Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart

Pfaffenwaldring 5B

D-70569 Stuttgart

Master Thesis

Melodies of the Majority Language: Investigating Prosodic Phenomena in Majority English

Tanishq Quraishi

Studiengang:	M.Sc. Computational Linguistics
Prüfer*innen:	Prof. Sebastian Padó
	Prof. Sabine Zerbian
Betreuer:	Yulia Zuban
Beginn der Arbeit:	06.08.2024
Ende der Arbeit:	06.02.2025

Erklärung (Statement of Authorship)

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und dabei keine andere als die angegebene Literatur verwendet habe. Alle Zitate und sinngemäßen Entlehnungen sind als solche unter genauer Angabe der Quelle gekennzeichnet. Die eingereichte Arbeit ist weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen. Sie ist weder vollständig noch in Teilen bereits veröffentlicht. Die beigefügte elektronische Version stimmt mit dem Druckexemplar überein.

[1]

(Tanishq Quraishi)

[1] Non-binding translation for convenience: This thesis is the result of my own independent work, and any material from work of others which is used either verbatim or indirectly in the text is credited to the author including details about the exact source in the text. This work has not been part of any other previous examination, neither completely nor in parts. It has neither completely nor partially been published before. The submitted electronic version is identical to this print version.

Abstract

This thesis is a comparative study of majority and monolingual English speaker groups on three key prosodic phenomena, namely pitch accents, boundary tones and intonational phrases. To that end, the data employed is extracted from a multilingual corpus consisting of comparable speaker data with the necessary annotations. The nature of the dataset is on of semi-spontaneous narratives elicited in a semi-naturalistic manner. Three binomial generalized linear mixed effects models and one linear mixed effect model encode fixed and random effects in order to analyse the influence and interaction of various factors such as bilingualism, gender and formality on linguistic intonation. The statistical analysis performed on semi-spontaneous speech, reveals a significant influence of formality on higher monotonal pitch accent usage. The influence of gender is significant on the likelihood of a speaker using a high boundary tone in their utterance. The role of bilingualism and formality influence pitch accent placements on content words, and lastly formality impacts the length of an intonational phrase, where formal contexts consist of shorter phrases.

Contents

List of Figures

List of Tables

Listings

1 Introduction

Heritage language speakers are speakers who grow up in a bilingual or multilingual home with a heritage language, in addition to the majority language dominant in the larger society around them (Wiese et al., 2022, p.210). The present study investigates speakers from the heritage speaker cohort, and therefore it is important to learn more about them in this introduction. Heritage speakers are typically considered native speakers of both their languages, although the term native speaker is often debated in the field, with little to no convergence on a definition (Wiese et al., 2022, p.210). The terms heritage speaker and heritage language have originally been used with reference to larger minorities, for e.g. immigrants from Latin-America, Russia or South Korea to the United States of America (Kupisch et al., n.d., p.348).

Heritage speakers may be exposed to the majority language either as simultaneous acquirers of two first languages or when they begin preschool (Pashkova, 2024, pg.4). One could understand the space heritage speaker occupy by contextualizing them to bilinguals. Heritage speakers pose a specific case of bilinguals of a minority or heritage language and a majority language (Böttcher & Zellers, 2024, p.4) and they fall somewhere in the middle of second language learners and monolingually raised speakers on a native-speaker continuum (Wiese et al., 2022).

Naturally, as heritage speakers become familiar with the majority language (also referred to as the dominant language), they begin to employ it in far more communicative situations than their heritage language, and the latter is relegated to the four walls of their home. In order to understand what a communicative situation is, consider the following example: asking a friend at university regarding an upcoming lecture juxtaposed with a request to your boss for vacation. The two communicative situations are distinct and there is a pre-determined demand for two varying repertoires for each situation. Each situation

is characterized by the relationship held between the speakers, the location such as university or a place of work and other such day-to-day factors. Thus, a communicative situation is determined by the setting of a social activity and different situational characteristics, characterized largely by what speakers consider relevant (Wiese, 2021, p.6).

Comparative studies on heritage speakers have historically used data from monolingual speakers, to understand convergences or, more predominantly, divergences in heritage speaker grammar. Monolingual speakers as a cohort have traditionally been considered native speakers of their language, formally known to be the gold standard against which linguistic competence must be measured. However, in the present landscape of linguistic research, monolinguals and heritage speakers are treated with parity, wherein monolinguals simply represent a control group, in order to test various linguistic phenomena occurring in heritage speaker languages (Wiese, 2021, pg.2). Several studies demonstrate not only the influence of the heritage language on the majority language, but also the other way around and they are briefly discussed next.

Dehé's (2018) study investigates the intonation of polar questions in North American heritage Icelandic and compares it to the intonation of polar questions in Icelandic spoken in Iceland as well as North American English spoken in Manitoba, Canada. Their findings reveal that intonational features characteristic of Icelandic polar questions are present to a considerable extent in the heritage variety of the language (Dehé, 2018, p.251). This suggests that speakers of North American heritage Icelandic have retained salient prosodic patterns from their heritage language (Dehé, 2018, see overview)

Interestingly, the analysis also indicates that intonational features typical of North American English polar questions are frequently observed in the heritage data. Furthermore, the study identified a tendency for intonational characteristics of Icelandic polar questions to emerge in the English polar questions produced by speakers with an Icelandic heritage background, more so than in the English of those without such a linguistic back-

ground. This points to bidirectional prosodic influence, where the majority language exerts influence on the heritage language. These findings thus contribute to the growing body of research on the linguistic features of heritage languages, highlighting how language contact and bilingualism can shape the prosodic development of a heritage variety.

1.1 The Influence of Socio-demographic Factors on Language Use

Non linguistic, socio-demographic factors influence language use. Polinsky (2018) gives an overview of Kiezdeutsch, a way of speaking that emerged among young people in multiethnic urban areas of Germany. The word Kiezdeutsch (literally "Hood German," where hood refers to an ethnic enclave in a larger city), refers to the variety of German used by younger speakers in large cities. The language itself is characterized by informal, everyday communication regardless of the speaker's ethnic background (Polinsky, 2018, p.148).

Languages encode the notion of formality in many ways. Pragmatics are closely intertwined with cultural norms and social practices, and can be found in V(ous)-T(u) contrast in languages like French used to indicate level of formality, politeness and intimacy (Böttcher & Zellers, 2024, p.4). Similarly in German, the "du" (second person singular pronoun) form may be used with friends and peers, while the "sie" (second person singular and plural pronoun) form is considered a polite form used in formal contexts, for instance with elders and those in positions of authority. Typically, in formal contexts speakers are assumed to be more careful regarding speech planning; they probably try to be as precise as possible, and they may also plan their productions carefully in regards to speech style and register (Böttcher & Zellers, 2024, p.15).

Previous majority language research (experimental or not) rarely considers register variation, which is also an integral part of everyday language use (Pashkova, 2024, p.11).

As aforementioned, register variation would stem from varying communicative situations that demand different registers for a given situation. In the investigation of clause type distribution, Pashkova et al. (2024) found similarities between English dominant German heritage speakers and English monolinguals speakers. They patterned similarly in associating certain clause types with certain registers- both speaker groups produced more independent main clauses in written narratives than spoken ones, more coordinate main clauses in informal written narratives than informal spoken ones, and more subordinate clauses in formal written narratives than informal written ones (Pashkova, 2024, p.39).

In a study done on the same language constellation as this thesis (heritage-Russian speakers who are English dominant and monolingual English speakers) researchers examined filler particles (words or phrases like ”um,” ”uh,” or ”like” that often don't have a significant meaning but function to fill pauses or connect phrases in speech) produced by German and Russian heritage speakers in majority English narratives as compared to English monolingual speakers (Böttcher & Zellers, 2024). While the results indicated several similarities between heritage speakers and their monolingual counterparts (in frequency of filler particles, their duration and segmental form), a difference in the pitch levels of filler particles was found. Since Russian heritage speakers were found to produce significantly larger slopes in rising intonation contours than German heritage speakers and English monolingual speakers, the authors suggest that this difference might be due to cross-linguistic influence from heritage Russian (Böttcher & Zellers, 2024).

In an attempt to understand Russian intonation, moving from a purely phonetic analysis to one that incorporates social and situational factors, Kachkovskaia et al. (2022) investigate the variability of peak alignment in Russian rise-fall tunes (H*L), a common melodic pattern in non-final intonational phrases and yes-no questions. The study found that in general, female speakers shift the peaks more often than maled (Kachkovskaia, Zimina, Portnova, & Kocharov, 2022, p.864). This study opens up the question of gender

as a socio-demographic factor influencing linguistic phenomena.

1.2 Research Goals

The term intonation refers to a means for conveying information in speech which is independent of the words, and central to intonation is variation in speaking pitch, where intonation is often thought of as the use of pitch over the domain of the utterance (Nolan, 2020, p.385). The word prosody is convenient as a more general term to include patterns of pitch, timing, loudness, and voice quality (Nolan, 2020, p.385). However, the proposed research carried out uses intonation, prosody and occasionally melody interchangeably. There is evidence to believe that there exists prosodic variation between heritage speakers and monolingual speakers in the majority language. As illustrated through the studies in the introduction, the area of morpho-syntax is well studied in several languages, but the phonetics of heritage speaker speech are not well represented in heritage language research (Polinsky, 2018, an overview).

In this thesis, a comparative analysis has been undertaken between two speaker groups: Russian immigrants in the United States of America, who acquired English as a majority language, and monolingual speakers of English. Through the deployment of (generalized) mixed linear models, various factors and their impact are analysed on English-dominant and monolingual speaker data. Since less is known about gender as a variable that affects prosody (Clopper & Smiljanić, 2011, p.3), this research also includes an assessment of gender as a factor to further contribute to the understanding of prosodic patterns in English-dominant heritage speakers.

The aim of this research is to describe trends in the majority language of heritage speakers, and assess whether those trends align or mis-align with monolingual speakers. The primary research objective is to compare specific prosodic phenomena in majority and

monolingual English, within the scope of the following:

- inventory and type of pitch accents (monotonal and bitonal pitch accents)
- inventory and type of boundary tones (high and low boundary tones)
- number of pitch accents on content words
- length of intonational phrases

By examining global intonational features, the proposed research aims to provide a more comprehensive understanding of prosodic patterns across a speaker's language repertoire. The above prosodic phenomena are explained at length in the Background section. This research also follows the presentation in Zerbian et al. (2024), reviewed in detail in the literature review chapter. The expectation across all four research questions and hypotheses for this thesis is that majority English speakers will pattern similarly as they also happen to be heritage Russian speakers (the main speaker group assessed in Zerbian et al. (2024)).

1.3 Research Questions and Hypotheses

1. R.Q.1. How do majority English speakers differ from monolingual English speakers in their production of monotonal pitch accents?
    (a) H.1. Majority English speakers are likely to produce more monotonal pitch accents than their monolingual counterparts.
2. R.Q.2 What is the effect of gender on high boundary tones?
    (a) H.2. Female speakers are more likely to produce high boundary tones than male speakers.

3. R.Q.3 What is the likelihood of a pitch accent on a content word by a speaker considering formality and bilingualism as factors?

    (a) H.3. The formal mode is more likely to elicit a pitch accent on a content word than the informal mode. Bilingualism will influence accented content words.

4. R.Q.4 What factors emerge significant in their impact on the length of intonational phrases?

    (a) H.4. Formality will impact the length of intonational phrases.

1.4 Structure of this Thesis

In the background section of this thesis, key prosodic phenomena such as pitch accents, boundary tones and intonational phrases are described in detail along with examples. Following that, the literature review chapter discusses critical pieces of research that have contributed to the field of heritage languages and relevant to this body of work, particularly Zerbian et al. (2024) which is the primary motivation of this thesis. Next, the materials section outlines the main corpus used in this thesis, including its collection and creation. The methodology section covers the extraction of the majority and monolingual English data include data preparation steps and model design. The statistical modeling used to identify factors influencing prosodic variation and their interaction, its analysis and the results are in the results and analysis section. This research is concluded with a discussion of the results, limitations of this research and suggestions for future work.

2 Background

2.1 Key Prosodic Phenomena

The introduction contextualised heritage speakers with respect to bilinguals and native speakers. Heritage language speech is distinct from that of monolingual speakers as well as L2 speakers on suprasegmental contrasts, even though heritage speakers pattern differently on different dimensions, sometimes falling closer to one group or the other and sometimes being distinct from both native and L2 speakers (Polinsky, 2018, p.148). To that end, it is important to understand these suprasegmental features, their grammatical function and types. Intonational tones appear either on (or near) accented syllables, in which case they are (intonational) pitch accents, or at the edges of prosodic constituents, like the intonational phrase, in which case they are boundary tones (Gussenhoven, 2004, p.22). The intonation of a string can be characterized as a series of L (low) and H (high) tones, organized into pitch accents, phrase accents and boundary tones (Gussenhoven, 2004, 67).

