
clearinfo

form Parameters
  comment Directory that contains the original stimuli
  sentence inDirectory /Users/thalesbuzan/TOBI ANNOTATION
  sentence filetype wav
  comment Name of the output csv file with the measurements
  sentence outFile measurements
  choice speaker_gender: 1
        button Female
        button Male
endform
```

```
Create Strings as file list... fileList 'inDirectory$'*.TextGrid

fileappend 'inDirectory$''outFile$'.txt sound name,
	...segment,start,end,duration,overall_dB,overall_f0,
	...50_cursor,50_dB,50_f0,50_F1,50_F2,50_F3,
	...'newline$'

#loop for all files
nFiles = Get number of strings
for d from 1 to nFiles


  #read in sound file (wav) and textgrid
  select Strings fileList
  fileName$ = Get string... d
  dotInd = rindex(fileName$, ".")
  soundName$ = left$(fileName$, dotInd - 1)
  printline Processing file 'soundName$'...
  Read from file... 'inDirectory$''soundName$'.'filetype$'
  Read from file... 'inDirectory$''soundName$'.Textgrid



  #Get relevant information from the textgrid (exclude non-labeled segments) and do
measurements

  select TextGrid 'soundName$'
  number_of_intervals = Get number of intervals... 1
```

```
select Sound 'soundName$'
if speaker_gender$ == "Female"
	To Formant (burg)... 0.0025 5 5500 0.025 50
else
  To Formant (burg)... 0.0025 5 5000 0.025 50
endif

select Sound 'soundName$'
To Pitch... 0.01 75 600

select Sound 'soundName$'
To Intensity... 100 0 yes

for s from 1 to 'number_of_intervals'
  select TextGrid 'soundName$'
  segment$ = Get label of interval... 1 s
  if segment$ <> ""
		startSeg = Get starting point... 1 s
	endSeg = Get end point... 1 s
	durSeg = 'endSeg' - 'startSeg'
	cursor_50 = 'startSeg' + ('durSeg' * 0.5)

		select Intensity 'soundName$'
	    overall_dB = Get mean... 'startSeg' 'endSeg' dB
		dB_50 = Get value at time... 'cursor_50' Cubic

		if segment$ == "S" or segment$ == "V" or segment$ == "O"
			select Pitch 'soundName$'
			overall_f0 = Get mean... 'startSeg' 'endSeg' Hertz
			f0_50 = Get value at time... 'cursor_50' Hertz Linear
```

```
                select Formant 'soundName$'
                f1_50 = Get value at time... 1 'cursor_50' Hertz Linear
                f2_50 = Get value at time... 2 'cursor_50' Hertz Linear
                f3_50 = Get value at time... 3 'cursor_50' Hertz Linear

                fileappend 'inDirectory$''outFile$'.txt 'soundName$',
         ...'segment$','startSeg','endSeg','durSeg',
         ...'overall_dB','overall_f0',
         ...'cursor_50','dB_50','f0_50','f1_50','f2_50','f3_50',
         ...'newline$'
        else
                fileappend 'inDirectory$''outFile$'.txt 'soundName$',
         ...'segment$','startSeg','endSeg','durSeg',
         ...'overall_dB',,
         ...'cursor_50','dB_50',,,,,
         ...'newline$'
        endif


