



languages

The Effects of Cross-Language Differences on Bilingual Production and/or Perception of Sentence-Level Intonation

Edited by

Ineke Mennen and Laura Colantoni

Printed Edition of the Special Issue Published in *Languages*

**The Effects of Cross-Language
Differences on Bilingual Production
and/or Perception of Sentence-Level
Intonation**

The Effects of Cross-Language Differences on Bilingual Production and/or Perception of Sentence-Level Intonation

Editors

Ineke Mennen

Laura Colantoni

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editors

Ineke Mennen
University of Graz
Austria

Laura Colantoni
University of Toronto
Canada

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Languages* (ISSN 2226-471X) (available at: https://www.mdpi.com/journal/languages/special_issues/Cross-Language_Bilingual_Production_Perception_Intonation).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , <i>Volume Number</i> , Page Range.
--

ISBN 978-3-0365-7454-7 (Hbk)

ISBN 978-3-0365-7455-4 (PDF)

© 2023 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

Laura Colantoni and Ineke Mennen

The Effects of Cross-Language Differences on Bilingual Production and/or Perception of Sentence-Level Intonation

Reprinted from: *Languages* 2023, 8, 108, doi:10.3390/languages8020108 1

Bistra Andreeva and Snezhina Dimitrova

L1 Influences on Bulgarian-Accented German: Prosodic Units and Prenuclear Pitch Accents

Reprinted from: *Languages* 2022, 7, 263, doi:10.3390/languages7040263 7

Laura Colantoni, Gabrielle Klassen, Matthew Patience, Malina Radu and Olga Tararova

Perception and Production of Sentence Types by Inuktitut-English Bilinguals

Reprinted from: *Languages* 2022, 7, 193, doi:10.3390/languages7030193 25

Silvia Dahmen, Martine Grice and Simon Roessig

Prosodic and Segmental Aspects of Pronunciation Training and Their Effects on L2

Reprinted from: *Languages* 2023, 8, 74, doi:10.3390/languages8010074 51

Cyrille Granget and Elisabeth Delais-Roussarie

Does Japanese/German L1 Metrical and Tonal Structure Constrain the Acquisition of French L2 Morphology?

Reprinted from: *Languages* 2022, 7, 305, doi:10.3390/languages7040305 79

Fatima Hamlaoui, Marzena Żygis, Jonas Engelmann and Sergio I. Quiroz

Prosodic Transfer in Contact Varieties: Vocative Calls in Metropolitan and Basaa-Cameroonian French

Reprinted from: *Languages* 2022, 7, 285, doi:10.3390/languages7040285 101

Sam Hellmuth

Sentence Prosody and Register Variation in Arabic

Reprinted from: *Languages* 2022, 7, 129, doi:10.3390/languages7020129 129

Niamh Kelly

Change across Time in L2 Intonation vs. Segments: A Longitudinal Study of the English of Ole Gunnar Solskjaer

Reprinted from: *Languages* 2022, 7, 210, doi:10.3390/languages7030210 147

Ji Young Kim

Spanish-English Cross-Linguistic Influence on Heritage Bilinguals' Production of Uptalk

Reprinted from: *Languages* 2023, 8, 22, doi:10.3390/languages8010022 165

Ineke Mennen, Ulrich Reubold, Kerstin Endes and Robert May

Plasticity of Native Intonation in the L1 of English Migrants to Austria

Reprinted from: *Languages* 2022, 7, 241, doi:10.3390/languages7030241 191

Andrea Pešková

Intonation Patterns Used in Non-Neutral Statements by Czech Learners of Italian and Spanish: A Cross-Linguistic Comparison

Reprinted from: *Languages* 2022, 7, 282, doi:10.3390/languages7040282 219

Editorial

The Effects of Cross-Language Differences on Bilingual Production and/or Perception of Sentence-Level Intonation

Laura Colantoni ^{1,*} and Ineke Mennen ^{2,*}

¹ Department of Spanish and Portuguese, University of Toronto, Toronto, ON M5S 1K7, Canada

² Department of English, University of Graz, 8010 Graz, Austria

* Correspondence: laura.colantoni@utoronto.ca (L.C.); ineke.mennen@uni-graz.at (I.M.)

1. Overview

We put together this Special Issue with the goal of collecting state-of-the-art articles on the intonational patterns of different types of bilinguals, with a particular focus on understudied language pairings, and we believe that we have succeeded. The ten articles included here encompass a wide variety of languages (e.g., Arabic, German, Bulgarian, Inuktitut, English, Spanish, French, Norwegian, Czech, Italian, and Japanese) and bilingual contexts (e.g., Heritage Speakers, L2 learners, societal bilingualism).

The contributions of this issue, however, are not merely empirical. Several papers (Andreeva and Dimitrova, Granget and Delais-Roussarie, Kim, Mennen et al., Pešková) set out to test whether models, such as Mennen's (2015) LILt model or the prosodic transfer hypothesis (Goad and White 2004), could predict which aspects of intonation are more prone to cross-linguistic prosodic influence or which L1 prosodic structures interact with L2 morphology. Independently of the theoretical model used, all of the papers have contributed to bringing us a step closer to finding answers to the specific questions we proposed, such as: *Can we determine a hierarchy of difficulty or transferability? How does prosody interact with other components of the grammar, such as morphology or syntax, in a contact situation? Which aspects are more prone to bidirectional interference? Which changes in intonation make speakers sound foreign in their second (or first) language?* We will briefly discuss here how these questions were addressed, but we will first summarize the structures analyzed in the volume, the types of bilingual contexts, the specific language pairings studies, and the methodologies employed.

1.1. Structures Analyzed in This Special Issue

The papers included in this volume cover a rich variety of sentence types and prosodic structures. Several papers study broad focus declaratives exclusively (e.g., Hellmuth, Kelly, Andreeva and Dimitrova, Kim), while others simultaneously analyze a range of sentence types, including non-neutral statements (Pešková) or broad-focus declaratives compared to different types of canonical and non-canonical interrogatives (Colantoni et al., Mennen et al.).

A subset of contributions focuses on the prosody of specific syntactic structures, such as vocative calls (Hamlaoui et al.), yes-no questions (Dahmen et al.), or the type of inflection chosen in the expression of subject-verb agreement by L2 speakers (Granget and Delais-Roussarie). The range of syntactic structures analyzed is mirrored by the variety of prosodic structures. The contributions included here analyze the type (e.g., Kelly, Pešková) and realization (Colantoni et al., Dahmen et al., Hellmuth, Mennen et al.) of pitch accents and nuclear contours and/or the tune associated with a given structure, such as the vocative chant (Hamlaoui et al.). They also investigated the relative difficulty of acquiring prosody vs. segments (Kelly, Dahmen et al.) and the role that L1 prosodic structure may play in the acquisition of L2 morphology (Granget and Delais-Roussarie). This is indeed an impressive

Citation: Colantoni, Laura, and Ineke Mennen. 2023. The Effects of Cross-Language Differences on Bilingual Production and/or Perception of Sentence-Level Intonation. *Languages* 8: 108. <https://doi.org/10.3390/languages8020108>

Received: 7 April 2023
Accepted: 7 April 2023
Published: 17 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

range of syntactic and prosodic structures, which, as we will discuss in Section 2, allows us to further our understanding of bilingual prosody.

1.2. Types of Bilingual Contexts

The present Special Issue illustrates bilingual situations that vary along different axes, such as the age of onset of acquisition of each language and the typological diversity between the languages or varieties in contact. Some of the studies focused on early bilinguals (Colantoni et al., Hamlaoui et al., Kim) who were exposed to both languages at home or in school in a context of societal bilingualism. Others analyzed cases of adult L2 learning (Granget and Delais-Roussarie; Pešková, Dahmen et al.), including instances of speakers who lived in a L2 context for a prolonged time (Kelly; Mennen et al.). As such, the whole spectrum of proficiencies is reflected in these studies, from the A2 to the C2 level, as described by the CEFR (Common European Framework of Reference).

As concerns typological distance, studies range from looking at typologically close languages (e.g., two intonational languages, such as L1 English–L2 German) or prosodically and phylogenetically unrelated languages (e.g., Basaá–French in Hamlaoui et al.; Inuktitut–English in Colantoni et al.). A special case within this group is the study of two varieties of Arabic (Hellmuth), such as a local variety (Yemeni Arabic) and Modern Standard Arabic, which are in a diglossic relation. Both the variety of bilingual contexts and the inclusion of prosodically diverse languages will allow researchers to start painting a picture of how duration and intensity of contact interact with prosodic typology in predicting patterns of cross-linguistic influence.

1.3. Language Pairings

One of the goals of the present issue was to broadcast research on understudied language pairings, and we were impressed with the range of contributions received. Even the study that focused on a relatively well-studied pairing (i.e., Kim; English–Spanish) looked at a population (heritage speakers) and a structure (uptalk) that have not been previously explored. This study nicely fits within a group of papers that explored the contact between two intonational languages, such as L1 English–L2 German (Mennen et al.), L1 Bulgarian–L2 German (Andreeva and Dimitrova), L1 Italian–L2 German (Dahmen et al.), L1 Czech–L2 Italian or Spanish (Pešková), and the interaction between local and standard varieties of Arabic (Hellmuth).

Another group of studies focused on speakers whose L1 is a tonal language and who are acquiring an intonational language. Within this group, Kelly analyzed the acquisition of English pitch accents by a L1 Norwegian speaker, comparing two moments in time. Hamlaoui et al. investigated the realization of vocative contours in L2 speakers of Cameroonian French whose L1 is Basaá, which is a language with H and L tones, plus toneless moras.

Granget and Delais-Roussarie analyzed the possibility of prosodic transfer in the realization of verbal morphology in L2 French by comparing speakers of a mora-timed language (i.e., L1 Japanese) with those of a stress-timed language (i.e., German). The final study (Colantoni et al.) looked at the contact between English and Inuktitut, a language that has been described as having no lexical stress and where intonation is used as a cue for phrasing.

1.4. Variety of Methods

We also want to point out the variety of methods exemplified in this Special Issue. Although there is not a balance between studies focusing on perception vs. production (i.e., only one study analyzes both; Colantoni et al.), the production studies encompass a variety of participants, tasks, and data analysis techniques. Studies range from an in-depth analysis of the prosody of one participant, both across multiple interviews (Hellmuth) or at different points in time (Kelly), to a study with 52 participants (Pešková). Most of the studies include 20–30 participants.

The papers included in this Special Issue vary in the selection of tasks. A couple of studies incorporate the analysis of semi-spontaneous tasks, such as narratives (Granget and Delais-Roussarie; Kelly), sociolinguistic interviews (Hellmuth), or dialogues (Kim). A frequent task used to elicit naturalistic data is the Discourse Completion Task (Colantoni et al.; Hamlaoui et al.; Pešková). Reading tasks (Mennen et al.) and elicited imitation (Colantoni et al.) were also discussed.

In addition to the phonological analysis of pitch accents and boundary tones, studies include detailed analyses of local or global f_0 excursions (e.g., Colantoni et al.; Dahmen et al., Hellmuth, Kelly, Mennen et al.), f_0 changes (e.g., Kim, Colantoni et al., Kelly), pitch dynamism quotient (Kelly), and alignment of the start and/or end of rises (Andreeva and Dimitrova; Mennen et al.). Duration (Andreeva and Dimitrova, Hamlaoui et al., Kim), intensity (Hamlaoui et al.), and final laryngealization (Hellmuth) are also analyzed.

2. Finding Answers to Our Questions

The first question we sought to answer was whether it would be possible to determine a *hierarchy of difficulty or transferability*, and the papers included here provide novel and complementary answers. Some studies have expanded the scope of our question by looking at the relative difficulty of acquiring intonation vs. other components of the grammar. Kelly, for example, investigates whether the L1 Norwegian speaker that she analyzes would show a greater degree of cross-linguistic influence in the frequency and realization of pitch accents or in the voicing of /z/, and her results show that the speaker is closer to the L2 target in the segmental than in the prosodic realization. Colantoni et al. explore whether context would have an impact in the selection of the appropriate syntactic and prosodic structure by looking at the English production of L1 Inuktitut speakers. They find that participants have difficulty in the syntax–pragmatic interface (i.e., producing the type of question that is appropriate to the context) and in incorporating tonal movement in the prenuclear region, as indicated by the absence of tonal movement. The prenuclear region is clearly an area of difficulty for these bilinguals, and this is also the case for perception. Whereas rising boundary tones are systematically interpreted as a cue for question, a falling boundary tone is not categorically associated with statements, probably due to the significant tonal movement in the English prenuclear region. The most systematic answer to this question, though, comes from the papers that have tested the adequacy of Mennen’s (2015) LILt model. These show that a systematic analysis of the four dimensions (semantic, frequency, systemic, and realizational) allows us to answer where transfer will or will not occur. Realizational differences are reported in all the studies that have applied this framework (Mennen et al., Andreeva and Dimitrova, Kim; Pešková). Interestingly, Mennen et al. show that realizational differences come in many flavors, and in their data set, for example, alignment, rather than pitch range, turned out to be significant. Pešková, in a finding that resembles the one reported by Colantoni et al., observes that realization interacts with position, since her participants were more target-like with boundary tones than with pre-nuclear pitch accents. Differences in the frequency dimension were observed in two papers (Mennen et al., Andreeva and Dimitrova). Two papers document deviances in the semantic dimension, either by showing that bilinguals differ from monolinguals of each language (Mennen et al.) or by arguing that bilinguals have a reduced inventory because they are not exposed to melodies that are used in informal contexts by monolinguals (Hamlaoui et al.). Finally, systemic differences were found in just one study (Mennen et al.), which interestingly showed that all bilinguals transferred a pitch accent, which is absent in English but is frequently used in Austrian German, to their L1.

As mentioned, the question of the hierarchy of difficulty is intertwined with our second question, which concerns the *interaction between prosody and other components of the grammar*. We have discussed how context interacts with prosody and syntax in the selection of question types in the speech of L1 Inuktitut–L2 English bilinguals. Bilinguals overextended the use of do-support or inversion to mark questions to contexts in which declarative questions are expected (i.e., questions that are identical to statements in word

order, but which differ in prosody). If we follow the argument in Hamlaoui et al., who also study the role of context in the selection of a specific vocative contour, we could argue that frequency also plays a role here. Indeed, Hamlaoui et al. show that the vocative chant that is frequently used in Metropolitan French is absent in Cameroonian French. They explain this absence by arguing that the vocative chant only appears in informal context, and Cameroonian French speakers are rarely exposed to contexts in which this chant is appropriate. In the same vein, we could argue that L1 Inuktitut speakers are less exposed to contexts that require the use of declarative questions.

One of the papers in this Special Issue specifically investigates the interaction between prosody and morphosyntax in the acquisition of the French verbal morphology (Granget and Delais-Roussarie). The authors tested the prosodic transfer hypothesis, which states that L1 prosodic structures interact with L2 morphology in the acquisition process (Goad and White 2004). To achieve this goal, they compared the marking of verb–subject agreement in narratives by speakers whose L1 is Japanese (mora-timed language) with L1 German speakers (stress-timed language). If L2 morphosyntax is guided by the L1 prosodic template, then the insertion of syllables would be more likely found in the speech of L1 Japanese speakers when compared to speakers whose L1 is German. This is indeed what they found, since L1 Japanese speakers tended to insert what they called ‘dummy auxiliaries’ into their L2, consisting of a vowel (e.g., /a, e/) or a whole syllable (e.g., /so/).

A third question concerned the prosodic aspects that are more prone to *bidirectional interference*. Mennen et al. explicitly addressed this question and showed that, not only the L2, but also the L1 prosody, is affected in the speech of adult L2 learners who have lived for many years in the context where the L2 is spoken. Indeed, they observed signs of L2-induced influences on bilinguals’ L1 speech in the systemic, frequency, and semantic dimensions and also in some aspects of the realizational dimension. Although not specifically set to test the hypothesis of bidirectionality, Hellmuth observes a high degree of mixing between the three registers of Arabic studied. In particular, all registers shared the frequency of density of phrasing boundaries. She also found that there are prosodic aspects that are under the control of the speaker, such as features of the low variety, which are suppressed when speaking the high variety. This, in turn, can be interpreted as a sign of bidirectional mixing. Finally, Kim raises the question of whether the uptalk patterns observed in the speech of Spanish–English bilinguals are the result of bidirectional interference. This seems to be the case for some of the speakers. In most cases, however, heritage Spanish speakers can keep the two types of rises in declaratives (short rise for Spanish vs. long rise for English) apart.

Our last question concerned the extent to which changes in intonation make speakers sound foreign. While none of the contributions to this Special Issue tackled this question directly, a few papers investigated which aspects in the speech of bilinguals differentiate them the most from monolingual speakers. Andreeva and Dimitrova, for instance, concluded that L1 influence was more obvious in the frequency dimension than in other dimensions, and this was particularly the case for those L1 Bulgarian–L2 German speakers who were exposed to the L2 later in life. Mennen et al. showed that the extent of L2 influence on L1 intonation was most extensive in the systemic dimension of L1 English–L2 Austrian German speakers immersed in a L2-speaking environment. Finally, two papers (Dahmen et al., Kelly) examined whether L2 experience or training improved the relative extent of bilinguals’ deviances in intonation or segments. Kelly, in a longitudinal study of a L1 Norwegian speaker who moved to England, found that segment realization improved more over time than the realization of pitch accents. Dahmen et al. showed, in Italians learning German, that segment-oriented training improved the learners’ production of segments, and prosody-oriented training improved their production of prosody. However, prosody training was found to also be beneficial for the production of segments, whereas no positive effects on prosody were found for segmental training.

3. Contributions and Future Studies

The papers offer comprehensive and new answers to the questions proposed, and open multiple avenues for new research, of which we have identified a few. First, a clear picture that emerges from this volume is the valuable contribution of the LILt model (Mennen 2015) to systematize and move forward the research on the role of cross-linguistic influence in prosody. The application of the model, which adopts from L2 speech acquisition models variables, such as positional variability, highlighted the need for more research on the interaction between position in the prosodic phrase and cross-linguistic influence on the realization of pitch events. As Pešková showed, L2 speakers were more successful at matching the target in the realization of boundary tones than of pitch accents. This was consistent with Colantoni et al.'s finding that very little tonal movement was found in the realization of pre-nuclear accents. As Pešková argues, the role of perceptual saliency needs to be highlighted to account for these findings. This takes us to our second point: in general, more evidence from perception is needed both to support this claim and to align the research on bilingual intonation with L2 speech acquisition models (e.g., Best and Tyler 2007; Flege 1995; Flege and Bohn 2021). A third aspect that requires more attention, and which is characteristic of bilingual studies more generally, is the wide range of individual variability. Kim especially addresses this issue by showing how participants vary in rise duration and in IP-final deaccenting. Colantoni et al. also highlight the variability observed in perceptual accuracy and in sentence type selection. Both studies come to identical conclusions: language experience (at least in the way in which it was quantified in each of these studies) cannot always account for the patterns observed. Finally, several of the papers included here remind us of the importance of looking at the interaction between prosody and other components of grammar. This is a point that has theoretical (see Feldhausen et al. 2021) and pedagogical implications, since research is starting to show how instruction and training focused on prosody can facilitate the learning of other aspects, such as segmental accuracy (as shown in Dahmen et al., but see also Li et al. 2022).

Research on intonation and bilingualism is still in its infancy (Trouvain and Braun 2020). We hope that this Special Issue, together with other Special Issues (Mennen and de Leeuw 2014; Rao forthcoming; Face and Armstrong forthcoming) and volumes (Delais-Roussarie et al. 2015), continue to inspire future studies.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins, pp. 13–34.
- Delais-Roussarie, Elizabeth, Mathieu Avanzi, and Sophie Herment, eds. 2015. *Prosody and Languages in Contact: L2 Acquisition, Attrition, Languages in Multilingual Situations*. Berlin and Heidelberg: Springer.
- Face, Timothy L., and Meghan Armstrong, eds. forthcoming. Prosody in Shared-Linguistic Spaces in the Spanish-Speaking World in *Languages*. Available online: https://www.mdpi.com/journal/languages/special_issues/XZDM6TGWC9 (accessed on 6 April 2023).
- Feldhausen, Ingo, Caroline Féry, and Frank Kügler, eds. 2021. Phonology-Syntax Interface and Recursivity in *Languages*. Available online: https://www.mdpi.com/journal/languages/special_issues/phonology_syntax (accessed on 6 April 2023).
- Flege, James. 1995. Second-language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issue in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 229–73.
- Flege, James Emil, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Ratree Wayland. Cambridge: Cambridge University Press, pp. 3–83.
- Goad, Heather, and Lydia White. 2004. Ultimate attainment of L2 inflection: Effects of L1 prosodic structure. *ESUROLA Yearbook* 4: 119–45. [CrossRef]
- Li, Peng, Florence Baills, Lorraine Baqué, and Pilar Prieto. 2022. The effectiveness of embodied prosodic training in L2 accentedness and vowel accuracy. *Second Language Research*. [CrossRef]

- Mennen, Ineke. 2015. Beyond segments: Towards a L2 intonation learning theory. In *Prosody and Languages in Contact*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin and Heidelberg: Springer, pp. 171–88.
- Mennen, Ineke, and Esther de Leeuw. 2014. Beyond segments: Prosody in SLA. *Studies in Second Language Acquisition* 36: 183–94. [[CrossRef](#)]
- Rao, Rajiv, ed. forthcoming. Prosody and Immigration in *Languages*. Available online: https://www.mdpi.com/journal/languages/special_issues/DTB64LM303 (accessed on 6 April 2023).
- Trouvain, Jürgen, and Bettina Braun. 2020. Sentence Prosody in a Second Language. In *The Oxford Handbook of Language Prosody*. Edited by Carlos Gussenhoven and Aaju Chen. Oxford: Oxford University Press, pp. 605–18.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

L1 Influences on Bulgarian-Accented German: Prosodic Units and Prenuclear Pitch Accents

Bistra Andreeva ^{1,*} and Snezhina Dimitrova ^{2,*}¹ Department of Language Science and Technology, Saarland University, 66123 Saarbrücken, Germany² Department of English and American Studies, Sofia University, 1504 Sofia, Bulgaria

* Correspondence: andreeva@lst.uni-saarland.de (B.A.); snezhina@uni-sofia.bg (S.D.)

