THE PROSODY OF MONTEVIDEO SPANISH: AN INTONATIONAL, RHYTHMIC AND TEMPO DESCRIPTION*

Vanina Machado and Laura Escobar
University of Toronto

1. Introduction

Montevideo Spanish (MS) is a Rioplatense variety spoken in Uruguay which has roots in Italian and Castilian Spanish, and an influence from other languages such as Guarani, Quechua, and Portuguese (Bertolotti and Coll 2014). As the literature on the prosody of MS remains sparse compared to other Spanish varieties, this paper seeks to provide the first multidimensional prosodic description of this Uruguayan Spanish variety. Considering F0 and duration interact with each other to give a language its prosodic structure (Steffman and Jun 2013), our approach utilizes the analysis of intonational, rhythmic, and speech rate patterns using tonal and durational measures. Furthermore, we investigate potential intragenerational variability. With this, we provide a comprehensive description that will serve not only to uncover MS' own "tonada" but also to establish a benchmark for future comparative studies with other Spanish varieties.

This study intends to understand: (1) what characterizes MS intonation, rhythm, and speech rate, (2) how different generations of MS monolingual speakers differ prosodically and what variables best exemplify these differences, (3) if existing analytical prosodic models can effectively account for this Spanish variety. To address these questions, we examine prenuclear and nuclear pitch configurations, deaccenting, peak placement, and peak height difference, F0 and durational rhythmic patterns, as well as durational speech rate across three main sentence-types: yes/no questions (YN), tag questions (TQ), and statements (SS). Participants were recorded performing different production tasks that range from spontaneous to controlled to ensure that the collected data encompasses a spectrum of naturalistic speech instances.[1]

2. Research context and previous studies

Prosody refers to "nonsegmental aspects of abstract linguistic structure, such as a particular type of constituent structure and the presence or absence of accents, that are […] systematically reflected in the phonetic rendition of utterances" (Nootebom 1999: 1). It includes the study of intonation, lexical stress, rhythm, and fluency (Colantoni and Steele 2015) and plays a crucial role in the perception and comprehension of spoken language, conveying emotions, intentions, and pragmatic meanings (Xu 2019). While it has been

*We want to thank Prof. Colantoni and Prof. Rafat for their invaluable support. Also to Sylvia Costa for generously providing us with a testing space in Montevideo, Uruguay. Finally, we wish to thank Miguel Barreto for his assistance with the preliminary descriptive statistics.

[1] Data was originally gathered for the first author's dissertation (Machado forthcoming).

Actes du congrès annuel de l'Association canadienne de linguistique 2023.
Proceedings of the 2023 annual conference of the Canadian Linguistic Association.

previously studied in other Spanish varieties (i.e., Hualde and Prieto 2015, Prieto and Roseano 2010, Colantoni and Gurlekian 2004, Mariano and Romano 2010), there remains a gap in the literature concerning MS. This section highlights significant findings on Spanish prosody, and underscores the need for further research into MS' system. Furthermore, it highlights the importance of multidimensional analytical approaches, as well as a consideration for sociolinguistic factors when describing the prosody of a language.

2.1 Prosody: basic notions and models

The Autosegmental Metrical Model (AM) rooted in intonational phonology (Pierrehumbert 1980) has been a prominent framework in intonational research. It recognizes role of prosody in conveying prominence and phrasing by associating tones with prosodic constituents and coordinating their timing. Within this framework, Spanish has been described as having a rich nuclear inventory that, together with different peak alignment patterns, serve to differentiate utterance types and pragmatic contexts. Other factors like initial peak height are also indicate utterance distinctions in Spanish, with prenuclear peaks usually being higher in questions than in statements (Face 2001). Additionally, Spanish intonation displays noticeable dialectal variation (Aziz et al. 2023).

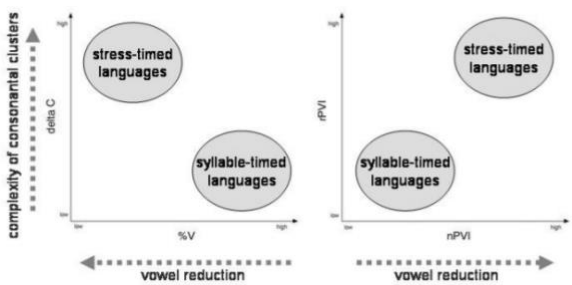


Figure 1. Rhythmic class correlations with Ramus et al. (1999) and Grabe and Low (2002) measures (Dellwo 2010).

Similarly, other researchers have developed metrics rooted in various acoustic correlates that measure, for instance, the duration of vocalic and consonantal intervals (Ramus et al. 1999, Grabe and Low 2002, Dellwo 2006, Bertinetto and Bertini 2008). Metrics like vocalic proportion (V%), vocalic (ΔV) and consonantal variability (ΔC), VarcoV and VarcoC, Pairwise Variability Index (PVI), and Control and Compensation Index (CCI), have been shown to identify rhythmic types and represent language-specific properties (Dellwo 2010) (Figure 1). Many romance languages like French, Romanian, and Spanish are considered to be syllable-timed (Nespor et al. 2011).[2] Despite this, the hypothesis that grouping and prominence are the universal principles of rhythm (Arvaniti 2009) has led to

[2] In this study, we opt for the term "syllable-based" as it is deemed to provide a more precise representation of the rhythmic nature of a language (Simões 2022).

analyzing the role of other key factors in terms of rhythm, such as F0 (Tilsen and Arvaniti 2013). Within this AM-based framework, Jun (2014) proposed the concept of "macro rhythm", considered as a sequence of H/L alternation within an Intonation Phrase, calculating the degree of variation and shape of tonal units in a phrase (Jun 2013) with measures like Macro frequency (number of f0 peaks per number of prosodic words). Spanish has been described by Jun (2014) as having a strong macro rhythm (near 1).

Tempo has been demonstrated to play a crucial role in understanding rhythm (Lehiste 1970). Among methods for quantifying speech rate that observe tempo in terms of duration, the CV interval ratio (Dellwo 2010), which measures consonantal and vocalic intervals per second, is significantly different between stress- and syllable-based languages. Syllable-based languages like Italian were observed to be perceptually regular, often exceeding 11 CV intervals per second (range of 9-16), while stress-based ones like English were perceived as irregular (usually below 11).