In order to transcribe these significant prosodical units, ToBI as a system, and was first developed to standardise prosodic annotations for English (Silverman et al., 1992, an overview). Overtime, the system, which stands for Tones and Break Indices has been adapted for other languages. Tones can be classified in many segments with the help of ToBI parameters. All H tones could be additionally upstepped (i.e., pitch range is expanded compared to a preceding high tone, indicated by ˆ) or downstepped (i.e., pitch range is compressed compared to a preceding high tone, indicated by !) (Zerbian, Zuban, & Klotz, 2024, p.8). All examples and annotations of pitch accents and boundary tones in this paper belong to the ToBI design system.

In languages like English and Dutch, there is gradience in the way words are emphasized. The height of the pitch peak corresponds to its perceived emphasis (Gussenhoven,

2004, p.52). Take for example the following sentences in English annotated for pitch accents and boundary tones using the TobI system of annotation:

1. *Ketchup is a vegetable.*
   L* L-L%
2. *Ketchup is a vegetable?*
   L* L-H%

In this example, a high boundary tone at the end of the utterance is elicited when the statement is posed as a question. The statement itself has a L-L%, indicating a low pitch throughout the utterance, with a final fall. Human beings are tuned to understanding and differentiating between content that is seemingly, visually identical with the help of suprasegmental features that are superimposed on these two utterances. To that end, the key prosodic phenomena that merit closer examination are highlighted.

Pitch Accents are local prominences in the intonation of an utterance, typically associated with a stressed syllable (Ladd, 2008, p.54). They are characterized by a change in the fundamental frequency (F0) of the voice, which is perceived as a change in pitch (Pierrehumbert, 1980, p.31). Pitch accents may be used to mark important information, focus or emphasis in an utterance (Beckman & Pierrehumbert, 1986, p.285). English typically consists of monotonal and bitonal pitch accents (Gussenhoven, 2004, p.124). Examples of common pitch accent types in English include high (H*), low (L*), rising (L+H*) and falling (H*+L), where the * aligns with a stressed syllable (Pierrehumbert, 1980, p.23). It is important to note that pitch accents differ from commonly known lexical pitch occuring in languages such as Japanese (Ladd, 2008, p.49). According to Pierrehumbert (1980)'s notation, a bitonal pitch accent such as L+H*, refers to a leading low accent followed by a high starred tone.

Boundary Tones are pitch movements that occur at the end of intonational phrases, marking the end of a prosodic unit (Pierrehumbert, 1980, p.16) and are associated with the final syllable or word of the intonational phrase (Ladd, 2008, p.88). Boundary tones can be low (L%), high (H%) or rising (LH%) and they convey information about the speaker's intent, such as continuation, finality or question (Pierrehumbert, 1980, p.82, 106).

Intonational Phrases are the largest prosodic units in spoken language, consisting of one or more intermediate phrases (Beckman & Pierrehumbert, 1986). Intonational phrases are bounded by strong prosodic breaks and are characterized by a coherent intonation contour (Ladd, 2008, p.19). The length of an IP is typically measured in terms of the number of words or syllables it contains (Pierrehumbert, 1980, p.64). Typically, the boundaries of an intonational phrase are marked by boundary tones and they may also contain one or more pitch accents (Pierrehumbert, 1980, p.22).

Table 1: Example of pitch accent, boundary tone and intonational phrase annotations from the RUEG corpus

norm	**word_pa**	**word_bt**
hello	L+H*	H-H%
calling	L+H*	None
in	None	None
connection	!H*	None
with	None	L-L%
a	None	None
case	H*	None
number	None	None
F16	L*	None
F16	L+^H*	H-H%
providing	L*	None
a	None	None
witness	L*	None
report	L+H*	H-H%

The autosegmental-metrical (AM) theory provides a basis for mapping between categorically distinct elements and continuous acoustic structures (Ladd, 2008, p.43). The AM

theory explains the relationship between pitch accents and boundary tones, where the two phenomena are grouped as tonal events allowing intonation to be analyzed as sequences of discrete intonational events and the two main types of tonal events are pitch accents and edge tones (Zerbian et al., 2024, p.3). Pitch accents are prominence-lending pitch excursions associated with metrically strong syllables while edge tones are tones associated with the periphery of a prosodic domain (Ladd, 2008, p.47). In AM approaches, intonational contours can be divided into large units known as intonation phrases (IPs) (Nance, Kirkham, & Groarke, 2018, p.281). The relationship among the three phenomena may be understood by centralizing an intonational phrase. This large chunk of information consists of pitch accents within it, to stress or focus on different parts of the phrase, while the boundary tone signals some form of break, pause or finality to the given intonational phrase.

In Table 1, an example from the RUEG corpus is generated to illustrate the relationship between pitch accents and boundary tones, given a set of utterances. Since an intonational phrase is delimited by a boundary tone, the following intonational phrases can be found in the table-"hello", "calling in connection with", "a case number F16 F16" (possibly a repetition by the speaker), "providing a witness report". Furthermore, the word "connection" is annotated for !H*, which entails that the pitch range is compressed before a high tone.

The motivating factor for choosing prosody as a domain involves the sentiment that prosodic distinctions, as opposed to other grammatical domains are much more subject to change, variation, and loss (Polinsky, 2018, p.116). Creating a tonal inventory, which is a set of distinct pitch patterns or tones in a language provides insightful information regarding the language. Assessing the placement of pitch accents will aid in determining linguistic differences or similarities between speaker groups in their conveyance of meaning. Furthermore, an intriguing area of investigation is factors such as a speaker's heritage language, gender or other characteristics and whether they influence the usage of the afore-

mentioned prosodic phenomena. Substantial research has delved into canonical prosodic manifestations across various English dialects, from Belfast English to Standard British English (Gussenhoven, 2004, p.54). However, the exploration of how heritage speakers may diverge from or converge towards the canonical prosodic forms used by monolingual speakers is an important space for further probing.

3 Literature Review

> Heritage speakers do not pattern like native speakers; rather, they consistently get recognized as a separate group. (Polinsky, 2018, p.116)

The example in the introduction involved two situations, one where the speaker asks a friend at university regarding an upcoming lecture, and two, when a speaker requests their boss for vacation. Closely-proximal instances of a sociolinguistic variable are more likely to occur under similar social-contextual circumstances than two instances that are further apart, and thus are more likely to have matching outcomes (Tamminga, Ahern, & Ecay, n.d., p.2). It is safe to assume that language use would pattern similarly by a speaker at university with their peers. Existing research has found a profound influence of register on linguistic choices so much so that it has been recommended that majority language studies that use (semi) naturalistic data would benefit from specifying the register of elicited data and/or including it into the statistical analysis (Pashkova, 2024, p.152, 154). The influence of formality as a sociolinguistic variable is important to highlight.

3.1 Influence of Formality on Speaker Intonation

As illustrated in the introduction, heritage speakers are mostly exposed to informal language at home through oral interactions with family members and often have limited access to formal registers in their heritage language (Wiese et al., 2022, p.224). In the formal contexts speakers are assumed to be more careful regarding speech planning; they probably try to be as precise as possible, and they may also plan their productions carefully in regards to speech style and register (Böttcher & Zellers, 2024, p.15). It is therefore reasonable to assume that similar outcomes are expected in the intonation of speakers in formal versus informal settings. In majority language use, non-canonical patterns were found to

be dominant in spoken and/or informal registers- specific to majority-English, it was found that spoken registers featured more non-canonical article choices than written registers in both monolingual and bilingual speakers (Wiese et al., 2022, p.225). This opens up a large space for investigation, particularly for the present study.

Findings from Pashkova (2024) illustrate that while there aren't many patterned differences across different linguistic phenomena (use of pronouns and null anaphora for subjects of coordinate finite clauses, the use of modified and non-modified referring expressions, the use of several clause types and the of left dislocations in the spoken mode) between majority and monolingual speakers in English, there are certainly variations when formality comes into play. It was found that Russian and Turkish heritage speakers produced more noun-headed noun phrases and fewer pronouns than English monolinguals in some informal narratives (Pashkova, 2024, p.152). Aside from sociolinguistic and sociodemographic (such as gender, described later) reasons, there is stronger evidence for the intonational phenomena chosen to be investigated in this research. Below, the central study off which this research was born out of is discussed.

3.2 Intonation in Monolingual and Heritage Russian in U.S

The research objective is primarily informed by findings in Zerbian et al. (2024), who investigate the inventory and frequency of pitch accents and boundary tones, and the length of intonational phrases in heritage and monolingual Russian. The primary contributions made by their paper are:

1. The creation of a Russian subcorpora known as RuPro from the RUEG corpus, rich with prosodic annotations including intonational phrases, pitch accents and boundary tones in monolingual and heritage Russian.

2. A comparative analysis of the aforementioned prosodic phenomena between her-

itage and monolingual Russian speakers.

As for pitch accents, heritage speakers accent content words more frequently than monolingually raised speakers (Zerbian et al., 2024, p.20). With regards to boundary tones, it was found that there is an equal distribution of high and low boundary tones across both speaker groups. Formality was found to be a decisive factor in the length of intonational phrases. Despite these variations, no evidence was found for a reduced tonal inventory among heritage speakers, contrasting earlier assumptions about bilingual simplification. Overall, Zerbian et al. (2024) suggest that the speaker groups are similar to each other in some aspects of intonation (e.g. in the inventory of pitch accents and boundary tones). However, monolingual Russian speakers produce longer intonational phrases in informal narrations and heritage speakers show no difference in IP length between formal and informal settings. Results are analysed from the lens of register differentiation and prosodic density across speaker groups.

The following thesis proposal is in cooperation with the Research Unit "Emerging Grammars in Language Contact Situations: A Comparative Approach" (RUEG). A similar task will be undertaken for majority and monolingual English, investigating the aforementioned prosodic phenomena. However, prosodic annotations for pitch accents and boundary tones are readily available in the case of English data, and do not require the creation of a RuPro like corpus. For the purpose of this thesis, intonational phrases were manually identified with the help of word boundary tones. In order to further extend this line of inquiry, a comparative analysis is carried out on three major prosodic phenomena occurring in majority and monolingual English, however the novel contribution of this research is the assessment of gender as a factor influencing intonation. The study is backed by powerful statistical models that can capture the influence and interaction of a wide range of factors.

The general expectation across all four research questions and hypotheses for this the-

sis is an alignment in the findings with the study on Russian intonation, particularly given that the majority English speaker cohort are also Russian heritage speakers.

3.2.1 The Investigation of Pitch Accents

In their study on Russian intonation, Zerbian et al. (2024) specifically found differences between the amount of monotonal and bitonal pitch accents produced by each speaker group. Across the groups, there were slightly more monotonal accents by heritage Russian speakers (55.6% of all pitch accents) than by their monolingual counterparts (48.2% of all pitch accents) (Zerbian et al., 2024, p.13). Their findings come forth, despite citing that some scholars consider the monotonal H* to occur rarely or close to never in Standard Russian. The paper defends its stance by considering the nature of the dataset used (spontaneous narrations), leading to a reduction in pitch variation. It is thus necessary to examine if there exists a similar difference between majority English and monolingual English speakers, especially considering that English typically does consist of monotonal and bitonal pitch accents (Gussenhoven, 2004, p.124). R.Q.1 and H.1 are based on these findings, with an inquiry into the heritage Russian speaker cohort who have evidence based reasons to also produce more monotonal pitch accents in their majority language than their monolingual counterparts.

3.2.2 The Investigation of Pitch Accents on Content Words

The corpus consists of part of speech annotations including function words, connectors, conjunctions, and interjections, however pitch accents are structurally not expected on these words and might be a consequence of speech planning or processing issues in unprepared speech (Zerbian et al., 2024, p.14). Following the presentation in Zerbian et al. (2024), the main area of interest is in identifying the frequency of use of structurally licit

pitch accents, i.e., on content words, i.e., nouns, proper nouns, adjectives, adverbs, verbs and numerals. In their research on heritage Russian and monolingual speakers, it was found that heritage speakers produce more pitch accents on content words than monolingual speakers. There was a main effect of formality, i.e., both speaker groups produced more pitch accents on verbs in informal communicative situations. R.Q.3. and H.3. are based on these findings.

3.2.3 The Investigation of Intonational Phrases

Zerbian et al. (2024) specifically found that intonational phrases are generally shorter in formal communicative situations than in informal ones. However, only monolingually raised speakers of Russian showed a register effect, by producing lengthier intonational phrases in the informal than in the formal mode. This finding yields an impetus to investigate whether there are possible differences in phrase lengths between English-dominant and monolingual English speakers, and the possible of influence of formality, leading to R.Q.4 and H.4.

The next section opens up research with reference to the second research question and corresponding hypothesis regarding gender.

3.3 Influence of Gender on Speaker Intonation

> People will project what they perceive as their social role in their voice, and gender roles are no exception (Gussenhoven, 2004, p.51).