    endif
  endfor


  #clean object list
  select all
  minus Strings fileList
  Remove
endfor
```

```
#clean object list
select all
Remove
printline --------------------------------
printline Processed all 'filetype$' files in 'inDirectory$'
fileappend 'inDirectory$''outFile$'.txt Processed all files in 'inDirectory$'
```

APPENDIX 4: Experiment 2 stimuli

TEST TRIALS

Sentence number	Context	Sentence
1	Your friends and family are planning a visit to North America. You are not sure if Mark has a US visa, so you ask your cousin:	Can Mark travel to America?
2	You are planning a trip to the beach the next day with your friends. You ask them:	Is it going to rain tomorrow?
3	You are doing a family gathering and want to take a photo. Everyone is looking for the camera and you ask:	Does Anne have the camera?
4	You are setting up a band with your friend. You need a pianist, so you ask them:	Can Maria play the clarinet?
5	You enter a grocery store and ask:	Does this store sell bananas?
6	You see two identical laptops on a table at the library, one of them is yours. As you are not sure which one is, you ask the person next to you:	Is this your computer?
7	Your friends and family are planning a visit to North America. Your cousin is not sure if Mark has a US visa, so you tell them:	Mark can travel to America.
8	Your friends are planning a trip to the beach the next day with you, but forecast is not looking good. You tell them:	It is going to rain tomorrow.
9	You and your friends are talking, and you are sharing what kind of instruments you can play. You tell them:	Maria can play the clarinet.
10	You work at a grocery store. You are asked if you sell bananas and reply:	This store sells bananas.
11	You see two identical laptops on a table at the library, one of them is not yours. The person next to you is not sure which one is theirs, you tell them:	This is my computer.
12	You are doing a family gathering and want to take a photo. Everyone is looking for the camera and you say:	Anne has the camera.

TRAINING TRIALS

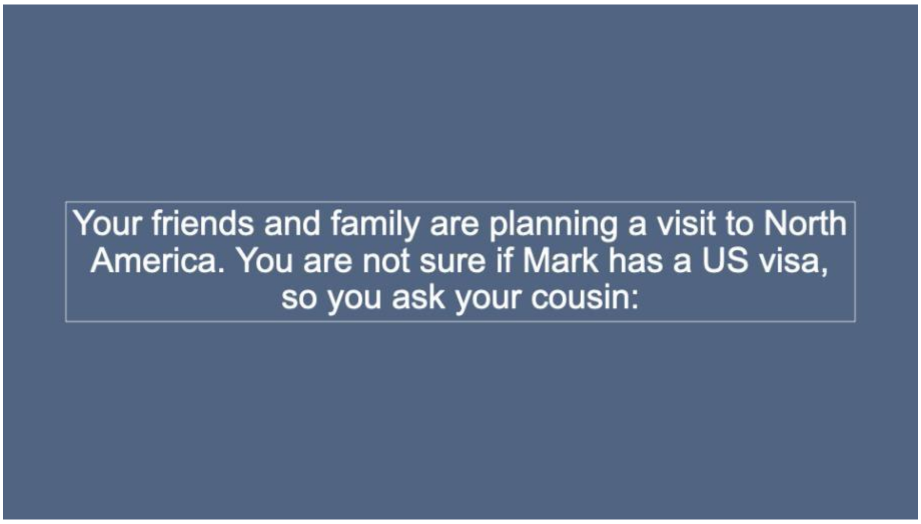
Sentence number	Context	Sentence
1	You are out on a walk when your mom calls you. She complains that she can't hear you very well. You go somewhere else and ask her:	Can you hear me now?
2	Your professor just told you about an upcoming trip they are going to do. You ask:	When are you going?
3	You are exiting the subway train and the lady in front of you let something fall out of her pocket. You try to call her saying:	Lady! You dropped something.

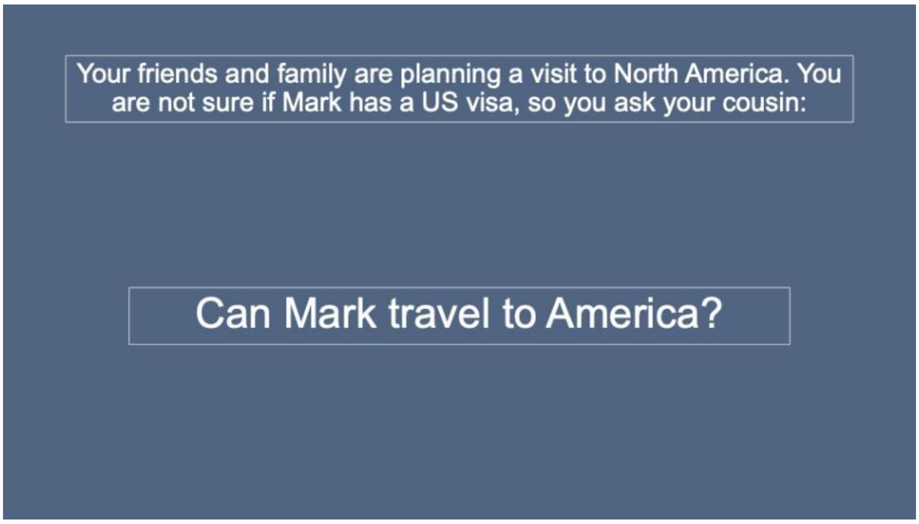
DISTRACTORS

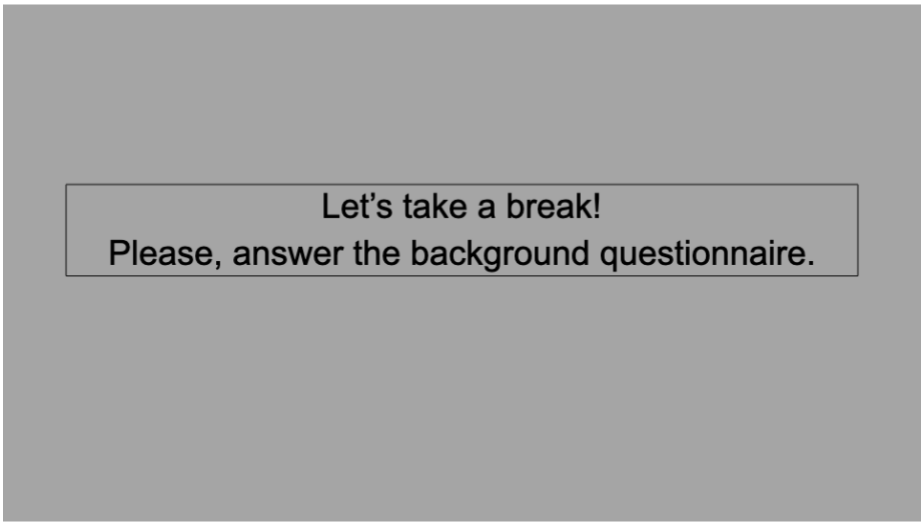
Sentence number	Context	Sentence
1	You work at a company. Your boss asks you if your co-worker has already arrived and you respond:	I don't believe he's arrived yet.
2	You are scheduling a meeting with your co-workers. One of them suggests that one week from now would be the best option and you say:	This time next week I'll be travelling abroad.
3	You're heading home when you decide to have some beer at a nearby pub. You realize the pub is almost closing. You ask the bartender:	Can I still order some beer?
4	You are a tour guide and before you and your group leave the hotel for your daily sightseeing, you ask them:	Have you got your passports with you?
5	You woke up this morning and you looked outside the window. The sky is completely dark. You say to yourself:	The weather is grim today.
6	You are at a restaurant, and you are not sure where the washroom is. You ask the waiter for confirmation:	Is the washroom this way?
7	You just met up with a new friend and you are talking about your favourite seasons. You say:	Summer is my favourite season.
8	You are at a candy shop with your friends trying some chocolates. You say:	This chocolate is good, isn't it?
9	You and your friends are deciding which restaurant to go to, and your friend picked the one down the street. You reply:	Yes! We should go!
10	You and your friend are in a crowded party. Eventually you lose him. When you finally find him, you shout:	Ryan! Over here!
11	You and your relatives are at a grocery store, and you see some fruits on sale. You say to them:	That is an amazing deal!
12	You are supposed to go to a party, but you are not sure if plans have changed. You want to clarify, and ask:	Are you sure about the party, Jason?