2.2 Montevideo Spanish

Uruguay is located within the Rioplatense dialectal region, which includes Buenos Aires (porteño) Spanish. MS is the variety of the capital and is considered the national prestige model (Carvalho 2014). Despite this, MS is particularly underexplored regarding prosody. Although there has been previous research in MS prosody (Cunha et al. 2008, Araujo 2013, Rebollo Couto et al. 2014, Brockhorst 2020), no previous work has provided a comprehensive description of the MS rhythmic system and speech rate patterns. A summary of relevant findings drawn from previous literature is presented in Table 1.

Table 1. Summary of previous findings on MS prosody.

Prenuclear YN	L*+H (Cunha et al. 2008; labeled as L*+>H in Araujo 2013)
Nuclear YN	H* (Cunha et al. 2008); L*+H (late L+H* in Rebollo Couto 2014); L*+¡H (L+>¡H* in Araujo 2013)
Boundary tone YN	¡HL% (Cunha et al. 2008); HL% or L% (Araujo 2013, Rebollo Couto et al. 2014).
Prenuclear SS	L*+H (Labeled as L*+>H in Araujo 2013)
Nuclear SS	L+H* or L* (Araujo 2013)
Boundary tone SS	L% (Araujo 2013)
TQ	n/a[3]
Rhythm	Shorter tonic, longer posttonic (Cunha et al. 2008). Intermediate duration of tonic syllable between Italian and Spanish; longer vowels in older generations (Brockhorst 2020).
Speech rate	n/a[4]

[3] No studies have examined TQ intonation in MS thus far. However, Valenzuela-Farias (2013) conducted a study on TQ intonation in Spanish, finding its first IP to be similar that of SS, with the presence of rising or rise-falling patterns in the IP of tag words.

[4] The mean articulation rate (syllables/s with no pauses) in Spanish is approx. 7 (Santiago and Mairano 2022).

2.3 Hypotheses

H1. I) The MS intonational system will use nuclear peak alignment and peak height with phonological function instead of relying only on pitch accent (Araujo 2013) to mark differences within and across sentence types. II) Durational metrics will confirm MS syllable-based nature, and tonal macro frequency values will establish it a strong macro rhythmic language. III) MS will exhibit syllable-based compatible rates, better represented by CV rates than syllables per second. **H2.** There will be visible generational differences like in Brockhorst (2020), especially evident in vocalic rhythmic and speech rate values. **H3.** Most models and measures will capture MS prosody, demonstrating that tonal and durational measures can both describe its system and sociolinguistic variation.

3. Methodology

3.1 Participants

The study involved 30 female monolingual speakers of MS, all of whom were lifelong residents of Montevideo City. They were categorized into three generational groups: Generation 3 (gen3, ages 18-30), Generation 2 (gen2, ages 31-59), and Generation 1 (gen1, ages 60+). The grouping criteria was based on local lifestyle, age-specific attributes linked to social norms and identity, language use, and socio-historical events.

3.2 Stimuli

The stimuli consisted of 20 yes/no questions, 20 statements, 20 tag questions, and 20 distractors (wh-questions and imperatives) per task, all in Spanish (Table 2). The number of stimuli is consistent across all tasks, only distractors vary.

Table 2. Stimuli examples.

YN	SS	TQ	Distractors
¿**Ven**den ce**bo**llas? ‘Sell.3PL.INF.PRS onions?’	**Ven**den ce**bo**llas ‘Sell.3PL.INF.PRS onions’	**Ven**den ce**bo**llas, ¿**cier**to? /**¿no?** ‘Sell.3PL.INF.PRS onions, right/no?’	¿**Có**mo **va**mos? ‘How to go.1PL?’

Each participant produced 160 isolated Spanish utterances, with word length, syllable structure, and stress patterns controlled. Utterances consisted of two prosodic words (PW) (prenuclear and nuclear). To observe peak alignment and pitch change across utterances, words were two to four syllables long and stressed on the penultimate syllable (the unmarked stress pattern). The structures of tonic syllables consisted of those found in syllable-based languages (CV, CVC, CVV, VC, and CCV(C)). To analyze continuous pitch contours, tonic syllables mostly included voiced consonants followed by low or mid vowels.

3.3 Tasks

Participants were tested locally and compensated for their time and participation. They completed a series of speech production computer-based tasks ranging from spontaneous to controlled (with five-minute breaks in between). Labvanced (Finger et al. 2017) was used for the design of tests and used as a presentation tool while participants were recorded. The experiment began with a Sociolinguistic Interview (EOQ) (adapted from Rato and Machado 2024), conducted by a non-Montevidean Uruguayan female interviewer. Participants then engaged in four speech production tasks: a Monologue (MG) to evaluate proficiency; a Contextualized Task (CT), to generate structured yet natural responses to 80 fictional scenarios; a Robot Task (RT), repeating after a robot with flat intonation; and a Shadowing Task (ST), repeating after MS monolinguals (mixed random order).[5]

4. Analysis

Audios were examined using Praat (Boersma and Houven 2001) and utterances were segmented and annotated in three tiers: syllables, number of peaks, and interconsonantal and intervocalic intervals to report both F0 and durational patterns. Intonational patterns were analyzed using the Autosegmental Metrical Model (Pierrehumbert 1980) by labelling prenuclear and nuclear pitch accents, and boundary tones with SP_ToBI (Aguilar et al. 2009).[6] F0 peak height was also measured to observe F0 range and F0 height difference between prenuclear and nuclear peaks. As for rhythm and tempo, we calculated (a) the level of Macrorhythm by measuring Macro frequency (Macr_freq) (Jun 2014); (b) interval variability measuring V%, ΔC (Ramus et al. 1999), VarcoV, VarcoC (Dellwo 2006), nPVI, rPVI, (Grabe and Low 2002), CCI (Bertinetto and Bertini 2008) in Correlatore (Mariano and Romano 2010); and (c) speech rate, by measuring CV-rate, V-rate, C-rate and syllable-rate (Dellwo 2010). Finally, for categorical nominal variables, we ran CHAID analyses in DataTab (DataTab Team 2023), while for numerical continuous variables, Linear Mixed-effects models (LMM) or a Generalized Mixed-effects (GLMM) models in R (R Core Team 2022) using the lme4 package (Bates et al 2014).