Abstract: This study investigates the L1 influence on the use of accentual patterns, choice of prenuclear pitch accent types and their realization on L2 prosody. We use Mennen’s LILt model as a framework for our analysis. We recorded ten Bulgarian female speakers of German and ten female native German speakers who read Aesop’s fable *The North Wind and the Sun*. We found that the tendency for the Bulgarian native speakers to use more pitch accents than German native speakers is transferred to the L2 German of the Bulgarian learners. L*+H was the most frequent prenuclear pitch accent used by all groups. We also found that the Bulgarian learners stressed more function words and tolerated more stress clashes than the native German speakers. When speaking German, under the influence of the statistical regularities that relate to prosodic word patterns in their mother tongue, Bulgarian learners phrased their L2 speech into a higher number of shorter prosodic words, and therefore realized more pitch accents and aligned the high tonal target earlier than the native speakers. Concerning the variable alignment of the high target, we propose the prosodic word or the two-syllable window as the tentative candidate for an anchorage region. Our findings can be explained with respect to age of learning, as proposed by LILt’s general theoretical assumptions.

Keywords: Bulgarian; German; Bulgarian-accented German; intonation; prenuclear pitch accents; prosodic word; anchorage domain

Citation: Andreeva, Bistra, and Snezhina Dimitrova. 2022. L1 Influences on Bulgarian-Accented German: Prosodic Units and Prenuclear Pitch Accents. *Languages* 7: 263. <https://doi.org/10.3390/languages7040263>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 8 April 2022

Accepted: 13 September 2022

Published: 14 October 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The suprasegmental characteristics of L2 speech have for a long time been ignored by educators and researchers alike. The former have tended to focus on the segmental system (the vowels and the consonants) of the foreign language, on the assumption that mastering the individual sounds is crucial, if not sufficient, for efficient communication in the L2 (e.g., Eckert and Barry 2002; Baker 2006; among many others). The latter have for a long time ignored investigation into L2 prosody, not least because of the lack of sound and consistent methodology for the contrastive study of suprasegmental features in speech (see Ulbrich and Mennen 2015 for an overview). Some of the most popular L2 learning models, such as the Speech Learning Model (Flege 1995, 2007; Flege and Bohn 2021), the Native Language Magnet model (Kuhl 1991, 1992, 2000), and the Perceptual Assimilation Model (Best 1995; Best and Tyler 2007) focus almost exclusively on the segmental level.

An important step towards the development of a comprehensive model of L2 prosody acquisition is the L2 Intonation Learning Theory (LILt) put forward by Mennen (2007, 2015). The theory attempts to offer an extensive account of the major suprasegmental problems experienced by L2 learners, especially those in the area of intonation. Mennen draws an important distinction between phonological representation and phonetic implementation. She hypothesizes that L2 learners first acquire the phonological patterns in the foreign language, and only afterwards try to master the phonetic implementations of those patterns.

Mennen distinguishes four dimensions along which L2 intonation may deviate. The first of these—the systemic dimension—deals with the inventory of structural prosodic

elements and their distribution. The categorical elements can be pitch accents, accentual units of different size, or boundary phenomena. This dimension also involves the ways in which structural elements such as pitch accents combine with one another—for example, what combinations of High (H) and Low (L) pitch targets are admissible in a given language. In addition, it also looks at tune–text association (Ladd 2008), that is, the way the tune is mapped onto the segmental string. The second dimension of the LILt model—the realizational, or phonetic, dimension—is concerned with the phonetic implementation of the categorical elements of the system: this may involve the actual alignment of pitch accents, their scaling (i.e., their relative height), and their shape, or slope, e.g., shallow vs. steep rises or falls. The third dimension in Mennen’s LILt model is the semantic one: it deals with the ways in which the systemic elements are used to signal intonation functions. The fourth and final dimension of LILt—the frequency dimension—takes into account how often the structural elements are used. Each of LILt’s four dimensions makes it possible to establish cross-language similarities and differences in intonation, on the basis of which predictions can be made about intonation deviations from the target norm which are likely to occur in L2 learners’ speech. Based on existing L2 research, LILt also puts forward some general theoretical assumptions concerning L2 intonation acquisition in relation to learners’ proficiency levels, ages of arrival and learning, L1 background, speaking style, etc. LILt can thus help predict “the relative difficulty learners would experience with certain L2 intonational parameters or dimensions, and to shed light on the principles, which govern the acquisition process of intonation such as the rate and order in which parameters of intonation develop in a L2” (Mennen 2015, p. 178).

Therefore, the LILt model provides a sound basis for the present study whose aim is to identify cross-linguistic differences and similarities in the use of accentual patterns, choice of prenuclear pitch accent types and their realization in the read speech of Bulgarian and German native speakers and Bulgarian learners of German, as well as to predict where deviations from the native norm are likely to occur.

1.1. Previous Comparative Research on Bulgarian and German Intonation

Until the end of the 20th century, comparative research on Bulgarian and German was carried out with the aim of establishing the main differences between the segmental systems of the two languages, predicting the problematic areas for Bulgarian learners of German and designing practice materials for use in the language classroom. Amongst the few suprasegmental features which were investigated was question intonation, which turned out to be a particularly problematic area. Comparative work on question intonation in the two languages focused on three topics. First, the goal has been, by means of perception testing and acoustic analysis, to identify and systematize the features and parameters which are similar or different in the intonation of questions in Bulgarian and German, and thus to find the specific structural properties of the ‘Yes–No’ questions (Simeonova 1997) and the ‘Information’ questions (Simeonova 1986; Grigorova 1996). Second, the functional load of questions in Bulgarian and German is pursued from the perceptual standpoint, specifically for cases in which the communicative goal of the question can be determined by intonation alone (Grigorova 1994, 1997; Misheva and Grigorova 1997). Third, based on the research carried out, a series of textbooks for Bulgarians (Simeonova 1972) and for Bulgarian advanced students of German (Simeonova 1985; Simeonova 2000; Simeonova et al. 2000) was published.

More recently, Bulgarian-accented German has been studied by Andreeva (2017) and Andreeva and Dimitrova (2022a, 2022b). Andreeva (2017) analyzed the prosodic marking of information structure by highly proficient Bulgarian speakers of German and compared the non-native patterns with production patterns in native German and native Bulgarian, drawing particular attention (a) to their similarities and differences in the semantic and realizational dimension and (b) to the specific contribution of global and local cues signalling the information structure. Additionally, she aimed to test whether Flege’s SLM can be applied to aspects of the prosodic domain that are influenced by

information structure. With respect to the semantic dimension, Andreeva found that in both L1 and L2 German, given material always correlates with de-accentuation in post-nuclear position. In prenuclear position, Bulgarian speakers produce considerably more prenuclear accents than German speakers. Bulgarian speakers use more H* accents and German speakers use more L+H* to mark narrow focus in their L1. With respect to the realizational dimension, Andreeva found that Bulgarian speakers do not align the peak in narrow focus in a consistent manner in the target language. Since vowel length is not contrastive in Bulgarian segmental phonology, and since phonological duration has an impact on peak alignment, durational uncertainty may lead to peak alignment uncertainty. Bulgarian speakers of German transfer the differences/similarities in global tempo and spectral cues from their native into their target language. They also exploit F0-related novel cues to differentiate between focus conditions. They use later peak alignment, greater peak excursion, a greater amount of F0 change in the nuclear-accented syllable and suppress F0 in the prenuclear interval to establish a (greater) difference between contrastive and non-contrastive focus. A novel cue related to duration is established to mark the difference between narrow and broad focus in the target language: Bulgarian L2 speakers of German produce the accented vowel with longer duration in narrow focus (which was not the case in the L1 German data).

Andreeva and Dimitrova (2022b) investigated some prosodic characteristics of Bulgarian-accented L2 German compared to L1 German and L1 Bulgarian. All F0-related long-term distributional measures (mean and median pitch level, pitch span and pitch variation) in the speech of the Bulgarian learners of German were lower than in their L1, but higher than those of the native German speakers. These results corroborate the findings in Andreeva et al. (2015, 2014), who report the use of a wider pitch range and higher variability in two Slavic languages (Bulgarian and Polish) compared to two Germanic languages (German and English). With regard to the duration-related parameters, Andreeva and Dimitrova (2022b) found that the Bulgarian speakers used a slower articulation rate, more IPs and pauses and more pitch accents in their L2 than the native speakers. The strong correlation which was found between the L1 and L2 speaking rates of the Bulgarian speakers is evidence that the L1 speaking rate can indeed predict the speaking rate in L2, which is in accordance with the findings in Bradlow et al. (2017).

Building on investigations by Mennen et al. (2012), Andreeva and Dimitrova (2022a) also measured specific tonal targets in the F0 contour which are linguistic in nature, but which long-term distributional measures fail to capture. These linguistic measures are tonal landmarks (local maxima and minima) associated with prominent or non-prominent syllables and initial and non-initial peaks in intonation phrases. They found that the Bulgarian speakers of German realized the majority of the linguistically relevant targets in a way which was very similar to the respective realizations of these targets in their mother tongue.

In sum, the results in Andreeva and Dimitrova (2022a, 2022b) suggest that L2 speech is influenced by L1 prosody with respect to both F0-related and duration-related features.

1.2. Prosodic Word Patterns in Bulgarian

Descriptions of Bulgarian prosody have included units formed on the basis of the presence of stress. They have been called ‘phonetic words’ (Stojkov 1966; Misheva 1991) or ‘accentual-rhythmic units’ (Tilkov 1981). Stojkov’s ‘phonetic word’ is characterized by the presence of one stressed syllable which can either be preceded or followed by a number of unstressed syllables. Tilkov’s ‘accentual-rhythmic unit’ is also characterized by the presence of a single stress and, like Stojkov’s phonetic word, can include both proclitic and enclitic syllables. In both Tilkov’s (1981) and Misheva’s (1991) analyses of corpora containing about 35,000 units, those comprising three syllables appear to be the most frequent, followed by four- and two-syllable units. Stress tends towards the middle of the units and is viewed as ‘organizing’ the unstressed syllables in them. In autosegmental-metrical terms, both the

accentual-rhythmic unit and the phonetic word correspond to the prosodic word, which consists of a content word and its clitics.

Bulgarian has an unbounded weight-insensitive lexical stress system. Lexical stress is not fixed for any lemma, but its position changes when various affixes are added. According to Bulgarian statistical surveys (Misheva 1991; Kotova and Yanakiev 2001), the most often lexically stressed syllable is the penultimate one. So, the penultimate position can possibly be considered as the default or regular position of stress in the language. Andreeva et al. (2019) show that stress on the penultimate syllable is also predominant in the prosodic word in Bulgarian spontaneous speech (see Figure 1).

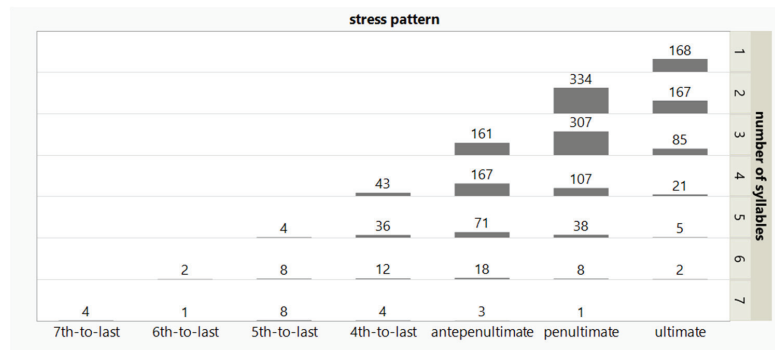


Figure 1. Stress patterns per number of syllables within the prosodic word (adapted from Andreeva et al. 2019).

As can be seen in Figure 1, the most frequent pattern in prosodic words consisting of two or three syllables is stress on the penultimate syllable. In prosodic words of more than three syllables, stress on the antepenultimate syllable predominates. There are some occurrences of prosodic words consisting of seven syllables which do not follow this pattern, but their number is negligible. In German, a language with a bounded weight-sensitive lexical stress system, the stress can be assigned to one of the last three full syllables of a prosodic word (see Domahs et al. 2008 and references therein).

The role of the prosodic word within the prosodic hierarchy in Bulgarian has remained outside the scope of previous research. Therefore, one of the aims of the present investigation is to shed light on its role in tonal alignment.

1.3. Aspects of Bulgarian and German Intonation: A Comparison within the LILt Model

The present study considers the L1 influence on the prosody of the Bulgarian-accented German of speakers at medium proficiency level within Mennen’s (2015) LILt model. In the systemic dimension, Bulgarian and German have been described as having the same inventory of structural phonological elements: two prosodic constituents in the prosodic hierarchy—the intermediate phrase and the intonational phrase—six pitch accents (L*, H*, L*+H, L+H*, H+L*, H+!H*), two phrase accents (L– and H–), and one initial and two final boundary tones (%H, L%, H%) (Grice et al. 2005; Andreeva 2007). Regarding the frequency dimension, L*+H is the most frequent pitch accent in prenuclear position in both languages (Dimitrova and Andreeva 2017; Baumann et al. 2021). Other prenuclear accent types which are used less frequently in the two languages are H* and L+H*. Baumann et al. (2021) report limited use of L* in prenuclear position in German as well. With respect to the semantic dimension, it has been demonstrated that in the production (Braun 2006) and the processing (Braun and Biezma 2019) of prenuclear L*+H and L+H*, German informants prefer the former in contexts that trigger a contrastive topic interpretation. Using the concept of informativeness, which they define as relating to “both the information status of a referring expression and its role as part of a specific focus domain”, Baumann et al.

(2021) conclude that informativeness does not affect the choice of prenuclear accent type, although they find a stronger (but non-significant) tendency for contrastive topics to be produced with L*+H rather than L+H*. No comparable data on the pragmatic-semantic interpretation of prenuclear accents in Bulgarian are available.

Concerning the realizational and/or systemic dimension, three phenomena are of interest: the realization of the prenuclear L*+H, the distribution of accents and the tolerance towards stress clashes. Dimitrova and Jun (2015) report on the variable alignment of the high trailing tone in the prenuclear L*+H in Bulgarian, which in their data was sometimes realized as far to the right as the second posttonic syllable. They suggest that the H tone may be a phrasal accent. Dimitrova and Andreeva (2017) argue that the H target is not separated by a fixed interval from the starred tone, as postulated by Pierrehumbert's invariance hypothesis (Pierrehumbert 1980). In German, the H target of the rising prenuclear accent aligns with the vowel in the post-accented syllable (Atterer and Ladd 2004; Mücke et al. 2009).

As regards accentuation, Andreeva (2017) found that Bulgarian speakers produce considerably more prenuclear accents than German native speakers, tending to accent nearly every content word.

With respect to stress clashes, they occur when two syllables bearing primary stress are adjacent in the same phonological domain, for example, in a phonological phrase (e.g., [θɜ:ti:n 'men]_{PH}). Dimitrova (1998) found that, in Bulgarian, adjective + noun phrase clashes were tolerated in more than 56% of her test items, in which there was a choice between two alternative (standard) stress patterns, one of which allowed avoidance of the clash. In German, Wagner and Fischenbeck (2002) report that in compounds stress clash resolution through stress shift is relatively rare and an alternative strategy is destressing of the secondary accent. However, Karen et al. (2011) found both perceptual and production experimental evidence that stress shift operates on a regular basis within and beyond word boundaries in order to prevent stress clashes and hence rhythmically irregular structures. Rieger and Piontek (2015) found cases in a German radio news corpus where pitch accents are shifted from the noun to the adjective in order to prevent a focus-internal accent clash.

Taking into consideration the similarities and differences between Bulgarian and German prosody discussed so far, we use Mennen's (2015) LILt model and its assumptions regarding age of learning and language exposure to predict deviations. The research questions we set out to answer were the following:

Regarding the frequency dimension

- Do Bulgarian L2 speakers of German produce more pitch accents than German L1 speakers?
- Is L*+H the most frequent pitch accent in prenuclear position in both L1 and L2 German?

Regarding the realizational dimension

- Do Bulgarian L2 speakers of German tolerate more stress clashes than German L1 speakers?

Regarding the systemic dimension

- Does the prosodic word constitute an anchorage domain both in Bulgarian L1 and German L2 for the trailing tone of the L*+H pitch accent?

2. Materials and Methods

To answer our research questions, we recorded ten Bulgarian speakers of German at B2 level of proficiency, according to the Common European Framework of Reference for Languages, and ten German native speakers as controls. All speakers were female university students of comparable age (average 20.7 years) and spoke the respective standard language varieties. The Bulgarian participants were all foreign learners who started learning the language at the age of 13 at a German-language-medium school in Bulgaria. They had received between 5 and 7 years of German tuition and had some knowledge of the

phonetics and phonology of German. Since prosodic deviation in L2 can be due to learners' different levels of proficiency, different ages at which they started learning the language, different amount of experience with the L2, different speaking styles, etc., as suggested by LILt, we chose a homogeneous group of speakers with respect to these variables.

The material recorded was Aesop's fable *The North Wind and the Sun*, with the Bulgarians reading the text in Bulgarian, as well as in German (see Appendices A and B). We obtained three data sets: (a) ten recordings of the fable by speakers of Bulgarian as L1 (BG_L1), (b) ten recordings of the same Bulgarian speakers reading the fable in German as L2 (DE_L2) and (c) ten recordings of the fable by speakers of German as L1 (DE_L1).

2.1. Measurements

First, syllable, prosodic word and phrase boundaries, as well as pauses, were segmented, and lexically stressed syllables were labeled manually in Praat. Second, all accented syllables were marked and counted, including those in lexical words with double prominence and in prominent function words.

2.1.1. Pitch Analyses

We labeled linguistically relevant tonal landmarks, slightly modifying the method proposed by Patterson (2000) and Mennen et al. (2012). We only labeled tonal landmarks aligned with stressed (L*, H*) and unstressed syllables (L, H), as well as final lows (FL) and highs (FH). Then, we marked the relevant prenuclear pitch accents based on the inventories proposed for Bulgarian (Andreeva 2007; Andreeva et al. 2016; Dimitrova and Jun 2015) and German (Grice et al. 2005). The pitch accents, accented syllables and intonation phrases were labeled by careful auditory inspection carried out by the two authors working together. Occasional disagreements were resolved after discussion and repeated listening. An example of the labeling is provided in Figure 2. For the DE_L2 data set, we do not claim that our ToBI labeling represents underlying (phonological) categories. The annotation used for this data set rather represents a systematization of the tonal landmarks (as defined above) according to the ToBI labeling conventions proposed in the literature (e.g., Silverman et al. 1992). The F0 values corresponding to the L* and H targets of the manually labeled L*+H were obtained in semitones relative to 1 Hz using Praat scripts. We also calculated the span between the low and the high target of these pitch accents.

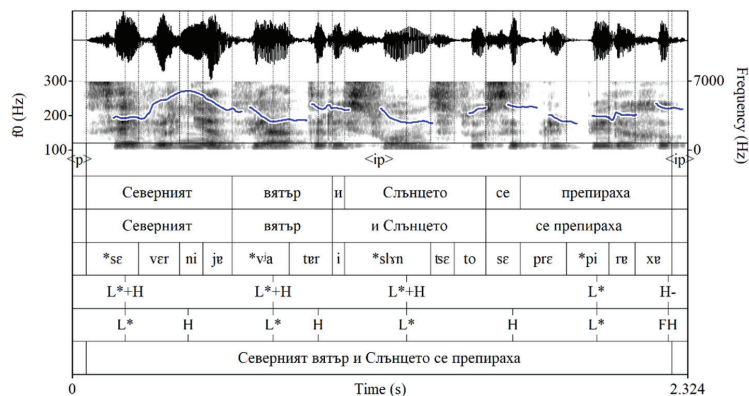


Figure 2. The utterance *Северният вятър и Слънцето се препираха* ('The North Wind and the Sun were disputing'), pronounced as a single intermediate phrase by a Bulgarian speaker. Labeling of the data: tier 1—intonation and intermediate phrase boundaries; tier 2—word boundaries; tier 3—prosodic word boundaries; tier 4—syllable boundaries (accented syllables marked with *); tier 5—ToBI labeling; tier 6—tonal landmarks; tier 7—Bulgarian text.

2.1.2. Temporal Features

The durations of the IPs, pauses and pitch-accented syllables were extracted per reading, speaker and native/target language using Praat scripts. In addition, we calculated the articulation rate (AR) for those IPs in which the L*+H pitch accent occurs. AR was computed as the number of canonical syllables divided by the duration of the respective IP. As a measure of peak alignment, the absolute temporal distance from the F0 peak to the accented syllable onset was calculated. In order to compensate for the influence of segmental durations on peak alignment, these absolute measures were converted to relative measures, taken as a proportion of syllable durations. We also counted the distance between the L and the H target and between the H target and the end of the prosodic word in terms of number of intervening syllables.

3. Results

3.1. Stress and Accentuation

In order to answer our research question regarding the frequency and realizational dimension, namely, whether the Bulgarian speakers of German produce more pitch accents and tolerate more stress clashes than the German native speakers, we first analyzed the stress and accentuation in the three data sets. Table 1 summarizes the total number of words, content (CW) and function (FW) words and syllables in the texts, as well as the mean number of all accented syllables and of the accented syllables in the content and function words.

Table 1. Words and syllables in the three data sets: exploratory statistics.

	BG_L1	DE_L1	DE_L2
No. words	91	108	108
No. content words	56	52	52
No. function words	35	56	56
No. syllables	200	180	180
Mean No. of accented syllables	63.5	50.8	74.5
Mean No. of accented syllables of CWs	55.2	46.5	58.6
Mean No. of accented syllables of FWs	8.3	4.3	15.9

The Bulgarian text of the fable *The North Wind and the Sun* consists of 91 words, of which 56 are content words and 35 are function words, giving a total of 200 syllables. Most content words in Bulgarian are single-stressed. However, three adverbs and two adjectives (one of them twice) occurred in their comparative or superlative forms, which in Bulgarian can be pronounced either with single or with double stress (Tilkov and Boyadzhiev 1977, p. 160). Lexically stressed syllables have the potential to receive a pitch accent (Lehiste 1970). The average number of accented syllables in the readings of the Bulgarian speakers was 63.5 (55.2 on content words, including double stress on some occurrences of the above-mentioned adverbs and adjectives, and 8.3 on function words, such as *те*—‘they’, *този*—‘this one’, *беше*—‘was’).

