The Effects of Syntactic Dependencies and Speech Tempo on Macro-Rhythm

Torres, Catalina ; Morand, Marie-Anne ; Sauppe, Sebastian ; Bickel, Balthasar

Originally published at:
Torres, Catalina; Morand, Marie-Anne; Sauppe, Sebastian; Bickel, Balthasar (2024). The Effects of Syntactic Dependencies and Speech Tempo on Macro-Rhythm. In: Nineteenth Australasian International Conference on Speech Science and Technology, Melbourne, Australia, 2 December 2024 - 5 December 2024. ASSTA, 157-161.

The Effects of Syntactic Dependencies and Speech Tempo on Macro-Rhythm

Catalina Torres[1], *Marie-Anne Morand*[2], *Sebastian Sauppe*[1], *Balthasar Bickel*[1]

[1]University of Zurich, [2]University of Fribourg

catalina.torres@ivs.uzh.ch, marie-anne.morand@unifr.ch,
sebastian.sauppe@psychologie.uzh.ch, balthasar.bickel@uzh.ch

Abstract

Macro-rhythm is a parameter that distinguishes the prosodic profiles of languages at the intonational level. However, no conclusive quantification of macro-rhythm differences between languages has been proposed and there are to date no dedicated studies investigating the regularity and variability of macro-rhythm within a single language. This study experimentally examines the global tonal patterns of Swiss German to test (i) internal effects related to word order and (ii) articulation rate effects on global tonal patterns. We find that macro-rhythm is affected by language-internal syntactic variation and by intra-speaker variability (in terms of speech tempo).

Index Terms: macro-rhythm, prosody, word order, verb clusters, crossed/nested/adjoined dependencies, Swiss German

1. Introduction

This study analyses the relationship between intonational phonology and syntactic dependencies by examining global tonal patterns within the autosegmental-metrical (AM) approach [1, 2, 3]. The goal is to investigate the correspondence between intonational phonology and syntactic linearisation in Zurich German, a variety of German spoken in Switzerland allowing for variable word order in complex verb clusters. In AM, tonal space refers to an interval where the high (H) tones are at the top level and the low (L) tones are at the bottom level of a band of fundamental frequency (F0) values [2]. Variations in tonal space, such as intonational boundary tones, can serve as a phonological representation of prosodic constituency [4] or interact with other linguistic domains, such as syntactic structure. For example, a major syntactic break can be associated with an intermediate phrase boundary [5, 6]. In addition to signaling prosodic prominence and boundary marking, prosodic macro-rhythm [3] has also been proposed as a third parameter distinguishing the prosodic profiles of languages. Macro-rhythm is the temporal organisation of speech, perceived by the regular occurrence of tonal events in the F0 contour (i.e., rising and falling tonal movements, see Figure 1). At the intonational level, macro-rhythm differs in three dimensions [3]: (i) the *number of possible accents*, with larger accent inventories predicting more tonal variability and therefore lower degrees of macro-rhythm; (ii) the *type of most common accents*, with rising or falling accents producing higher degrees of macro-rhythm than level accents; and (iii) the *frequency or domain of accents*, with more accents in a phrase resulting in a higher degree of macro-rhythm, because every smaller domain (e.g., every prosodic word) receives an accent.

Cross-linguistic studies on macro-rhythm have focused on F0 measures taken from speech corpora and used different calculations, with mixed results [7, 8, 9]. While nPVI calculates the variability in the distance of intervals between F0 peaks and valleys, MacR_Var focuses on the standard deviations of rising or falling slopes, peak-to-peak distance and valley-to-valley distance per Intonation Phrase. However, these measures do not account for possible variation arising from the use of different syntactic structures within a language or from intra-speaker variability in the rate of articulation. The effects arising from the number of possible accents may be less important in predicting macro-rhythm compared to the type of the most common accent and their frequency domain [8]. Here, we examine the influence of syntax on the number of possible accents per phrase, while also considering the most common accents in Swiss German. Although boundary synchronisation between prosodic and syntactic constituents has been documented, little is known about how syntax interacts with macro-rhythm and phrase-level global tonal patterns. Rising and falling tonal movements play an important role in attention orienting. For example, neurophysiological studies have shown that rises in amplitude of pure sine tones are linked to auditory looming effects [10] and that tonal rises, in particular, guide auditory attention when listening to speech [11]. We examine the relationship between verb clusters and macro-rhythm testing whether word order has an effect on intonational patterning.