The influence of gender was briefly illustrated in the introduction with a study on social variability of peak alignment in Russian (Kachkovskaia et al., 2022). The study compares the nature of "misplaced" melodic peaks with the phenomenon of uptalk (Kachkovskaia et al., 2022, p.862). One of the most prominent differences in gendered speech has been

found in utterance final rises. An utterance final rise, high rising terminal (HRT) or uptalk is a contour that rises and then keeps on rising until the end of the phrase to the uppermost reaches of a speaker's commonly used pitch range (Ladd, 2008, p.125). In order to investigate whether there exists dialectal intonational variation in the frequency and phonetic properties of utterance-final rising intonation, Armstrong et al. (2015) employed a mixed effects model on naturalistic narrative data. Contrary to expectation, the study found no regional differences in uptalk in American English, produced by Southern Californian and Massachusetts speakers. Since uptalk is characterized by a rising pitch specifically at the end of declarative sentences and was commonly associated with female speech (Clopper & Smiljanić, 2011, p.238), contrary to stereotype, the result data showed no significant effect of gender for frequency of rise use between females and males, however, significant gender differences emerged when SoCal females produced both the longest and steepest rises (Armstrong, Piccinini, & Ritchart, 2015, p.4).

Since there is longstanding speculation regarding the impact of gender, and contradictory findings, in their paper, Nance et al. (2018), looked to find evidence suggesting sociolinguistic variation in Liverpool English intonation, a dialect in the Urban Northern British (UNB) group. To that end, the authors collected data from 5 male and 4 female speakers from Liverpool. Participants were asked to read sentences that were designed to elicit specific intonational patterns, in addition to a spontaneous speech task, wherein a commentary on a silent Mr. Bean video (only the read sentences were analyzed in this study) was required. In their analysis of pitch accents and boundary tones, Nance et al. (2018) utilized an adapted version of ToBI known as GlaToBI. The study employed a logistic mixed effects regression model with contour variant as outcome variable, sentence type and gender as fixed effects, and speaker and accented word as random intercepts.

In terms of gender-based variation, they found significant differences between male and female speakers in the proportion of L* H-L% contour (Nance et al., 2018, p.287).

Nance et al. (2018) explain outliers in their result, suggesting that participants in their study "were reacting to the fairly formal context of reading sentences from a computer screen" (Nance et al., 2018, p.289) suggesting that certain study designs automatically yield a more formal atmosphere for participants. It would therefore, also benefit to test for interaction between gender and formality for the present study, given that the nature of elicitation for the data used, is similar to the elicitation technique used by them.

Falling in line with variationist research, gender differences were also found on the distribution of pitch accents and phrasal boundary tone combinations (Clopper & Smiljanić, 2011). Clopper et al. (2011) found that gender and dialect influenced pitch accent preferences. In order to arrive at this conclusion, the study analyzed read speech from 20 speakers (10 male, 10 female) aged 18–25, evenly split between two dialect regions-Midland and Southern English. Using the ToBI transcription system, prosodic features were coded and analyzed with mixed-effects regression models to identify patterns related to dialect, gender, and the passage itself. It was found that, for instance, Southern female speakers preferred rising pitch accents (L*+H), indicating a potential gender-specific prosodic style. A main effect of gender was also observed for the L-H% phrasal-boundary tone combination (z = -2.2, p = .027) , where female speakers across both dialects were more likely to use rising boundary tones such as L-H%, often associated with continuation or uncertainty (Clopper & Smiljanić, 2011, p. 8–9). The study concludes that gender identity is in part encoded through prosodic patterns and suprasegmental cues (Clopper & Smiljanić, 2011, p.3).

3.3.1 The Investigation of Boundary Tones

Similar to the approaches by Nance et al. (2018) and Clopper et al. (2011), this research also undertakes mixed effect modelling in order to investigate the influence of socio-

demographic factors on certain prosodic phenomena. However, the main gap in research is that the investigation of gender variation in English were primarily conducted among monolingual English speakers in their native language (Clopper & Smiljanić, 2011, pg.4), leaving an open query space for majority English heritage speakers. Secondly, the participants were a homogeneous group of people hailing from similar socioeconomic background (Clopper & Smiljanić, 2011, p.4), lacking the demographic diversity of participants in the present study. An observable shortcoming across the literature on gender variation in intonation is the sample size, often averaging around 14 speakers (Clopper & Smiljanić, 2011, p.238) (Nance et al., 2018, p.282) (Armstrong et al., 2015, p.2) (Kachkovskaia et al., 2022, p.862), underscoring the need for a study with larger, diverse samples to draw more robust conclusions. The present thesis grounds itself in existing work, therefore, R.Q.2 and H.2 are based on above findings. There is an expectation that females will produce higher boundary tones, given findings regarding the frequent usage of the rise L-H%, and phenomena such as uptalk.

4 Materials

> Corpus linguistics should not be concerned with distribution of language elements as an end in itself, rather what that distribution may signify (Chafe, 1992, p.271).

4.1 The RUEG Corpus

This study uses data that belongs to a large body of linguistic data known as the RUEG corpus which is named after the creators of the corpus, the Research Unit for "Emerging Grammars in Language Contact Situations: A Comparative Approach". The corpus contains linguistic productions from 720 speakers, of which 349 are adults and 371 are adolescents. Two of the main contributions of this corpus, that also motivate its usage for this thesis, are firstly, the possibility of matching comparisons with monolingual repertoires in countries of residence, in ancestral / "heritage" countries and across all registers. Secondly, the corpus enables large-scale comparative study for the same heritage language with different majority languages, different heritage languages with the same majority language and lastly, the same language as heritage versus as a majority language. The RUEG corpus contains language data from English, German, Greek, Russian, and Turkish speakers, enabling detailed comparisons across heritage and majority language contexts. Depending on a study' research question, the RUEG corpus lends itself to cross-linguistic research. The corpus, including RUEG instruments like stimula, elicitation order and training clip are open access (Wiese et al., n.d.). The breadth of linguistic repertoires in the RUEG corpus broadly cover:

1. Heritage and majority speakers
2. Formal and informal settings

3. Written and spoken mode

As part of the data collection, RUEG followed the Language Situations methodology (explained below) and spoken productions were recorded and transcribed. The corpus is lemmatised and contains several annotation layers, including but not limited to part-of-speech (POS) tags, dependency tags and prosodic annotations for English and Russian data. The data is accessible through ANNIS, a browser based search and visualization architecture for complex, multi layer linguistic corpora (Krause, 2019). ANNIS is a graph-based query system for deeply annotated text corpora of annotated nodes and edges forming a variety of linguistic structures, such as constituent or dependency syntax trees, coreference, rhetorical structure and parallel alignment edges, span annotations and associated multimodal data (Wiese et al., n.d.). The ANNIS interface is mainly comprised of the search form on the left and the results on the right. ANNIS search works in tandem with AQL-ANNIS Query Language and filters the annotations according to the search criteria.

4.1.1 Speakers

The speaker groups are broadly divided into heritage speakers and monolingual speakers. As stated in the introduction, heritage speakers were defined as speakers who had grown up with a family language in addition to the country's majority language (Wiese, 2021, p.4). Therefore, the RUEG corpus consists of the following participant groups:

- Monolingual speakers of German, English, Russian, Greek and Turkish
- Heritage speakers of Russian, Greek and Turkish in the US and Germany
- Heritage speakers of German in the US

The main condition ascertaining the status of a monolingual speaker is that they must use

only one language at home, i.e. the majority language of the country they were born in. The two age groups in the study are adolescents (14–18 years), and adults (22–35 years).

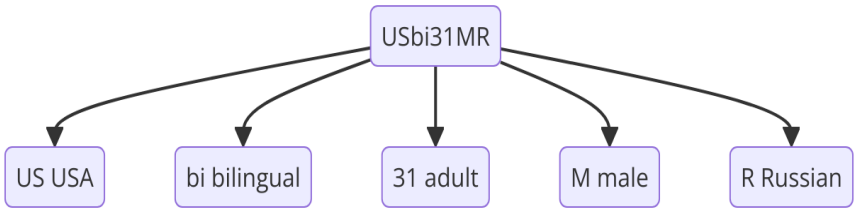

Figure 1: Encoded Speaker ID

As illustrated in the figure 1, the speaker ID has been encoded as such, with information regarding their country of origin, bilingual status, age, gender and language. This creates a unique ID for each speaker in the dataset. There are 30 participants per group.

4.1.2 Language Situations Methodology

To gather naturalistic language data across different communicative contexts, the data collection team building the RUEG corpus employed the Language Situations methodology. This methodology involves a controlled elicitation task where participants watch a non-verbal video created specifically for data collection (Wiese et al., 2022, p.211). The following description below is an overview; see (Wiese, 2020).

In the video depicting a fictitious car accident, the scene begins with two people present in a parking area, with the man bouncing a ball. Across from them, a woman with a dog unloads groceries from her car. As two cars approach the lane, the man loses control of the ball, which bounces in front of the oncoming car. The dog runs into the lane toward the ball, and the woman drops her groceries. Suddenly, the first driver hits the brakes, causing the second car to bump the rear end. In the end, the man with the ball helps the woman pick up her groceries, while the two drivers exit their cars, with one of them calling the

police (as indicated by the close-up of the emergency number on his phone).

Five versions of this video were developed for use in Germany, Greece, Russia, Turkey, and the United States (Wiese et al., 2022, p.212). To support cross-linguistic comparisons, these versions were reported to be identical except for the emergency number displayed in each country. As part of data elicitation, participants were then asked to recount the incident in both formal and informal contexts, simulating everyday communication. They are asked to imagine themselves as a witness to this event and to act out describing it in different communicative situations, yielding register-differentiated productions(Wiese, 2020, p.3).

1. Formal Contexts: For the formal mode, participants were asked to leave a voice message to a fake police hotline in addition to writing a formal report to the police. These tasks were conducted in an office-like setting to reinforce the formality of the situation. The formal environment was further supported by having the interviewer dressed in formal attire and behaving in a professional manner. An example statement submitted by a participant in the formal mode: ”Okay, this is for case F16, I just witnessed a car accident and this is to provide testimony of what had happened...”

2. Informal Contexts: For the informal mode, participants were prompted to leave a WhatsApp voice message as well as write a text to a friend. These tasks were conducted in a casual environment where the interviewer dressed informally and engaged in prior small talk (10-15 minutes) with the participant to create a relaxed, friendly atmosphere. Snacks and drinks were provided to increase the comfort level, encouraging participants to produce more spontaneous responses (Pashkova, 2024, p.34). An example statement submitted by a participant in the informal mode: ”Hey, so I just saw this very nearly terrible car accident, right in the middle of the street

just now—there were these two people..."

What makes the Language Situations methodology trump over other data elicitation methods such as grammaticality judgments or picture matching, is that naturalistic interviews involve the least amount of control, even if interviewees may consciously choose conversational topics that they find easy to talk about (Kupisch et al., n.d., p.350). This method appears suitable for the present study because the elicitation task required participants to narrate an incident they observed, resulting in semi-spontaneous speech. Previously, this methodology has proven to be useful, given the systematic comparability of the data elicited with this set-up, particularly in capturing bilingual repertoires (Wiese, 2020, p.5). Language Situations methodology provides a structured yet flexible approach, allowing for the collection of semi-naturalistic data mirroring real-world communicative situations and falls right in the middle of a spectrum that includes naturally occurring data on one end and a controlled experiment on the other. The diversity in communicative contexts—formal versus informal, spoken versus written—offers a rich foundation for analyzing register variation and bilingual language use.

4.1.3 Corpus Organization

RUEG is a multi-layer corpus architecture with multiple tokenisations. To understand the organisation of data in ANNIS, there are two broad annotation categories of relevance. The first is node annotations and the second is meta annotations as illustrated in table 2 and 3. The name column illustrates the category and the example column showcases a sample of what the output would look like. This feature may be extracted through the ANNIS user interface directly or manually inputted into the AQL query box.

For the purpose of a short overview of what an AQL query looks like, the following example illustrates the query written to extract PA on content words and its correspond-

Table 2: RUEG-EN-SPOKEN Meta Annotations in ANNIS

Sr. No	Name	Example
1	speaker-bilingual	node & meta::speaker-bilingual=”yes”
2	setting	node & meta::setting=”informal”
3	speaker-gender	node & meta::speaker-gender=”female”
4	speaker-id	node & meta::speaker-id=”USbi31MR”

ing metadata information. The query was used to retrieve nodes annotated with `norm`, `word_pa`, and `pos`, while including specific metadata. The metadata annotations, specified using the `metakeys` parameter, were `speaker-id`, `speaker-bilingual`, `speaker-gender`, `setting`. To look for specific sequences, search terms may be separated based on a predefined list of operators that specify the relation between the elements. In this example, _i_ is applied when one annotation covers a span identical to or larger than another. The operator examines whether one annotation fully contains the span of another annotation (the i stands for 'includes'), thus capturing normalized tokens, word level pitch accents and part of speech tags. In addition to the query, a set of predefined AQL parameters may be inputted while extracting specific data.