The German text of the fable consists of 108 words, of which 52 are content words and 56 are function words¹. The total number of syllables is 180. The average number of accented syllables realized by the German native speakers is 50.8 (46.5 on content words and 4.3 on function words, such as *wer*—‘who’, *seinen*—‘his’, *sollte*—‘should’). The Bulgarian learners of German on average realized 74.5 accents (58.6 on content words and 15.9 on function words). This tendency for overproduction of pitch accents in L2 speech has been observed in several L2 varieties for learners at different proficiency levels (e.g., Archibald 1997; Rasier and Hiligsmann 2007; Avesani et al. 2015). We also observed that the L2 speakers of German realized two accents on some compounds. For example, “North Wind” is a compound in German (*Nordwind*) and an adjective + noun phrase in Bulgarian (*северният вятър*). The native German speakers in our data always realized the word with a single pitch accent on the first element of the compound (see Figure 3a), while the L2

German speakers used two pitch accents on the two parts of the compound (see Figure 3b) in 21 out of the 40 realizations (52.5%). They showed the same tendency to use two pitch accents when pronouncing the word *Augenblicken*—‘moments’. Bulgarian linguists in the field of word formation share the opinion that compounding is an atypical word-formation process for the Bulgarian language (and for Slavic languages in general) and point out its poor productivity (Radeva 2007, p. 57). It is more common in Germanic languages. This provides one possible explanation for the use of two pitch accents in the L2 by the Bulgarians.

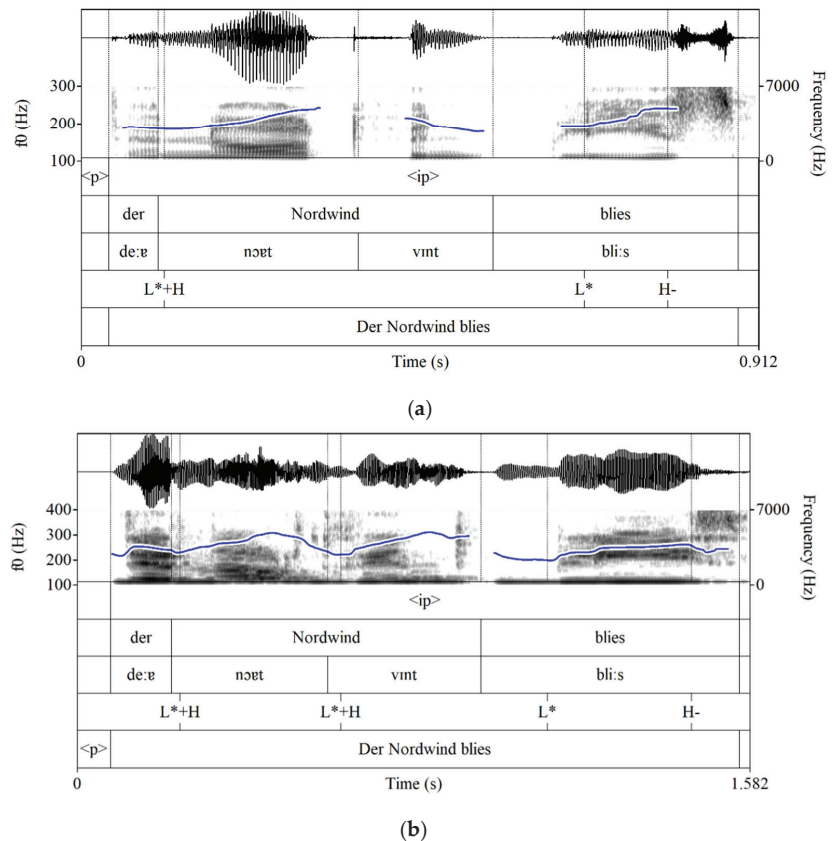


Figure 3. The phrase *der Nordwind blies* (‘the North Wind blew’) pronounced by (a) a native German speaker without stress clash and (b) a Bulgarian speaker of German with stress clash.

Another explanation comes from the greater tolerance in Bulgarian to stress clashes. Dimitrova (1998) found that in Bulgarian sentences with potential stress clash, the clash was tolerated in 56.3% of the cases. We found a similar amount of stress clashes—53% (59 realizations out of 110 potential ones in all readings) in our BG_L1 data. In DE_L2, realized stress clashes amounted to 53% of the 100 potential cases. In addition, we found 38 more clashes due to accent on a function word. In the DE_L1 data set, on the other hand, stress clashes constituted only 15% (15 realizations) of the 100 potential cases. This confirms previous findings and provides evidence that the greater tolerance of stress clashes in Bulgarian is transferred to the L2.

Last but not least, the average number of pitch accents on function words in the DE_L2 data set is considerably higher (15.9) than in DE_L1 (4.3) or in BG_L1 (8.3). This confirms the tendency to overuse pitch accents in the L2, commented on above.

We next analyze the types, frequencies and realizations of the prenuclear pitch accents in the three data sets.

3.2. Prenuclear Pitch Accent Types

In order to find out whether L*+H is the most frequent pitch accent in prenuclear position in both L1 and L2 German, we analyzed the pitch accent types used by the speakers in prenuclear position. We found five different pitch accent types, namely, L*+H, (!)H*, L*, L+H* and H+!H*. As can be seen in Figure 4, the choice of the pitch accent types and their relative frequency is comparable for the three data sets. However, as can be seen from Table 2, the Bulgarian speakers realized about 1.4 times more prenuclear accents in their Bulgarian readings (45.1 pitch accents on average per reading) and 1.5 times more accents in their German readings (47.6 pitch accents on average) than the German speakers (32.7 pitch accents on average). Thus, the tendency for Bulgarian speakers to use more prenuclear pitch accents than German speakers is carried over to their L2 as well.

Table 2. Number of prenuclear pitch accents in the three data sets.

	Data Set		
	BG_L1	DE_L1	DE_L2
No. prenuclear pitch accents	451	327	476

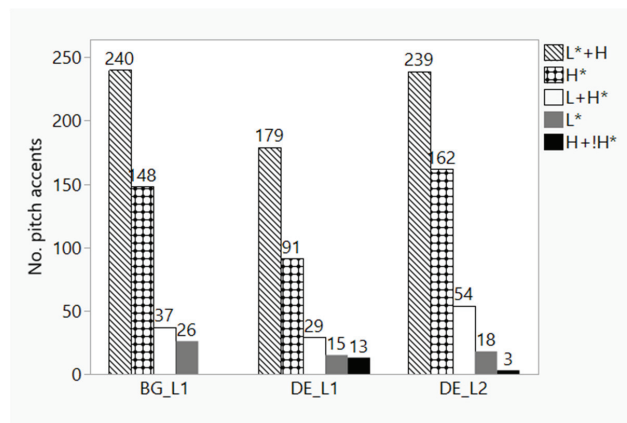


Figure 4. Number of pitch accents in the three data sets.

In all three data sets, the predominant prenuclear pitch accent is L*+H: 240 occurrences (53%) for BG_L1, 179 (55%) for DE_L1 and 239 (50%) for DE_L2. The predominant use of L*+H in the DE_L1 data set is in line with the findings of [Baumann et al. \(2021\)](#), who report that sentence topics are consistently marked by rising prenuclear accents and not even given items are deaccented. Our findings also confirm Truckenbrodt’s claim that L*+H is the neutral prenuclear accent type in German ([Truckenbrodt 2002](#)). However, the frequency of use of the different prenuclear accent types in our data differs from that in [Baumann et al. \(2021\)](#).

In [Baumann et al.’s \(2021\)](#) study, the second most frequently used prenuclear pitch accent in (L1) German was L+H* in 17.8% of all cases. In our data, it is H*: 91 occurrences (28%) for DE_L1 and 162 (34%) for DE_L2. In the Bulgarian L1 data set, H* was used 148 times (33%).

In our data, L+H* was the third most frequently used prenuclear pitch accent: it occurred 29 times (9%) in DE_L1 and 54 times (11%) in DE_L2. In BG_L1, it was used 37 times (8.2%). However, it must be noted that this pitch accent type was used predominantly by only two of the Bulgarian speakers both in their L1 and their L2 readings, which constitutes evidence of L1 transfer.

The differences in the choice of prenuclear pitch accent types in our data and in the data reported by Baumann et al. (2021) could be due to the different types of text used in the two experiments: while Baumann and colleagues analyzed separate read sentences, some of which were intended to elicit contrastivity, our speakers read a continuous text. The only prenuclear pitch accent type which is not found in all three data sets is H+!H*. It occurs only in our two German data sets: 13 times (4%) in DE_L1 and three times (1%) in DE_L2.

3.3. Realization of the Prenuclear L*+H

Concerning our research question about the prosodic word as an anchorage domain for the trailing tone of the L*+H pitch accent, we next focus on a comparison of the realizations of L*+H in the three data sets.

3.3.1. Alignment of the High Target with Respect to the Prosodic Word

As mentioned in Section 1.2, the role of the prosodic word has been neglected in research on Bulgarian intonation. On the other hand, the concept of the prosodic word (Tilkov’s ‘accentual rhythmic unit’, Misheva’s ‘phonetic word’) is to be found in virtually all descriptions of accentuation above the word level by Bulgarian scholars. In this study, one of our aims is to investigate if the variability of the high tonal target alignment described in Section 1.3 can be explained with reference to the prosodic word. It is also our purpose to explore the potential role of the prosodic word in the realizations of L*+H in L2 German. Following Welby and Lævenbruck (2005, 2006), we postulate an anchorage domain for the trailing high tone of the L*+H, where it aligns with an unstressed syllable in the region up to the right boundary of a prosodic word.

Table 3 summarizes the realizations of the H target within and outside of the prosodic word in the three data sets. The Bulgarian speakers align the H target within the prosodic word, both in their L1 and in their L2 readings, more often than the L1 German speakers. However, the differences between the three data sets are relatively small.

Table 3. Realizations of the H target relative to the end of the prosodic word in the three data sets.

Position of the H Target	Data Set		
	BG_L1	DE_L1	DE_L2
Within the prosodic word	206 (93.6%)	118 (88.1%)	216 (99.0%)
Outside the prosodic word	14 (6.4%)	16 (11.9%)	2 (1.0%)

To explore the reasons for these differences (even though they are small), we analyzed the structure of the underlying prosodic words in the Bulgarian and German material used in the present experiment (Figure 5). In Bulgarian, two-, three- and four-syllable-long prosodic words predominate, which is in line with earlier findings (Tilkov 1981; Misheva 1991). There are very few instances of prosodic words containing five and six syllables, and no prosodic words of only one syllable. In German, on the other hand, the number of prosodic words containing one² to five syllables is almost equal, whereas prosodic words containing six syllables are fewer.

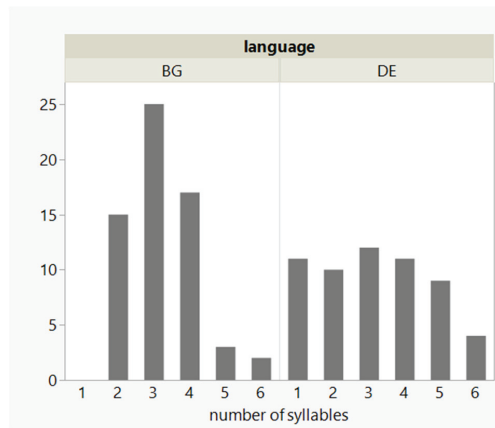


Figure 5. Number of syllables in the prosodic word in Bulgarian (left) and German (right).

Comparing the position of stress in the prosodic words in our material, it turns out that, in Bulgarian, penultimate and antepenultimate stress predominates, which is also in line with previous findings, whereas, in German, the distribution of the different stress patterns is more varied (Figure 6). Thus, on the systemic level, there are differences between the Bulgarian and the German material in terms of prosodic word structure.

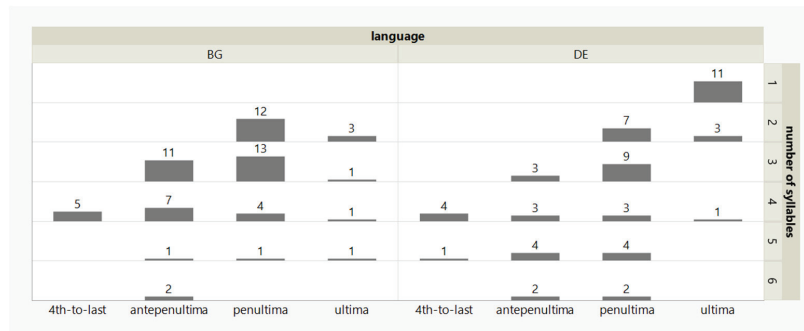


Figure 6. Stress position in the prosodic word in Bulgarian (left) and German (right).

In terms of realization, a closer look at the alignment of H with reference to the length of the prosodic word and the position of stress reveals that most cases of alignment of the high target outside the prosodic word can be explained by taking into account stress position. In BG_L1, DE_L1 and DE_L2, the default alignment of the H target is within the prosodic word. If stress is on the penultimate or antepenultimate syllable, the H target of the pitch accent can be aligned with the first syllable of the next prosodic word, if it is unstressed. In both languages, this can be a function word which can cliticize with a following or a preceding host (see [Tilkov 1981](#) for Bulgarian) and, in Bulgarian, may be accompanied by vowel coalescence as well (e.g., *слънцето е*—‘the sun is’—/ slɤntsɛto ε/ > [slɤntsɛto̯ɛ]). Although we found differences in the structure of the prosodic word in our Bulgarian and German material, the smallest number of cases of H alignment outside the prosodic word was observed in DE_L2.

The above results indicate that the prosodic word does not provide an optimal explanation for the variability of the H target alignment of the prenuclear rising pitch accent and is therefore an unlikely anchorage region.

3.3.2. Height and Alignment of the Low and High Target

We next investigate the height and alignment of the low and high target of the L*+H prenuclear pitch accent with respect to the accented syllable.

With respect to pitch height, for both targets we observed the highest values for BG_L1, the lowest values for DE_L1 and intermediate values for DE_L2 (see Figure 7). In other words, the speakers from the three data sets used different register, namely, higher in BG_L1, lower in DE_L1 and intermediate in DE_L2.

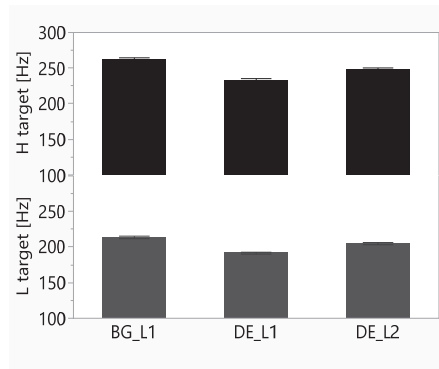


Figure 7. Mean pitch height (in Hz) of the L target (lower panel) and H target (upper panel) for the three data sets. Error bars represent standard errors.

With respect to the alignment of the low target, we found that the low target is aligned later in BG_L1 (about 57% from the beginning of the accented syllable) than in the German data sets (about 48% from the beginning of the accented syllable in both DE_L1 and DE_L2). It has been reported that German speakers align the L target within the consonant or even in the vowel of the accented syllable (Atterer and Ladd 2004; Mücke et al. 2009), which is in accordance with our findings. In our data, about half of the accented syllables in Bulgarian (56%) had a voiceless onset, whereas, in the German data, they were 11.8% in DE_L1 and 16.3% in DE_L2. Thus, the difference in the phonological make-up of the accented syllables can explain the later alignment of the low tonal target in Bulgarian found in our data (see Figure 8 lower panel).

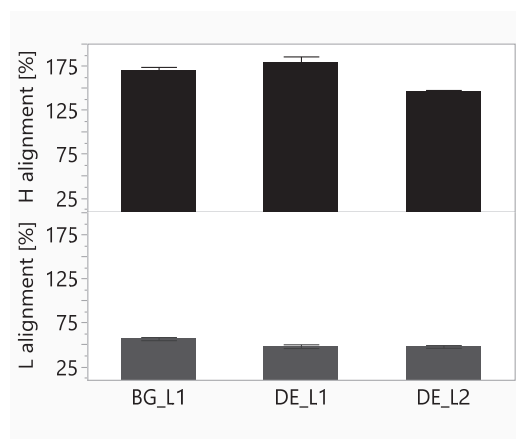


Figure 8. Mean alignment values (in %) of the L target (lower panel) and H target (upper panel) relative to the syllable onset for the three data sets. Error bars represent standard errors.

For the alignment of the H target we also found that the peak is aligned earlier in DE_L2 (145%, i.e., slightly before the middle of the post-accented syllable) than in the L1 data sets (170% vs. 179% for BG_L1 and DE_L1, respectively, i.e., within the second half of the post-accented syllable). Given that it has been reported previously that speech tempo influences the alignment of tonal targets (e.g., [Silverman and Pierrehumbert 1990](#)), we checked articulation rates in our data. We measured the articulation rate in all IPs in which prenuclear L*+H occurred. It turned out that the DE_L2 speakers whose articulation rate was the slowest also aligned the peak earlier with respect to the syllable onset, which contradicts previous findings that slower speech tempo results in later alignment (see [Figure 8](#) upper panel).

Within the prosodic word, we also observed variation in the alignment of the high tonal target outside the post-accented syllable in the three datasets. Therefore, we measured the distance between the two tonal targets in terms of number of intervening syllables. This is shown in [Table 4](#), where 0 indicates that the two targets are in adjacent syllables, 1 indicates that there is one syllable in between and 2 shows that there are two syllables which separate L from H. There are very few instances of more than two syllables separating the two targets; therefore, we report those together.

The Bulgarian speakers realized the H target in the syllable immediately following the accented syllable in 89.9% of cases in their German readings, unlike the native German speakers, who used such realizations in 75.9% of the L*+H prenuclear accents. In BG_L1, we observed the smallest proportion of realizations of L*+H with no intervening syllables between the two targets (72.5% of cases). From these data, we can conclude that, in more than 90% of cases (95.9% in BG_L1, 91.1% in DE_L1 and 99.5% in DE_L2), the H target is aligned within a window of two post-accented syllables. This window predicts the position of the high tonal target slightly better than the prosodic word.

Table 4. Number of intervening syllables between L* and H (in % in parenthesis) in the three data sets.

No. Intervening Syllables	Data Set		
	BG_L1	DE_L1	DE_L2
0	161 (72.5%)	110 (75.9%)	196 (89.9%)
1	52 (23.4%)	22 (15.2%)	21 (9.6%)
2	9 (4.1%)	10 (6.9%)	1 (0.5%)
3 or 4	0	3 (2.0%)	0

4. Discussion

In this study, we examined the choice of prenuclear pitch accent types, their distribution and the realization of the default prenuclear L*+H pitch accent in the read speech of Bulgarian and German native speakers and Bulgarian learners of German, drawing particular attention to their similarities and differences in the systemic, realizational and frequency dimension ([Mennen 2015](#)). The main question we asked is whether and to what extent the native language affects the non-native prosody of Bulgarian speakers of German at a medium level of proficiency.

Regarding the frequency dimension, namely, whether Bulgarian L2 speakers of German produce more pitch accents than German L1 speakers, we found that the Bulgarian learners of German used 1.5 times more pitch accents than the German native speakers, which confirms our expectations. The explanation for this is twofold. On the one hand, [Andreeva and Dimitrova \(2022a, 2022b\)](#) found that the Bulgarian learners produced more intonation phrases than the German native speakers (27.2 vs. 18.9) and, as a result, more nuclear accents. On the other hand, both speaker groups realized pitch accents on function words in prenuclear position as well. Optional accentuation on function words in German has been reported by [Bögel \(2021\)](#), [Kügler \(2018\)](#) and [Zerbian and Böttcher \(2019\)](#), among others. Despite the fact that there are fewer function words in the Bulgarian text which we used for data collection, the Bulgarian speakers used twice as many accents on them in

their native language compared to the accents put on function words by the native German speakers. Again, transferring the tendency to often accent function words to their German L2, the Bulgarian learners realized the pitch accent on a function word 3.7 times more often than the native German speakers.

Regarding the frequency dimension, from the prenuclear pitch accent types found in the three data sets, namely L*+H, H*, L+H*, L* and H+!H*, L*+H was the most frequently used, which is in line with Truckenbrodt (2002) and Baumann et al. (2021). What is more, compared to the German native speakers, the Bulgarians realized about 1.4 times more prenuclear accents in their Bulgarian readings and 1.5 times more prenuclear accents in their German readings. Thus, the tendency for the Bulgarian speakers to use more prenuclear pitch accents in Bulgarian than German speakers in German is transferred to the L2 German of the Bulgarian learners as well.

Our research question regarding the realizational dimension on whether Bulgarian L2 speakers of German tolerate more stress clashes than German L1 speakers was answered positively. We found that more than half of the potential cases of stress clash were tolerated in both BG_L1 and DE_L2, while in DE_L1 such cases constituted only 15% of the underlying stress clashes. In DE_L2, we found additional clashes due to additional accent placement on a function word. Stress clash tolerance provides further evidence of L1 transfer in the readings of Bulgarian learners of German.