1.1. Zurich German

In German-speaking Switzerland, people usually use at least two language varieties of German regularly: (Swiss) Standard German and a Swiss German dialect. In contrast to other German-speaking countries, there is no standard-dialect continuum [12]. The sociolinguistic situation can thus be described as *diglossia* [13, 14, 15] with rather clear-cut boundaries between the standard and the vernacular varieties. Previous studies of prosodic features in Swiss German focused on temporal aspects, such as speech rhythm and speech tempo [16, 17, 18], and little is known about intonational patterns. Compared to (northern) Standard German, Zurich German shows a larger overall F0 range, with a greater number of pitch movements and a default pitch accent (L*+H) consisting of a low–rising contour, and a slower speech rate [19]. Thus, Zurich German can be characterised as having a stronger macro-rhythm, compared to Standard German with its medium degree of macro-rhythm [3].

1.2. Complex verb clusters in Zurich German

This study examines adjacent and non-adjacent syntactic dependencies (i.e., relations between sentence elements) and how they are prosodically marked. Dependencies can be expressed in different ways: Either so that dependency arcs cross each other (as in the pattern *ABAB*, where the *A*s and the *B*s are non-adjacent, cf. Figure 2-1) or so that the elements and their dependencies are nested within each other (as in *ABBA*, where the *B*s are adjacent, but not the *A*s, cf. Figure 2-2) [20]. A third possible realisation are adjoined dependencies (as in *AABB*, where each dependency arc is immediately closed, cf. Figure 2-3).

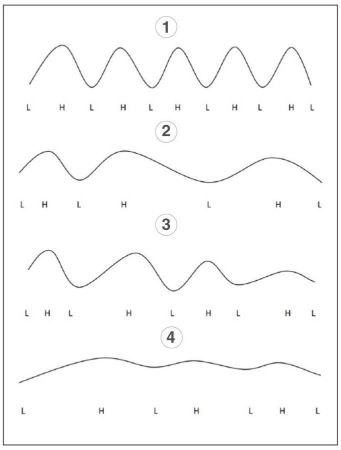

Figure 1: Schematic pitch contours showing differences in the temporal (1–2) and frequency domain (3–4) of macro-rhythm (adapted from [7]). Contour 1 exhibits H and L turning points that are more evenly distributed in time and thus has a stronger macro-rhythm than contour 2. Contour 3 exhibits larger differences in frequency between successive H and L turning points and thus creates a stronger macro-rhythm than contour 4.

Verb clusters (or verbal complexes) are verb phrases that contain several verbs. In German, these verbs are usually a finite auxiliary or modal verb (*will* 'want' in 2), a main verb in the infinitive form (*geh* 'give' in Figure 2), and its object if it is transitive (*en Ring* 'a ring' in Figure 2). While any lexical verb can function as the infinitive main verb, there is only a small, closed class of auxiliaries and modals that can function as finite verb in verbal complexes. Non-adjacent dependencies bear great relevance in linguistics because they have played a major role in Formal Language Theory (FLT), which describes which computational properties grammars of human languages need to have [21, 22]. In the Chomsky hierarchy in FLT, syntactic structures with crossed dependencies require more complex computations to be describable, thus requiring context-sensitive grammars [21]. Crossed dependencies are cross-linguistically rare and have so far only been attested in Dutch and Swiss German [23, 24]. We take advantage of the flexibility of word order in the Swiss German verbal complex to explore whether there are prosodic correlates of syntactic structure that signal which type of dependency a currently uttered sentence has. This is possible because nested, crossed, or adjoined dependencies can be used to express the same meaning (Figure 2). Signaling the difference between crossed and nested dependencies could, for example, be helpful to listeners, who are known to form expectations about the upcoming linguistic input [25, 26] and thus would potentially benefit from cues about the dependency type.