5. Results

5.1 Intonation

5.1.1 Tonal configurations

Prenuclear pitch accents (pPA) in MS typically consist of L*+H across all three sentence types, as illustrated in the Sankey diagram in Figure 3 below (left). This pattern is

[5] Robot speech was synthesized in Praat using a flat-intonation-resynthesizer script (Vicenik 2011).

[6] The tonal inventory we used for labeling pitch accents and boundary tones drew upon various models describing intonational phonology for Spanish (Hualde 2003, Jun and Fletcher 2014, Prieto and Roseano 2010, Sosa 1999). Our labeling captures both F0 behavior and peak position.

characterized by the presence of a posttonic peak that starts rising during the tonic syllable and fully realizes in the following one (black arrow in Figure 2). A low occurrence of L+H* and H* is seen mainly in SS but is not significant overall.[7] When analyzing the percentage of pPA per sentence type in each generation (Sankey diagram in Figure 3) the prevalence of L*+H across groups is confirmed. The younger generation displays greater prenuclear diversification, although not at significant levels of production.

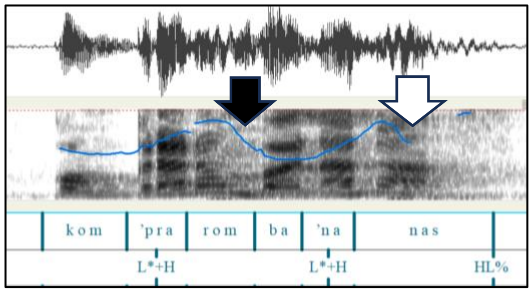


Figure 2. Initial and final L*+H in MS YN.

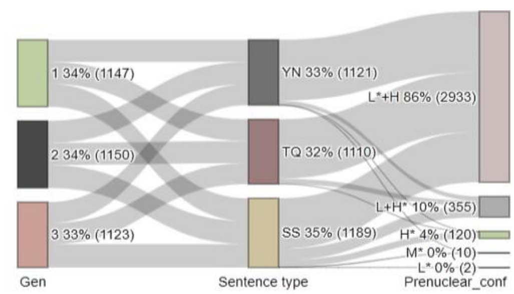


Figure 3. Generational pPA distribution.

While MS exhibits very limited prenuclear production, nuclear pitch accents (nPA) in MS form a more diverse and variable tonal inventory. The inventory is mainly composed of four pitch accent types, demonstrating its higher level of tonal complexity. L+H*, L*, L*+H and L*+¡H are the main configurations.

Table 3. Prevalent nPA in MS per sentence type.

Sentence	Nuclear Pitch Accent (%)
YN	L*+H (57.75%)> L*+¡H (36.54%)
Ss	L* (72.95%)> L+H* (19.54%)
TQ	L+H* (87.27%)> L* (6.41%)

[7] MS speakers sporadically use prenuclear L+H* in declaratives to mark early focus, especially in two-syllable words like “venden” or “trae,”. This is not in line with the expected neutral utterance.

A clearer distinction arises when examining nPA by sentence types. As shown in Table 3, MS YN sentences exhibit a more variable nuclear production than SS and TQ, which have a single overwhelmingly preferred configuration. Polar questions normally feature L*+H (white arrow in Figure 2), with the second most common type being upstepped L*+¡H (+1.5 ST higher than prenuclear) (Table 4). In contrast, SS primarily exhibit a sustained nuclear L*, followed in frequency by L+H* (Figure 4). Conversely, for TQ, MS speakers mainly produce L+H*.

A Chi Squared Automatic Interaction Detection (CHAID) analysis of generational patterns reveals clear trends (appendix, CHAID tree 2): in YN, nuclear L*+¡H dominates among older individuals (56.2%) but decreases the younger the speaker (41% gen2, 9.34% gen3). This is due to younger generations increasingly producing nuclear L*+H, which is particularly high in gen3. In SS, a generational L+H* decline can be observed, while L* remains dominant, peaking in gen2. Conversely, TQ maintains stable L+H* presence, especially in gen2, due to stronger L* representation in the other groups. Notably, nuclear M* is only present among younger speakers and usually in declaratives. These results expose significant differences in nPA distribution and highlight the contrast in intonation patterns between MS generations.

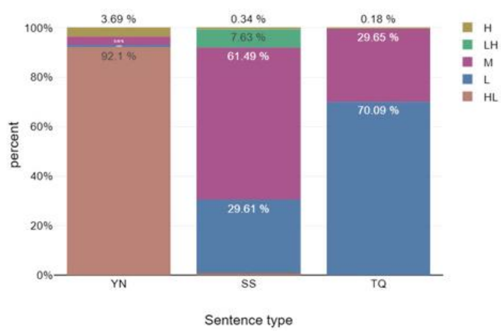


Figure 4. Bar chart with distribution of boundary tones within sentence types.

Finally, boundary tones in MS exhibit both monotonal and bitonal configurations, which vary depending on the sentence type (Figure 4). In YN sentences, the circumflex HL% is the most frequent choice. However, MS declaratives regularly feature a final mid tone (M%), with a much lower occurrence of a low L%. LH% is infrequently observed in statements, but when present, it typically conveys confirmatory meaning rather than neutrality. In TQ, intermediate boundary tones tend to be low (L%), or, in lower proportion, mid (M%), and their occurrence is influenced by the tonal configuration of the following tag word.

A CHAID analysis of BT (appendix, CHAID tree 3) confirmed that while YN in MS tend to end in HL%, the presence of other less common final tones reveals generational trends. While H% decreases significantly from gen1 to the younger generations, M% follows the opposite trend. Similarly, final M% in statements follows a steady growth from gen1 to gen3, which explains the decrease of production of L% in the younger generation.

This same trend can be seen in the production of final M% in TQ intermediate boundary tones, which explains the decrease of L% in gen3.

Table 4. Top prevalent nuclear configurations (nuclear pitch accents + boundary tones).

Sentence type	Nuclear configurations
YN	L*+H HL% (54%) > L*+¡H HL% (34%)
SS	L* M% (46%) > L* L% (22%) > L+H* M% (12%)
TQ	L+H* L% (63%) > L+H* M% (24.5%)

Table 4 summarizes the most frequent nuclear combinations of nPA and BT in MS for the three involved sentence types.[8] Upon analyzing nuclear combinations per sentence type and generation and in the understanding that all sentences combine almost exclusively with prenuclear L*+H, it becomes evident that the most produced final pattern for YN is L*+H HL%, which is preferred by gen3 (Table 4). L*+¡H HL% (Figure 6), follows in frequency and is produced mainly by first- and second-generation speakers. Other final tones like H% and M% are rare; while H% often results from truncation of an upstepped nuclear peak, M% appears following L*+H, particularly among younger generations. In contrast, SS are mostly produced with final L* M% (the preferred pattern by younger speakers) and, less frequently, with L+H* L%, common among older speakers.[9] Although LH% appears combined with L*, L+H* in statements, it does not constitute a regular pattern as seen in Figure 4.