Regarding the systemic dimension, our analysis shows that Bulgarian speakers use an anchorage domain both in Bulgarian L1 and German L2 for the trailing tone of the L*+H pitch accent. However, we did not find conclusive evidence about the exact region of the anchorage. Our analyses revealed that the H target occurs more often within the prosodic word in the BG_L1 and DE_L2 data set than in the DE_L1 data set, and that the H target is aligned earlier in DE_L2 compared to BG_L1 and DE_L1. However, we also found counterexamples in which the H spreads to the first or second syllable of the next prosodic word in the three data sets, although their amount for the DE_L2 data was negligible (only two cases). The finding that the Bulgarian speakers of German align the trailing tone significantly earlier than the native speakers, in spite of the fact that they have a significantly slower articulation rate compared to the native speakers, is surprising and contradicts previous findings that a slower speech tempo results in later alignment. We explain these findings in terms of L1 transfer: when speaking German, under the influence of the statistical regularities that relate to prosodic word patterns in their mother tongue, Bulgarian learners of German phrase their L2 speech into a higher number of shorter prosodic words, and therefore realize more pitch accents and align the high tonal target earlier than the native speakers.

Since we found additional variation of the trailing tone alignment within the prosodic word which cannot be explained in terms of prosodic word structure (stress position and number of syllables between the stressed syllable and the end of the prosodic word), we focused on the number of syllables between the two tonal targets of the prenuclear pitch accent. It turns out that the H is aligned within a window of two post-accented syllables in 95.9% of the cases in BG_L1, 91.1% in DE_L1 and 99.5% in DE_L2. This window predicts the position of the high tonal target slightly better than the prosodic word. However, it must be borne in mind that our results are based on read speech data, which does not control for many possible factors that can cause variability. Several studies have suggested that the specification for the alignment of tonal targets is a function of speech tempo, phonological vowel length, syllabic structure and segmental effects (intrinsic vowel duration, vowel quality, consonant voicing, etc.), adjacency to word and intonational boundaries, proximity to other tones, as well as dialectal background (Arvaniti et al. 1998; Jilka and Möbius 2007; Ladd et al. 2000; Möbius and Jilka 2007; Mücke et al. 2009; Prieto and Torreira 2007; Silverman and Pierrehumbert 1990, among others).

Turning back to the anchorage domain, it should be noted that the two-syllable window coincides with the default pattern for the prosodic word structure in Bulgarian, in which stress is on the penultimate syllable in two- and three-syllable prosodic words

and on the antepenultimate syllable in four-syllable prosodic words. Thus, the question of whether the prosodic word or the two-syllable window provides a better explanation for the variability of the alignment of the high trailing tone remains open. Evidence from more specifically designed experiments is needed to confirm or reject our predictions.

In conclusion, our results suggest that the L2 speech of Bulgarian learners of German at intermediate level is influenced by L1 intonation features in cases when there are differences between the mother tongue and the target language. In our case, this may be due to the age of learning. The LILt assumes that the age of first (regular) exposure to an L2 is an important factor in predicting overall success in acquiring L2 intonation (Mennen 2015, p. 180). The ten Bulgarian speakers of German all started learning the language relatively late at the age of 13 at a German-language-medium school in Bulgaria and were at B2 level of proficiency according to the Common European Framework of Reference for Languages when they took part in the experiment. Moreover, they reported relatively limited exposure to the L2 and almost no immersion in a German-speaking environment. Our results provide further evidence that ‘the earlier the better’ also applies to intonation learning, as suggested by LILt. Thus, in addition to the four dimensions put forward by LILt, its general assumptions provide a very useful basis for any investigation of L2 intonation.

Author Contributions: Conceptualization, B.A. and S.D.; methodology, B.A. and S.D.; formal analysis, B.A. and S.D.; investigation, B.A. and S.D.; resources, B.A. and S.D.; data curation, B.A. and S.D.; writing—original draft preparation, B.A. and S.D.; writing—review and editing, B.A. and S.D.; visualization, B.A. and S.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Bulgarian National Science Fund, project No. Кп-06-H40/11 from 12.12.2019 ‘Prosodic aspects of Bulgarian in comparison with other languages with lexical stress’.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Due to confidentiality restrictions, the data analyzed in this study are not publicly available. They are available on request from B.A.

Acknowledgments: We thank Christoph Gabriel and Ulrike Dohmas for advice on prosodic unit boundaries in German, and Kirstin Kolmorgen, Anna Spasiano, and Hanna Zimmermann for technical support.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Северният вятър и Слънцето

Северният вятър и Слънцето се препираха кой е по-силен, когато един пътник, завит в топла дреха, мина покрай тях. Те решиха, че този, който пръв накара пътника да си свали дрехата, ще се счита по-силен от другия. Тогава Северният вятър започна да духа с всичка сила, но колкото по-силно вятърът духаше, толкова по-плътнo пътникът увиваше дрехата около себе си. Най-после Северният вятър прекъсна усилията си. Тогава Слънцето започна да грее силно и пътникът веднага свали дрехата си. И така, Северният вятър беше принуден да признае, че Слънцето е по-силно от него.

Appendix B

Nordwind und Sonne

Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam. Sie wurden einig, dass derjenige für den Stärkeren gelten sollte, der den Wanderer zwingen würde, seinen Mantel abzulegen. Der Nordwind blies mit aller Macht, aber je mehr er blies, desto fester hüllte sich der Wanderer in seinen Mantel ein. Endlich gab der Nordwind

den Kampf auf. Nun erwärmte die Sonne die Luft mit ihren freundlichen Strahlen, und schon nach wenigen Augenblicken zog der Wanderer seinen Mantel aus. Da musste der Nordwind zugeben, dass die Sonne von ihnen beiden der Stärkere war.

Notes

- ¹ The smaller number of function words in the Bulgarian text in comparison with the German one can be attributed, among other things, to the post-positioned article in the language.
- ² Note, however, that in our German data there were no L*+H prenuclear pitch accents realized on monosyllabic prosodic words.

References

- Andreeva, Bistra. 2007. Zur Phonetik und Phonologie der Intonation der Sofioter-Varietät des Bulgarischen. Ph.D. thesis, Institute of Phonetics, University of the Saarland, Saarbrücken, Germany. PHONUS 12.
- Andreeva, Bistra. 2017. Prosodic Encoding of Phrasal Prominence and Information Structure in a Second Language: When Bulgarian and German Prosody meet. *Contrastive Linguistics XLIII* 1: 59–83.
- Andreeva, Bistra, and Snezhina Dimitrova. 2022a. Prosodic characteristics of Bulgarian-accented German. In *Elektronische Sprachsignalverarbeitung 2022, Tagungsband der 33. Konferenz (Sønderborg)*. Dresden: TUD Press, pp. 220–27.
- Andreeva, Bistra, and Snezhina Dimitrova. 2022b. The influence of L1 prosody on Bulgarian-accented German and English. *Proceedings of Speech Prosody 2022*: 764–68. [CrossRef]
- Andreeva, Bistra, Bernd Möbius, Grażyna Demenko, Frank Zimmerer, and Jeanin Jügler. 2015. Linguistic Measures of Pitch Range in Slavic and Germanic Languages. Paper presented at the Interspeech 2015: Speech Beyond Speech: Towards a Better Understanding of the Most Important Biosignal, Dresden, Germany, September 6–10; pp. 968–72. [CrossRef]
- Andreeva, Bistra, Grażyna Demenko, Bernd Möbius, Frank Zimmerer, Jeanin Jügler, and Magdalena Jastrzebska. 2014. Differences of Pitch Profiles in Germanic and Slavic Languages. *Proceedings of Interspeech 2014*: 1307–11. [CrossRef]
- Andreeva, Bistra, Snezhina Dimitrova, Christoph Gabriel, and Jonas Grünke. 2019. The intonation of Bulgarian Judeo-Spanish spontaneous speech. In *Proceedings of the 19th International Congress of Phonetic Sciences*. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Australia: Australasian Speech Science and Technology Association Inc., pp. 3827–21.
- Andreeva, Bistra, William J. Barry, and Jacques Koreman. 2016. Local and Global Cues in the Prosodic Realization of Broad and Narrow Focus in Bulgarian. *Phonetica* 73: 256–78. [CrossRef] [PubMed]
- Archibald, John. 1997. The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics* 35: 167–82. [CrossRef]
- Arvaniti, Amalia, Dwight Robert Ladd, and Ineke Mennen. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* 26: 3–25. [CrossRef]
- Atterer, Michaela, and Dwight Robert Ladd. 2004. On the phonetics and phonology of “segmental anchoring” of F0: Evidence from German. *Journal of Phonetics* 32: 177–97. [CrossRef]
- Avesani, Cinzia, Giuliano Bocci, Mario Vayra, and Alessandra Zappoli. 2015. Prosody and information status in Italian and German L2 intonation. In *Il parlato in [italiano] L2: Aspetti Pragmatici e Prosodici/[Italian] L2 Spoken Discourse: Pragmatic and Prosodic Aspects*. Edited by Marina Chini. Milano: Franco Angeli (Materiali Linguistici, 71), pp. 93–116.
- Baker, Ann. 2006. *Ship or Sheep? Student's Book. An Intermediate Pronunciation Course*, 3rd ed. Cambridge: Cambridge University Press.
- Baumann, Stefan, Jane Mertens, and Janina Kalbertodt. 2021. The influence of informativeness on the prosody of sentence topics. *Glossa: A Journal of General Linguistics* 6: 95. [CrossRef]
- Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 171–204.
- Best, Catherine, and Michael Tyler. 2007. Non-native and second-language speech perception. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: Benjamins, pp. 13–34.
- Bögel, Tina. 2021. Function Words at the Interface: A Two-Tier Approach. *Languages* 6: 197. [CrossRef]
- Karen, Bohn, Richard Wiese, and Ulrike Dohmas. 2011. The status of the rhythm rule within and across word boundaries. Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China, August 17–21; pp. 332–35.
- Bradlow, Ann, Midam Kim, and Michael Blasingame. 2017. Language independent talker-specificity in first-language and second language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *The Journal of the Acoustical Society of America* 141: 886–99. [CrossRef]
- Braun, Bettina. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49: 451–93. [CrossRef]
- Braun, Bettina, and Maria Biezma. 2019. Prenuclear L*+H activates alternatives for the accented word. *Frontiers in Psychology* 10: 1993. [CrossRef]
- Dimitrova, Snezhina. 1998. Stress shift in English, Bulgarian English and Bulgarian. Paper presented at the International Conference 70 Years of English and American Studies in Bulgaria, Sofia, Bulgaria, October 1–3; pp. 30–34.

- Dimitrova, Snezhina, and Bistra Andreeva. 2017. Prenuklearni akcenti v Balgarskia ezik. In *Novi Napravlenia v Sinhronmata i Diahromata Fonetika i Fonologia*. Edited by Madlen Danova, Snezhina Dimitrova and Mira Kovacheva. St. Kliment Ohridski: Sofia University, pp. 152–66.
- Dimitrova, Snezhina, and Sun-Ah Jun. 2015. Pitch accent variability in focus production and perception in Bulgarian declaratives. In *Proceeding of the 18th ICPHS*. Glasgow: The University of Glasgow, Paper 0832.1-5. Available online: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPHS2015/Papers/ICPHS0832.pdf> (accessed on 18 January 2021).
- Domahs, Ulrike, Richard Wiese, Ina Bornkessel-Schlesewsky, and Matthias Schlesewsky. 2008. The Processing of German Word Stress: Evidence for the Prosodic Hierarchy. *Phonology* 25: 1–36. [CrossRef]
- Eckert, Hartwig, and William Barry. 2002. *The Phonetics and Phonology of English and Pronunciation: A Coursebook*. Trier: WVT.
- Flege, James Emil. 1995. Second language speech learning: Theory, findings and problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*. Edited by Winfried Strange. Timonium: York Press, pp. 233–72.
- Flege, James Emil. 2007. Language contact in bilingualism: Phonetic system interactions. In *Laboratory Phonology*. Edited by Jennifer Cole and José Ignacio Hualde. Berlin: Mouton de Gruyter, pp. 353–80.
- Flege, James Emil, and Ocke-Schwen Bohn. 2021. The revised speech learning model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Rtree Wayland. Cambridge: Cambridge University Press, pp. 3–83.
- Grice, Martine, Stefan Baumann, and Ralf Benz Müller. 2005. German Intonation in Autosegmental-Metrical Phonology. In *Prosodic Typology*. Edited by Sun-Ah Jun. Oxford: Oxford University Press, chp. 2. pp. 430–58.
- Grigorova, Evelina. 1994. Perceptivno izsledvane na komunikativnata funkcia na intonaciata pri frazi s leksikalnio-gramatichna struktura na chasten vapros v nemski i balgarski ezik. *Contrastive Linguistics* XIX: 5–14.
- Grigorova, Evelina. 1996. “Atomistische” Untersuchung der Intonation von W-Fragen mit verschiedener kommunikativer Funktion im Deutschen und im Bulgarischen. *Contrastive Linguistics* XXI: 5–14.
- Grigorova, Evelina. 1997. Kontrastiv-perzeptive Untersuchung der funktionalen Belastung der Intonation zur Unterscheidung verschiedener Kommunikativer Vorhaben am Beispiel Deutsch/Bulgarisch. In *Kontrastive Intonationsstudien Deutsch/Bulgarisch. Forum Phonetikum* 62. Edited by Ruska Simeonova and Hans-Walter Wodarz. Frankfurt/Main: Hector, pp. 31–60.
- Jilka, Matthias, and Bernd Möbius. 2007. The influence of vowel quality features on peak alignment. Paper presented at the INTERSPEECH 2007, 8th Annual Conference of the International Speech Communication Association, Antwerp, Belgium, August 27–31; pp. 2621–24.
- Kotova, Natalia, and Miroslav Yanakiev. 2001. *Grammatika bolgarskogo jazyka dlja vladejuštih ruskim jazykom*. Moscow: MGU.
- Kügler, Frank. 2018. Optional accentuation of pronouns in German. Paper presented at the 9th International Conference on Speech Prosody 2018, Poznań, Poland, June 13–16; pp. 671–75.
- Kuhl, Patricia K. 1991. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93–107. [CrossRef] [PubMed]
- Kuhl, Patricia K. 1992. Speech prototypes: Studies on the nature, function, ontogeny and phylogeny of the “centers” of speech categories. In *Speech, Perception, Production and Linguistic Structure*. Edited by Yoh’ichi Tohkura, Eric Vatikiotis-Bateson and Yoshinori Sagisaka. Tokyo: IOS Press, pp. 240–64.
- Kuhl, Patricia K. 2000. A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America* 97: 11850–57. [CrossRef] [PubMed]
- Ladd, Dwight Robert. 2008. *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press.
- Ladd, Dwight Robert, Ineke Mennen, and Astrid Schepman. 2000. Phonological conditioning of peak alignment of rising pitch accents in Dutch. *Journal of Acoustic Society of America* 107: 2685–96. [CrossRef]
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge: MIT Press.
- Mennen, Ineke. 2007. Phonological and phonetic influences in non-native intonation. In *Non-Native Prosody. Phonetic Description and Teaching Practice (Trends in Linguistics. Studies and Monographs (TiLSM) 186*. Edited by Jürgen Trouvain and Ulrike Gut. Berlin: Mouton de Gruyter, pp. 53–76.
- Mennen, Ineke. 2015. Beyond segments: Towards an L2 intonation learning theory (LILt). In *Prosody and Languages in Contact: L2 Acquisition, Attrition, Languages in Multilingual Situation*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin: Springer, pp. 171–88.
- Mennen, Ineke, Felix Schaeffler, and Gerard Docherty. 2012. Cross-language differences in fundamental frequency range: A comparison of English and German. *Journal of the Acoustical Society of America* 131: 2249–60. [CrossRef]
- Misheva, Anastassia. 1991. *Intonacionna sistema na balgarskia ezik. [Intonation system of Bulgarian]*. Sofia: Balgarska Akademija na Naukite.
- Misheva, Anastassia, and Evelina Grigorova. 1997. Kontrastive Untersuchung des Einflusses der Muttersprache auf die Perzeption von suprasegmentalen Charakteristika der Rede am Beispiel Deutsch/Bulgarisch. In *Kontrastive Intonationsstudien Deutsch/Bulgarisch, Forum Phonetikum* 62. Edited by Ruska Simeonova and Hans-Walter Wodarz. Frankfurt/Main: Hector, pp. 61–159.
- Möbius, Bernd, and Matthias Jilka. 2007. Effects of syllable structure and nuclear pitch accents on peak alignment: A corpus-based analysis. Paper presented at the 16th international Congress of Phonetic Sciences, Saarbrücken, Germany, August 6–10; pp. 1173–76.

- Mücke, Doris, Martine Grice, Johannes Becker, and Anne Hermes. 2009. Sources of variation in tonal alignment: Evidence from acoustic and kinematic data. *Journal of Phonetics* 37: 321–38. [CrossRef]
- Patterson, David. 2000. A Linguistic Approach to Pitch Range Modelling. Ph.D. dissertation, University of Edinburgh, Edinburgh, UK.
- Pierrehumbert, Janet. 1980. *The Phonetics and Phonology of English Intonation*. Cambridge: MIT Press.
- Prieto, Pilar, and Francisco Torreira. 2007. The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35: 473–500. [CrossRef]
- Radeva, Vasilka. 2007. *V sveta na doumite*. (In *the World of Words*). Sofia: UI “Sv. Kliment Ohridski”.
- Rasier, Laurent, and Philippe Hiligsmann. 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Francaise* 28: 41–66.
- Riester, Arndt, and Jörn Piontek. 2015. Anarchy in the NP: when new nouns get deaccented and given nouns don't. *Lingua* 165: 230–53. [CrossRef]
- Silverman, Kim, and Janet Pierrehumbert. 1990. The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Edited by John Kingston and Mary Beckman. Cambridge: Cambridge University Press, pp. 72–106.
- Silverman, Kim, Mary Beckman, John Pitrelli, Mori Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. TOBI: A standard for labeling English prosody. Paper presented at the Second International Conference on Spoken Language Processing, ICSLP 1992, Banff, AL, Canada, October 13–16; pp. 867–70.
- Simeonova, Ruska. 1972. *Übungsbuch zur Deutschen Aussprache. Audiolingualer Kurs für Bulgaren*. Sofia: Nauka i izkustvo.
- Simeonova, Ruska. 1985. *Gesprochenes und Geschriebenes Deutsch. Korrekativer Kurs für Bulgarische Germanistikstudenten*. Sofia: Universitätsverlag.
- Simeonova, Ruska. 1986. Zur Intonation der Ergänzungsfrage im Deutschen und im Bulgarischen. *Kwartalnik Neofilologiczny* 33: 307–23.
- Simeonova, Ruska. 1997. Spezifische Besonderheiten der Entscheidungsfrage im Deutschen und im Bulgarischen. In *Kontrastive Intonationsstudien Deutsch/Bulgarisch*. *Phorum Phonetikum* 62. Edited by Ruska Simeonova and Hans-Walter Wodarz. Frankfurt/Main: Hector, pp. 1–30.
- Simeonova, Ruska. 2000. *Grundzüge einer Kontrastiven Phonetik und Phonologie Deutsch/Bulgarisch*, 2nd ed. Sofia: Svjat.
- Simeonova, Ruska, Haritina Kostova-Dobrevva, and Evelina Grigorova. 2000. *Gesprochenes Deutsch. Korrekativer Kurs für Fortgeschrittene*. Sofia: Svjat, Nauka.
- Stojkov, Stojko. 1966. *Uvod vav Fonetikata na Balgarskia ezik*. [Introduction to Bulgarian Phonetics]. Sofia: Nauka i izkustvo.
- Tilkov, Dimitar. 1981. *Intonacijata v Balgarskia ezik*. [Bulgarian Intonation]. Sofia: Narodna prosveta.
- Tilkov, Dimitar, and Todor Boyadzhiev. 1977. *Balgarska Fonetika*. [Bulgarian phonetics]. Sofia: Nauka i izkustvo.
- Truckenbrodt, Hubert. 2002. Upstep and embedded register levels. *Phonology* 19: 77–120. [CrossRef]
- Ulbrich, Christiane, and Ineke Mennen. 2015. When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism* 20: 1–28. [CrossRef]
- Wagner, Petra, and Eva Fischenbeck. 2002. Stress perception and production in German stress clash environments. Paper presented at the Speech Prosody 2002, Aix-en-Provence, France, April 11–13; pp. 687–90.
- Welby, Pauline, and Hélène Løevenbruck. 2006. Anchored down in Anchorage: Syllable structure and segmental anchoring in French. *Italian Journal of Linguistics/Rivista di Linguistica, Pacini Editore S.p.A* 18: 74–124.
- Welby, Pauline, and Hélène Løevenbruck. 2005. Segmental “anchorage” and the French late rise. Paper presented at the Interspeech 2005: The 9th Annual Conference on Speech Communication and Technology, Lisboa, Portugal, September 4–8; pp. 2369–72.
- Zerbian, Sabine, and Marlene Böttcher. 2019. Stressed pronouns in mono- and bilingual German. Paper presented at the 19th International Congress of Phonetic Sciences, Melbourne, Australia, August 5–9; pp. 2640–45.

Article

Perception and Production of Sentence Types by Inuktitut-English Bilinguals

Laura Colantoni ^{1,*}, Gabrielle Klassen ¹, Matthew Patience ¹, Malina Radu ¹ and Olga Tararova ²

¹ Department of Spanish and Portuguese, University of Toronto, Toronto, ON M5S 3K1, Canada; gabrielle.klassen@mail.utoronto.ca (G.K.); matt.patience@utoronto.ca (M.P.); malina.radu@utoronto.ca (M.R.)

² Department of Languages and Cultures, University of Western Ontario, London, ON N6A 3K7, Canada; otararov@uwo.ca

* Correspondence: laura.colantoni@utoronto.ca

Abstract: We explore the perception and production of English statements, absolute yes-no questions, and declarative questions by Inuktitut-English sequential bilinguals. Inuktitut does not mark stress, and intonation is used as a cue for phrasing, while statements and questions are morphologically marked by a suffix added to the verbal root. Conversely, English absolute questions are both prosodically and syntactically marked, whereas the difference between statements and declarative questions is prosodic. To determine the degree of crosslinguistic influence (CLI) and whether CLI is more prevalent in tasks that require access to contextual information, bilinguals and controls performed three perception and two production tasks, with varying degrees of context. Results showed that bilinguals did not differ from controls in their perception of low-pass filtered utterances but diverged in contextualized tasks. In production, bilinguals, as opposed to controls, displayed a reduced use of pitch in the first pitch accent. In a discourse-completion task, they also diverged from controls in the number of non-target-like realizations, particularly in declarative question contexts. These findings demonstrate patterns of prosodic and morphosyntactic CLI and highlight the importance of incorporating contextual information in prosodic studies. Moreover, we show that the absence of tonal variations can be transferred in a stable language contact situation. Finally, the results indicate that comprehension may be hindered for this group of bilinguals when sentence type is not redundantly marked.