1.3. Research questions

The variable word order in Swiss German verb clusters offers the possibility to test how intonational structure is realised in relation to syntax, while keeping the semantic information constant. Considering the reported relevance of rising tonal movements for guiding perception [11], it is conceivable that the intonational structure of complex syntactic structures is organised to facilitate their parsing. In this study, we therefore ask the following questions:

(1) dass	d'Maria	em Manu	en Ring	will	geh
that	Maria	the.DAT Manu	a ring	want	give
that	Maria	the.DAT Manu	object	verb 1	verb 2

(2) dass	d'Maria	em Manu	en Ring	geh	will
that	Maria	the.DAT Manu	a ring	give	want
that	Maria	the.DAT Manu	object	verb 2	verb 1

(3) dass	d'Maria	em Manu	will	en Ring	geh
dass	Maria	the.DAT Manu	want	a ring	give
that	Maria	the.DAT Manu	verb 1	object	verb 2

Figure 2: Three verb cluster orders and their dependency structure in Zurich German for the clause '. . . that Maria wants to give Manu a ring'. (1) crossed, (2) nested, (3) adjoined.

R1 What are the global tonal patterns of utterances with the three different word orders in the verb cluster (crossed, nested, adjoined)?

R2 Do these global tonal patterns differ?

R3 If global tonal patterns differ, how do they differ?

2. Materials and Methods

Stimulus sentences were created by a native speaker of Swiss German, specialist of Zurich German. Subsequently, these sentences were checked for comprehensibility by a native speaker of Zurich German. Sentences considered difficult to comprehend were discarded before data collection. To create stimuli that allowed the recognition of intonation contours, the sentences contained as many sonorant consonants as possible. A set of 55 'dass' frame sentences were created and the three possible verb clusters were used (cf. Figure 2). To test whether sentence length has an effect on the intonational contour we created stimuli of varying length. The sentences vary in the number of syllables they contain, in a scale of five steps each sentence becomes longer by one additional syllable. This count only includes the syllables of the words that can be accented. This means, that in the region of interest, we have contours with 5 to 10 syllables. As there is no standard orthography for Zurich (or Swiss) German, our stimuli were made using a writing system that could be used, for instance, when chatting on social media and which had been tested in another study before.

2.1. Participants

Ten speakers (5 female) of Zurich German were invited to the study. Speakers were 20-33 years old (mean = 24.2 years) and reported growing up in the canton of Zurich, as well as living in the city of Zurich (or near by) at the time of recording.

2.2. Recordings

Participants were recorded at a self-selected normal and fast speech rate and compensated with study credit. Before recording, the participants obtained a printed copy containing all the sentences so that they could become familiar with them and ask questions if anything was unclear. They were seated in front of a screen in a sound-attenuated booth and were prompted with a written sentence they had to read out loud, using ProRec software (Mark Huckvale, University College London). If they felt they were not fluent or misread the sentence, the recording was repeated. The recordings were made at a sampling rate of 44.1 kHz and 16 bit, using a Røde 1000 large diaphrame condensor microphone.

2.3. Analysis procedures

To obtain a transcription at the utterance level, a bash script was used. Text files and associated WAV files were used for forced alignment of the speech signal via the web interface of the Munich Automatic Segmentation System using the CH (German Dieth) language model. For forced alignment, the web service G2P was used to convert the orthographic text input into a canonical phonological transcript [27]. The resulting files were used to compute a phonetic segmentation and labelling based on the speech signal and a phonological transcript in webMAUS [28]. The alignment process provided a TextGrid for each utterance in which all words were segmented and marked with boundaries. Data was further processed using Praat 6.3.06 [29]. There were a total of 1556 utterances in the corpus[1].

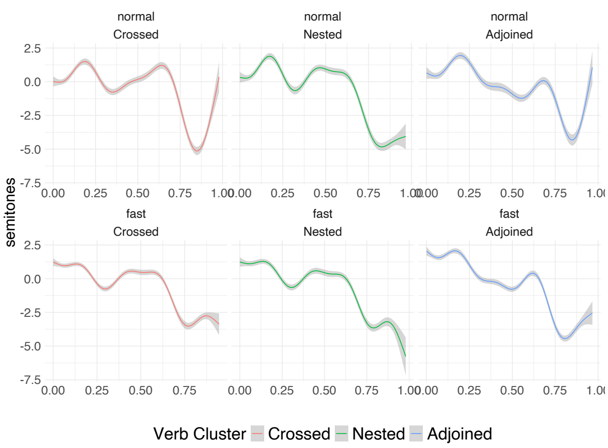

Figure 3: Smoothed fundamental frequency trajectories of three verb clusters in semitones of normal (top) versus fast (bottom) speech rate. The start of the word *Manu* represents the onset.