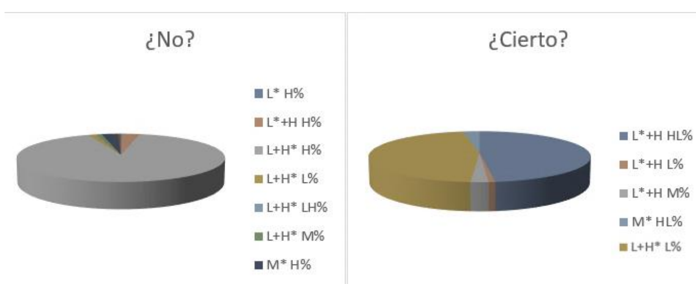


Figure 5. Pie charts with distribution of tonal configurations for tag words.

Finally, TQs first IP are typically produced with nuclear L+H* L%, alternating with M%, which is more frequent in younger speakers. These two configurations combine with either L+H* H% in monosyllabic tag words (“¿no?”) or L*+H HL% and L+H* L% in disyllabic tag words like “¿cierto?” (Figure 5). These variations highlight the prosodic differences

[8] As MS utterances usually share the same prenuclear pitch accent, the description of more global configurations does not address the combination of nuclear configurations with prenuclear.

[9] Notably, prenuclear L+H* pattern combines with these two nPAs but in confirmatory pragmatic contexts, where speakers emphasize the prenuclear word to reaffirm their statement.

between disyllabic and monosyllabic tag words, with “¿no?” being less variable than “¿cierto?”.

5.1.2 Deaccenting (absence of F0 peak)

The overall percentages of deaccenting in MS are notably higher in nuclear position (25.2%) compared to prenuclear (3.9%). When compared across sentence types, prenuclear and nuclear deaccenting mostly occurs in statements and is considerably more pronounced in nuclear position (10.7% prenuclear, 74.3% nuclear). This result is consistent with the observations detailed in section 4.1.1, which emphasize the presence of nuclear L* and M* in declarative sentences. Consequently, while SS may occasionally feature both prenuclear and nuclear peaks, often represented as L*+H L+H* (right of Figure 6), the prevailing pattern typically produces only a prenuclear peak, often in the form of L*+H L* (left of Figure 6). It is worth nothing that TQ also exhibit nuclear deaccenting, although in much lower proportion, in the form of a low plateau in the tonic syllable.

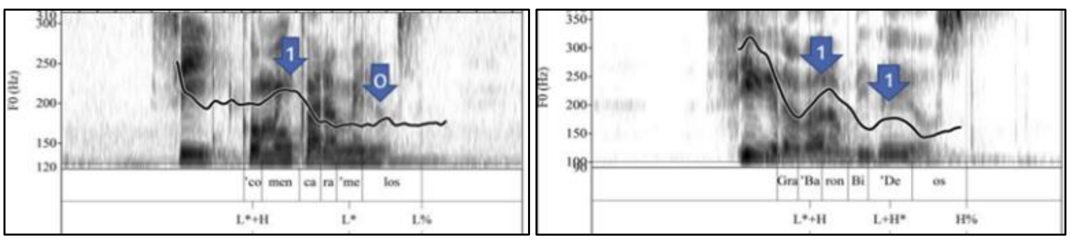


Figure 6. Gen3 SS with nuclear deaccenting (left) vs. gen1 SS with nuclear peak (right).

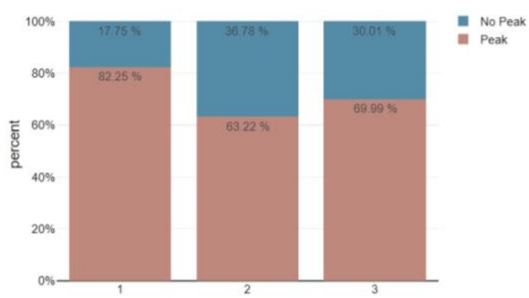


Figure 7. Bar chart showing percentage of deaccenting (in blue) in MS generations.

The percentages of nuclear deaccenting across generations are higher in younger speakers (Figure 7). However, deaccenting reaches its highest value in the second generation, indicating a non-linear generational continuum.

5.1.3 Peak position

Table 5 below provides an overview of peak position patterns in MS across different sentence types, categorized by pitch accent position. Results show that whether the peak falls within the tonic syllable (tonic) or realizes within the following syllable (posttonic) is primarily influenced by the sentence type.

Table 5. Overview of peak position patterns in MS.

Sentence type	Prenuclear Peak	Nuclear Peak
YN	Posttonic (93.67%)	Posttonic (96.3%)
SS	Posttonic (90.58%)	Tonic (78.43%) > Posttonic (20.59%)
TQ	Posttonic (92.21%)	Tonic (92%)
Tag word	n/a	Tonic (90.5%) *¿no?* Posttonic *¿cierto?* (94.33%)

Notably, prenuclear peaks tend to be posttonic across all sentence types and all generations, aligning with the frequent production of prenuclear L*+H. In nuclear position, however, the patterns exhibit more variability: while nuclear peaks in polar questions are principally posttonic, in statements and tag questions they are tonic, frequently forming nPA L+H*. In the IP of tag words, tonic peaks associate with monosyllabic structures, and posttonic with disyllabic ones.

When examining the nuclear peak position across different generations (appendix, CHAID tree 4), a trend emerges in the production of tonic peaks within statements. The percentages exhibit a gradual rise from the youngest to the oldest speakers, demonstrating a strong presence of tonic patterns in the oldest generation. This direct relationship, however, is not observed in tag questions: gen2 produces more tonic peaks than gen1 and gen3, yet the two oldest groups approach each other. Interestingly, the gen1 exhibits a reduced yet frequent amount of nuclear posttonic peaks in both statements and tag questions, while the third one only in the latter.