Keywords: intonation; prosody; L2 speech; bilingualism; L2 acquisition; phonetics; production; perception; English; Inuktitut

Citation: Colantoni, Laura, Gabrielle Klassen, Matthew Patience, Malina Radu, and Olga Tararova. 2022. Perception and Production of Sentence Types by Inuktitut-English Bilinguals. *Languages* 7: 193. <https://doi.org/10.3390/languages7030193>

Academic Editors: Juana M. Liceran and Raquel Fernández Fuertes

Received: 24 February 2022

Accepted: 6 July 2022

Published: 25 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Post-lexical or intonational uses of pitch can be transferred from one language to another in a contact situation (Queen 2001, 2012; Colantoni and Gurlekian 2004), which raises the question of whether the absence of pitch movement is also susceptible to crosslinguistic influence (CLI). To tackle this question, it is crucial to find a contact situation in which one of the languages has a restricted use of pitch. Thus, in this paper we direct our attention to Inuktitut, an Eskimo-Aleut language spoken in Eastern Canada,¹ which has been in contact with English since the 16th century (Dorais 2010), and which has been described as having no stress (Fortescue 1983; Shokeir 2009; Arnhold et al. Forthcoming) and a very limited use of intonation (Massenet 1980; Fortescue 1983; Shokeir 2009). We focus on the perception, interpretation, and production of the three sentence types listed in examples (1)–(3).

- | | |
|--------------------------------|------------------------|
| (1) Statements (S) | Peter bought a piano. |
| (2) Absolute questions (AQ) | Did Peter buy a piano? |
| (3) Declarative questions (DQ) | Peter bought a piano? |

We have chosen to begin our descriptive enterprise with sentence types since there is a high degree of consensus that if prosody is used linguistically, we should expect it to be used to mark sentence type (Gussenhoven 2004), independently of the acoustic correlates that a particular language may use. How is prosody, in particular intonation, used to mark sentence type in English in the examples above? First, it is the only cue to distinguish (1) from (3). Whereas statements typically end with a falling contour, declarative questions end with a raising contour (e.g., Pierrehumbert 1980; Bartels 1999). Rising contours are also characteristic of absolute questions, at least in Canadian English, the variety studied here (Hedberg and Sosa 2002; Hedberg et al. 2017; Patience et al. 2018). Second, statements and questions not only differ in their realization of nuclear contours, but there is increasing evidence suggesting that, in Canadian English, questions (both AQs and DQs) are marked by a different initial pitch accent when compared to statements (L+H* vs. H*, respectively) and by a higher pitch peak (Saindon et al. 2017b; Patience et al. 2018). In this sense, English would resemble a wide variety of languages that mark interrogativity with a higher initial pitch when compared to statements (Face 2007; Petrone and Niebuhr 2014; Sicoli et al. 2015). Third, although prosody is crucial in differentiating Ss and DQs, it is redundant in signaling the distinction between (2) and (1) or (3). Indeed, AQs are syntactically marked by inversion (e.g., *Are you coming to the party?*) or do-support as in (2). Finally, although AQs and DQs are prosodically similar, they differ syntactically and pragmatically. Crucially, as opposed to AQs, DQs cannot be used in out-of-the-blue contexts (Gunlogson 2002). DQs express surprise or incredulity (Truckenbrodt 2011) and have a mirative interpretation (Peterson 2016).

As opposed to English, Inuktitut is an agglutinative and polysynthetic language (Johns 2010; Fortescue 2017) that marks sentence type with verb suffixes. The verb has ten moods (Dorais 2010), among which we find the Declarative, Indicative, Interrogative, and Imperative-Optative. The Declarative has an evidential reading; namely, it signals that what the speaker says occurred (Dorais 2010, p. 78)—as in (4). The Indicative, instead, expresses a general situation, as in (5). The Interrogative mood marks that the speaker is asking a question, as in (6). Although descriptions of other types of questions (with the exclusion of Wh-questions) are rather limited, previous literature indicates that confirmation questions use the same mood as absolute questions but may be signaled by additional lengthening (Fortescue 1983; Massenet 1980), a rising contour (Fortescue 1983) or a particle (Fortescue 1983). The Imperative-Optative mood, which is used to request an action from the hearer (Smith 1977), is illustrated in (7). Imperative sentences can have an exclamatory force. Very little is known about the prosodic realization of such utterances, but Fortescue (1983) indicates that “commands and exclamations tend to have the highest pitch on the final segment” (Fortescue 1983, p. 115). Finally, Inuktitut has a wide variety of particles, many of which can be used to mark evidentials. For example, the particle *-mmarik-*, which is usually placed word-medially, can be added to verbs and adjectives to indicate credibility (Spalding 1979, p. 97), as in (8). Thus, Inuktitut uses a rich morphology to mark sentence type, rather than syntax, as English does. Indeed, although the default word order is SOV, word order is variable and may vary depending on the information structure (e.g., Fortescue 1984).

(4) taku-vutit (Dorais 2010, p. 283)

See INTR-2 SG DECLARATIVE
You see (something)

(5) taku-jutit (Dorais 2010, p. 284)

See INTR-2 SG INDICATIVE
You see (something)

(6) taku-viit (Dorais 2010, p. 284)

See INTR-2 SG INTERROGATIVE
Do you see (something)?

(7) taku-git (Smith 1977, p. 15)
See INTR-2 SG IMPERATIVE
Look!

(8) qai-mmarik-tuq (Spalding 1979, p. 97)
Come-really-3 SG INDICATIVE
He really comes

Instrumental and experimental descriptions of Inuktitut prosody are not abundant, but the existing ones clearly suggest that Inuktitut is a language that does not mark lexical stress (Fortescue 1983; Shokeir 2009; Arnhold et al. Forthcoming), and that tonal variations are restricted to the end of the utterance (Massenet 1980; Fortescue 1983; Shokeir 2009; see also Thalbitzer 1904, p. 141). Massenet (1980) analyzes the variety spoken in Resolute Bay and concludes that declaratives have a rise, associated with the penultimate syllable, followed by a fall. Absolute yes-no questions are signaled by a rise associated with the antepenultimate syllable, which is followed by a fall in the penultimate and a rise in the final syllable (HLH contour). Questions are also marked by vowel lengthening. Fortescue's (1983) overview of twelve Eskimo varieties shows that dialects differ in their rhythmic patterns (syllable vs. mora time), in the syllable to which the tonal movement is associated, and in whether interrogatives end with a fall or with a rise. Of the varieties surveyed in his study, the two closest to the variety analyzed here are characterized by a fall in declaratives, and either a sustained pitch or a sharp rise in interrogatives, which are also signaled by vowel lengthening (i.e., a lengthening of the final vowel, as illustrated in (6)). Shokeir's (2009) autosegmental metrical analysis of multiple narratives produced by Inuktitut speakers confirms, to a large extent, the conclusions of previous work. First, she showed that tonal movements are restricted to the last two syllables in the utterance. Second, rising contours (LH) have the basic meaning of continuation and can be used to hold a turn (Figure 1, top). Third, falling contours (HL) have the basic cross-linguistic meaning of finality and can thus signal the end of a turn (Figure 1, bottom). Fourth, rising contours may be found in interrogatives, but she concludes that the most consistent acoustic correlate of interrogative utterances is vowel lengthening. Finally, she highlighted the fact that Inuktitut does not show the declination patterns characteristic of most of the world languages, and this is also illustrated in Figure 1. Given these prosodic characteristics, Inuktitut could be classified as an edge-prominence language (Jun 2014; see Arnhold 2014 for West Greenlandic), in which tonal events are located at the end of a domain.

Although the systematic differences between the languages at the prosodic and morphosyntactic level may hinder any type of prosodic convergence, Inuktitut and English have been in contact since the 1500s (Dorais 2010, chp. 5), and thus, sociolinguistic conditions lead us to hypothesize that Inuktitut prosodic features may be transferred to English.^{2,3} Indeed, most of the population of Nunavut, where our participants are from, is bilingual (Allen 2007; Dorais 2010; Statistics Canada 2019). There is still a large percentage of speakers who claim Inuktitut as their first language and this percentage is higher than any other aboriginal language in Canada (Allen 2007). Moreover, a series of political decisions, such as the creation of Nunavut in 1999, the Official Languages Act (1988) and the Inuit Language Protection Act (2008), have resulted in the promotion of positive attitudes towards the language (Dorais 2010).⁴ Education has also played a role in language maintenance. The introduction of education in Inuktitut up to Grade 2 or 4 has allowed children to develop their writing skills in their first language, whereas the absence of a comprehensive curriculum in Inuktitut (Aylward 2010; Dorais 2010) yields an increasing use of English in later grades. As is the case in most bilingual communities, though, there is a large degree of individual variation in terms of proficiency and language use. Patterns of language use not only vary across the Arctic region (Dorais 2010, pp. 226–7), but depend on the specific demographic conditions of each individual, as we will see when we discuss the participants' profiles (Table 1).

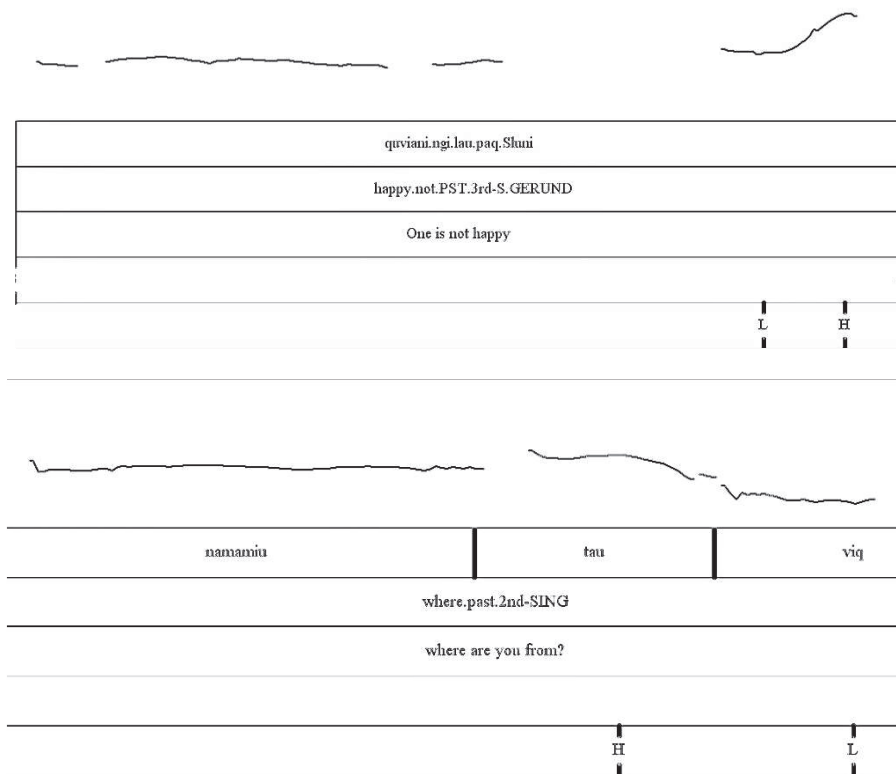


Figure 1. Realization of intonational contours in Inuktitut. Top: Rising contour to hold a turn; female speaker from Baker Lake (Shokeir 2009, p. 21). Bottom: Falling contour to indicate the end of a turn; female speaker from Iqaluit (Shokeir 2009, p. 24).

Thus, if there is an influence from Inuktitut into English, we expect to observe an overall difference between bilinguals and English monolinguals. These differences are predicted to be larger in the perception, interpretation, and production of tonal movements at the beginning rather than at the end of the utterance (see also Section 3). Since previous research on bilingual intonation has suggested that differences between monolinguals and bilinguals are modulated by the type of task (Grabe et al. 2003; Ortega-Llebaria and Colantoni 2014), our secondary goal is to analyze whether group differences are smaller in tasks that tap auditory rather than contextualized perception, and in imitation rather than in contextualized production tasks. In the next section, we review the literature on the perception and production of sentence types in bilinguals. This is followed by our research questions and hypotheses in Section 3, and our methodology in Section 4. In Section 5, we summarize our results, first for perception and then for production, and then we compare the perception-production results. We discuss our findings in Section 6, and briefly conclude in Section 7.

2. The Prosody of Sentence Types in Bilinguals

2.1. Cross-Linguistic Influence and Sentence Types

Regarding intonation, more is known about production than perception in language contact situations. This represents a striking contrast with the Second Language Acquisition literature, where theoretical models derive their primitives from perception (e.g., Flege 1995; Flege and Bohn 2021; Best 1995; Best and Tyler 2007). Different scenarios have been

studied, which include situations of stable social bilingualism, migratory languages, and heritage speakers. In addition, studies have factored in language typology (contact between typologically similar and different languages), as well as the possibility of bidirectional influence (Delais-Roussarie et al. 2015) and language attrition. The overall picture suggests that intonation is permeable to the influence of language contact, with the possibility of one (Muntendam and Torreira 2016) or both languages (Mennen 2004; Delais-Roussarie et al. 2015; Queen 2012; Dehé 2018) being affected. Some prosodic structures are more susceptible than others (Delais-Roussarie et al. 2015), and some positions in the contour are also more prone than others to being affected by language contact.

Studies involving the perception of sentence types have concentrated on the role of cross-linguistic influence (CLI) and have mostly focused on English, either as the L1 or L2. Beginning with studies that investigated the perception of a foreign language, there is evidence that one's L1 influences foreign language perception of sentence types. For example, Cruz-Ferreira (1983) tested Portuguese and English speakers on their perception of Ss and DQs in English and Portuguese, respectively. She found differences in the identification of sentence types, particularly in those that were characterized by a low-rise, showing that sentence type identification in a foreign language is influenced by the L1. Similarly, Liu and Rodríguez (2012) looked at the identification and discrimination of final contours in English statements and yes/no questions by monolingual English and Chinese speakers. They found that the groups differed in the processing of contours, since in Chinese there is an interaction of intonation and lexical tone.

Closer to our study are investigations that have analyzed the role of CLI in the perception of sentence types in early and late bilinguals. These studies have used a variety of methodologies, such as gating paradigms (Marasco 2020), identification tasks (Radu et al. 2018; Patience et al. 2020) or imitation of resynthesized stimuli (Zárate-Sánchez 2015) and have yielded mixed results. L1 English-L2 Spanish speakers were the focus of two studies (Zárate-Sánchez 2015; Marasco 2020). Marasco (2020) investigated the perception of initial boundary tones and prenuclear peaks in advanced learners, whereas Zárate-Sánchez (2015) analyzed the perception of prenuclear accents and final boundary tones in learners of different proficiencies (beginners and advanced), as well as heritage speakers. While Marasco (2020) found no evidence of CLI in the perception of pre-nuclear accents (i.e., controls and learners were equally accurate at distinguishing statements from yes-no questions), Zárate-Sánchez (2015) reported group differences in the perception of prenuclear accents, but not in boundary tones. Beginner learners were not able to detect alignment differences that corresponded to broad and narrow focus patterns, as expected from CLI, while all the other groups did. However, heritage speakers and advanced learners shifted between the two categories at an earlier point than native speakers. Radu et al. (2018) explored the identification and comprehension of statements, yes-no questions and declarative questions in L1 Spanish-L2 English advanced learners and found no evidence of CLI in the perception of either low-pass filtered or isolated stimuli. CLI was restricted to the interpretation of the different question types (see Section 2.2). Finally, Patience et al. (2020) found that L1 Mandarin speakers could use intonation to identify questions from statements in a low-pass filtered task; however, when the statements were presented as isolated utterances (not low-pass filtered), the L1 Mandarin speakers had difficulty distinguishing between Ss and DQs, suggesting that they paid more attention to the syntax than the intonation. The authors interpreted this as evidence of CLI, given that Mandarin yes-no questions are marked more reliably with syntax than with prosody. Some potential evidence for positive CLI was also observed in the L1 Mandarin-L2 English learners. When utterances were presented following a context that prompted either an AQ, DQ, or S, the L1 Mandarin speakers performed similarly to the L1 English controls, and outperformed L1 Spanish speakers. The authors attributed this to positive transfer, given that an AQ-DQ pragmatic contrast is marked prosodically in Mandarin, but not in Spanish.

In summary, perception studies offer mixed results, but suggest that group differences are larger with prenuclear accents than with boundary tones. Moreover, the findings reveal

that CLI related to the syntax and pragmatics of sentence types may play a more influential role than prosody, although this is dependent on the L1 of the L2 learners.

Production studies are not only more abundant, but they have also investigated a wider range of language pairings, including many typologically different languages. Once again, the evidence supports CLI, but different outcomes have been reported, such as convergence (Colantoni and Gurlekian 2004; Barnes and Michnowicz 2015), hybrid (Lai 2018) or mixed patterns (Queen 2001, 2012), overgeneralizations and hypercorrections (Santiago and Delais-Roussarie 2012). Moreover, changes due to CLI have been documented for different parts of the utterance. For example, changes in the alignment of high tones in prenuclear accents have been reported for Spanish declaratives in contact with Italian (Colantoni and Gurlekian 2004; Barnes and Michnowicz 2015) and for Spanish and English declaratives in English-Spanish bilinguals (Zárate-Sández 2015). Nuclear contours have been reported to display patterns of convergence in declaratives in Spanish in contact with Italian (Colantoni and Gurlekian 2004), Spanish in contact with Catalan (Simonet 2011) and Yami in contact with Mandarin (Lai 2018). Moreover, nuclear contours in interrogatives may be even more susceptible to change than nuclear contours in declaratives, as shown by Alvord (2007), who looked at the realization of final contours in declaratives and polar interrogatives in the Spanish of three generations of Cuban Spanish-English bilinguals. In addition to convergence studies, others have observed the emergence of mixed patterns, specifically the use of the same contour in both languages, albeit with a different pragmatic distribution (Queen 2001, 2012).

Crucially for our study, there is evidence of CLI from substratum indigenous languages, such as Quechua, into Indo-European languages, such as Spanish. Similar to Inuktitut, Quechua is an agglutinative language that uses morphemes to mark sentence type (e.g., Cerrón Palomino 1988; O'Rourke 2009) or information structure (Sánchez 2008). Evidence of influence from Quechua into Spanish has been reported in peak alignment patterns in prenuclear accents in broad and narrow focus declaratives (O'Rourke 2012), as well as in the lack of use of f0 cues to mark narrow (O'Rourke 2012) or contrastive focus (Muntendam and Torreira 2016). Overall, these studies show that Quechua-Spanish bilinguals use different peak alignment patterns and have a restricted use of pitch in Spanish when compared to monolingual controls. If, as Colantoni and Sánchez (2021) suggest, this is a result of different patterns of module interactions across languages according to which languages that have a rich morphological layer tend to have a restricted use of pitch to mark sentence types or information structure, we should expect to see a more restricted use of intonation in the English spoken by L1 Inuktitut speakers.

2.2. Role of Task Type in Modulating CLI in Contact Situations

Bilinguals have been reported to perform differently across tasks, particularly when tasks are not culturally appropriate in one of the languages (e.g., Sánchez 2008; Kiser 2014) or demand the use of skills that bilinguals may not have in one language (e.g., reading in the language in which they have not been educated; Tsimpli 2014). These task effects have also been observed in intonation studies (Barnes and Michnowicz 2015; Colantoni et al. 2016). These studies, however, did not explore whether access to contextual information was responsible for the differences observed. Studies that have examined the role of access to contextual information have yielded consistent results both in perception (Grabe et al. 2003; Radu et al. 2018; Patience et al. 2020) and perception-production (Ortega-Llebaria and Colantoni 2014). Grabe et al.'s (2003) pioneering study showed that three groups of English, Mandarin and Peninsular Spanish speakers did not differ in their discrimination of falling and rising contours in non-speech stimuli (i.e., frequency modulated sine-waves), but differed in their perception of falling contours when they listened to the utterance *Melanie Maloney* (whose final syllable was manipulated to generate 11 different stimuli with rising and falling contours) produced by a Scottish English speaker. Ortega-Llebaria and Colantoni (2014) studied the effect of access to meaning in the perception and production of English corrective stress by English controls and two groups of L2 speakers (L1 Mandarin

and L1 Spanish). Once again, learners diverged from controls more in perception and in production in tasks whereby they either answered or produced utterances appropriate to a context. Interestingly, the contextual effect was modulated by CLI. Spanish participants were outperformed by Mandarin learners, which is expected given that Mandarin resembles English more closely than Spanish in the prosodic marking of corrective stress. Radu et al. (2018) analyzed the perception and interpretation of Ss, AQs and DQs by L1 Spanish-L2 English speakers, using a variety of tasks. They found that learners did not differ from controls in tasks that tapped auditory processing, but they diverged from controls in contextualized tasks in which they had to choose the sentence type that appropriately completed a given context. Patience et al. (2020), which was based on the same methodology as Radu et al. (2018), found that Mandarin speakers behaved similar to controls, outperforming L1 Spanish speakers. As mentioned, the relative success of the L1 Mandarin speakers was attributed to positive CLI, given that Mandarin also contrasts prosodically between AQs and DQs. Note that these results mirror the findings in Ortega-Llebaria and Colantoni (2014), given that they also found that the Mandarin speakers outperformed the Spanish speakers, due to similarities in the prosody of the structure under examination.

In summary, previous work investigating the role of task has found that CLI (including positive CLI) is more prevalent as access to contextual meaning increases. As a result, we expect to find the same results in the speakers of the present study. We outline our specific hypotheses in the next section.

3. Research Questions and Predictions

Based on the research on the role of CLI and access to contextual information in the perception and production of intonation reviewed in the previous section, we formulate two research questions followed by the corresponding predictions.