We are interested in the intonational contours at the phrasal level. Figure 3 shows F0 contours for the three verb clusters. To obtain the F0 contours, 150 consecutive measurements were taken per utterance, starting with the onset of the word *Manu* (cf. Figure 2), up to the end of the utterance. As described in §2, we created our speech materials to allow for a continuous F0 contour with minimal devoiced segmental material (which could not be entirely avoided). However, there are unavoidable phonetic perturbations that are not intended when the F0 movements are planned during speech production and which can lead to inaccurate F0 measurements. As the F0 contour can be affected by short-term perturbations caused by segmental characteristics such as junctures between consonants and vowels, variations in voice quality, or F0 tracking errors, we decided to use an automated method for detecting F0 measurement jumps based on sample-to-sample differences [30]. Utterances produced with intervening pauses were discarded to avoid effects resulting from pitch reset in relation to prosodic boundaries. To test whether forced alignment provided a reliable segmentation of words, a subset of the data was hand corrected (31% of the corpus). The onset and offset time stamps of the word *Manu* were queried and statistically evaluated using Wilcoxon rank sum tests. The mean onset in forced-aligned words (840 ms) was not significantly different from that of the hand-corrected sample (836 ms) ($p = 0.6$, effect size $r = 0.015$). A similar result was obtained for the offset ($p = 0.4$, effect size $r = 0.026$), confirming the reliability of forced-aligned utterances.

2.3.1. Statistical analysis

The F0 trajectories are analysed with Generalised Additive Mixed Models (GAMMs) in R [31], using the `mgcv` [32]. We include syntax and sex as parametric predictors and additional random effects for speaker and sentence ID. Following [33], we created new (factor smooth) variables, representing the interaction between verb cluster type and the number of syllables per clause and representing the interaction between verb cluster type and speech rate.

3. Results

The results of the GAMMs show that the tonal global patterns systematically differ between the three verb clusters.

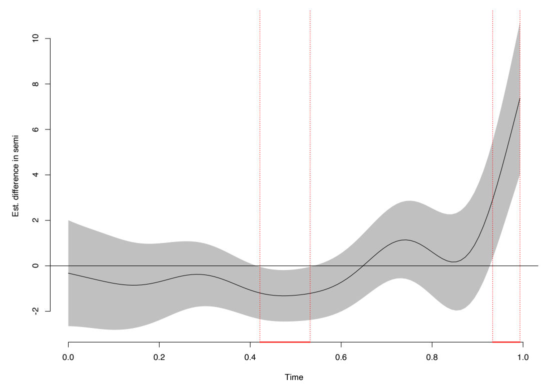

Figure 4: Difference smooths contrasting crossed versus nested verb clusters in short utterances. The pointwise 95%-confidence interval is shown by the shaded band.

Additionally, we find a statistically significant difference for the manipulation of speech rate. As shown in Figure 3, contours at a normal speech rate show more expanded tonal movements (i.e., stronger macro-rhythm), while the F0 range is narrower at faster speech rates (i.e., weaker macro-rhythm). Figures 4–6 show the fitted difference smooths between verb cluster types. When the estimated difference is significantly different from zero (i.e., when the 95% confidence interval for the difference between smooths does not include 0), this is indicated by a red line on the x-axis and vertical dotted lines on the y-axis. First, the onset word *Manu* (held constant across all stimuli) does not show any significant difference. However, the tonal patterns differ significantly towards the middle and end portions of the contours. Figure 4 shows the estimated difference between short crossed and nested verb clusters. This illus-

[1]For two female participants recorded in a pilot there were only 48 frame sentences (288 in total).