5.1.4 F0 peak height difference (PHD)

The violin plot in Figure 8 below illustrates the balanced concentration of MS peak height difference, as data is divided between positive and negative values, with a median of 0 (-13 +14.3) (Figure 8).

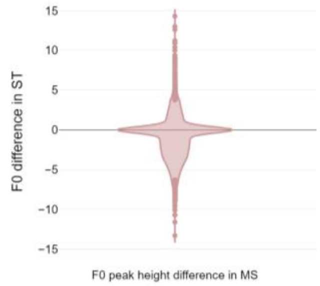


Figure 8. Violin plot of MS PHD.

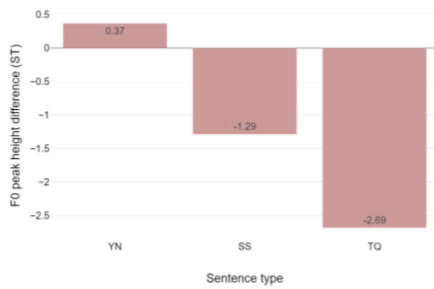


Figure 9. Bar chart of PHD per sentence type.

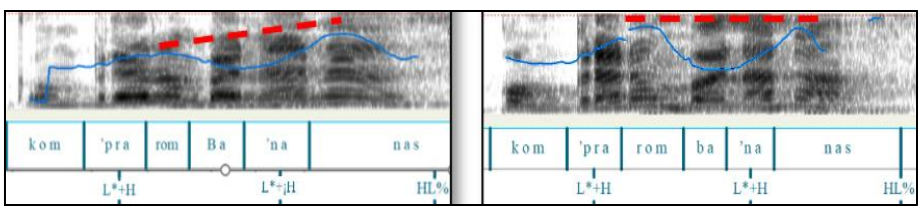


Figure 10. YN by gen1 (left) with positive PHD vs. by gen3 with 0 PHD (right).

Upon observation of F0 height values per sentence type, a clear positive mean PHD stands out in YN (Figure 9). This can be attributed to the presence of peaks of equal height and higher nuclear than prenuclear peak, which frequently correspond to L*+H or L*+¡H (Figure 10). Notably, the influence of nuclear upstep, a phenomenon exclusive of MS polar questions, is significantly reflected in this positive value. In contrast, PHD decreases considerably in SS and even more in TQ, as the initial peak tends to be higher than the nuclear peak, shaping a tonal downtrend throughout the utterance (Figure 6).

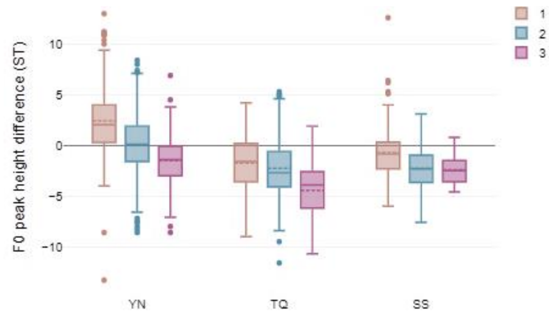


Figure 11. Box plots depicting generational PHD within sentence types.

F0 peak height differences per sentence type across generations (p <.001) reveal a continuum, with values steadily decreasing from the oldest to the youngest generations (Figure 11). As depicted in the boxplots in Figure 10, this is especially noticeable in YN, where gen1 F0 peak height difference values are all positive, while gen3's are negative and gen2's, intermediate.

5.2 Rhythm

5.2.1 Tonal

The average macro frequency level in MS stands at 0.72, providing evidence for the typical presence of more than one peak per utterance. An analysis of values per sentence type shows that its highest levels are usually seen in YN and the lowest in SS and TQ, which is in line with the high percentage of nuclear deaccenting observed for the latter.

The ANOVA results in the mixed model revealed significant differences in means per generation (F=70.34, p<.001) and Post hoc tests (Bonferroni) confirmed p<0.05 for pairwise group comparisons (1 - 2, 1 - 3, 2 - 3). As depicted by the box plots in Figure 12, macro frequency levels decline with each successive generation. Nevertheless, the eldest generation demonstrates more significant variation, reaching the highest maximum levels recorded.

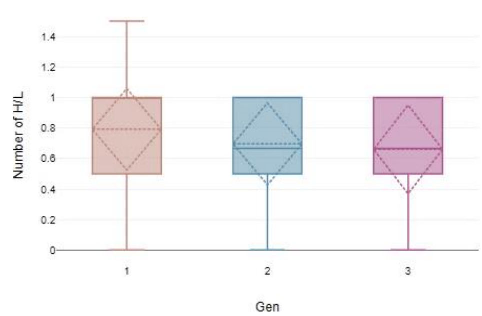


Figure 12. Boxplots of Macro Frequency levels in MS generations.

5.2.2 Durational

A closer analysis of MS' rhythmic durational values in comparison with previously described languages in Mariano and Romano (2010), reveals insightful findings.

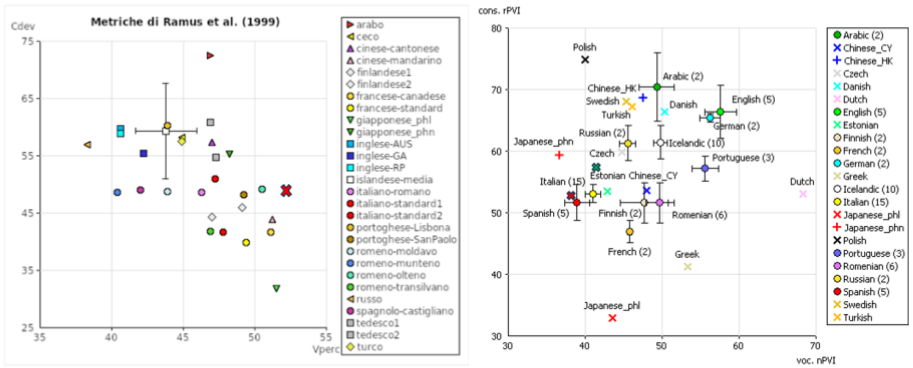


Figure 13. Recreation of Figures 8 and 11 in Mairano and Romano (2010) with MS included.