Is there evidence of CLI in the bilinguals' perception and production of sentence types?

Based on previous descriptions of Inuktitut, which revealed that tonal movement is mostly restricted to nuclear position and rising and falling contours are not strictly associated with sentence types, we predict that bilinguals will be less accurate than English monolinguals at identifying statements and questions, when exposed to low-pass filtered stimuli or when the syntactic structure is identical (i.e., Ss vs. DQs). In production, the experimental group should differ from the control group to a larger extent in prenuclear than in nuclear position since there is little tonal movement in the L1 of the former group. Finally, bilinguals are expected to produce a larger number of rising contours in Ss than controls, given that final rises are not strictly associated with questions in their L1.

Is CLI modulated by task type?

Based on previous research (see Section 2.2), we expect to see larger between-group differences in more contextualized as opposed to more controlled tasks. In perception, bilinguals are expected to have difficulty identifying the appropriate context for AQs and DQs. In production, we expect bilinguals to resemble controls' pitch patterns more closely in controlled tasks than in contextualized tasks. We also expect bilinguals to have difficulty choosing the appropriate question type in the contextualized task (i.e., variable production of AQs in DQ-prompting contexts), given that sentence type is encoded in the morphology rather than in the syntax in their L1.

4. Methods

4.1. Participants

The study includes 16 English controls (12 females, 4 males) with a mean age of 24 (range: 18–30). All controls were born and raised in Canada and were studying or had completed a university degree. The bilingual group includes 13 participants (10 females, 3 males). Table 1 summarizes the bilingual participants' profiles.

Table 1. Participants' profiles. Notes: AoA = Age of onset of Acquisition of English. English Use = Self-reported percentage use of English in daily life. Self-rating: A = Advanced; NN = Near Native.

Participant	Gender	Age	Education	AoA	English Use	Self-Rating
I01	F	20	College	6	25	A
I02	M	35	College	3	37.5	A
I03	F	26	College	0	50	NN
I04	F	25	College	0	55	A
I05	F	19	College	0	58.3	NN
I06	F	22	College	8	46.6	A
I07	F	71	University	6	87.5	A
I08	M	23	College	11	62.5	A
I09	F	55	College	6	65	A
I10	F	57	College	9	53.8	A
I11	M	47	Secondary	6	97.5	A
I12	F	25	Secondary	6	80	A
I13	F	28	College	6	73.3	NN

At the start of the testing session, participants completed a background questionnaire detailing several aspects of their language experience and abilities. All bilingual speakers were exposed to Eastern Canadian Inuktitut at home where either one (3/13) or both parents spoke Inuktitut. They constitute a fairly homogenous dialectal group (Dorais 2010, p. 19), but they represent different “speech areas” (Dorais 2010) within this dialect (North and South Baffin: N = 9; Nunatsiavut: N = 3; Aivilik: N = 1). This means that all participants were born and raised in areas in which Inuktitut is currently in contact with English (rather than with French). As mentioned in the Introduction, most participants (N = 9) came from an area where bilingualism has been expanding since the 70s, but where Inuktitut is still both an official language and the language of the home.

It is important to highlight that the groups exhibit several differences regarding their education (only one bilingual participant completed a university degree) and mean age (bilinguals are older than controls). The bilingual group also exhibits some variability in terms of the age of onset of English acquisition. In the sample (Table 1), we have three simultaneous bilinguals and one participant who was exposed to English before entering school, whereas the rest were exposed to English upon entering the school system or slightly later. The amount of English used daily also varies. Table 1 presents the mean proportion of English use by speaker, which is the result of averaging the proportion of English used at home, at work, in school and in social situations. Although the mean proportion of English use is 61%, there is a wide range, with some participants reporting to use English only 25% of the time and others using almost exclusively English. Given the variability in our sample, in addition to the group results, we will present individual results both for perception and production.

4.2. Materials

The data reported here include perception and production experiments, and within each category, we developed tasks that manipulated the degree of access to contextual information, ranging from no access to contextual information to perceiving and producing utterances appropriate to a context (see Table 2 for a summary).

The perception component of the experiment included three tasks. In the first task, or intonation only task (IO), participants heard a low-pass filtered stimulus out of context. In the second task, participants heard isolated unaltered utterances that contained segmental

and intonation information (SI task). In the third task, participants heard a scenario followed by three utterances, only one of which was appropriate to the context (C task).

Table 2. Study design.

	Tasks	Task-Type	Conditions	Number of Trials
Perception	Intonation only (IO)	Forced-choice identification (low-pass filtered stimuli) (question/statement/exclamation)	Sentence type (S/AQ/DQ)	10 items × 3 types = (30) + 25 distracters
	Intonation and lexical information (SI)	Forced-choice identification (question/statement/exclamation)	Sentence type (S/AQ/DQ)	10 items × 3 types = (30) + 25 distracters
	Context (C)	Sentence preference (choice of statement/non-inverted question/inverted question)	Sentence type (S/AQ/DQ)	6 items × 3 types = 18
Production	Sentence Imitation	Sentence Repetition	Sentence type (S/AQ/DQ)	10 items × 3 types = (30) + 25 distracters
	Context	Completion task	Sentence type (S/AQ/DQ)	6 items × 3 types = 18

Our production experiment included two tasks that also varied according to the amount of contextual information. In the first task, participants heard an utterance in isolation and were asked to repeat it (Sentence Imitation task—SI). In the second task, they heard the same scenarios used in the perception task, but, this time, participants were asked to produce an utterance appropriate to the context (C task).

The stimuli used for the IO and SI perception tasks, and the SI production task consisted of 10 utterances for each sentence type (AQ, DQ, S) and 25 distractors, which included Wh-questions and exclamations. The stimuli were recorded by a Canadian female speaker using a Marantz solid-state recorder PMD-661 and a unidirectional lavalier microphone. The stimuli were digitized using a 22,000 sample-rate and a 16-bit resolution. All of the stimuli were checked for naturalness and potential reading errors by all authors.

The stimuli in the C task consisted of six scenarios, as in (9), per sentence type and no distractors.⁵ These scenarios were selected from a larger set of scenarios piloted, given that they prompted appropriate responses in monolingual and L2 speakers of English alike. In the perception task, after hearing the scenario, participants heard three utterances only one of which was appropriate to the context. In the production component, participants had to produce a phrase appropriate to the context. Materials for the contextualized tasks were recorded by the same Canadian female speaker who recorded the other stimuli using the same equipment described above.

(9) C task

Context (S):

Mary is on vacation in Toronto and really wants to see a racoon. One of her friends knows of a place with a bunch of trees where racoons live and takes Mary there to see if she can finally see one. Soon after they arrive, a racoon shows up and Mary's friend says, "Look . . ."

(a) **This is a racoon.** (b) *This is a racoon?* (c) *Is this a racoon?*

Context (DQ):

Before coming to Toronto from Australia, Mary heard about racoons, looked at some pictures and thought they were cute little things. One evening, she was eating outside with friends and saw a mid-sized animal crossing the street and thought it was a dog. Her friends commented that it was a racoon, and she asked . . .

(a) *This is a racoon.* (b) **This is a racoon?** (c) *Is this a racoon?*

Context (AQ):

Mary is from Australia, and she has never seen a raccoon in her life. When she got to Toronto, she spent hours in the evening trying to spot one. One evening, she is sitting outside with a bunch of friends and she sees something that she believes may be a raccoon. She points at the animal and asks . . .

(a) *This is a racoon.*

(b) *This is a racoon?*

(c) *Is this a racoon?*

The stimuli used were acoustically analyzed to determine whether the target sentence types were produced with the intended characteristics. Table 3 summarizes the acoustic characteristics of the target stimuli used in the IO and SI perception tasks, as well as in the SI production task.

Table 3. Acoustic analysis of the perception (SI and IO tasks) and production stimuli (SI task). Mean max F0 values in the first pitch accent and the nuclear contour, and F0 excursion in the first pitch accent and nuclear contour (values in semitones).

Sentence Type	Pitch Accent (Max F0)	Nuclear Contour (Max F0)	F0 Excursion (Pitch Accent)	F0 Excursion (Nuclear Contour)
AQ	97	101	6	16
DQ	98	104	9	20
S	98	95	6	9

The stimuli for the three sentence types used in these decontextualized tasks clearly differed in the realization of the nuclear contour (pitch excursion: DQ > AQ > S) and partially differed in the realization of the first pitch accent, which had a larger pitch excursion in DQs than in the other sentence types. Most importantly, the prosodic characteristics of the stimuli used are consistent with those reported in previous descriptions of American English (e.g., Bartels 1999).⁶

Finally, Table 4 displays the characteristics of the stimuli used in the C task (Perception only). Once again, the three sentence types differed in the degree of pitch change in the nuclear contour (DQ > AQ > S), although the pitch excursion was smaller in this task than in the others. Similar F0 maximum values were obtained for the first pitch accent in the three sentence types, but here, as opposed to the other tasks, the largest pitch excursion was produced in Ss.

Table 4. Perception stimuli used in the C task. Mean max F0 values in the first pitch accent and the nuclear contour, and F0 excursion in the first pitch accent and nuclear contour (values in semitones).

Sentence Type	Pitch Accent (Max F0)	Nuclear Contour (Max F0)	F0 Excursion (Pitch Accent)	F0 Excursion (Nuclear Contour)
AQ	93	101	2	14
DQ	93	103	1	18
S	97	92	4	4

4.3. Procedure and Data Analysis

The perception and production tasks reported in this paper are part of a larger project in which we analyzed other structures (e.g., attachment ambiguity) and included additional L1 groups (Spanish and Mandarin). Thus, we had two testing sessions which were one week apart. Perception and production components of each task were divided into the two testing sessions and participants were randomly assigned to start either with the perception or the production component.

The perception tasks were administered using SuperLab pro. In the IO and SI tasks, participants listened to the stimulus, and then pressed one of the three colored keys on the keypad corresponding to Statement, Question or Exclamation. This last response was included since DQs could be interpreted as exclamations and there were exclamations among the distractors. In the C task, participants listened to the scenario and then heard three possible options that would complete the scenario, either a statement, a DQ or an AQ.

Participants only listened to each stimulus once. After having heard the last option, they had to press one of the three keys on the keypad. Before testing began, we included a short practice session.

The production portion of the experiment was administered via PowerPoint, and responses were recorded with the same equipment used to prepare the stimuli and analyzed with Praat (Boersma and Weenink 2017). In the SI task, participants listened to a stimulus and were asked to repeat it. In the C task, participants listened to the scenario and then had to produce an utterance that would complete each scenario. In both cases, participants were allowed to listen to the stimulus more than once. In all cases, practice sessions were introduced at the beginning of each task.

Perception data were analyzed for accuracy. In the production data (Figure 2), we identified the first pitch accent and the nuclear contour (i.e., last pitch accent and boundary tone). We labeled each tonal event using the ToBI system (Beckman and Ayers Elam 1997) and measured the maximum and minimum f0 (in semitones) associated with each tonal event. We then calculated the pitch change (i.e., the f0 maximum minus the f0 minimum) over the first pitch accent and the nuclear contour. Labeling was conducted by one of the authors and then checked by a second author.

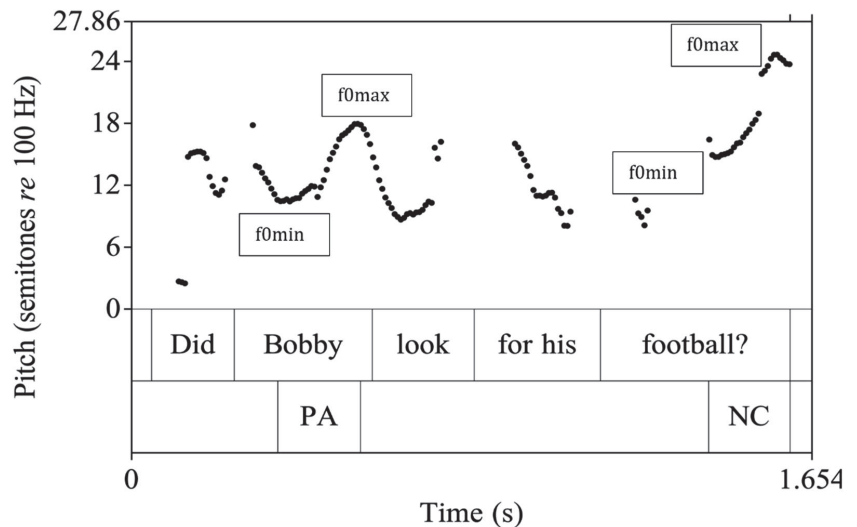


Figure 2. Example of labeling of pitch accent (PA) and nuclear contour (NC).

Statistics were calculated with R Core Team (2013). We used a combination of linear mixed effects models and binomial mixed effect models, with treatment coding contrasts for our categorical variables. In all of the statistical analyses, for the sentence type variable, “AQ” was the reference level; for language, “English” was the reference level; and for task, the reference level was the C task. The values that we display in the results of our statistical tables therefore reflect the listed value with that of the reference level. We will provide details about the specific models in each of the results sections.

5. Results

5.1. Perception

Table 5 displays the mean accuracy by task and shows that bilinguals had a lower proportion of accurate answers across tasks and sentence types than controls. However, except for DQs in the C task, responses were always above chance.

Table 5. Proportion of accurate responses by Task and Sentence Type organized by language group.

Task/Sentence Type	IO			SI			C			Mean Accuracy
	S	AQ	DQ	S	AQ	DQ	S	AQ	DQ	
English	83	84	86	93	97	83	87	93	84	88
Inuktitut	76	75	73	80	91	76	67	84	33	73

To determine whether group differences were statistically significant and to understand whether such differences were larger in contextualized than in de-contextualized tasks, we fitted a generalized binomial mixed effects model with accuracy (Accurate; Non-accurate) as the dependent variable, Task (C, SI, IO), Sentence Type (AQ, DQ, S) and Language (English, Inuktitut) as fixed factors, and Participant and Item as random factors (random intercepts). We also tested models with two and three-way interactions. Model comparisons using the AIC criterion revealed that the model which best fitted the data (i.e., the one with the lowest AIC value = 1805.4) was the one that included all the fixed factors and a three-way interaction. Results of this model are reported in Table 6, confirming that the number of non-accurate responses was significantly higher in the experimental than in the control group. Bilinguals also were less accurate in DQs when compared to AQs, but as expected, the non-accurate responses with DQs were lower in the SI task than in the C task.

Table 6. Binomial mixed effects model with Language, Task and Sentence Type as fixed effects and Language*Task*Sentence Type interaction (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). Reference values: Non-accurate, English, Task C, Sentence type AQ.

Fixed Effects	Estimate	SE	z Value	Pr (> z)
(Intercept)	-2.98	0.51	-5.85	<0.001 ***
Language (I)	1.01	0.62	1.63	<0.001 ***
Task (IO)	1.12	0.53	2.11	0.030 *
Task (SI)	-0.98	0.69	-1.41	0.150
Sentence (DQ)	1.00	0.57	1.72	0.080
Sentence (S)	0.79	0.58	1.34	0.170
Language (I)*Task (IO)	-0.44	0.65	-0.68	0.490
Language (I)*Task (SI)	0.30	0.82	0.36	0.710
Language (I)*Sentence (DQ)	1.81	0.67	2.67	0.007 **
Language (I)*Sentence (S)	0.37	0.68	0.55	0.580
Task (IO)*Sentence (DQ)	-1.15	0.69	-1.67	0.090
Task (SI)*Sentence (DQ)	1.11	0.82	1.85	0.170
Task (IO)*Sentence (S)	-0.79	0.69	-1.14	0.250
Task (SI)*Sentence (S)	0.30	0.85	0.35	0.720
Language (I)*Task (IO)*Sentence (DQ)	-1.49	0.85	-1.17	0.080
Language (I)*Task (SI)*Sentence (DQ)	-2.59	0.99	-2.62	0.008 **
Language (I)*Task (IO)*Sentence (S)	-0.36	0.83	-0.43	0.660
Language (I)*Task (SI)*Sentence (S)	-0.42	1.00	-0.42	0.67

Results of post-hoc pairwise Tukey-adjusted comparisons revealed that, in the IO task, there were no significant between-group differences for any of the sentence types tested. In the SI task, instead, bilinguals were less accurate than controls in DQs ($\beta = -2.64$; $SE = 0.38$; z ratio = -4.00 ; $p = 0.007$) and in Ss ($\beta = -2.37$; $SE = 0.66$; z ratio = -3.56 ; $p = 0.03$). Controls also were less accurate with DQs than with AQs ($\beta = -2.11$; $SE = 0.58$; z ratio = -3.62 ; $p = 0.3$). Finally, in the C task, bilinguals displayed a higher number of non-accurate responses than controls in DQ-prompting contexts ($\beta = -2.83$; $SE = 0.51$; z ratio = -5.50 ; $p < 0.0001$); their accuracy was also lower in this context than with AQ- ($\beta = -2.81$; $SE = 0.49$; z ratio = -5.67 ; $p < 0.0001$) and S-prompting contexts ($\beta = 1.64$; $SE = 0.44$; z ratio = 3.68 ; $p = 0.02$). No within-group differences were found in the control group.

An analysis of the response patterns (Figure 3), particularly in the C task, revealed that bilinguals differed from controls in their responses to DQ- and S-prompting contexts. As concerns the former, bilinguals were twice as likely as controls (33% vs. 15%, respectively) to choose AQ as a possible answer, although DQ was still the most frequently chosen response (59%). The proportion of non-target-like responses was smaller in the S- than in the DQ-prompting contexts (33% vs. 41%, respectively), and DQs and AQs were chosen as a response at a similar rate (15% to 18%, respectively). Thus, we can partially answer our first research question; namely, bilinguals differed from controls in their identification of sentence types, displaying a higher number of non-accurate responses than controls across tasks, particularly in the C task (see RQ2).

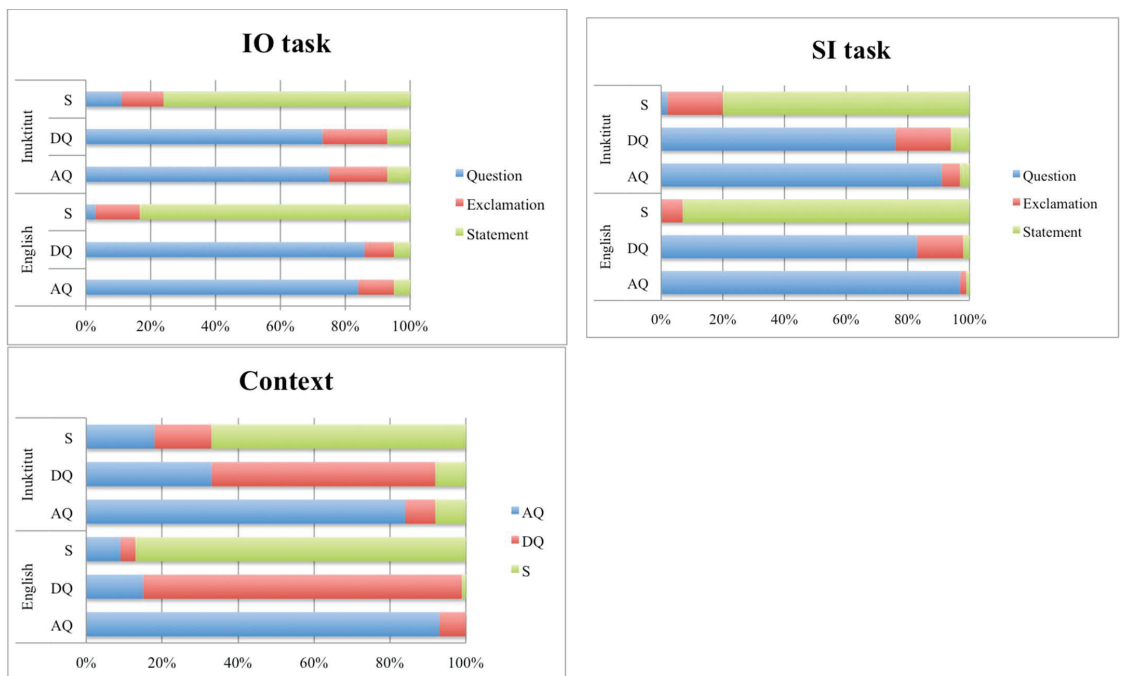


Figure 3. Proportion of response type by Task (IO = Intonation only; SI = segments and intonation; C = Contextualized) and Sentence Type. Results are organized by language group.

The results reported above reflect the behavior of both groups, but bilinguals have diverse language histories and their behavior is highly variable, so it is crucial to explore to what extent individuals mirror the group behavior. As seen in Figure 4 (see also Table 5), 10/13 speakers displayed accuracy values that were within one SD from the mean. Two speakers (I04 and I12) were above that threshold and one participant (I11) was clearly below one SD from the mean. Demographic variables may account in part for these results; I04 was a simultaneous bilingual, with college education who used both languages in equal proportions. I12 was the same age as I04, and, although she was exposed to English when she entered the school system, she reported using English most of the time. I11's behavior, however, is difficult to explain with the information available, since his language learning profile was similar to I12's and he was the participant who reported using English the most (Table 1). In the next section, we will compare these findings to the production results to better understand if his lower accuracy is a consequence of the perception tasks used, or if it reflects his overall performance.