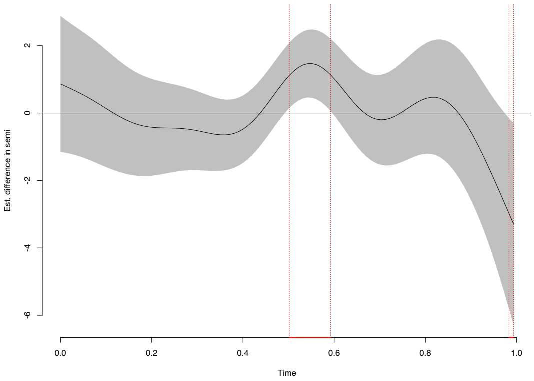

Figure 5: Difference smooths contrasting crossed versus adjoined verb clusters in short utterances. The pointwise 95%-confidence interval is shown by the shaded band.

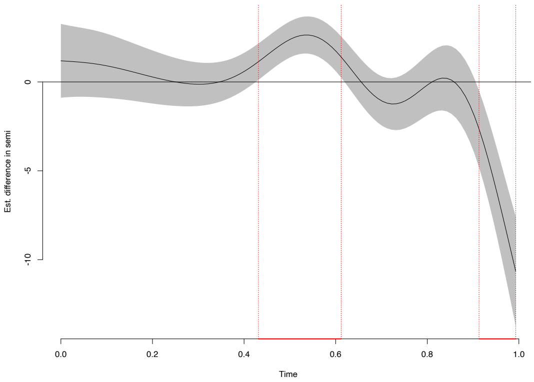

Figure 6: Difference smooths contrasting nested versus adjoined verb clusters in short utterances. The pointwise 95%-confidence interval is shown by the shaded band.

trates a significant difference towards the midpoint and the end of the utterances, whereby the contour in the crossed clusters is lower in the middle but higher at the end compared to the nested condition. Figure 5 shows the estimated difference between the crossed and adjoined verb clusters whereby the intonational contour of crossed verb clusters is higher at the midpoint and towards the end compared to nested verb clusters. Figure 6 shows the estimated difference between the adjoined and nested verb clusters. In this case, the intonation contour in the nested condition appears to be higher at midpoint but lower towards the end compared to the adjoined contours. Taken together, the contours vary due to word order, despite all sentences sharing the same lexical material. This is evident in the short utterances in which all content words in the verb cluster are monosyllabic (see Figure 2). Additionally, the differences in global tonal pattern are mostly robust for further comparisons between crossed versus nested and nested versus adjoined. However, in longer utterances comparisons between crossed and adjoined are not significantly different.

4. Discussion and Conclusion

Although macro-rhythm has been proposed as an additional parameter to distinguish languages prosodically [3], it has proved to be difficult to measure in acoustic terms. We investigated two factors that may have an influence on macro-rhythm within a single language, focusing on the role of language-internal word order alternations and on speech tempo variation. We find that an interaction between syntax and prosody modulates macro-rhythm and that speech tempo affects measures of macro-rhythm through the magnitude of tonal movements. It is likely that, similar to rhythm metrics [34], the (ir-)regularity of macro-rhythm is related to perceptual rather than acoustic factors. We note that in our data, global tonal effects seem to be driven by the predictability of lexical material, whereby less predictable content words are realised with rising tonal movements and more predictable verbs are deaccented. In a comparison of the intonational contours of three different verb clusters in Zurich German (crossed, nested, and adjoined), we find differences between all of them. As shown in Figure 3, within the portion of the dependencies (starting with *Manu*, Figure 2), we find three tonal rises in the crossed and adjoined contours at a global level. In comparison, the nested verb clusters show only two tonal rises. Additionally, the nested contours in Figure 3 also show deaccentuation of the sentence-final finite verb (*will* 'want', Figure 2). Although the current methodology does not allow us to reconstruct the exact position of segments, the controlled nature of the stimulus material makes it possible to estimate the most important points in the utterance, where we observe significant variation.