13	Your roommate forgot something cooking in the kitchen. You tell them:	The water is boiling.
14	You are in a class and your professor mentioned an activity that you have never heard before. You ask a classmate:	Is this the new assignment?
15	A friend of yours is telling you about a wrong order they received at a restaurant. You seem shocked and say:	What did they give you?!
16	You are at a friend's house, and they notice Mark is missing. Your friend asks you where he is, and you say:	He is outside watching the sunset.
17	Your roommate mentioned they are going to watch a movie tonight. You are not sure about the time, so you ask:	When are you going to the movies?
18	You just came back from a play you have seen. Your friend asks you how it went, and you reply:	The play was amazing!
19	You and your friends are at the library studying. One of them asks about the closing time. You reply:	I'm not so sure, Max.
20	Your friend hiked the Himalayas, and you are curious on the reason behind the trip. You ask them:	Why did you travel there?
21	You broke a glass in the kitchen, and you need help from the people you live with. You say to them:	Come here! Please!

22	You are going to a friend’s house party, but your roommate is not quite sure how far it is. You say:	She lives twenty minutes South.
23	You just won the lottery and you do not know where to go for the holiday with your friends. You ask them:	Where should we go?
24	You are walking around a new building, and you are thirsty. You go to the front desk and say:	I am looking for a water fountain.

APPENDIX 5: PowerPoint examples of the production task in Experiment 2







APPENDIX 6: Linguistic background questionnaire

Spanish & Portuguese
UNIVERSITY OF TORONTO

Subject Number :

Study :

Participant language background questionnaire

A. Personal Information

- Sex: ☐ Male ☐ Female
- Year of Birth: ______________________
- Place of Birth: City ______________ Country ______________________
- Occupation: ______________________
- Highest Level of Schooling: ☐ Secondary ☐ College/Professional ☐ University
- If you were not born in Canada, at what age did you move here? ______________________

B. First Language

What is your first language? ______________________

What is the first language of: your mother? ______________ your father? ______________

Did you learn your first language from birth? ☐ Yes ☐ No

- If you answered 'No' to the question above, please explain:

  ____

Which language(s) did you speak at home as a child? ______________________

Is your first language the language with which you are the most comfortable? ☐ Yes ☐ No

- If you answered 'No' to the question above, please explain:

  ____

C. Education & Language Use

Which language(s) were you formally educated in? Where (i.e. country)?

Primary/Elementary School ______________________

High School ______________________

University ______________________

Which language(s) do you use (Indicate approximate percentage, e.g. 0, 50, 100%):

At school ______________________

At home ______________________

At work ______________________

In social situations ______________________

D. Second Languages

	Second Languages A.	B.
At what age did you begin to learn your 2nd language?		
Where did you learn your 2nd language? Give place and years.		
Were your teachers native speakers of this language?		
Did you learn this language as a subject or was it the principal medium of instruction?	☐ Subject ☐ Medium of Instruction	☐ Subject ☐ Medium of Instruction
Have you ever spent time in an area where this language was the native language?	Where? How long?	Where? How long?
Approximately how many hours a week do you use this language? Specify for each of speaking, listening and reading.	Speaking : _______ hrs Listening : _______ hrs Reading : _______ hrs	Speaking : _______ hrs Listening : _______ hrs Reading : _______ hrs

- *Please rate your linguistic ability in each of your second languages in the following areas by checking the appropriate answer.*

	Beginner	Intermediate	Advanced	Near-Native
READING				
Language A				
Language B				
WRITING				
Language A				
Language B				
SPEAKING				
Language A				
Language B				
LISTENING				
Language A				
Language B				
OVERALL COMPETENCE				
Language A				
Language B				

Do you know any other second languages? Please specify: ____

APPENDIX 7: Coding Experiment 2