Query:

Listing 1: AQL Query for PA on Content Words

```
norm & word_pa & pos & #1 _i_ #2 _i_ #3
```

Parameters

Listing 2: AQL Parameters for PA on Content Words

```
metakeys=speaker-id, speaker-age-group,
speaker-bilingual, speaker-gender, setting
```

This query retrieves all nodes where the specified annotations (`norm`, `word_pa`, and

`pos`) are present, while the `metakeys` parameter ensures the inclusion of metadata information such as speaker ID, age group, bilingual status, gender and the formality setting.

5 Methodology

5.1 Corpus Extraction

The main language constellation relevant to this thesis is English as a majority language in the United States spoken by monolingual speakers, and by bilingual speakers with Russian as a heritage language. As for the Language Situations methodology, for the purpose of this study, formality was contrasted as data from the oral witness report to the police (formal spoken) and a voice message to a friend (informal spoken) regarding the fake car accident. The version of the corpus used to extract English data for this thesis is "RUEG-EN-SPOKEN_PREVIEW-2021-07-09" (hereon, refered to as RUEG-EN-SPOKEN) which contains spoken English data with an annotated prosodic layer. The target data for each prosodic phenomena was extracted from ANNIS using AQL. In the case of RUEG-EN-SPOKEN, the listed metadata annotations are filtered for what centrally concerns this research topic, this includes unique speaker identification numbers, gender (male, female or diverse) and the setting (formal or informal). The final data outputted from the original corpus was grouped on the basis of prosodic phenomena, namely pitch accents, pitch accents on content words, boundary tones and intonational phrases. All the data extracted had fixed meta annotations selected, based on the scope of this research and its questions. The relevant meta annotations to this study are described in table 2. For every prosodic phenomena occurring as a "node", all 4 of the meta annotations present in table 2 were selected, namely bilingualism (majority (yes) or monolingual (no)), setting (formal or informal), gender (male, female, diverse) and speaker-id.

Notable to the final data are intonational phrases as well as pitch accents on content words. Since the RUEG-EN-SPOKEN data does not include annotated intonational phrases, boundary tones at the word level proved useful in segmenting communicative

units. As illustrated in table 3 serial number 3 (pos) and 6 (word_bt), namely, the normalised word and its corresponding word boundary tone (if present) was used to extract IP. In the case of pitch accent on content words, again, serial number 3 (norm) along with 4 (pa) and 5 (pos), were extracted.

Table 3: RUEG-EN-SPOKEN Node Annotations in ANNIS

Sr. No	Name	Example
1	bt	bt="L-L%"
2	norm	norm="the"
3	pa	pa="H*"
4	pos	pos="NOUN"
5	word_bt	word_bt="H-H%"
6	word_pa	word_pa="L+H*"

5.1.1 Data Preparation

The English subcorpus extracted for the purpose of this thesis, includes approximately 20,000 normalized tokens. To ensure that the data is complete, clean and relevant, all null or missing values were discarded. Since the annotators were encouraged to be honest in their labeling, uncertain labels such as question marks were also discarded. With respect to gender, during the data collection process, one participant identified as diverse in the monolingual speaker group from the US. This speaker's data was left out from the study because there exists no comparable speaker or counterpart in the majority language speaker group in the dataset. There were 86 speakers contributing to the data for each prosodic phenomena, including pitch accents, boundary tones, pitch accents on content words and intonational phrases. After the data preparation, there are 85 speakers in total, with 53 speakers identifying as female for each experiment.

Additionally, following Zerbian et al. (2024), pitch accents on content words were

extracted for all part of speech tags. Then, 6 specific parts of speech were filtered using the Pandas library during code development. Pitch accents on function words, connectors, conjunctions and interjections were dropped because not only are they structurally not expected on these words, but they may also be a consequence of speech planning or processing issues in unprepared speech (Zerbian et al., 2024). The six categories are noun, proper noun, adjective, adverb, verb and numeral.

5.1.2 Tonal Label Processing

Upon the recommendation of Prof. Dr. Sabine Zerbian and Yulia Zuban, a minor label processing task was undertaken during the preprocessing of the data. The motivation behind this was for the following reasons:

- To reduce the number of labels to relevant labels, within the purview of this research
- Clean obvious typos and approach the prosodic phenomena from a broad annotation scope rather than narrow

In order to understand broad versus narrow labeling, the following example can be taken into consideration. The counts of L+ˆH* has been conflated with the counts of L+H*. While both labels convey similar linguistic information, L+H* is a rather coarse label, lacking the upstep information. In order to process the data, three major operations were conducted on PA and BT, in order to implement the recommended changes. The mapping was structured as a list of tuples and each pair specified a variant label mapped to its chosen form. An example of this mapping is provided in Listing 3.

1. Merge: The merge operation retained the original annotated label but was merged with another label during frequency counts. For example the counts of !H-L% were combined with the counts of H-L%.

2. Correct: Typos with a clear indication of the right label, such as H** was corrected to H*. A sub operation of correct plus merge was involved. For instance, ˆH*H was corrected to ˆH* and merged with the count of H*.

3. Discard: Obvious typos such as *?, incorrect labels like L-H% as a pitch accent annotation and ambiguous labels such as L*H* were entirely discarded from the dataset.

PA Merge Mappings

Listing 3: PA Merge Mappings

```
pa_merge_mappings = {"!H*" : "H*",
                     "L+^H*" : "L+H*",
                     "^H*": "H*",
                     "L*+^H" : "L*+H",
                     "^H*+!H" : "H*+!H",
                     "^H+L*" : "H+L*",
                     "^H*+L" : "H* + L",
                     "L+!H*" : "L+H*",
                     "!H*+L" : "H* + L",
                     "L+^*H" : "L+H*",
                     "^H*" : "H*"
}
```

5.1.3 Intonational Phrase Boundaries Creation

For the purpose of their research, Zerbian et al. (2024) brought in manual annotators to segment IP boundaries in the Russian corpus, and reported a reasonable inter-rater agreement score for the task (Zerbian et al., 2024, p.11). As stated earlier, within the scope of this thesis, intonational phrase boundaries are automatically detected at the end of a corresponding boundary tone, since tunes for an intonation phrase are comprised of one or more pitch accents and a boundary tone (Pierrehumbert, 1980, p.22). For the scope of intonational phrases, L and H which indicate intermediate phrases are discarded. The example of the phenomena is explained with a corresponding table in the background section (see Table 1).

5.2 Generalized Linear Mixed Models for Social Sciences and Linguistic Data

After the data preparation, the question regarding an appropriate model choice arises. Data in the field of social sciences is often organized in complex ways, perhaps clustered, hierarchical or correlated. They are rarely normally distributed because they are often elicited in naturalistic or semi-naturalistic ways, in order to emulate real world situations (Jolly, n.d.; Bono, Alarcón, & Blanca, n.d.; Wiese, 2020). Much like the present research, new approaches contribute to the growing importance of using powerful statistical methods, where more and more studies are investigating the relevance of non-linguistic factors such as socio-demographic ones (Coupé, n.d., p.2) on linguistic phenomena.

In order to address a wide range of types of predictor variables through linear combinations, generalized linear mixed models have proven to be useful for social sciences such as psychology (Bono et al., n.d., p.2), where one or multiple predictor variables (influencing factors) are encoded as fixed effects. The main benefit of generalized linear models (GLM) is that they allow the response variable to follow a distribution other than Gaussian (also called normal distribution). What makes a generalized linear model a generalized linear mixed model (GLMM), is when random effects are included (Coupé, n.d., p.7).

There are various ways in which a model's quality can be tested. In the field of psychology, the most commonly used method for parameter estimation is maximum likelihood estimation (MLE) (Bono et al., n.d., p.12). This estimation method is used to determine the model fit (Bono et al., n.d., p.8). To assess model quality, criteria such as log-likelihood may be employed. The log-likelihood, which is always negative, provides a measure of how well the model fits the data, the closer it is to zero, the better the model's fit (Bono et al., n.d., p.4).

Tamminga et al. (2016) discuss the issue of non-independence in linguistic data, par-

ticularly in research using naturalistic speech. Such research aims to link observed intra-speaker variability in spontaneous speech with findings from more controlled experimental studies (Tamminga et al., n.d., p.1). One key source of non-independence is the collection of multiple data points from the same individual, which is necessary for studying intra-speaker variation (Tamminga et al., n.d., p.1). While their work focuses on intra-speaker variability, the broader issue of non-independence also applies to inter-speaker variability. Given this, the incorporation of random effects in the models for this paper ensures a more accurate representation of inter-speaker differences while accounting for potential dependencies in the data. Per speaker, this data used for this research also includes multiple data points per speaker.

Overall, generalized linear mixed models possess credit to their name in assessing naturalistic data, including predictor variables as well as random effects. For all four of the hypotheses stated in this research, a linear mixed model would perhaps perform well in answering the influence of factors such as formality, bilingualism and gender on key prosodic phenomena, while keeping the individual speaker, as a random effect, in its estimation.

5.3 Modeling Design for Majority and Monolingual English

There are four specific experiments conducted for each prosodic phenomena in this research. At the onset of this study, the idea was to follow the methodology adopted by Zerbian et al. (2024), however the modelling for their research was deployed in R. In order to continue using Python as a preferred approach, a small caveat i.e. the limitation of the statsmodels library, was overcome by adopting pymer 4 0.8.2, in the following manner.

The statsmodels libary could not compile a binomial generalized mixed linear model along with random effects (as of version 0.14.2). Thus, pymer4 has been used for all the

models in this paper. It is an interface to R that connects to Python (Jolly, n.d.). The main model used for the binomial generalized linear mixed model is Lmer, a function that is typically used for multi-level models estimated using glmer() in R and this configuration requires R to be installed on the system, along with specific R packages such as lme4 (Jolly, n.d.). See the Appendix for more details on code implementation. lmer() takes as inputs the dataset and a formula specifying the predicted variable, the fixed effects and the random effects of the desired model, and outputs the estimates for the various parameters of this model (Bono et al., n.d., p.5).

In the case of the present thesis, the fixed effects range from bilingualism (speak group), gender (male or female) to formality (formal or informal), while the random effect is the individual speaker. The fixed effects are encoded based on the research question.

Experiment	**Model Type**	**Family**	**Estimation & Model Fit**
Exp.1 Pitch Accents	GLMM	Binomial	Maximum Likelihood Log-Likelihood
Exp.2 Boundary Tones	GLMM	Binomial	Maximum Likelihood Log-Likelihood
Exp.3 Pitch Accents on Content Words	GLMM	Binomial	Maximum Likelihood Log-Likelihood
Exp.4 Intonational Phrase Length	LMM	Gaussian	Maximum Likelihood Log-Likelihood

Table 4: Summary of Model Types, Family, and Estimation Methods

There are four experiments in this thesis and three out of four experiments in this research follow the same model choice. Table 4, consists of an overview regarding all the important methodological decisions for modeling. Since the number of trials (1 or 0), i.e. the presence or absence, are known for pitch accents, boundary tones and pitch accents on content words, the GLMM of choice is a binomial model. The models are fit using maximum likelihood estimation for all four experiments, since it is the only correct approach when comparing models with different fixed effects (Coupé, n.d., p.5), particularly

when pitted against restricted/residual maximum likelihood (REML) which is the other alternative provided by the chosen modeling library.

In the case of the fourth experiment, assessing the impact of various variables on the length of an intonational phrase is not binomial in nature, since the response variable is ”length of an IP”. The model of choice is a linear mixed model, again fit with maximum likelihood estimation. The fixed effects are bilingualism and formality. The random effect is speaker.

5.3.1 Experiment 1: Pitch Accents

A binomial generalized linear mixed model is fit to the data using the pymer4 library. The response variable is a monotonal versus bitonal pitch accent. The model tests for the likelihood of a monotonal pitch accent, with fixed effects of speaker group and gender, and the speaker ID is added a random effect. The interaction of formality and gender, bilingualism and gender, and bilingualism and formality are also tested. Table 5 provides a tabulated overview of the aforementioned details.

After data preparation, the effects are tested on 85 speakers, out of which 62 are majority English speakers and 53 identify as female speakers. The model is fit using maximum likelihood estimation (MLE), while the model family is binomial. Model quality is assessed using the log-likelihood.