Out of the three verb clusters studied, the nested structure shows a lower macro-rhythm. Important differences arise in the region of the verb clusters (excluding the tonal rise in Manu, see Figure 2). We find that intonational differences are modulated by the height of the contour (where we predict L*+H pitch accents) on the object and the infinite verb 2 (*geh* 'give', Figure 2) while the finite verb (*will* 'want', Figure 2) shows deaccentuation. One possible explanation for these differences is that the number of verbs that can occur as the finite verb in complex verbal clusters is limited and therefore more predictable. In contrast, the object and infinite verb in these constructions can be freely chosen and are thus less predictable. If we consider the relevance of rising tonal movements in guiding auditory perception when listening to speech material [11], this can explain why the less predictable speech material is accented (object, infinite verb) while the more predictable speech material (restricted set of possible finite verbs) is not. In the related variety of Bernese German [35], deaccentuation was reported for words out of focus. This is in line with our observation that the more informative material is realised with a rising intonation whereas the less informative material is deaccented. The experimental investigation shows the effects of syntax and speech tempo on global tonal patterns. Our results show that macro-rhythm is sensitive to language internal structure provided by word order, whereby some structures can lead to an increased number of tonal movements. Additionally, we show that variations of speech tempo have an effect on the contour, whereby a fast speech tempo leads to a narrower F0 range. These results can help us understand why previous measures led to mixed results. First, we find that variable word order of otherwise equal lexical material leads to variation in global tonal patterns. Thus, not only the number of words per utterance but also how these words are syntactically structured plays an important role. Second, although intra-speaker variation has been acknowledged as a possible confound [9, 8], its role had not been demonstrated before. Taken together, we show that variable word order in verb clusters influences global tonal patterns and that speech tempo modulates pitch range in Zurich German.

5. Acknowledgements

We thank Sabrina Ryffel and Giorgia Sironi for help during data collection and Pelin Durtschi for her support in hand correcting segmentation, as well as Chantal Oderbolz for support with the creation of stimuli. We also thank the Linguistic Research Infrastructure (LiRI) at the University of Zurich for technical support. This work was funded by the University of Zurich and the Swiss National Center for Competence in Research "Evolving Language" (Swiss National Science Foundation agreement number 51NF40_180888, B.B.). S.S. is supported by the Swiss National Science Foundation Grant PZ00P1_208915.

6. References

[1] J. Pierrehumbert, "The phonology and phonetics of English intonation," Ph.D. dissertation, 1980.

[2] D. R. Ladd, *Intonational phonology*. Cambridge University Press, 2008.

[3] S.-A. Jun, *Prosodic typology: by prominence type, word prosody, and macro-rhythm*. Oxford University Press, 2014, pp. 520–540.

[4] H. Truckenbrodt, F. Sandalo, and B. Abaurre, "Elements of Brazilian Portuguese intonation," *Journal of Portuguese Linguistics*, vol. 8, no. 1, pp. 75–114, 2009.

[5] M. D'Imperio and A. Michelas, "Pitch scaling and the internal structuring of the Intonation Phrase in French," *Phonology*, vol. 31, no. 1, pp. 95–122, 2014.

[6] C. Torres, J. Fletcher, and G. Wigglesworth, "Fundamental frequency and regional variation in Lifou French," *Language and Speech*, vol. 65, no. 4, pp. 889–922, 2022.

[7] L. Polyanskaya, M. G. Busà, and M. Ordin, "Capturing cross-linguistic differences in macro-rhythm: The case of Italian and English," *Language and Speech*, vol. 63, no. 2, pp. 242–263, 2020.

[8] C. Kaland, "Bending the string: intonation contour length as a correlate of macro-rhythm." in *Interspeech*, 2022, pp. 5233–5237.

[9] C. Prechtel, "Macro-rhythm in English and Spanish: Evidence from radio newscaster speech," *Speech Prosody 2020*, pp. 675–679, 2020.

[10] D. R. Bach, H. Schächinger, J. G. Neuhoff, F. Esposito, F. D. Salle, C. Lehmann, M. Herdener, K. Scheffler, and E. Seifritz, "Rising sound intensity: an intrinsic warning cue activating the amygdala," *Cerebral Cortex*, vol. 18, no. 1, pp. 145–150, 2008.

[11] M. Lialiou, M. Grice, C. T. Röhr, and P. B. Schumacher, "Auditory processing of intonational rises and falls in German: rises are special in attention orienting," *Journal of cognitive neuroscience*, vol. 36, no. 6, pp. 1099–1122, 2024.

[12] U. Ammon, "Dialektschwund, Dialekt-Standard-Kontinuum, Diglossie: Drei Typen des Verhältnisses Dialekt-Standardvarietät im deutschen Sprachgebiet," in *Standardfragen: Soziolinguistische Perspektiven auf Geschichte, Sprachkontakt und Sprachvariation*, J. Androutsopoulos and E. Ziegler, Eds. Lang, 2003, pp. 163–171.