The interaction between V% (52.9) and ΔC (47) places MS among syllable-based languages (red cross in left chart of Figure 13). When examining these values individually, MS aligns with Castilian Spanish in consonantal deviation, yet its vocalic percentage exceeds the upper limits reported by Ramus et al. (1999). Similarly, PVI values situate it within the syllable-timed languages in Grabe and Low (2002); CrPVI (54.4) and VnPVI (39.30) set MS in an intermediate position between Italian and Spanish (red cross in right chart of Figure 13). The interplay between Varcos (VarcoV 50.68; VarcoC 51.25), however, situates MS in a transitional zone between syllable- and stress-based languages in Dellwo (2006), notably approaching English. Yet, when we assess them independently, MS VarcoC approaches Italian, a syllable-based language.[10]

Table 6. PVI values across generations of MS.

MS Generation	CrPVI	VnPVI
Gen1	53.83	42.48
Gen2	52.42	36.27
Gen3	56.89	36.16

When looking at durational rhythmic values across generations, differences can be clearly observed in PVI values; Table 6 shows that while consonantal rPVI gradually increases towards the third generation, vocalic nPVI decreases.

[10] No specific language associations can be drawn from CCI values or Vdev.

5.3 Speech rate

Data analysis for CV-rate, V-rate, C-rate, and syllable-rate (articulation) reveals that an MS speaker produces an average of 10.29 CV segments/s and 5.36 syllables/s.

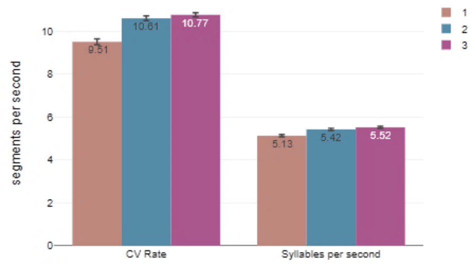


Figure 14. Generational CV- and syllable-rates.

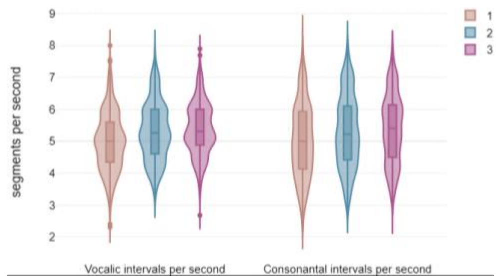


Figure 15. Generational V- and C-rates.

Results demonstrated statistically significant differences between MS generations for all speech rate variables (p=<.001). While a notable trend of growth emerges across MS generations in both measures (Figure 14), the gradual increase is notable in CV rate, where a significant shift can be seen between the oldest generation and the rest. When observing V- and C-rates independently, the cross-generational rise is constant in the latter; in vocalic rate, second and third generations converge, both surpassing gen1's average (Figure 15). Interestingly, while the oldest group exhibits more variability (variance of 6.18) reaching the lowest and highest registered rates, gen3's dispersion is smaller (variance of 3.63).

6. Discussion and conclusions

An analysis of intonational variables in section 4.1 suggests the MS intonational system relies mainly on peak placement, peak height difference and boundary tones to mark differences within and across utterance types that are neutral. Prenuclear and nuclear inventories in MS are mainly composed of low-high bitonal pitch accents, highlighting its low complexity. While prenuclear production remains consistent, nuclear configurations exhibit great variability, primarily driven by distinctive peak position and peak height behavior. More specifically, polar questions exhibit posttonic peaks in both positions, creating identical initial and final low-high pitch accents that results in final HL%, typical of Rioplatense (Rebollo Couto et al. 2014). The common presence of equal height peaks and higher nuclear ones in YN (often forming L*+¡H) explains its positive peak height difference, setting it apart from the other utterance types. Conversely, statements follow a tonal downtrend due to nuclear deaccenting or the presence of a lower tonic-aligned peak, both cases ending in M%. Similarly, the main IP of tag questions primarily exhibit low nuclear tonic-aligned peaks shaping a tonal fall that ends in a L% and combines with a variety of patterns that mark confirmation or agreement in the tag word. The lower nuclear peaks compared to prenuclear ones in SS and TQ display inverse tonal patterns than YN, explaining their distinct negative PHD values.

In terms of rhythm, tonal and durational results in section 4.2 partially align with predictions, revealing interesting new insights. While it could be said that MS is a strong macro rhythmic language in terms of its low number of possible pitch accents (all low-high bitonal) and uncommon presence of level tones in pitch accents, we cannot state the same in terms of (a) macro frequency level, as it does not approach 1 due to the recurrent absence of nuclear peaks in statements, and of (b) boundary tones, which highly vary and frequently form level tones (rules i to iii in Jun (2012)). More transparently, durational metrics verify its syllable-based nature; vocalic and consonantal PVI values effectively place MS within syllable-based languages (Grabe and Low 2002) and in an intermediate position between Spanish and Italian, which is expected given the strong Italian influence on Uruguayan Spanish vocalic duration (Brockhorst 2020). Interestingly, results in 4.3 are consistent with these findings: MS presents a mean CV-rate that falls within the anticipated range and threshold of syllable-based languages (Dellwo 2010). Interestingly, the notably low mean articulation rate observed in MS place it below the average values of other Spanish varieties like Mexican or Madrilian (Santiago and Mairano 2022). Further perceptual research would be needed to confirm whether MS is perceived as a fast or slow variety of Spanish.

Across generations, significant differences emerge in MS prosody: (1) older speakers tend to use upstepped pitch accents in polar questions, in contrast to younger speakers; (2) positive peak height difference values are exclusive to the first generation, reflecting their more frequent production of higher nuclear peaks; (3) Nuclear deaccenting is more common among the younger generation, especially in statements, and (4) younger speakers exhibit lower vocalic durational variability, which was expected due to the longer vocalic durational patterns seen in Bronckhorst's (2020) in older MS speakers.
These cross-generational differences not only confirm the dynamic nature of MS' prosodic system but also signal changes in progress, possibly indicating a gradual loss of the Italian-like Rioplatense prosody and a shift towards a less fluctuating tonal and durational system.

In conclusion, our hypotheses were mostly confirmed. Evidence suggests MS is a syllable-based language with a dynamic intonational and rhythmic system that employs various strategies to convey meaning, possibly undergoing change. The fact that most measures returned transparent results allowed us not only to describe its prosody but also to account for generational variation. Finally, this paper shows that adopting a multidimensional analytical approach, diverging from a single uniform model, can enrich our comprehension of the intricacies of the prosodic system of a language and test their complementary nature. With this first complete description of MS prosody, we hope to motivate further research that continues to explore its similarities and differences with its sister language, Porteño Spanish. Further studies should focus on exploring MS prosody in the context of speech perception.