5.3.2 Experiment 2: Boundary Tones

Similarly, a GLMM is used to test the influence of factors on high and low boundary tones. The effects are tested on 85 speakers, out of which 62 are majority English speakers and 53 identify as female speakers. The response variable is the choice of high or low boundary tone. The model predicts the probability of using a high boundary tone based on

Table 5: **Experiment 1**: Monotonal vs. Bitonal Pitch Accents Model Overview

Feature	**Description**
Formula	`pa_type_binary ~ bilingual_contrast + formality_contrast + gender_contrast + interaction_bilingual_formality + interaction_bilingual_gender + interaction_formality_gender + (1|speaker_id)`
Model Type	Binomial Generalized Linear Mixed Model
Response Variable	`pa_type_binary` (Monotonal = 1, Bitonal = 0)
Predictors	• `bilingual_contrast`: Monolingual (-1) vs. Bilingual (+1) • `formality_contrast`: Informal (-1) vs. Formal (+1) • `gender_contrast`: Male (-1) vs. Female (+1) • `interaction_bilingual_formality`: Interaction between bilingualism and formality • `interaction_bilingual_gender`: Interaction between bilingualism and gender • `interaction_formality_gender`: Interaction between formality and gender
Random Effect	`speaker_id` accounts for variability across speakers

the predictor variables and their interactions.

The model is fit using maximum likelihood estimation (MLE), while the model family is binomial. Model quality is assessed using the log-likelihood. The model predicts the binary outcome based on the main effects of bilingualism, formality, and gender, as well as the interaction effects. The random effect of unique speaker IDs is included to account for variability between speakers.

Table 6: **Experiment 2**: High vs. Low Boundary Tones Model Overview

Feature	**Description**
Formula	`boundary_tone_binary ~ bilingual_contrast + formality_contrast + gender_contrast + interaction_bilingual_formality + interaction_bilingual_gender + interaction_formality_gender + (1|speaker_id)`
Model Type	Binomial Generalized Linear Mixed Model
Response Variable	`boundary_tone_binary` (High BT = 1, Low BT = 0)
Predictors	• `bilingual_contrast`: Monolingual (-1) vs. Bilingual (+1) • `formality_contrast`: Informal (-1) vs. Formal (+1) • `gender_contrast`: Male (-1) vs. Female (+1) • `interaction_bilingual_formality`: Interaction between bilingualism and formality • `interaction_bilingual_gender`: Interaction between bilingualism and gender • `interaction_formality_gender`: Interaction between formality and gender
Random Effect	`speaker_id` accounts for variability across speakers

5.3.3 Experiment 3: Pitch Accents on Content Words:

The response variable is the presence (1) or absence (0) of a pitch accent on a content word. The effects are tested on 85 speakers, out of which 62 are majority English speakers and 53 identify as female speakers. The predictor variables are speaker group and formality. The interaction of speaker group and formality is also tested. The model is fit using maximum

likelihood estimation (MLE), while the model family is binomial. Table 7 includes all the details.

Table 7: **Experiment 3**: Pitch Accents on Content Words Model Overview

Feature	**Description**
Formula	`word_pa_binary ~ bilingual_contrast × formality_contrast + (1|speaker_id)`
Model Type	Binomial Generalized Linear Mixed Model
Response Variable	`word_pa_binary` (presence or absence of a pitch accent on content words)
Predictors	• `bilingual_contrast`: Monolingual (-1) vs. Bilingual (+1) • `formality_contrast`: Informal (-1) vs. Formal (+1) • `bilingual_contrast:formality_contrast`: Interaction term
Random Effect	`speaker_id` accounts for variability across speakers

Model quality is assessed using the log-likelihood. It is important to note that 11243 words were retained from 11169 words after discarding the irrelevant data.

5.3.4 Experiment 4: Intonational Phrases

The response variable is the length of the intonational phrase which is calculated based on the number of words per IP. The effects are tested on 85 speakers, out of which 62 are majority English speakers and 53 identify as female speakers. The model of choice is a linear mixed model (LMM), fit with maximum likelihood estimation. The fixed effects are bilingualism and formality. The random effect is speaker. The predictor variables are speaker group and formality. The interaction of speaker group and formality is also tested.

Table 8: **Experiment 4**: Intonational Phrase Length Model Overview

Feature	**Description**
Formula	`IP_length ~ bilingual_contrast * formality_contrast + (1|speaker_id)`
Model Type	Gaussian Linear Mixed Model
Response Variable	`IP_length` (Length of Intonational Phrases)
Predictors	• `bilingual_contrast`: Monolingual (-1) vs. Bilingual (+1) • `formality_contrast`: Informal (-1) vs. Formal (+1) • `bilingual_contrast:formality_contrast`: Interaction term
Random Effect	`speaker_id` accounts for variability across speakers

The random effect is the individual speaker. Since the length of an intonational phrase depends on the number of words, it is important to note that 25802 words were retained from 26403 words after data cleaning. For a quick overview, refer to table 8

6 Results and Analysis

6.1 Experiment 1: Pitch Accents

6.1.1 Pitch Accent Inventory and Distribution

While assessing the pitch accent tonal inventory and distribution, it appears that majority speakers produce more pitch accents than monolingual speakers.

Table 9: **Experiment 1**: Pitch Accent Inventory across Speaker Groups.

Pitch Accent Type	**Overall Count**	**Bilingual Count**	**Monolingual Count**
H*, !H*, ^H*, L+^H, ^H	5214	3932	1282
L*	4107	3211	896
L+H*, L+^H*, L+!H*, L+^*H	2224	1750	474
L*+H	911	720	191
H+L*, ^H+L*	143	107	36
H*+!H, ^H*+!H	62	55	7
H*+L, ^H*+L, !H*+L	55	52	3
H+!H*	54	35	19
H* + L	4	3	1
L+*H	1	1	0
Total	12775	9866	2909
Average PA/Word	0.49	0.506	0.473
Number of Words	25631	19477	6154

As seen in table 9, not only are there approximately 2.7 times more bilingual speakers than monolingual speakers, but the assessment of pitch accent distribution also needs to be taken into consideration with regards to the number of words produced per group. After data cleaning, out of the 25,631 (normalized) words produced, bilingual speakers produce 19,477.

6.1.2 Pitch Accents: Monotonal vs. Bitonal Pitch Accent Model Results

The pitch accent generalized linear mixed model investigates the likelihood of a bitonal pitch accent, with fixed effects of speaker group , formality and gender, and the speaker ID as a random effect. The following pitch accents were chosen to be binomially distributed between monotonal and bitonal accents:

- Monotonal Pitch Accents: ”H*”, ”L*”
- Bitonal Pitch Accents: ”L+H*”, ”L*+H”, ”H+L*”, ”H*+!H”, ”H*+L”, ”H+!H*”, ”H*+L”

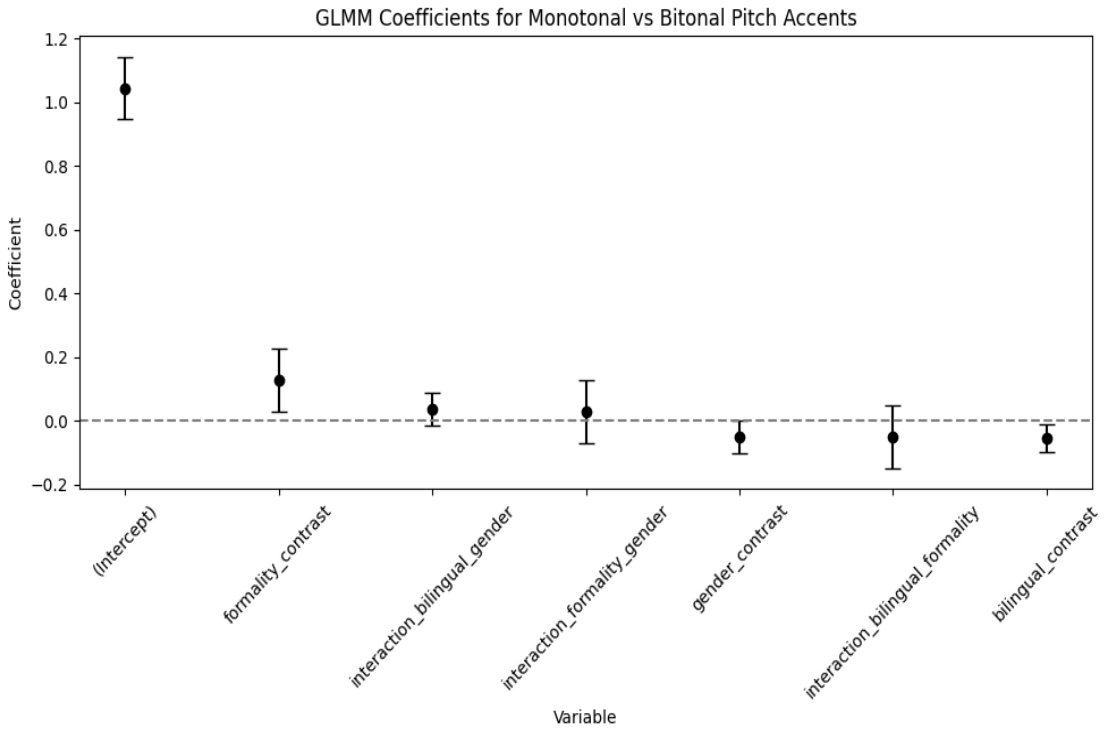

Figure 2: **Experiment 1** Visualization of the pitch accent model results

Figure 2 visualizes the estimated fixed effects coefficients from the GLMM model

predicting monotonal vs. bitonal pitch accents. The Y-axis represents the log-odds coefficients, with positive values indicating a greater likelihood of producing a monotonal pitch accent. The model was fit using maximum likelihood estimation, with a log-likelihood of -6608.009. Speakers are more likely to use monotonal pitch accents in formal settings $(z = 4.849, p =< 0.001)$, as seen in Figure 3. There is a weak indication that bilingualism might interact with formality in affecting pitch accent choice because this effect is only marginally significant $(p = 0.051)$ and above the threshold of significance. Overall, no interaction effects were found. The random effect variance for speakers 0.113 and its standard deviation 0.335 suggest that individual differences among speakers could contribute to the variation in pitch accent choice, but pitch accent use is expected to be determined by predictor variables and not individual variation. The relatively small variance indicates that the majority of variability is explained by fixed effects.

It is noteworthy to mention that a marginally higher volume of monotonal pitch accents were also found in heritage Russian speakers (Zerbian et al., 2024, p.13), with no mention of formality as an effect. Zerbian et al. (2024) claim that the high frequency of the H* monotonal accent (which is otherwise uncommon in Standard Russian), could have been found because the RUEG corpus consists of unprepared spontaneous speech as opposed to read and prepared speech. The analysis of the findings on English monotonal usage is as follows-

- Nature of the dataset: Since both papers use the RUEG corpus involving the same data elicitation method, speaker groups could pattern similarly overall in higher monotonal pitch accent production, in both languages.

- Difference in annotators: The RuPro corpus was separately created and annotated while the English data in RUEG-EN-Spoken was annotated by a different group of annotators, which could have resulted in formality playing a larger role in English

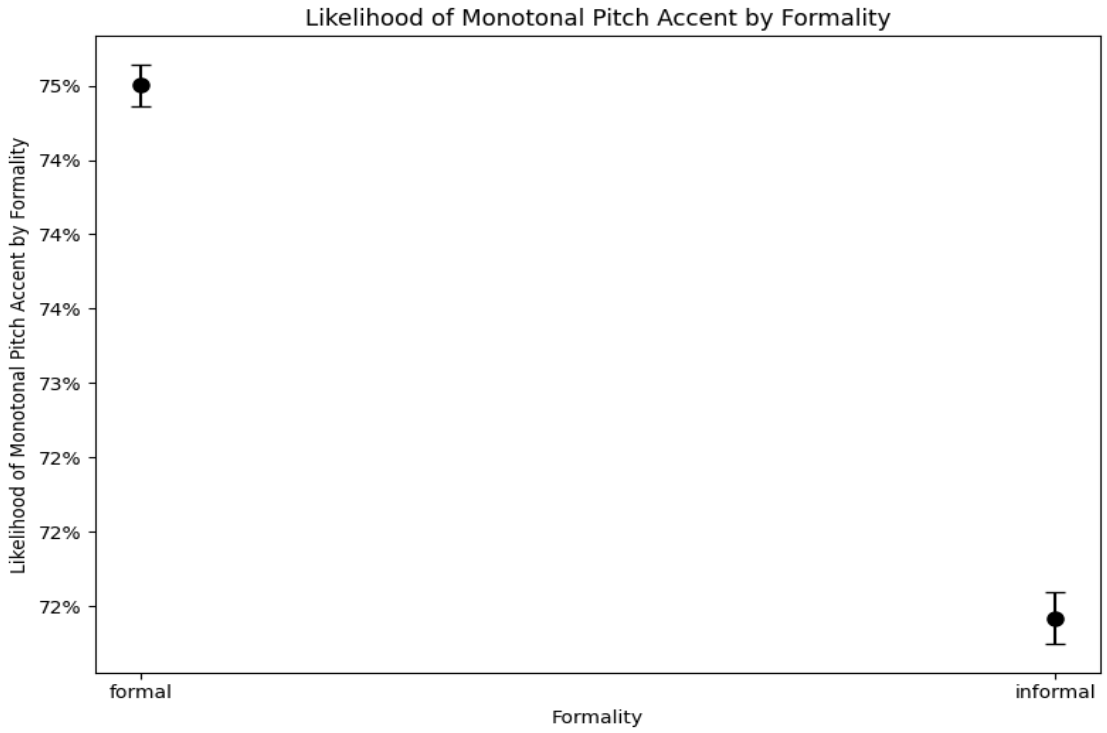

Figure 3: **Experiment 1** Visualization of the likelihood of monotonal pitch accent based on formality

than in Russian.