[13] C. A. Ferguson, "Diglossia," *word*, vol. 15, no. 2, pp. 325–340, 1959.

[14] J. A. Fishman, "Bilingualism with and without diglossia; diglossia with and without bilingualism," *Journal of Social Issues*, vol. 23, no. 2, pp. 29–38, 1967.

[15] G. Kolde, *Sprachkontakte in gemischtsprachigen Städten: vergleichende Untersuchungen über Voraussetzungen und Formen sprachlicher Interaktion verschiedensprachiger Jugendlicher in den Schweizer Städten Biel/Bienne und Fribourg/Freiburg i. Ue.* Steiner, 1981.

[16] A. Leemann, V. Dellwo, M.-J. Kolly, and S. Schmid, "Rhythmic variability in Swiss German dialects," in *Proceedings Speech Prosody 2012*, 2012, pp. 607–610.

[17] U. Zihlmann, "Vowel and consonant length in four Alemannic dialects and their influence on the respective varieties of Swiss Standard German," *Wiener Linguistische Gazette*, vol. 86, pp. 1–46, 2020.

[18] M.-A. Morand, M. Bruno, S. Schwab, and S. Schmid, "Syllable rate and speech rhythm in multiethnolectal Zurich German: A comparison of speaking styles," in *Proceedings Speech Prosody 2022*, 2022, pp. 337–341.

[19] J. Fleischer and S. Schmid, "Zurich German," *Journal of the International Phonetic Association*, vol. 36, no. 2, pp. 243–253, 2006.

[20] M. H. de Vries, K. M. Petersson, S. Geukes, P. Zwitserlood, and M. H. Christiansen, "Processing multiple non-adjacent dependencies: evidence from sequence learning," *Philosophical Transactions of The Royal Society B*, vol. 367, no. 1598, pp. 2065–2076, 2012.

[21] G. Jäger and J. Rogers, "Formal language theory: refining the Chomsky hierarchy," *Philosophical Transactions of The Royal Society B*, vol. 367, no. 1598, pp. 1956–1970, 2012.

[22] W. T. Fitch and A. D. Friederici, "Artificial grammar learning meets formal language theory: an overview," *Philosophical Transactions of The Royal Society B: Biological Sciences*, vol. 367, no. 1598, pp. 1933–1955, 2012.

[23] S. M. Shieber, "Evidence against the context-freeness of natural language," *Linguistics and Philosophy*, vol. 8, no. 3, pp. 333–345, 1985.

[24] J. Reese, *Swiss German: The Modern Alemannic Vernacular in and around Zurich*. München: Lincom Europa, 2007.

[25] G. R. Kuperberg and T. F. Jaeger, "What do we mean by prediction in language comprehension?" *Language, Cognition and Neuroscience*, vol. 31, no. 1, pp. 32–59, 2016.

[26] K. D. Federmeier, "Thinking ahead: The role and roots of prediction in language comprehension," *Psychophysiology*, vol. 44, no. 4, pp. 491–505, 2007.

[27] U. D. Reichel and T. Kisler, "Language-independent grapheme-phoneme conversion and word stress assignment as a web service," *Studientexte zur Sprachkommunikation*, pp. 42–49, 2014.

[28] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, 2017.

[29] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (version 6.3.06)[computer program]." 2023.

[30] J. Steffman and J. Cole, "An automated method for detecting F0 measurement jumps based on sample-to-sample differences," *JASA Express Letters*, vol. 2, no. 11, p. 115201, 11 2022. [Online]. Available: https://doi.org/10.1121/10.0015045

[31] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: https://www.R-project.org/

[32] S. N. Wood, "mgcv: GAMs and generalized ridge regression for R," *R news*, vol. 1, no. 2, pp. 20–25, 2001.

[33] M. Wieling, "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English," *Journal of Phonetics*, vol. 70, pp. 86–116, 2018.

[34] A. Arvaniti, "Rhythm, timing and the timing of rhythm," *Phonetica*, vol. 66, no. 1-2, pp. 46–63, 2009.

[35] J. Fitzpatrick-Cole, "The alpine intonation of Bern Swiss German," in *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 1, 1999, pp. 941–944.