References

Araujo, Deborah Silva de. 2013. Declarative and yes/no question intonation in Montevideo Spanish. Master's thesis, Stony Brook University.

Aziz, Jacob, Vanina Machado, Natasha Swederski, Celina Valdivia, Ryan Stevenson, Yasaman Rafat, and Rajiv Rao. 2022. Intonation patterns in Argentinean and Venezuelan Canadian heritage speakers of Spanish: Parental and English influences. *Heritage Language Journal* 2(1): 60–90.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1): 1–48.

Bertinetto, Pier Marco, and Chiara Bertini. 2010. Towards a unified predictive model of natural language rhythm. In *Prosodic Universals: Comparative studies in rhythmic modeling and rhythm typology*, ed. Michela Russo, 43–78. Rome: Aracne.

Bertolotti, Virigina, and Magdalena Coll. 2014. *Retrato lingüístico del Uruguay*: *un enfoque histórico sobre las lenguas en la región*. Montevideo: Universidad de la República.

Boersma, Paul, and Vincent van Houven. 2001. Praat: A system for doing phonetics by computer. *Glot International* 5(9/10): 341–345.

Brockhorst, Tobias. 2020. Ritmo y contacto interlingüístico: La duración vocálica en el español uruguayo. Master's thesis, Universidad Humboldt de Berlín.

Colantoni, Laura, and Jorge Gurlekian. 2004. Convergence and intonation: Historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition* 7(2): 107–119.

Colantoni, Laura, Jeffrey Steele, and Paola Escudero. 2015. *Second language speech*. Cambridge: Cambridge University Press.

Cunha, Claudia, Glaucia F. De Santos, Leticia Rebollo Couto, and Maristela da S. Pinto. 2008. Focus and intonational marking in boundaries dialects: Brazilian Portuguese and Uruguayan Spanish in yes/no questions. In *Proceedings from Speech Prosody 2008*, 551–554.

Datatab Team. 2023. Datatab: Online statistics calculator. DATAtab e.U. Graz: Austria.

Dellwo, Volker. 2006. Rhythm and speech rate: A variation coefficient for deltaC. In *Language and language processing*, ed. Pawel Karnowski and Imre Szigeti, 231–241. Berlin: Perter Lang.

Dellwo, Volker. 2010. Influences of speech rate on the acoustic correlates of speech rhythm: An experimental phonetic study based on acoustic and perceptual evidence. Doctoral dissertation, Universität Bonn.

Face, Timothy. 2001. Focus and early peak alignment in Spanish intonation. *Probus* 13(2): 223–246.

Grabe, Esther, and Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* 7(1982): 515–546.

Hualde, Jose Ignacio. 2003. Remarks on the diachronic reconstruction of intonational patterns in Romance with special attention to Occitan as a bridge language. *Catalan Journal of Linguistics* 2: 181–205.

Hualde, Jose Ignacio, and Pilar Prieto. 2015. Intonational variation in Spanish: European and American varieties. In *Intonation in Romance,* ed. Sonia Frota and Pilar Prieto, 350–391. Oxford: Oxford University Press.

Jun, Sun-Ah. 2012. Prosodic typology revisited: adding macro-rhythm. In *Proceedings from Speech Prosody 2012*, 535–538.

Jun, Sun-Ah. 2014. Prosodic typology: by prominence type, word prosody, and macro-rhythm. In *Prosodic typology II: The phonology of intonation and phrasing*, ed. Sun-Ah Jun, 520–539. Oxford: Oxford University Press.

Jun, Sun-Ah, and Janet Fletcher. 2014. Methodology of studying intonation: From data collection to data. In *Prosodic typology II: The phonology of intonation and phrasing*, ed. Sun-Ah Jun, 493–519. Oxford: Oxford University Press.

Machado, Vanina. 2024. The prosody of Uruguayan-Brazilian border Spanish: Cross-generational and social variation. Doctoral dissertation, University of Toronto.

Mairano, Paolo, and Antonio Romano. 2010. Un confronto tra diverse metriche ritmiche usando Correlatore. In *La dimensione temporale del parlato*, ed. Stefan Schmid, Franziska M. Schwarzenbach, and Joho D. Studer, 79–100.

Nooteboom, Sieb. 1999. The prosody of speech: Melody and rhythm. In *The handbook of phonetic sciences*, ed. William J. Hardcastle and John Laver, 640–673. Chichester, UK: Blackwell.

Pierrehumbert, Janet. 1980. The phonology and phonetics of English intonation. Doctoral dissertation, UILC Bloomington.

Prieto, Pilar, and Paolo Roseano. 2010. *Transcription of intonation of the Spanish language*. Munich: Lincom Europa.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramus, Franck, Marina Nespor, and Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3): 265–292.

Rato, Anabela, and Vanina Machado. 2024. Perception and production of stops by Canadian heritage speakers of European Portuguese. In *The phonetics and phonology of heritage languages*, ed. Rajiv Rao, 343–360. Cambridge: Cambridge University Press.

Rebollo Couto, Leticia, Priscilla Ferreira de Sá, and Natalia Dos Santos Figueiredo. 2014. Actitudes lingüísticas y entonación: acentos tonales y enunciados interrogativos en el español de Buenos Aires y Montevideo. In *Argumentos cuantitativos y cualitativos en sociolingüística: Segundo coloquio de cambio y variación lingüística,* 171–198.

Santiago, Fabian, and Paolo Mairano. 2022. Spaniards articulate faster than Mexicans: Temporal patterns in two varieties of Spanish. *Spanish in Context* 19(2): 244–264.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Price Colin Wightman, Patti Pierrehumbert, and Janet Hirschberg. 1992. ToBI: A standard for labeling English prosody. *ICSLP* 2: 867–870.

Simões, Antonio. 2022. *Spanish and Brazilian Portuguese pronunciation: The mainstream pronunciation of Spanish and Brazilian Portuguese, from sound segments to speech melodies*. Singapore: Springer Nature.

Sosa, Juan Manuel. 1999. *La entonación del español: Su estructura fónica, variabilidad y dialectología*. Madrid: Cátedra.