- Difference in languages: Standard Russian has not been known to have the H* monotonal accent, while English is expected to show both mono- and bitonal accents.

- Formality: Formal speech arguably requires a little more thought and intention than informal speech, which could result in a more ”prepared” style of speaking (though spontaneous), with a conscious effort to focus on topics of relevance.

6.2 Experiment 2: Boundary Tones

6.2.1 Boundary Tone Inventory and Distribution

The overall inventory of boundary tones is similar across speaker groups. At first glance, the difference in usage between high and low boundary tones, with a significant preference for low tones across groups and formality settings, reflect a tendency towards statements and declaratives, as seen in table 10. High boundary tones, though less frequent, in both majority and monolingual English, show slightly higher usage in formal contexts, such as highlighting or emphasizing specific points as was the case in participants reporting an accident they witnessed to police. Table 10 shows the results per speaker group. The following boundary tones are distributed between high and low boundary tones and are included in the tabulated counts:

- High boundary tones: 'H-H%', 'L-H%', 'H-', 'H-^H%', 'L-^H%', '!H-H%', '^H-H%'
- Low boundary tones: 'L-L%', 'H-L%', 'L-', '!H-L%', '^H-L%'.

Table 10: **Experiment 2**: Number of boundary tones across speakers.

Speaker Group	**H%**	**L%**
Majority English	1497	4738
formal	893	2635
informal	604	2103
Monolingual English	450	1666
formal	265	967
informal	185	699

6.2.2 Boundary Tones: High vs. Low Boundary Tone Model Results

As seen in figure 4, the binomial generalized linear mixed effects model reveals that gender is the only significant predictor, with increased odds of the outcome of females producing a high boundary tone $(z = 3.33, p =< 0.001)$. The figure visualizes the estimated coefficients from the GLMM model predicting high vs. low boundary tones. The Y-axis represents log-odds coefficients, where positive values indicate a greater likelihood of producing a high boundary tone. The model intercept of random effect is significantly negative, indicating that in the absence of strong predictor effects, the likelihood of producing a high boundary tone is relatively low, which tallies with table 10, given the overall weight of low boundary tones in the data.

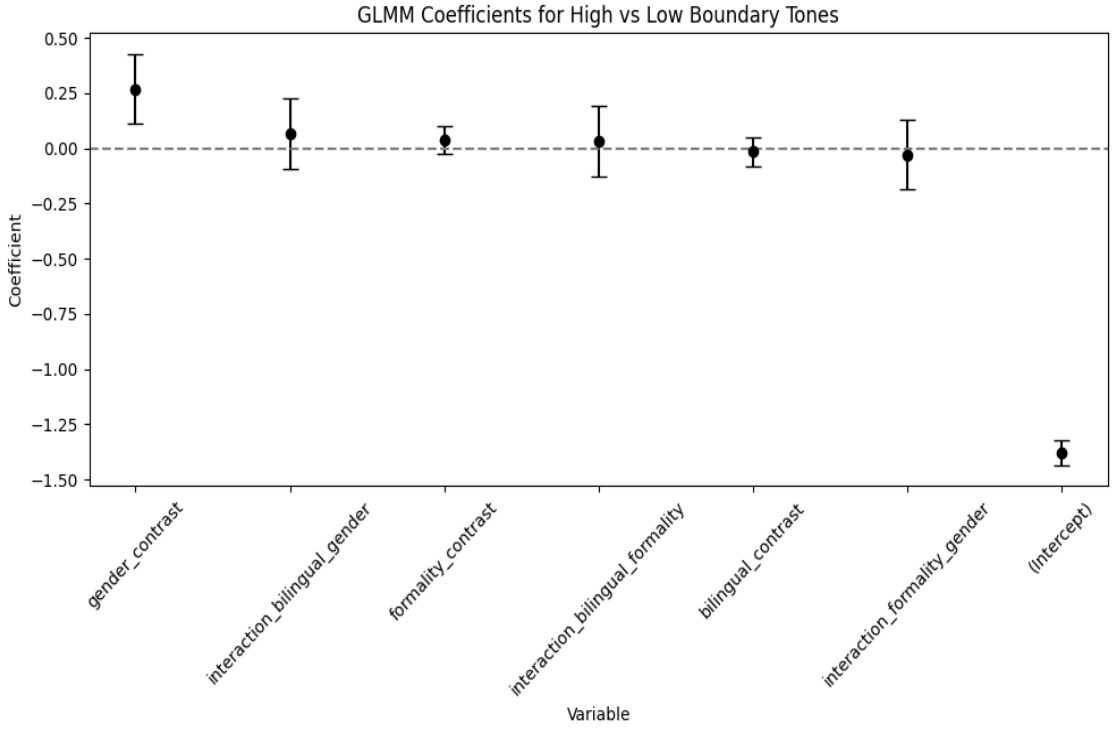

Figure 4: **Experiment 2** Visualization of the boundary tones model results

The model was fit using maximum likelihood estimation, with a log-likelihood of -4353.59. Out of the four experiments, this is the best fitting model, given that the number of observation are the highest, with 11169 observations. This does not come as a surprise, as previous research on gendered speech have employed mixed effect models as well. Neither bilingualism nor formality has a statistically significant effect on the outcome, suggesting these factors may not play a major role. This finding falls in line with existing research, regarding gendered speech. One could argue that since all the high boundary tones are grouped as H%, the inclusion of a rise could have contributed to this result. Existing research on the rise, characterized differently across various studies as uptalk, high rising terminal or in the case of Russian as "misplaced" melodic peaks contextualises the findings from experiment 2. The random effect variance 0.336 and standard deviation 0.58 reflect that individual speaker differences contribute to the model. Compared to all 4 experiments, the random effect has the highest variance for boundary tones. Inter-speaker variability, however a small influence is found due to a overall high number of low boundary tones produced, despite the assessment and expectation of high boundary tones.

Female speakers are more likely to end an utterance in a high boundary tone, as visualized in figure 5. When checked for the interaction between gender and bilingualism, the two factors further exact the initial hypothesis. Majority English female speakers produce the largest amount of high boundary tones compared to their female monolingual and male counterparts. This comes in light of findings regarding uptalk and misplaced peak alignments in Russian. The comparison between the nature of "misplaced" melodic peaks with the phenomenon of the uptalk (final tone rising at the end of declarative utterances) (Kachkovskaia et al., 2022, p.862) may tie into female speech, however, studies suggest various reasons gender-agnostic reasons for uptalk such as pragmatics. This leaves an open discussion space regarding high boundary tone usage which is outside the scope of this study.

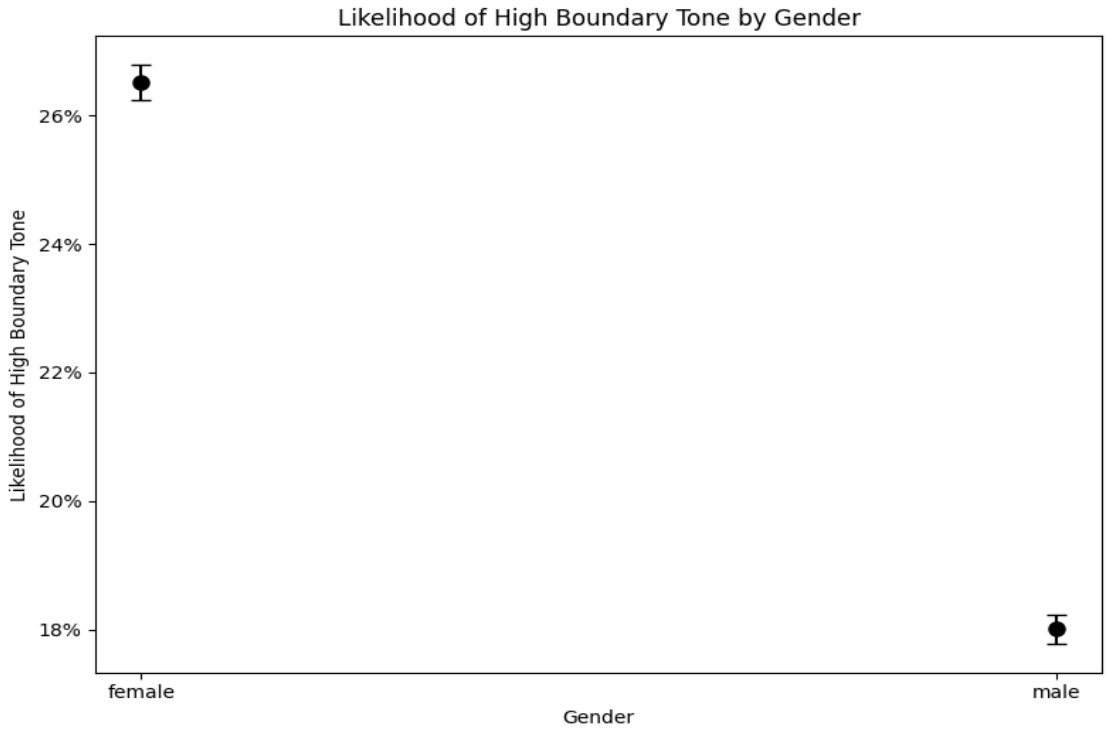

Figure 5: **Experiment 2** Visualization of the likelihood of a high boundary tone based on gender

6.3 Experiment 3: Pitch Accents on Content Words

Overall, majority English speakers produce more content words and pitch accents. As seen in Table 11, majority English speakers also accent their content words more frequently than monolingual speakers. Both groups pattern similarly in simply producing a pitch accent on a content word, with 83% of content words being accented by majority English speakers and 77% of content words being accented by monolingual speakers. Furthermore, formality impacts both speaker groups, where formal contexts elicit more pitch accents as compared to informal settings.

The model reflects similar results as the frequency statistics, revealing that English-

Table 11: **Experiment 3**: Number of pitch accents on content words across speaker groups by formality

Speaker Group	Number of Content Words	Number of PA on Content Words	PA/Content Word (std)
Majority English	8520	6991	0.82 (std. 0.385)
Formal	4753	3960	0.83 (0.373)
Informal	3767	3031	0.80 (0.397)
Monolingual English	2649	2033	0.77 (std. 0.424)
Formal	1536	1197	0.78 (0.415)
Informal	1113	836	0.75 (0.433)

dominant bilingual speakers are more likely to produce pitch accents than their monolingual counterparts ($z = 3.27, p = 0.001$). One possibility could be due to cross-linguistic influence from their heritage language Russian. The model was fit using maximum likelihood estimation, with a log-likelihood of -5401.109. Figure 6 visualizes the estimated coefficients from the GLMM model predicting the likelihood of pitch accents occurring on content words. The Y-axis represents log-odds coefficients, where positive values indicate a greater likelihood of a pitch accent on a content word. The results show that bilingual speakers are significantly more likely to produce pitch accents on content words compared to monolingual speakers, while formality has a smaller but positive effect. The random effect variance of speakers (0.099) and its standard deviation (0.315). Formal contexts reveal an increase in the likelihood of pitch accent production $(z = 3.346, p =< 0.001)$. As aforementioned, the influence of formality could possibly be due to a more deliberate and structured speaking style, with a motivation of focus or emphasis. Furthermore, the nature of the data elicited plays a role- recall that the data is from the oral witness report to the police. Naturally, the interaction between bilingualism and formality was also tested, but indicate no combined effect and is reflected in the figure, where its confidence interval includes zero.