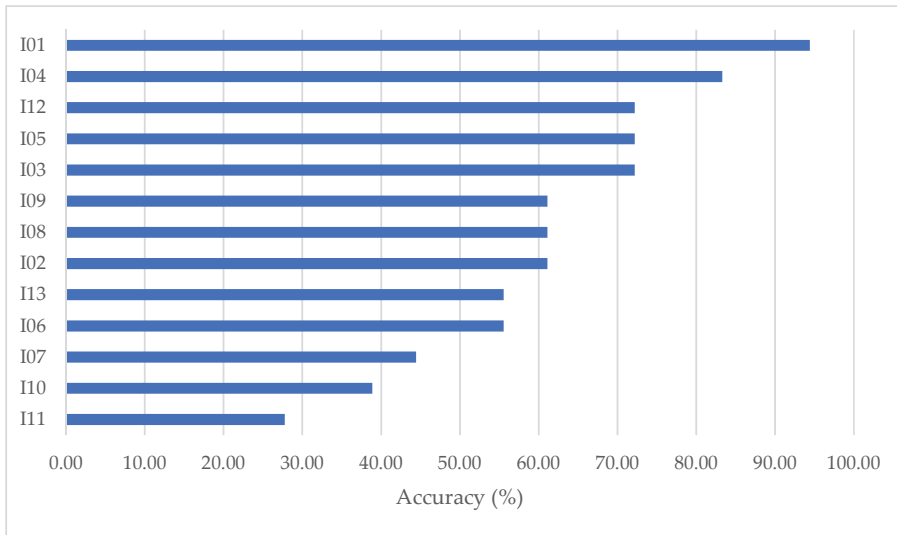


Figure 4. Percentage of accurate responses in all tasks combined by Inuktitut-English bilinguals.

5.2. Production

5.2.1. Accuracy

Before discussing pitch changes in pitch accents and nuclear contours, it is important to analyze the response accuracy, particularly in the C task, which allowed for open answers. We focus here on these results, which are displayed in Table 7, given that there were no repetition errors in the SI task. We treated any utterance that was not consistent with the contextual prompt as a non-target realization. For example, the use of a Wh-question or an inverted question in a context that prompted a DQ was treated as non-accurate, as was the use of a question in a context that was intended to prompt a statement.

Table 7. Percentage of target-like responses per Sentence Type in the contextualized production task organized by language group.

Sentence Type	AQ	DQ	S	Mean Accuracy
English	94	83	97	91
Inuktitut	87	64	85	79

Table 7 reveals that bilinguals were overall less accurate than controls, particularly in DQ-prompting contexts, where most of the non-target responses (86%) involved the production of an AQ. Results of a binomial mixed-effects model with Response (Accurate, Non-accurate) as the dependent variable, Language and Sentence Type as independent variables, and Participant and Item as random factors revealed that bilinguals did not differ from controls as a group (Table 8).⁷ Non-target responses were significantly higher in DQ-prompting contexts and post-hoc Tukey pairwise comparisons showed that this was the case for controls (AQ vs. DQ: $\beta = -1.71$; $SE = 0.59$; z ratio = -2.85 ; $p = 0.04$) and for bilinguals (AQ vs. DQ: $\beta = -1.70$; $SE = 0.59$; z ratio = -2.85 ; $p = 0.04$), but no differences were found in DQ accuracy between groups ($\beta = -0.88$; $SE = 0.72$; z ratio = -1.21 ; $p = 0.82$).

Table 8. Binomial mixed effects model with Language and Sentence Type as fixed effects (** $p < 0.01$; *** $p < 0.001$). Reference values: Non-accurate, English, Sentence type AQ.

Fixed Effects	Estimate	SE	z Value	Pr (> z)
(Intercept)	−3.95	0.70	−5.62	<0.0001 ***
Language (I)	0.88	0.72	1.21	0.22
Sentence (DQ)	1.70	0.59	2.85	0.004 **
Sentence (S)	0.04	0.63	0.07	0.94

Accuracy results, however, do not present an overall picture of participants’ behavior in this task. Whereas controls failed to produce an utterance appropriate to the context in a very small percentage of cases (AQ: 1%; DQ: 3%; S: 7%), bilinguals produced no responses or one-word responses in a larger proportion of contexts, particularly in DQ-prompting contexts (AQ: 8%; DQ: 26%; S: 15%). Individual results (Figure 5) reveal an interesting pattern; namely, there was a quasi-complementary distribution between non-target-like responses and the absence of response. Indeed, participants with the highest number of accurate responses did not produce utterances that were inappropriate to the context, but failed to produce an answer to some scenarios, whereas participants with the lowest accuracy tended to produce a response in all contexts.

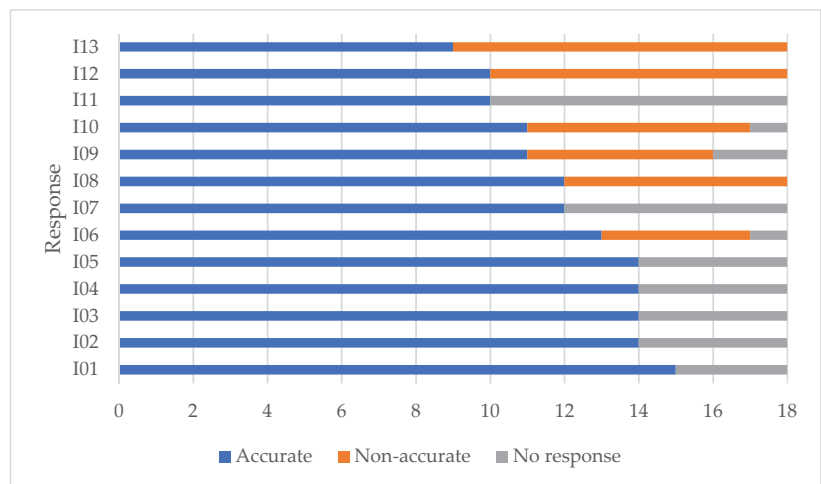


Figure 5. Accurate, non-accurate and no responses in the contextualized production task (bilinguals only). Note: total of contexts = 18.

Accuracy in production was equal (2/13) or higher than in perception for most bilingual participants (8/13), as illustrated in Figure 6. Moreover, all participants performed above chance in production, which was not the case in perception. Interestingly, participants who were exposed to English at home (i.e., I03, I04, I05) were the ones with the most consistent performance in perception and in production. As for the remaining participants, the overall higher accuracy in production may be attributed to the difficulty of the perception task, which tapped into more metalinguistic knowledge than the production task.

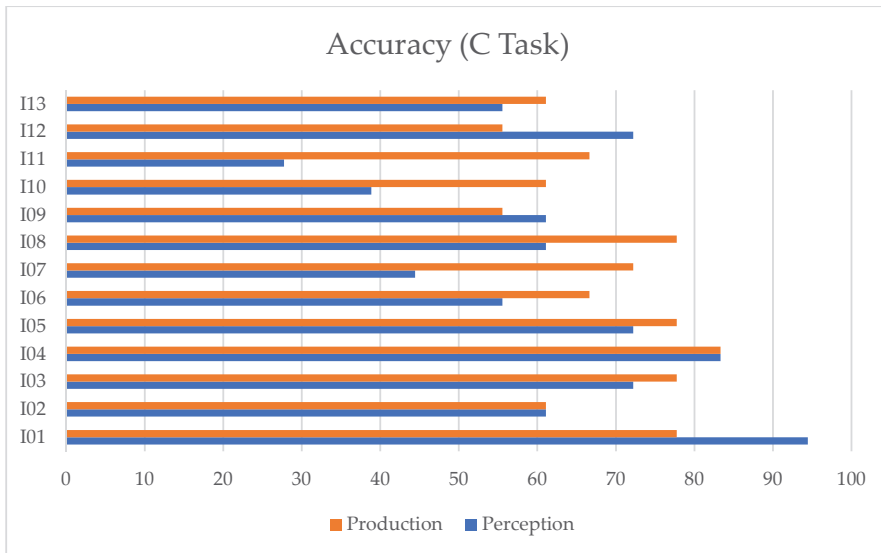


Figure 6. Accuracy in perception and production (C task only) by participant.

5.2.2. Phonetic Realization of Pitch Accents and Nuclear Contours

In this section, we analyze the patterns of pitch change in the first pitch accent and in the nuclear contours in both tasks. If there is an influence from Inuktitut into English, we expect to see very little pitch movement at the beginning of the utterance. Recall that we measured the f_0 maximum minus the f_0 minimum. Thus, if the first accent is a rising accent, we expect a positive difference, and if there is no pitch movement, we expect a value close to 0. Results displayed in Figure 7 suggest that the latter is the case. If we compare the patterns obtained for each group, we see that bilinguals have values that are close to 0 (C task (mean in ST): AQ = 0.8; DQ = 1.2; S = 0.4; SI task (mean in ST): AQ = 1.6; DQ = 1.4; S = 1.1) and that are relatively similar across sentence types and tasks. Controls, instead, showed larger pitch changes in questions than in statements (C task (mean in ST): AQ = 1.4; DQ = 3.4; S = 0.9; SI task (mean in ST): AQ = 4.8; DQ = 4.8; S = 1.9) and the amount of pitch change varied between tasks.

To determine the significance of pitch change in the first pitch accent, we ran a series of linear mixed effect models with pitch change (in semitones) as the dependent variable, Language and Sentence Type as the independent variables, and Participant and Stimulus as random factors. We also tested models with interactions. Here and elsewhere in this subsection, we will report the results of the best model according to the AIC criterion. In all cases, we compared the base model (only random effects) with models including only the independent variables or the independent variables plus the interactions. As for the first pitch accent, model comparisons revealed that the best model was the latter, and its output is reported in Table 9.⁸

Results showed a main effect of Sentence Type (larger pitch change in DQs than in other sentence types) and Task (larger pitch change in the SI than in the C task). Interactions between Language, Task and Sentence Type revealed that bilinguals had a smaller pitch change than controls in the SI task, in general, but pitch change was larger in this task in DQs and Ss when compared to those same sentences in the C task. Finally, post-hoc Tukey pairwise comparisons showed that controls had a larger pitch change in questions than in Ss (E,AQ vs. E,S: $\beta = 2.90$; $SE = 0.470$; $df = 49.2$; t ratio = 6.18; $p < 0.0001$; E, DQ vs. E,S: $\beta = 2.84$; $SE = 0.47$; $df = 49.1$; t ratio = 6.04; $p < 0.0001$) in the SI task, and between AQs and DQs ($\beta = -1.84$; $SE = 0.52$; $df = 158.9$; t ratio = 3.53; $p = 0.020$) and DQs and Ss

($\beta = 2.21$; $SE = 0.52$; $df = 143.4$; t ratio = 4.18; $p = 0.002$) in the C task. Bilinguals, instead, showed no significant differences across sentence types in both tasks. Figures 8 and 9 further show that the group tendencies hold for most of the individuals in the group, since bilinguals' values are closer to 0 and are similar across sentence types. It is important to remember, however, that fewer tokens of DQs were analyzed in the bilingual group in the C task because participants either failed to produce an analyzable utterance or produced an utterance that was not expected in that context (see Figure 5).

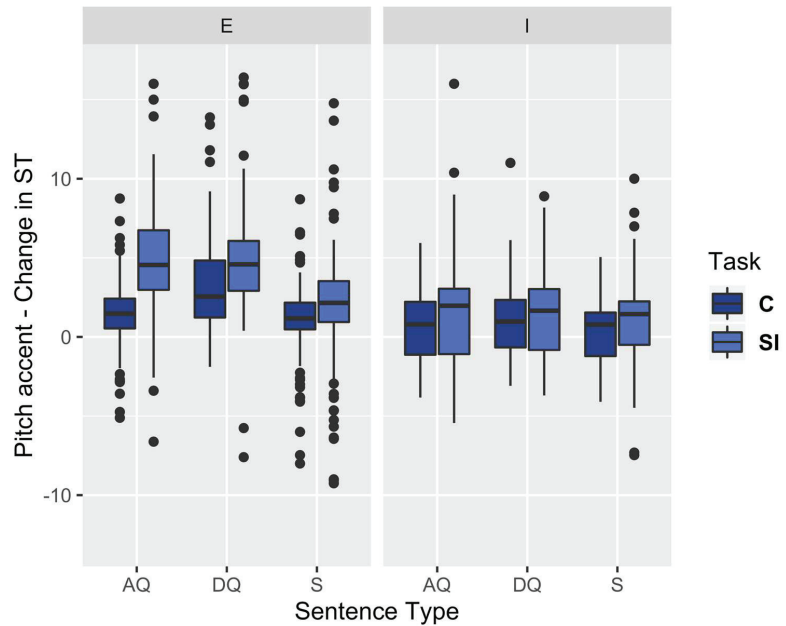


Figure 7. Boxplots displaying the pitch change (in semitones) over the first pitch accent in both tasks. Results organized by group.

Table 9. (Pitch accents). Linear mixed effect model with Language, Task and Sentence Type as fixed effects and Language**Sentence Type**Task interaction (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$).

Fixed Effects	Estimate	SE	df	t Value	Pr(> t)
(Intercept)	1.23	0.44	103.31	2.80	0.006 **
Language (I)	-0.54	0.59	100.51	-0.91	0.360
Sentence (DQ)	1.84	0.50	135.90	3.63	0.0004 ***
Sentence (S)	-0.37	0.50	110.77	-0.74	0.460
Task (SI)	3.55	0.36	1154.20	9.63	<0.001 ***
Language (I)*Task (SI)	-2.66	0.54	1157.55	-4.92	<0.001 ***
Language (I)*Sentence(DQ)	-1.87	0.69	1166.44	-2.68	0.007 **
Language (I)*Sentence (S)	0.15	0.63	1162.57	0.24	0.800
Sentence (DQ)*Task (SI)	-1.89	0.53	1161.08	-3.55	<0.001 ***
Sentence (S)*Task (SI)	-2.52	0.53	1174.89	-4.75	<0.001 ***
Language (I)*Sentence (DQ)*Task (SI)	1.74	0.83	1163.46	2.10	0.030 *
Language (I)*Sentence (S)*Task (SI)	2.29	0.79	1160.42	2.87	0.004 **

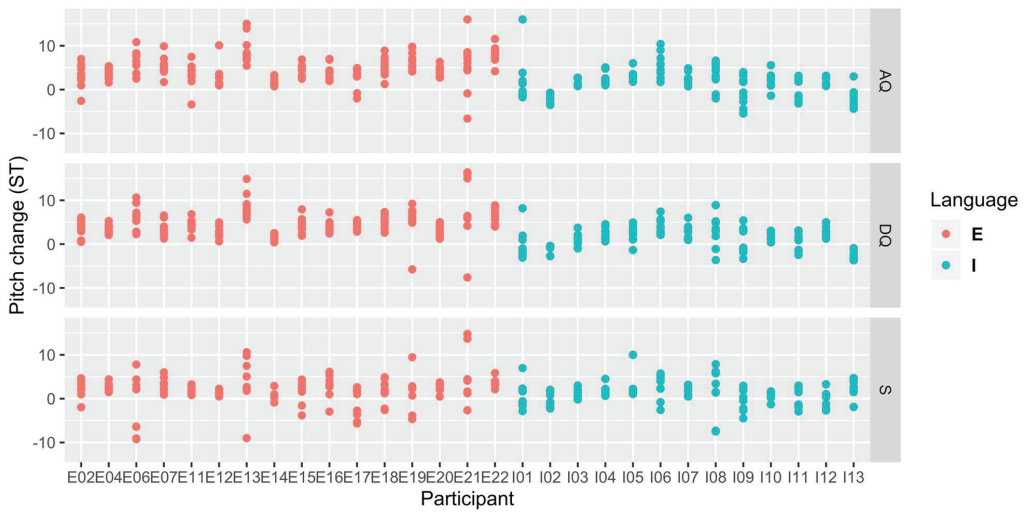


Figure 8. Pitch change in the first pitch accent (SI task) in each sentence type by participant.

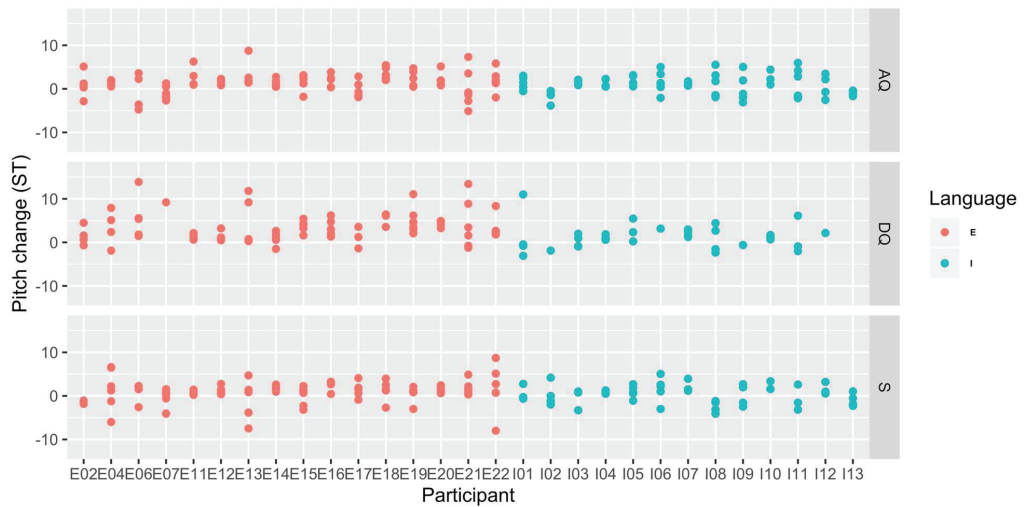


Figure 9. Pitch change in the first pitch accent (C task) in each sentence type by participant.

Individual results revealed that, in both tasks, some participants (e.g., I09) had consistently lower pitch change, whereas other participants (e.g., I06, I07) had consistently larger pitch changes. Other participants had a relatively large pitch change in the SI task, but a small pitch change in the C task (e.g., I08).

We now turn to the analysis of nuclear contours. Figure 10 displays the results obtained in both tasks for bilinguals and controls. Groups appear to resemble each other more closely in the realization of nuclear contours than in the realization of pitch accents (Figure 7).

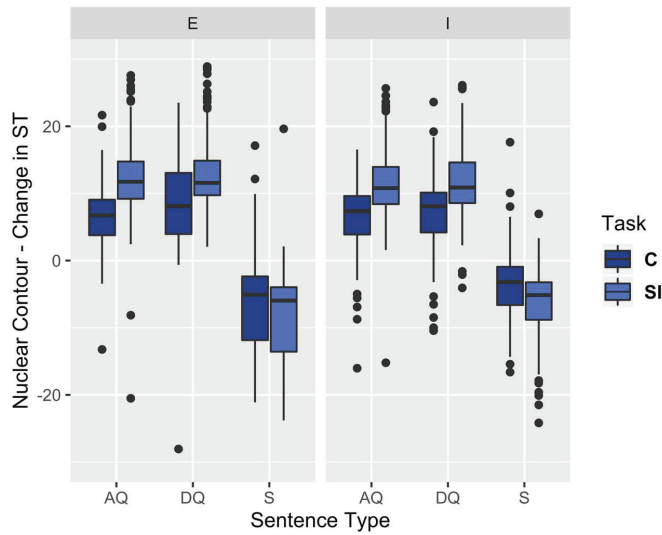


Figure 10. Boxplots displaying the pitch change (in semitones) over the nuclear contour in both tasks. Results organized by group.

To investigate whether there were any significant differences, we ran a series of linear-mixed effects models following the same procedure described for pitch accents. Once again, the best model (Table 10) was that with the three-way interaction.⁹ Results showed the expected difference in the realization of Ss when compared to questions, and as was the case with pitch accents, the task effect was also significant, revealing a larger pitch change in the SI task than in the C task, probably due to imitation. Groups only significantly differed in their realization of the nuclear falls in Ss, with bilinguals showing a smaller pitch change than controls (see Figure 10).

Table 10. (Nuclear contours): Linear mixed effect model with Language, Task and Sentence Type as fixed effects and Language*Task*Sentence Type interaction (* $p < 0.05$; *** $p < 0.001$).

Fixed Effects	Estimate	SE	df	t Value	Pr(> t)
(Intercept)	6.82	0.79	132.84	8.55	<0.001 ***
Language (I)	−0.55	1.14	164.58	−0.48	0.620
Sentence (DQ)	1.66	0.96	203.67	1.72	0.080
Sentence (S)	−13.41	0.98	203.62	−13.55	<0.001 ***
Task (SI)	5.54	0.78	1074.10	7.08	<0.001 ***
Language (I)*Task (SI)	−0.46	1.16	1144.69	−0.40	0.680
Language (I)*Sentence (DQ)	−1.27	1.49	1158.93	−0.85	0.390
Language (I)*Sentence (S)	3.45	1.39	1157.00	2.48	0.010 *
Sentence (DQ)*Task (SI)	−1.36	1.12	1087.04	−1.21	0.220
Sentence (S)*Task (SI)	−7.24	1.16	1133.28	−6.23	<0.001 ***
Language (I)*Sentence (DQ)*Task (SI)	1.42	1.77	1154.12	0.80	0.420
Language (I)*Sentence (S)*Task (SI)	−0.77	1.73	1152.74	−0.44	0.650

Results of post-hoc Tukey pairwise comparisons confirmed that both groups had the same patterns in the realization of nuclear contours; namely, rises in AQs and DQs did not differ significantly between groups and between tasks, whereas questions differed from statements in both tasks.

5.3. Summary of Results

Table 11 offers a qualitative summary of our perception and production results:

Table 11. Qualitative summary of the results obtained in the Perception and production experiments. Note: n.s. = non-significant difference.

Experiment	Parameter	Group	Task	Sentence Type
Perception	Accuracy	E > I	IO < SI, C	SI (both groups): DQ < AQ, S C (I only): DQ < AQ, S
	Accuracy	n.s.	–	DQ < AQ, S
Production	Pitch accent	E > I	SI > C	E: DQ > AQ > S (C task); DQ, AQ > S (SI task) I: n.s.
	Nuclear contour	I < E (Ss only)	SI > C	Questions > Ss

6. Discussion

6.1. Research Questions and Hypothesis Evaluation

We begin by returning to our first research question: *Is there evidence of CLI in the bilinguals’ perception and production of sentence types?* We found that, in perception, and as opposed to our prediction, groups did not differ in the IO task, where participants had to identify low-pass filtered stimuli, but did differ in the other two tasks in the direction predicted (i.e., with Ss and DQs). In the SI task, both groups were less accurate with DQs, but bilinguals, as opposed to controls, were also less accurate with Ss. In the C task, however, only bilinguals were less accurate in DQ contexts. This suggests that bilinguals associate meaningless intonation contours with sentence types, as monolinguals do, in patterns that resemble those observed in other studies with different language pairings (e.g., Grabe et al. 2003; Radu et al. 2018). However, when syntactic and contextual information are present, these take precedence over prosody, as expected due to CLI. As we mentioned, in Inuktitut, these sentence types are marked by different morphemes rather than by different intonation contours. In our study, when syntactic information was present (i.e., in AQs), bilinguals were as accurate as controls. However, when syntactic information was not informative (i.e., Ss and DQs), they were less accurate.