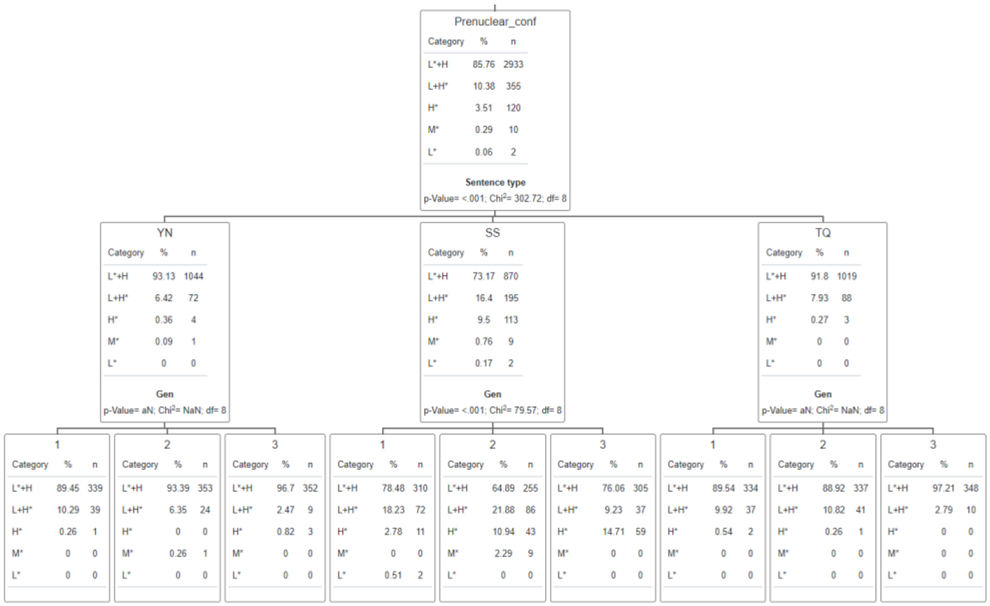
Steffman, Jeremym, and Sun-Ah Jun. 2019. Listeners integrate pitch and durational cues to prosodic structure in word categorization. In *Proceedings of the Linguistic Society of America*, ed. Patrick Farrell, 1–15. Washington, DC: Linguistic Society of America.

Tilsen, Sam, and Amalia Arvaniti. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, 134(1): 628–639.

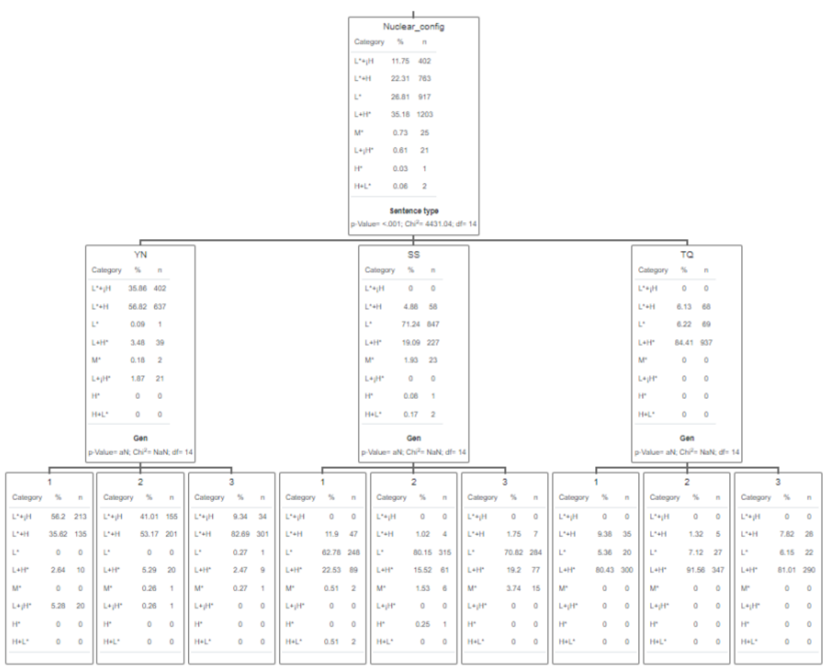
Vicenik, Chad. 2011. The role of intonation in language discrimination by infants and adults. Doctoral dissertation, University of California Los Angeles.

Appendix

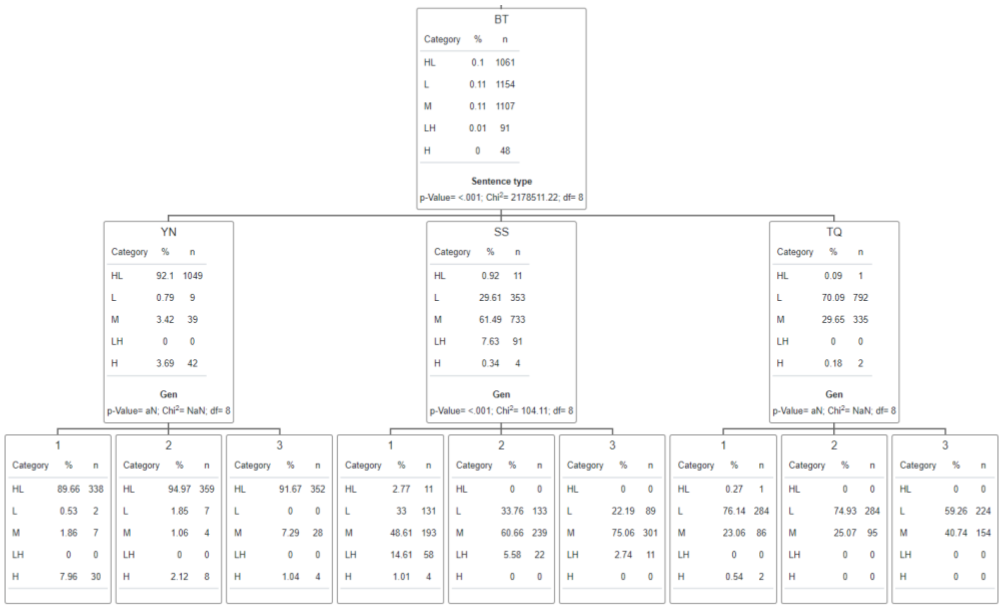
CHAID tree 1. Prenuclear pitch accents.



CHAID tree 2. Nuclear pitch accents.



CHAID tree 3. Boundary tones.



CHAID tree 4. Nuclear peak position.