```
// preload audios
var preload_audios = {
  type: 'preload_audios',
  auto_preload: true
 };
 // create timeline
 var timeline = [];
 // survey

 // create form
 var form = {
  type: "survey-html-form",
  preamble: "<strong>Before we start, we'd like to know more about you!</strong>",
  html: `
  <p>What are your initials? <input name='initials' type='text' placeholder='e.g. MJ'
maxlength='4' required /></p>

  <p>Which country are you from? <input name='from_country' type='text'
placeholder='e.g. Canada' required /></p>

  <p>Where do you live?
  <select id='live_country' name='live_country' required
onchange='populateStatesAndCities()'>
  <option value=''>Select your country</option>
  <option value='Canada'>Canada</option>
  <option value='United States'>United States</option>
  <option value='United Kingdom'>United Kingdom</option>
  <option value='Brazil'>Brazil</option>
  <option value='Other'>Other</option>
  </select>
  <span id='state_input'></span>
  <span id='city_input'></span>
  <span id='other_country_input'></span>
  </p>

  <p>How old are you?
  <select name='age' required>
  <option value=''>Select</option>
  ${[...Array(38)].map((_, i) => `<option value='${i+18}'>${i+18}</option>`).join('')}
  </select>
  </p>

  <p>Which is your first language?
  <input name='first_language' type='text' placeholder='e.g. English' required />
  </p>

  <p>Do you speak any other languages?
```

```
  <input type='radio' name='other_language' value='Yes' id='other_lang_yes'
required
onclick='document.getElementById(\"other_languages_input\").style.display=\"block\"
;'> Yes
  <input type='radio' name='other_language' value='No' id='other_lang_no'
onclick='document.getElementById(\"other_languages_input\").style.display=\"none\"
;'> No
  </p>