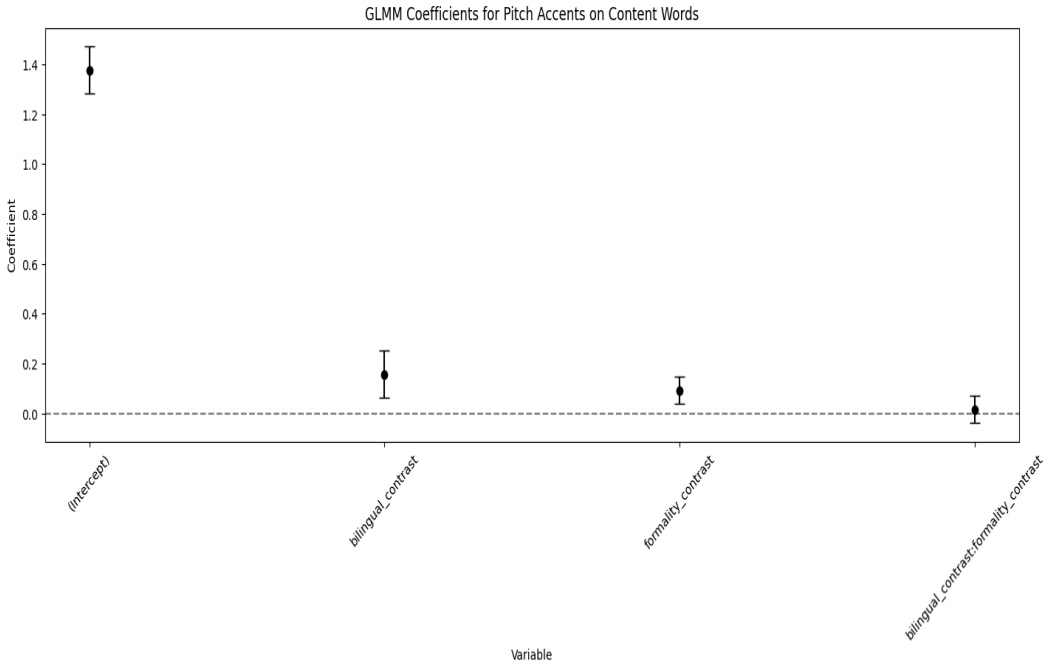

Figure 6: **Experiment 3** Visualization of the pitch accents on content words model results

Zerbian et al. (2024) also report one main effect of speaker group, where Russian heritage speakers produce more pitch accents on content words. However, they report a further analysis on pitch accent on verbs (which is also a content word), revealing that heritage Russian speakers accented verbs more frequently in informal settings (Zerbian et al., 2024, p.16). What would be interesting to investigate is which direction this influence heads in, or whether there is a bidirectional relationship and whether heritage speakers treat verbs in Russian and in their majority language of English, similarly or differently when formality comes into play.

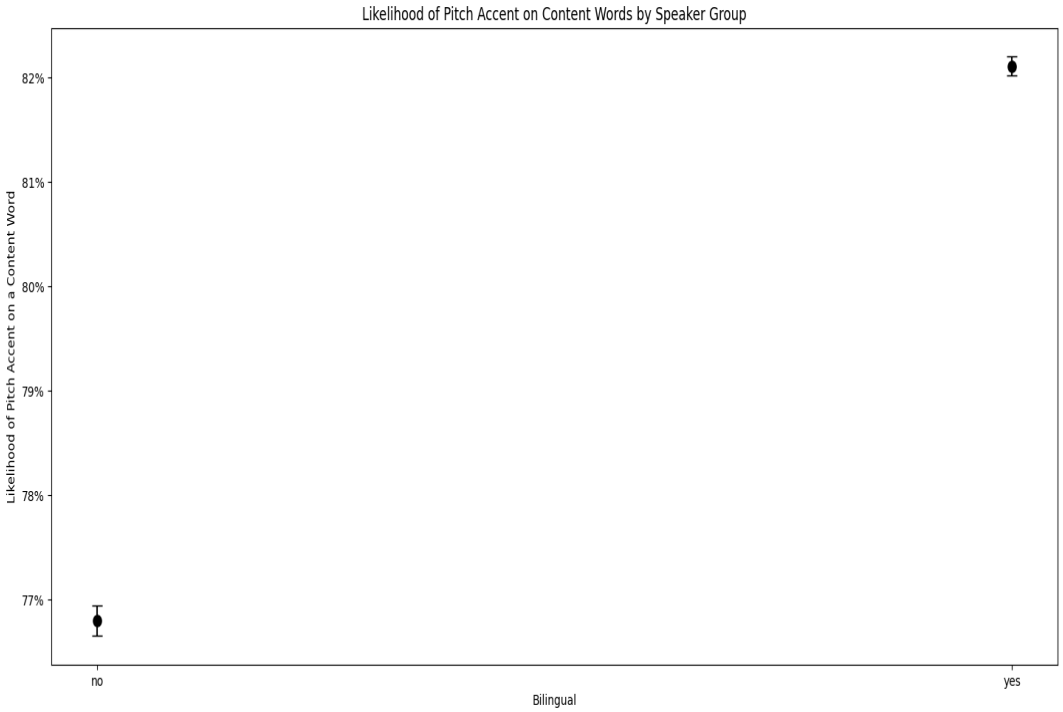

Figure 7: **Experiment 3** Visualization of the likelihood of a pitch accent on content words based on speaker group

6.4 Experiment 4: Intonational Phrases

Overall, bilinguals produce higher number of words and higher number of intonational phrases, as seen in table 12. These counts are similar to findings in Russian, where heritage Russian speakers produced a larger number of words as well as IPs (Zerbian et al., 2024, p.14).

The linear mixed likelihood model was fit with a log likelihood of -53206.04, while the variance (0.250) and standard deviation (0.500) indicate moderate variability in length across different speakers. As seen in figure 8, the estimated coefficients from the model predicting IP length are displayed on the Y-axis represents the coefficients, where positive

Table 12: **Experiment 4**: Length of IPs across speaker groups by formality.

	Number of Words	Number of IPs	Number of Words/IP
Bilingual English	19,477	5576	3.49 (std 1.89)
Formal	10,876	3,155	3.45 (std 1.87)
Informal	8,601	2,427	3.54 (std 1.93)
Monolingual English	6,154	1773	3.47 (std 1.89)
Formal	3,531	1,031	3.42 (std 1.81)
Informal	2,623	754	3.48 (std 1.97)

values indicate an increase in IP length. The x axis consists of the predictor variables and the interaction effects.

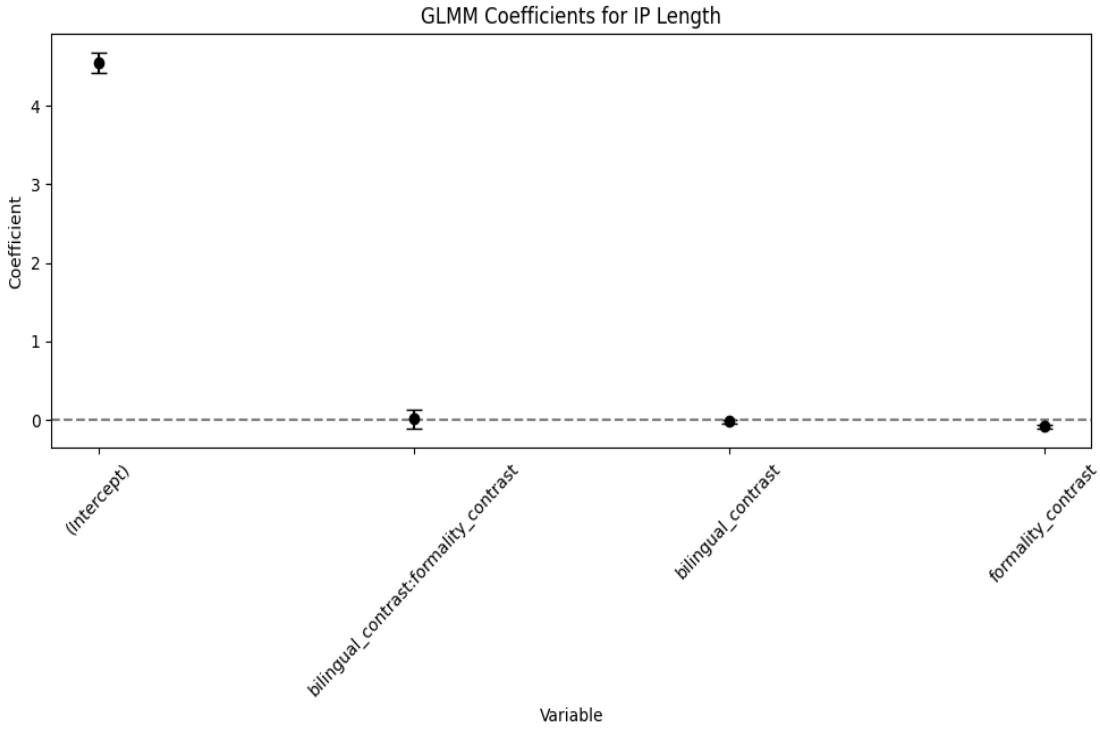

Figure 8: **Experiment 4** Visualization of the intonational phrase model results

The random effect variance for speakers (0.250) and its standard deviation (0.500)

suggest that some speakers naturally produce longer or shorter phrases than others. Bilingualism does not significantly influence IP length ($t = -0.29, p = 0.773$). This indicates that monolingual and bilingual speakers produce intonational phrases of similar lengths, regardless of their language background. The interaction effect between bilingualism and formality is also non-significant ($t = 0.80, p = 0.422$). The intonational phrase model is the worst fitting model compared to the other 3. For context, the log-likelihood of the best fitting model (experiment 2) is -5401 given 11169 observations, while the ip model log-likelihood is-53206.04 with 25631 observations. Although it is difficult to compare the two models, since they are both different from one another, the ratio of the model fit to the number of observations indicates that the intonational phrase model could be much better. The reason for this could be the complexity of the data. One suggestion is employing residual maximum likelihood estimation rather than maximum likelihoo estimation.

The main effect is of formality. The model results reveal that formality has a significant effect on intonational phrase length, where formal contexts result in shorter intonational phrases compared to informal contexts ($t = -5.93, p < 0.001$). The dot plot with error bars for formality is in figure 9. This suggests that speakers adjust their speech rhythm in formal settings, potentially due to increased planning or structured discourse. This finding falls in line with findings on Russian, where speakers of Russian showed a register effect, by producing lengthier intonational phrases in the informal than in the formal mode.

6.5 Summary of the results

Across all four experiments, there are some points that need to be stated, in order to understand the results and analysis with context. For each experiment, there are a total of 85 speakers, out of which 62 are majority English speakers and 53 identify as female speakers. A higher number of majority speakers as opposed to monolingual speakers, as well as

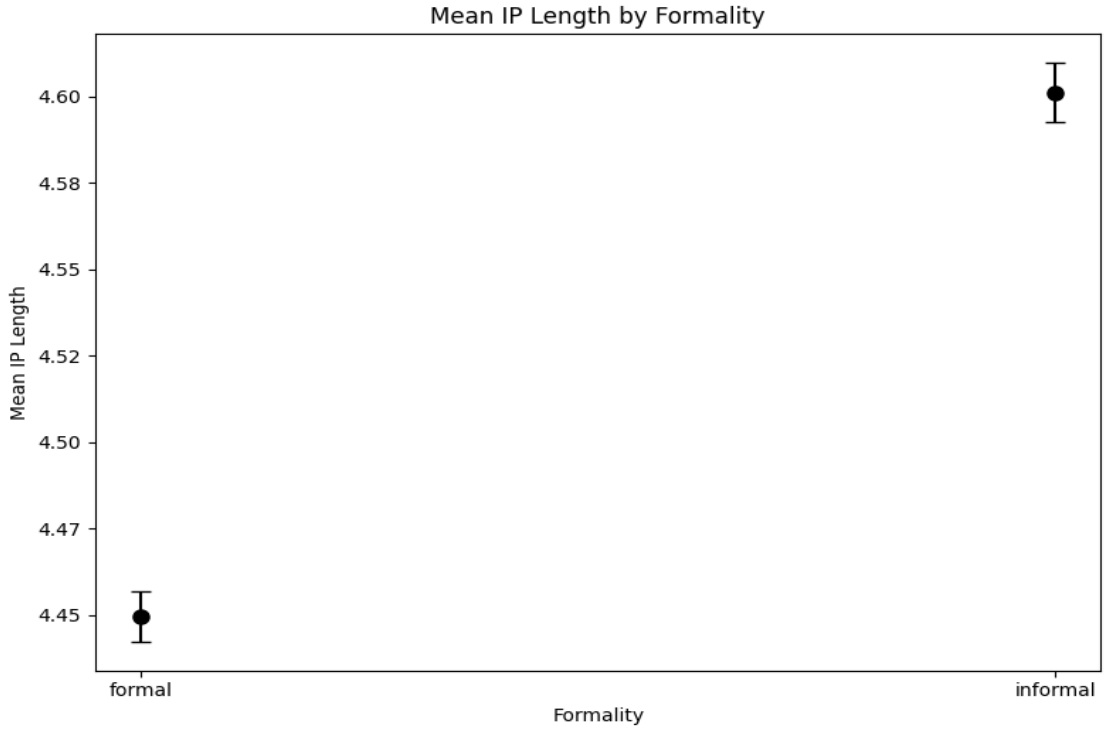

Figure 9: **Experiment 4** Visualization of the length of intonational phrase based on formality

a higher number of female speakers as compared to male speakers could play a role in the distribution of the prosodic items. Overall, majority English speakers also produced more words, resulting in more pitch accents, boundary tones and intonational phrases.

Regarding the random effects, the lowest variance (0.099) is observed in Experiment 3, which examines pitch accents on content words, suggesting minimal individual speaker variation. The highest variance (0.336) is found in Experiment 2, which focuses on boundary tones, indicating greater individual speaker variability, implying that speaker differences in the realization of pitch accents on content words are relatively small, whereas variation in boundary tones is more pronounced.