  <div id='other_languages_input' style='display:none;'>
  <p>If yes, what language(s)? <input name='other_languages_list' type='text'
placeholder='e.g. Spanish, French'></p>
  </div>
  `,
  button_label: "OK",
  data: {
   type: "form"
  },
  on_load: function() {
   window.populateStatesAndCities = function() {
    var country = document.getElementById('live_country').value;
    var state_input = document.getElementById('state_input');
    var city_input = document.getElementById('city_input');
    state_input.innerHTML = '';
    city_input.innerHTML = '';
    let stateLists = {
      'United States': ['Alabama',
       'Alaska',
       'American Samoa',
       'Arizona',
       'Arkansas',
       'California',
       'Colorado',
       'Connecticut',
       'Delaware',
       'District of Columbia',
       'Federated States of Micronesia',
       'Florida',
       'Georgia',
       'Guam',
       'Hawaii',
       'Idaho',
       'Illinois',
       'Indiana',
       'Iowa',
       'Kansas',
       'Kentucky',
       'Louisiana',
       'Maine',
       'Marshall Islands',
```

```
 'Maryland',
 'Massachusetts',
 'Michigan',
 'Minnesota',
 'Mississippi',
 'Missouri',
 'Montana',
 'Nebraska',
 'Nevada',
 'New Hampshire',
 'New Jersey',
 'New Mexico',
 'New York',
 'North Carolina',
 'North Dakota',
 'Northern Mariana Islands',
 'Ohio',
 'Oklahoma',
 'Oregon',
 'Palau',
 'Pennsylvania',
 'Puerto Rico',
 'Rhode Island',
 'South Carolina',
 'South Dakota',
 'Tennessee',
 'Texas',
 'Utah',
 'Vermont',
 'Virgin Island',
 'Virginia',
 'Washington',
 'West Virginia',
 'Wisconsin',
 'Wyoming'],

'Canada': ['Alberta',
 'British Columbia',
 'Manitoba',
 'New Brunswick',
 'Newfoundland and Labrador',
 'Northwest Territories',
 'Nova Scotia',
 'Nunavut',
 'Ontario',
 'Prince Edward Island',
 'Quebec',
 'Saskatchewan',
 'Yukon Territory'],
'United Kingdom': ['England',
```

```
        'Scotland',
        'Wales',
        'Northern Ireland'],
      };
      if (stateLists[country]) {
        let options = stateLists[country].map(state => `<option
value='${state}'>${state}</option>`).join('');
        state_input.innerHTML = `<select id='state_select' name='state' required
onchange='showCityInput()'><option value=''>Select
country/state/province</option>${options}</select>`;
      }
      if (country === "Other") {
        other_country_input.innerHTML = "<p>Please specify the country/region: <input
name='other_live_country' type='text' required></p>";
      }
     };
     window.showCityInput = function() {
      var country = document.getElementById('live_country').value;
      var state = document.getElementById('state_select') ?
document.getElementById('state_select').value: '';
      var city_input = document.getElementById('city_input');
      if ((country === 'United States' || country === 'Canada') && state && state !== '') {
        city_input.innerHTML = "<p>Which city? <input name='city' type='text'
placeholder='e.g. Toronto' required /></p>";
      } else {
        city_input.innerHTML = '';
      }
     };
   }
  };

  // create instructions
  var instructions = {
   type: 'html-button-response',
   stimulus:
   '<p><strong>INSTRUCTIONS</strong></p>' +
   '<p>You will listen to several audios. These audios are English sentences, but for
the purposes of this test all words have been removed - <strong>you will only hear
the melody of the sentence.</strong></p>' +
   '<p>For each audio there is one question in which you must answer if what you just
heard is a "QUESTION" or a "STATEMENT".</p>' +
   '<p>Remember to <strong>use headphones/earbuds</strong> during the
test.</p>' +
   '<p>We will begin with a training phase first.</p>' +
   '<p>Press <strong>OK</strong> to start.</p>',
   post_trial_gap: 500,
   choices: ['OK']
  };

  // create fixation
```

```
var fixation = {
 type: 'html-button-response',
 stimulus: '<p><strong></strong></p>',
 //choices: jsPsych.NO_KEYS,
 choices: ['+'],
 trial_duration: 1000
};

// training audio stimuli
var training_stimuli = [{
 stimulus: 'G1_02_30Q_purr.wav',
 data: {
   correct_response: 'QUESTION'
 }},
 {
   stimulus: 'G1_03_6Q_purr.wav',
   data: {
    correct_response: 'QUESTION'
   }},
 {
   stimulus: 'G1_02_21D_purr.wav',
   data: {
    correct_response: 'STATEMENT'
   }}];

// training block
var training_block = {
 timeline: [
   fixation,
   {
    type: 'audio-button-response',
    stimulus: jsPsych.timelineVariable('stimulus'),
    choices: ['QUESTION',
      'STATEMENT'],
    post_trial_gap: 700,
    trial_duration: 20000,
    data: jsPsych.timelineVariable('data'),
    on_finish: function(data) {
      let resposta = null;
      data.correct_response == 'QUESTION' ? resposta = 0: resposta = 1;
      if (data.response == resposta) {
       data.correct = true;
      } else {
       data.correct = false;
      }
    }
   }],
 timeline_variables: training_stimuli,
 randomize_order: false
};
```

```
// Add training instructions
var training_instructions = {
 type: 'html-button-response',
 stimulus:
 '<p><strong>TRAINING PHASE</strong></p>' +
 '<p>You will do a quick training to get used to the task.</p>' +
 '<p>For each audio, decide if you hear a "QUESTION" or a "STATEMENT".</p>' +
 '<p>Press <strong>OK</strong> to start training.</p>',
 choices: ['OK']
};

// put form in timeline
timeline.push(form);
// put instructions in timeline
timeline.push(instructions);
timeline.push(training_instructions);
timeline.push(training_block);

// create stimuli
var stimuli = [
 { stimulus: 'G1_02_12Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G1_08_22Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G1_08_4Q_purr.wav',  data: { correct_response: 'QUESTION' }},
 { stimulus: 'G1_08_8Q_purr.wav',  data: { correct_response: 'QUESTION' }},
 { stimulus: 'G1_03_24Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G1_04_30Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G1_04_35Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G1_02_15D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G1_03_4D_purr.wav',  data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G1_04_27D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G1_07_11D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G1_07_25D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G1_02_4D_purr.wav',  data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G1_03_15D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G3_02_4Q_purr.wav',  data: { correct_response: 'QUESTION' }},
 { stimulus: 'G3_07_30Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G3_09_8Q_purr.wav',  data: { correct_response: 'QUESTION' }},
 { stimulus: 'G3_08_35Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G3_02_28Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G3_09_22Q_purr.wav', data: { correct_response: 'QUESTION' }},
 { stimulus: 'G3_02_8Q_purr.wav',  data: { correct_response: 'QUESTION' }},
 { stimulus: 'G3_07_4D_purr.wav',  data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G3_08_27D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G3_10_31D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G3_12_15D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G3_11_27D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G3_07_21D_purr.wav', data: { correct_response: 'STATEMENT' }},
 { stimulus: 'G3_11_9D_purr.wav',  data: { correct_response: 'STATEMENT' }}
];
```

```
var practice_end_message = {
 type: 'html-button-response',
 stimulus:
   "<p><strong>Training is over.</strong></p>" +
   "<p>Ready? Let's begin with the task!</p>" +
   "<p>Press <strong>OK</strong> to continue.</p>",
 choices: ['OK']
};

timeline.push(practice_end_message);

// create test
var test = {
 type: 'audio-button-response',
 stimulus: jsPsych.timelineVariable('stimulus'),
 choices: ['QUESTION', 'STATEMENT'],
 response_allowed_while_playing: false, // ensure buttons are disabled until audio
ends
 trial_ends_after_audio: false,
 post_trial_gap: 700,
 trial_duration: null,
 data: jsPsych.timelineVariable('data'),

 on_load: function() {
   // Store reference to time when audio ends
   let audioEndedAt = null;
   const audio = document.querySelector('audio');
   const buttons = document.querySelectorAll('.jspsych-btn');

   // Just to be safe, disable all buttons immediately
   buttons.forEach(btn => btn.disabled = true);