Below, is an overview of the hypotheses that were successfully answered-

- Experiment 1: Pitch Accents (likelihood of a monotonal pitch accent). The effect of bilingualism, formality and gender were tested. Formality is a significant effect. This finding partly confirms H.1, bilingualism does not have an overall effect on the production of monotonal pitch accents.

- Experiment 2: Boundary Tones (likelihood of a high boundary tone). The effect of bilingualism, formality and gender were tested. Gender is a significant effect. This finding confirms R.Q.2 and H.2.

- Experiment 3: Pitch Accents on Content Words (likelihood of a pitch accent on a content word). The effect of bilingualism and formality were tested. Bilingualism and formality are significant effects. This finding confirms R.Q.3 and H.3.

- Experiment 4: Intonational Phrases (length of intonational phrase). The effect of bilingualism and formality were tested. Formality is a significant effect. This finding confirms R.Q.4 and H.4, where the formal mode consists of shorter intonational phrases.

7 Discussion

The study of heritage languages is an emerging field, but heritage languages have existed throughout human history. There have been heritage speakers as long as immigration has moved families across language borders and as long as bilingual communities have been divided into dominant and minority language settings (Polinsky, 2018, p.352). This research has assessed melodies of the majority language within a modern phonological framework, i.e. the auto-segmental metrical theory, that maps categorically distinct elements and continuous acoustic structures (Ladd, 2008, p.43). Not only can the same text have many different melodies, but the same melody can occur on many different texts (Pierrehumbert, 1980, p.8). The introduction gave an overview of the influence of sociodemographic factors influencing language use. From the Kiezdeutsch data it was concluded that in a bilingual dyad, both the dominant language and the weaker/heritage/minority language are affected (Polinsky, 2018, p.147). Within the scope of this research, the two broad parameters that has contributed to variation in the data are formality and gender. The third parameter is bilingualism.

After the data preparation, this paper reports a small number in terms of data loss. From the four experiments conducted to gather insights into majority language use, it was found that there are significant differences in the use of pitch accents and boundary tones between majority and monolingual English speakers, although no difference in the inventory. Majority English speakers produce more monotonal pitch accents than their monolingual counterparts in English. The reasons discussed in the results and analysis section refers to the nature of the dataset, because in spontaneous speech, pitch variation might also be less pronounced, and thus bitonal pitch accents are less easily detectable (Zerbian et al., 2024, p.13)

The novel contribution of this research is the influence of gender on the intonation

of majority English. It was found that female speakers are more likely to produce high boundary tones than male speakers, particularly majority English females. Since high boundary tones included a bunch of labels ending with H%, one of the reasons for the use of rising boundary tones such as L-H%, is often associated with continuation or uncertainty (Clopper & Smiljanić, 2011, p.8-9) which ties into the possibility of pragmatic choices made by female speakers.

It is unsurprising that bilingualism did not come forth as a major factor influencing variation in intonation, ascertaining the place that the heritage speaker holds on the native language continuum. Bilingualism as a factor influence three out of four experiments, but marginally so, and sometimes in an interaction effect but not as an individual factor. The most meaningful effect of bilingualism was found in experiment 3. In experiment three on pitch accents on content words, it was found that English- dominant bilingual speakers are more likely to produce pitch accents than their monolingual counterparts. The overproduction of accented content words in no way highlights divergence from canonical usage. Rather, it demonstrates a speaker's proficiency in their majority language and possible need to sound precise. Similar results were found in Russian, where heritage speakers accent content words more frequently than their monolingual counterparts (Zerbian et al., 2024, p.19).

Furthermore, register variation yet again proved itself as a crucial factor in influencing language intonation in the experiment on accented content words, which found that the formal mode is more likely to elicit a pitch accent on a content word than the informal mode across speaker groups. Formality also played a role in the length of intonational phrases. Overall, formal contexts resulted in shorter intonational phrases compared to informal contexts, falling in line with the expectation that formal speech requires more planning. Zerbian et al. (2024) also report one main effect of speaker group, where Russian heritage speakers produce more pitch accents on content words. The heritage speaker

cohort has repeatedly displayed a sense of precision in their productions. The findings from this research may be contextualized in light of other heritage speaker groups who are English dominant. Russian and Turkish heritage speakers produced more noun-headed noun phrases and fewer pronouns than English monolinguals in some informal narratives (Pashkova, 2024, p.152), suggesting that heritage speaker do not prefer to be general.

7.1 Outlook

According to previous research, the reasons for observed differences, most commonly cite cross linguistic influence of the heritage language on the majority language, where 9 out of 18 reviewed studies (by Pashkova (2024)), mention it either as the single cause of the differences or as one of several possible causes, while other causes include less exposure to the majority language, difficulty of certain structures for L2 learners, weakening of certain majority language constraints due to weakened heritage language constraints, cross-linguistic over-correction, communication with L2 speakers or input received from L2 speakers, simplification of language systems that include multiple options, internal dynamics of language change, language aptitude, and speaker bilingualism (Pashkova, 2024, p.25). One of the main limitations of this research is its quantitative nature. The aforementioned reasons citing differences are qualitative and require a nuanced understanding of linguistic structures and an in depth understanding of a speaker group's language background. Future work could assess findings from the present research along the parameters stated above.

While examining heritage speakers with one heritage language background is a feasible starting point for majority language research, it is crucial to compare the majority language of several heritage speaker groups using the same method and the same linguistic phenomenon to understand if the effects found in one group are generalizable to other

heritage speaker groups (Pashkova, 2024, p.27). Thus, a corpus such as RUEG could be effectively used in recreating a study such as this because it has demonstrated the value of comparable data. But research replication often comes with limitations- in terms of the present thesis, one aspect to bear in mind is that GLMMs include random effects that can change when the study is replicated (Bono et al., n.d., p.2).

The relevance of this research may be discussed from varying perspectives. As stated in the background chapter, prosodic distinctions, as opposed to other grammatical domains are much more subject to change, variation, and loss (Polinsky, 2018, p.116), and this research lends itself in the form of documenting majority and monolingual English intonation distribution and frequencies. In undertaking the assessment of majority English used by heritage speakers, the thesis also contributes to non-canonical prosodic forms. Since heritage speakers now make a large part of a the global population with rising immigration and movement numbers, it is important to build technology to include the nuances of this cohort's language use because understanding of prosody is critical in downstream applications such as spoken language understanding tasks and speech recognition systems (Silverman et al., 1992, p.1). Furthermore, this study contributes to the understudied space of heritage language research and successfully describes trends in the majority language of heritage speakers through the illustration of pitch accent and boundary tone inventories. Conclusively, it is safe to say that there is indeed evidence to believe that there exists prosodic variation between heritage speakers and monolingual speakers in the majority language.

8 Abbreviations

Note: If and when used for the sake of convenience, all abbreviations refer to the singular and plural forms of their respective phrases.

BT Boundary Tone

HL Heritage Language

HS Heritage Speaker

IP Intonational Phrase

ML Majority Language

MonoS Monolingual Speaker

PA Pitch Accent

9 Appendix

For the purpose of data extraction, the web-server version of ANNIS was utilized on version: 3.7.1.The main programming language used to model and test for statistical significance is Python 3.11.0 accessed through Visual Studio Code 1.96.2. Additionally, data preparation was conducted on Google Colab. Supporting libraries for the purpose of data preparation and visualization include Pandas 2.2.2, Numpy 1.24.3 and Matplotlib 3.9.2. The choice of software for version control is Github, accessed through git version 2.37.3.windows.1, on Microsoft Windows 10.0.26100.2605.

References

Armstrong, M., Piccinini, P., & Ritchart, A. (2015). The phonetics and distribution of non-question rises in two varieties of american english. In *Icphs*.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in japanese and english. *Phonology Yearbook*, *3*, 255–309.

Bono, R., Alarcón, R., & Blanca, M. J. (n.d.). Report quality of generalized linear mixed models in psychology: A systematic review. *Frontiers in Psychology*, *12*. Retrieved 2025-01-16, from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.666182/full doi: 10.3389/fpsyg.2021.666182

Böttcher, M., & Zellers, M. (2024, March). Do you say uh or uhm? A cross-linguistic approach to filler particle use in heritage and majority speakers across three languages. *Frontiers in Psychology*, *15*. Retrieved 2024-12-09, from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2024.1305862/full doi: 10.3389/fpsyg.2024.1305862

Chafe, W. (1992). The importance of corpus linguistics to understanding the nature of language. *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*, 271.

Clopper, C. G., & Smiljanić, R. (2011, 4). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of phonetics*, *39*(2), 237–245. Retrieved from https://doi.org/10.1016/j.wocn.2011.02.006 doi: 10.1016/j.wocn.2011.02.006

Coupé, C. (n.d.). Modeling linguistic variables with regression models: Addressing non-gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and

shape. *Frontiers in Psychology*, *9*. Retrieved 2025-01-23, from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.00513/full doi: 10.3389/fpsyg.2018.00513

Dehé, N. (2018, 8). The intonation of polar questions in North American ("Heritage") Icelandic. *Journal of Germanic linguistics*, *30*(3), 213–259. doi: 10.1017/s1470542717000125

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press. doi: 10.1017/CBO9780511616983

Jolly, E. (n.d.). Pymer4: Connecting r and python for linear mixed modeling. *Journal of Open Source Software*, *3*(31), 862. Retrieved 2025-01-08, from https://joss.theoj.org/papers/10.21105/joss.00862 doi: 10.21105/joss.00862

Kachkovskaia, T., Zimina, S., Portnova, A., & Kocharov, D. (2022). Social variability of peak alignment in russian rise-fall tunes. In *Speech prosody 2022* (pp. 862–866). doi: 10.21437/SpeechProsody.2022-175

Krause, T. (2019). *Annis: A graph-based query system for deeply annotated text corpora* (Ph.D. dissertation). Humboldt-Universität zu Berlin, Berlin, Germany.

Kupisch, T., Lein, T., Barton, D., Schröder, D. J., Stangen, I., & Stoehr, A. (n.d.). Acquisition outcomes across domains in adult simultaneous bilinguals with french as weaker and stronger language. *Journal of French Language Studies*, *24*(3), 347–376. Retrieved 2024-12-03, from https://www.cambridge.org/core/journals/journal-of-french-language-studies/article/acquisition-outcomes-across-domains-in-adult-simultaneous-bilinguals-with-french-as-weaker-and-stronger-language1/2D78D90B0BB6BED3796EF28F0752B722 doi: 10.1017/S0959269513000197

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Nance, C., Kirkham, S., & Groarke, E. (2018). Studying intonation in varieties of english: Gender and individual variation in liverpool. In N. Braber & S. Jansen (Eds.), *Sociolinguistics in england* (pp. 275–295). London: Palgrave Macmillan UK. doi: 10.1057/978-1-137-56288-3_11

Nolan, F. (2020). Intonation. In *The handbook of english linguistics* (p. 385-405). John Wiley & Sons, Ltd. doi: https://doi.org/10.1002/9781119540618.ch20

Pashkova, T. A. (2024). *Majority english of heritage speakers* (Unpublished doctoral dissertation). Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau.

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

Polinsky, M. (2018). *Heritage languages and their speakers* (Vol. 159). Cambridge University Press.

Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P., ... Hirschberg, J. (1992). Tobi: A standard for labeling english prosody. In *Icslp* (Vol. 2, pp. 867–870).

Tamminga, M., Ahern, C., & Ecay, A. (n.d.). Generalized additive mixed models for intraspeaker variation. *Linguistics Vanguard*, *2*. Retrieved 2025-01-23, from https://www.degruyter.com/document/doi/10.1515/lingvan-2016-0030/html doi: 10.1515/lingvan-2016-0030

Wiese, H. (2020). Language situations: A method for capturing variation within speakers' repertoires. *Methods in dialectology XVI*, *59*, 105–117.

Wiese, H. (2021). *Communicative situations as a basis for linguistic systems: Integrating linguistic multi-competence with grammatical structure.* (Working Papers in Urban Language and Literacies, Online Publication)

Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., ... Zuban, Y. (n.d.). *RUEG corpus.* Zenodo. Retrieved 2024-12-29, from https://zenodo

.org/record/3236069 doi: 10.5281/ZENODO.3236069

Wiese, H., Alexiadou, A., Allen, S., Bunk, O., Gagarina, N., Iefremenko, K., ... others (2022). Heritage speakers as part of the native language continuum. *Frontiers in Psychology*, *12*, 717973.

Zerbian, S., Zuban, Y., & Klotz, M. (2024). Intonational features of spontaneous narrations in monolingual and heritage russian in the u.s.—an exploration of the rueg corpus. *Languages*, *9*(1). Retrieved from https://www.mdpi.com/2226-471X/9/1/2 doi: 10.3390/languages9010002