   // Attach event handler
   if(audio){
    audio.addEventListener('ended', function() {
      audioEndedAt = performance.now();
      buttons.forEach(btn => btn.disabled = false); // Enable response
      // Save timestamp directly to the trial object for later access
      jsPsych.data.get().addToLast({audio_end_time: audioEndedAt});
    });
   }
 },

 on_finish: function(data) {
   // Find audio_end_time from trial data
   const audio_end_time = data.audio_end_time;
   if (audio_end_time !== undefined && data.rt !== null) {
    data.rt_adjusted = (data.start_time + data.rt) - audio_end_time;
    if (data.rt_adjusted < 0) data.rt_adjusted = 0; // in case of weirdness
```

```
    } else {
     data.rt_adjusted = null;
    }

    // Score response
    const correct_choice = data.correct_response === 'QUESTION' ? 0 : 1;
    data.correct = data.response == correct_choice;
  }
};

// create procedure
var procedure = {
 timeline: [fixation,
   test],
 timeline_variables: stimuli,
 randomize_order: true
};

// put procedure in timeline
timeline.push(procedure);

// create end_experiment
var end = {
 type: 'html-button-response',
 stimulus: '<p>End of the test.</p>' +
 '<p>Thanks for participating!</p>' +
 '<p>Click <strong>OK</strong> to end.</p>',
 trial_duration: 5000,
 choices: ['OK']
};
// put end in timeline
timeline.push(end);

// run experiment
jsPsych.init({
 timeline: timeline,
 preload_audio: [
 // training audio files
 'G1_02_30Q_purr.wav',
 'G1_03_6Q_purr.wav',
 'G1_02_21D_purr.wav',

 // test audio files
 'G1_02_12Q_purr.wav',
 'G1_08_22Q_purr.wav',
 'G1_08_4Q_purr.wav',
 'G1_08_8Q_purr.wav',
 'G1_03_24Q_purr.wav',
 'G1_04_30Q_purr.wav',
 'G1_04_35Q_purr.wav',
```

```
 'G1_02_15D_purr.wav',
 'G1_03_4D_purr.wav',
 'G1_04_27D_purr.wav',
 'G1_07_11D_purr.wav',
 'G1_07_25D_purr.wav',
 'G1_02_4D_purr.wav',
 'G1_03_15D_purr.wav',
 'G3_02_4Q_purr.wav',
 'G3_07_30Q_purr.wav',
 'G3_09_8Q_purr.wav',
 'G3_08_35Q_purr.wav',
 'G3_02_28Q_purr.wav',
 'G3_09_22Q_purr.wav',
 'G3_02_8Q_purr.wav',
 'G3_07_4D_purr.wav',
 'G3_08_27D_purr.wav',
 'G3_02_25D_purr.wav',
 'G3_12_15D_purr.wav',
 'G3_11_27D_purr.wav',
 'G3_07_21D_purr.wav',
 'G3_11_9D_purr.wav'
],
 use_webaudio: false,
                                                            });
```

APPENDIX 8: perception stimuli distribution for Experiments 1 and 1.2

AUDIO NUMBER	PARTICIPANT/SENTENCE NUMBER/SENTENCE TYPE/REPETITION NUMBER	L1
AUDIO 1	Participant E 10Q2	BP
AUDIO 2	Participant A 2D3	BP
AUDIO 3	Participant E 11Q3	BP
AUDIO 4	Participant B 5D1	BP
AUDIO 5	Participant C 9Q2	BP
AUDIO 6	Participant D 10Q2	AE
AUDIO 7	Participant F 1D1	BP
AUDIO 8	Participant D 11Q1	BP
AUDIO 9	Participant C 4D3	BP
AUDIO 10	Participant E 10Q2	BP
AUDIO 11	Participant H 7Q2	BP
AUDIO 12	Participant I 1D2	AE
AUDIO 13	Participant F 10Q2	BP
AUDIO 14	Participant J 8Q2	BE
AUDIO 15	Participant K 2D2	BE
AUDIO 16	Participant E 12Q3	BP
AUDIO 17	Participant J 3D1	BE
AUDIO 18	Participant G 9Q3	BP
AUDIO 19	Participant L 4D3	AE
AUDIO 20	Participant J 7Q3	BE
AUDIO 21	Participant K 11Q1	BE
AUDIO 22	Participant J 5D3	BE
AUDIO 23	Participant H 9Q2	BP
AUDIO 24	Participant M 8Q2	BP

AUDIO 25	Participant K 6D1	BE
AUDIO 26	Participant L 7Q2	AE
AUDIO 27	Participant C 11Q2	BP