Accuracy patterns in production differed from those found in perception. Group differences were not found to be statistically significant, and all bilingual participants performed above chance, which was not the case in perception. However, we found different response patterns in bilinguals when compared to monolinguals in the C task. First of all, several participants did not provide an answer, or, as predicted, produced an AQ in DQ-prompting contexts. The analysis of pitch change largely supported the prediction regarding differences in prenuclear accents. As expected from CLI, bilinguals displayed a smaller pitch change than controls across tasks and sentence types. Moreover, the pitch change hovered slightly above 0 (Figure 8), revealing almost no pitch movement, especially when compared to controls, whose average pitch change ranged from 5 STs in both question types in the SI task, to 3 and 1.5 STs in DQs and AQs, respectively, in the C task. In nuclear contours, the groups did not differ in the use of rising patterns, but bilinguals displayed a less sharp fall than controls in both tasks in Ss. Thus, evidence of CLI was observed in multiple dimensions, including difficulties in perception to determine differences between question types or by producing AQs or Wh-questions in DQ-prompting contexts. This is expected if we keep in mind that questions and statements are marked by morphology in Inuktitut, as opposed to English. Results obtained for pitch changes in prenuclear accents are consistent with previous studies that indicate that pitch is not a reliable cue to stress in the language (Fortescue 1983; Shokeir 2009; Arnhold et al. Forthcoming), and that tonal changes are restricted to the end of the utterance (Massenet 1980; Fortescue 1983; Shokeir 2009). Finally, the smaller pitch change observed in nuclear contours in Ss in our study may be attributed to the absence of declination observed in Inuktitut (Shokeir 2009).

Our second question was: *Is CLI modulated by task type?* We predicted larger differences in contextualized (perception and production) tasks than in tasks that had no access or limited access to contextual meaning. This prediction was partially supported in perception and in production. In perception, although bilinguals were overall less accurate than

controls, differences were restricted to the SI and C tasks, that is, in tasks that include either only lexical and syntactic information (SI) or contextual information (C). An interesting interaction between task and sentence type was observed in the C task, where only bilingual speakers exhibited significantly more non-target responses in DQ-prompting contexts than in the other two contexts. We attribute this effect to CLI, and we interpret this as a sign of either a reduced sensitivity to the contextual factors that yield a preference for non-inverted questions (which results in AQs being accepted in this context) or a reduced sensitivity to tonal cues, which would account for the choice of Ss as a preferred answer.¹⁰ A complementary explanation to the behavior of bilinguals in the contextualized perception task (Figure 8 shows a high degree of variability among participants) could be task difficulty. Support for such an explanation comes from production results in the C task, where we found no significant between-group differences in accuracy rate. Along these lines, we could speculate that our perception task tapped into metalinguistic knowledge, since participants had to understand the context and imagine what kind of sentence would complete it. Differences in performance between tasks that require skills that bilinguals may not be accustomed to performing in both languages have been previously observed in different types of bilingual populations (Sánchez 2008; Kiser 2014; Tsimpli 2014). In the production task, instead, participants were asked to engage in something that is common in their everyday interactions, which is to listen to what somebody says and react appropriately. In addition to the lack of significant differences, we also saw more consistent individual patterns in production. Indeed, none of the participants performed below chance (Figure 8).

Concerning the analysis of pitch change (Figures 7–10), differences between groups were not larger in the C than in the SI task for several reasons. First, in prenuclear accents, bilinguals displayed the same degree of pitch change across tasks and sentence types; namely the average pitch change across tasks was consistently close to 0, which suggests that they were not sensitive to the large pitch excursions (Table 3) in the SI stimuli. In contrast, controls displayed what we believe to be an imitation effect. Indeed, we observed a larger change in the SI task than in the C task. In the former, the average values for AQs-DQs and Ss were 4.8 STs and 1.9 STs, respectively. In the latter task, values were consistently lower; namely, the average pitch range in Ss was 0.9 STs, and a difference between AQs (1.4 STs) and DQs (3.4 STs) emerged. This is consistent with the large pitch excursions (Table 3) in SI task in our stimuli.

Regarding nuclear contours, group differences were restricted to the magnitude of the fall in Ss. Otherwise, bilinguals and controls showed a larger pitch change in the SI task than in the C task. Indeed, the average pitch change across sentence types in the SI task was 10 STs for bilinguals and 11STs for controls. In the C task, this change was reduced to 5STs for bilinguals and 7STs for controls. Once again, we interpret this task difference as an imitation effect, since the pitch change in nuclear contours in the SI stimuli (Table 3) was rather large. It is interesting to see that bilinguals adjusted their pitch change in nuclear contours (as controls did) in the SI, but adjustments were not observed in prenuclear accents, which is consistent with our predictions that bilinguals would be more sensitive to pitch changes in nuclear than in prenuclear position, since tonal changes are restricted to this position in Inuktitut. Arguably, pitch changes in final contours should also be more salient than at the beginning of the utterance, since pitch changes are much larger in nuclear positions than in prenuclear positions (see description of the stimuli in Tables 3 and 4), independently of the task. Moreover, and as summarized in the Introduction, there is agreement that nuclear contours are a cue to sentence type in North American English. However, evidence indicating that initial pitch differences are a cue to sentence type is much more recent and such differences were not consistently present in our own stimuli (see Tables 3 and 4). If we assume that imitation can be a proxy of perception, as has been argued by several scholars (e.g., Gussenhoven 2004; D’Imperio et al. 2014; Zárate-Sández 2015), we tentatively conclude that bilingual participants imitated the tonal movements that are meaningful in Inuktitut (i.e., final cues). Pitch changes at the beginning

of the utterance appear to have a purely paralinguistic meaning for participants. Further anecdotal evidence of the non-linguistic meaning of pitch variations for our participants were comments gathered during the testing process. Indeed, when performing the imitation task, participants would frequently laugh after finishing imitating an utterance.

6.2. Perception and Production

Results showed some interesting links between perception and production, as well as pathways for future research. As summarized in Figure 6, for most participants (i.e., 8/13) accuracy in production was higher than accuracy in perception, particularly in the C task, which is not the tendency in L2 and bilingual research. One explanation, which would account for the behavior of this sub-group, has to do with task demands. In perception, participants had to keep the context in mind, listen to the three possible matching options, and choose one. In addition to being more demanding for participants' memory and attention, this task required them to perform something that is absent from their daily lives, as opposed to the production task that prompted them to listen to a context and produce an appropriate response. Moreover, we can hypothesize that age factors (i.e., decline in auditory capacity due to aging) may account for the performance of two participants (I07; I10) who were the oldest in the sample.

The opposite trend (i.e., perception better than production) was observed in three participants (I01, I09, I12), and of those, only I01 was highly accurate in perception (indeed, this participant was the most accurate in our sample). We would expect a better performance in perception than in production for this participant (albeit her production was highly accurate) since she has been exposed to English in school but she mostly uses Inuktitut in her daily life. The other two participants, however, have little in common with I01, with the exception of their AoA and their gender.

Finally, the remaining two participants revealed a similar behavior in perception and production (I02, I04). These participants, however, differed in their accuracy rates. Whereas I04 was on average 83% accurate, I02's accuracy average was 61%. Interesting parallels emerge if we turn to previous literature. As was the case in previous studies (Grabe et al. 2003; Ortega-Llebaria and Colantoni 2014; Radu et al. 2018), participants did not differ when responding to stimuli with no linguistic content (IO task), as compared to tasks that had access to contextual meaning. Bilingual participants in our study were also highly accurate at imitating tonal changes in nuclear contours in the SI production task. As such, they resembled L1 Spanish-L2 English participants in Ortega-Llebaria and Colantoni (2014) who matched controls better in *f0* changes when the focalized element was in object position, where pitch is used in the L1. However, as opposed to participants in Ortega-Llebaria and Colantoni (2014)'s study, who were able to imitate tonal changes in focalized subjects and verbs, bilinguals in this study were not able to imitate the pitch change in prenuclear position, which suggests that the absence of tonal changes in Inuktitut is an entrenched feature in their L1.

6.3. Individual Variability

Given the characteristics of our population, it is important to turn briefly to patterns of individual variability, some of which have been highlighted throughout this study. While English and Inuktitut have been in contact for centuries and there is a high degree of social bilingualism, our participants (Table 1) differed along all the dimensions captured in our background questionnaire. Individual differences became especially apparent in the C task, both in perception and in production. Of the two participants who showed consistent patterns in perception and production (I04 and I02), only one of them (I04) was exposed to English from birth through one of her parents (Table 1). This participant had similar self-reported patterns of language use (i.e., she uses English approximately 50% of the time), but differed from I04 in her self-rating (Advanced as opposed to Near Native). The other two participants who were exposed to English at home (I03, I05) were also among those with the highest combined accuracy rate (75%). These participants resembled I04 in

their education and patterns of language use, but, once again, differed from her in their self-rating. Finally, I01, the participant with the highest average accuracy (86%) reported using English the least (25%) and began learning English at age 6. It is interesting to observe, though, that accuracy patterns do not seem to go hand in hand with patterns of pitch change in prenuclear accents. Of all the participants mentioned, only I05 produced differences that may be considered perceptible between sentence types, since her DQs are on average 1.5 STs higher than her Ss and one semitone higher than her AQs.

7. Conclusions

Our results confirm that, in cases of language contact, and given the appropriate demographic and social conditions, any pitch pattern can be transferred, including changes in alignment (Mennen 2004; Colantoni and Gurlekian 2004), in the size of the pitch excursion (e.g., Santiago and Delais-Roussarie 2012), in the frequency and use of pitch accents (Gut 2005; Queen 2001, 2012), and in the lack of tonal movements. Evidence of CLI was observed in perception and production. First and foremost, in perception, differences were not attested in the task without linguistic information, but emerged in the other two tasks, providing evidence of reduced sensitivity to tonal variations that signal sentence types. In production, and as in previous studies (e.g., Alvord 2007; Zárate-Sández 2015), we found positional asymmetries, with CLI being most evident in prenuclear position. The nonsignificant differences in pitch change across tasks and sentence types could be attributed to the fact that, in Inuktitut, tonal movements are restricted to the end of the sentence; tonal changes throughout the utterance do not encode grammatical information, this information being encoded by a rich morphology. Admittedly, tonal variations at the beginning of the utterance, albeit a cue for sentence type (see Saindon et al. 2017a), are redundant in English, since grammatical (changes in word order, *do*-support) and tonal information (final boundary tones) provide sufficient cues. We believe, however, that it was important to begin by analyzing sentence types to establish a descriptive basis for the uses of pitch in this bilingual population. We predict that this absence of tonal variations throughout the utterance will have consequences for the perception, interpretation, and production of other grammatical structures, such as corrective focus, where tonal movements in prenuclear position play a crucial role.

Finally, this study contributes to a growing literature that has shown that early (Queen 2001, 2012; Lleó et al. 2004; Rakow and Lleó 2011) and sequential bilinguals (Colantoni et al. 2016) exhibit CLI in their prosody. We are particularly interested in expanding our knowledge of the prosody of early and sequential bilinguals whose L1 is one of the many indigenous languages spoken in the Americas, given that most studies until now have focused on Spanish bilingualism (O'Rourke 2009, 2012; Muntendam and Torreira 2016). We have shown here that the English spoken by L1 Inuktitut speakers displays signs of CLI, and that not only tonal movements but also the absence of tonal variations can be transferred in a stable language contact situation.

Author Contributions: Conceptualization, all authors; methodology, all authors; software, all authors; validation, all authors; formal analysis, all authors; investigation, all authors; resources, all authors; writing—original draft preparation, L.C.; writing—review and editing, G.K., M.P., M.R. and O.T.; visualization, L.C.; supervision, L.C.; project administration, L.C.: funding acquisition, L.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a Social Sciences and Humanities Research Council (SSHRC) grant number 890-2011-0049.

Institutional Review Board Statement: This study was approved by the University of Toronto Research Ethics Board on 19 January 2016 (Protocol number: 00025928).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are not available to the public.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- 1 Inuktitut is spoken in Nunavut and in Labrador and Quebec. The degree of bilingualism varies in these areas.
- 2 In addition to English, Inuktitut has historically been in contact with several languages. For example, in the 17th century, an Arctic pidgin developed in Labrador because of the commercial contact with Basque, Breton and French speakers (for more details, see Dorais 2010, chp. 8).
- 3 Given the extensive contact, we can also hypothesize that English has influenced Inuktitut prosody, as was the case with the lexicon and some aspects of the morphology (Dorais 2010). As stated, our goal is to understand whether the absence of tonal variation can be transferred into an intonational language. For a discussion on which aspects of sentence prosody could be subject to bidirectional influence, we refer the reader to Colantoni and Sánchez (2021).
- 4 We are aware that there is a complex interaction between language attitudes, language maintenance and actual linguistic skills (see Miller 2017; Pérez Leroux et al. 2011 for a discussion).
- 5 No distractors were included in this task, since contexts, as opposed to sentence repetitions, introduced variability into the task. Additionally, the inclusion of contexts made the task longer, and, since in addition to the tasks described, we tested the perception and production of other structures, we tried to limit the testing time.
- 6 Although pitch excursions in nuclear contours are rather large, they were not considered unnatural by the authors, since they reflect this speaker's regular intonational patterns. Moreover, these sentences were listened to by two native English speakers who found them to sound natural for Canadian English.
- 7 Models with only random effects (AIC = 320.2) and with a two-way interaction (AIC = 317) were also tested but we report here the results of the best model according to the AIC criterion (AIC = 315).
- 8 The model selected had a lower AIC value (AIC = 5959) than the base model (AIC = 6146) and a model with the three independent variables (AIC = 6024).
- 9 The model with the three-way interaction had a lower AIC value (AIC = 7676) than either the base model (AIC = 7889) or the model with the three independent variables (AIC = 7760).
- 10 The larger pitch excursions in the initial accent in Ss in the stimuli (Table 4) may also contribute to explain our participants' behaviour.

References

- Allen, Shanley. 2007. The future of Inuktitut in the face of majority languages: Bilingualism or language shift? *Applied Psycholinguistics* 28: 515–36. [CrossRef]
- Alvord, Scott. 2007. Spanish Intonation in Contact: The Case of Miami Cuban Bilinguals (Florida). Ph.D. dissertation, University of Minnesota, Minneapolis, MN, USA.
- Arnhold, Anja, Emily Elfner, and Richard Compton. Forthcoming. Inuktitut and the concept of word-level prominence. In *Word Prominence in Languages with Complex Morphology*. Edited by Ksenia Bogomolets and Harry van der Hulst. Oxford: Oxford University Press.
- Arnhold, Anja. 2014. Prosodic structure and focus realization in West Greenlandic. In *Prosodic Typology II*. Edited by Sun-Ah Jun. Oxford: Oxford University Press, pp. 216–51. [CrossRef]
- Aylward, Lynn. 2010. The role of Inuit languages in Nunavut schooling: Nunavut teachers talk about bilingual education. *Canadian Journal of Education* 33: 295–328.
- Barnes, Hilary, and Jim Michnowicz. 2015. Broad focus declaratives in Veneto-Spanish bilinguals: Peak alignment and language contact. *Studies in Hispanic and Lusophone Linguistics* 8: 35. [CrossRef]
- Bartels, Christine. 1999. *The Intonation of English Statements and Questions: A Compositional Interpretation*. New York: Garland.
- Beckman, Mary, and Gayle Ayers Elam. 1997. Guidelines for ToBI labeling. Available online: https://www.ling.ohio-state.edu/research/phonetics/E_ToBI/ (accessed on 22 February 2022).
- Best, Catherine. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 107–26.
- Best, Catherine, and Michael Tyler. 2007. Nonnative and second-language speech perception. In *Language Experience in Second Language Learning: In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray Munro. Amsterdam: John Benjamins, pp. 13–44.
- Boersma, Paul, and David Weenink. 2017. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.30. Available online: <http://www.praat.org/> (accessed on 22 July 2017).
- Cerrón Palomino, Rodolfo. 1988. *Lingüística Quechua*. Cusco: Centro de Estudios Rurales Andinos Bartolomé de Las Casas.
- Colantoni, Laua, and Jorge Gurlekian. 2004. Convergence and intonation: Historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition* 7: 107–19. [CrossRef]
- Colantoni, Laura, and Liliana Sánchez. 2021. The Role of Prosody and Morphology in the Mapping of Information Structure onto Syntax. *Languages* 6: 207. [CrossRef]

- Colantoni, Laura, Natalia Mazzaro, and Alejandro Cuza. 2016. Task related effects in the prosody of Spanish heritage speakers. In *Intonational Grammar in Ibero-Romance: Approaches across Linguistic Sub-Fields*. Edited by Meghan Armstrong, Nicholas Henriksen and María del Mar Vanrell. Amsterdam: John Benjamins, pp. 3–23.
- Cruz-Ferreira, Madalena. 1983. Non-Native Comprehension of Intonation Patterns in Portuguese and in English. Unpublished Ph.D. thesis. Available online: https://www.academia.edu/922883/Non_native_Comprehension_of_Intonation_Patterns_in_Portuguese_and_in_English (accessed on 22 February 2022).
- Dehé, Nicole. 2018. The intonation of polar questions in north American (“Heritage”) Icelandic. *Journal of Germanic Linguistics* 30: 213–59.
- Delais-Roussarie, Elizabeth, Mathieu Avanzi, and Sophie Herment, eds. 2015. *Prosody and Languages in Contact: L2 Acquisition, Attrition, Languages in Multilingual Situations*. Berlin/Heidelberg: Springer.
- D’Imperio, Mariapaola, Rossana Cavone, and Caterina Petrone. 2014. Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology* 5. [CrossRef]
- Dorais, Louis-Jacques. 2010. *The Language of the Inuit: Syntax, Semantics and Society in the Arctic*. Montreal: McGill University Press.
- Face, Timothy. 2007. The Role of Intonational Cues in the Perception of Declaratives and Absolute Interrogatives in Castilian Spanish. *Estudios de Fonética Experimental* 16: 185–225.
- Flege, James. 1995. Second-language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issue in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 229–73.
- Flege, James, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Ratee Wayland. Cambridge: Cambridge University Press, pp. 3–83. [CrossRef]
- Fortescue, Michael. 1983. Intonation contours across Inuit dialects. *Inuit Studies* 7: 113–24.
- Fortescue, Michael. 1984. *West Greenlandic*. New York: Routledge.
- Fortescue, Michael. 2017. The Eskimo-Aleut Language family. In *The Cambridge Handbook of Linguistic Typology*. Cambridge: Cambridge University Press, pp. 683–706.
- Grabe, Esther, Burton Rosner, José Garcia-Albea, and Xiaolin Zhou. 2003. Perception of English intonation by English, Spanish, and Chinese listeners. *Language and Speech* 46: 375–401. [CrossRef]
- Gunlogson, Christine. 2002. Declarative questions. In *Proceedings of Semantics and Linguistic Theory XII*. Edited by Brendan Jackson. Ithaca: Cornell University.
- Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Gut, Ulrike. 2005. Nigerian English prosody. *English World-Wide* 26: 153–77. [CrossRef]
- Hedberg, Nancy, and Juan Manuel Sosa. 2002. The Prosody of Questions in Natural Discourse. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11–13; pp. 375–78.
- Hedberg, Nancy, Juan Manuel Sosa, and Emrah Görgülü. 2017. The Meaning of Intonation in Yes-no Questions in American English: A Corpus Study. *Corpus Linguistics and Linguistic Theory* 13: 1–48. [CrossRef]
- Inuit Language Protection Act. 2008. The Legislative Assembly of Nunavut. Available online: <https://www.clo-ocol.gc.ca/en/timeline-event/the-legislative-assembly-of-nunavut-adopts-the-official-languages-act-and-the-inuit> (accessed on 22 February 2022).
- Johns, Alana. 2010. Eskimo-Aleut languages. *Language and Linguistics Compass* 4: 1041–55. [CrossRef]
- Jun, Sun-Ah, ed. 2014. *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- Kiser, Michele. 2014. Analyzing Navajo Discourse: Investigating form and Function of Intonational Units in Referential Discourse. Ph.D. dissertation, University of New Mexico, Albuquerque, NM, USA.
- Lai, Li-Fang. 2018. Intonation in Contact: Prosodic Transfer and Innovation among Yami-Mandarin Bilinguals. Ph.D. dissertation, University of Pittsburgh, Pittsburgh, PA, USA.
- Liu, Chang, and Amanda Rodríguez. 2012. Categorical Perception of Intonation Contrasts: Effects of Listeners’ Language Background. *The Journal of the Acoustical Society of America* 131: 427–33. [CrossRef] [PubMed]
- Lleó, Conxita, Martin Rakow, and Margaret Kehoe. 2004. Acquisition of language-specific pitch accent by Spanish and German monolingual and bilingual children. In *Laboratory Approaches to Spanish Phonology*. Edited by Tim Face. Amsterdam: Mouton de Gruyter, pp. 3–27.
- Marasco, Olivia. 2020. Are You Asking Me or Telling Me? Production and Perception of Yes/No Questions and Statements in L2 Spanish. Ph.D. dissertation, University of Toronto, Toronto, ON, Canada.
- Massenet, Jean Marie. 1980. Notes sur l’intonation finale en inuktitut, in Inuktitut et Langues amérindiennes du Québec. *Cahier de Linguistique* 8: 195–214. [CrossRef]
- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63. [CrossRef]
- Miller, Lauren. 2017. The relationship between language proficiency and language attitudes: Evidence from young Spanish-English bilinguals. *Spanish in Context* 14: 99–123. [CrossRef]
- Muntendam, Antje, and Francisco Torreira. 2016. Focus and prosody in Spanish and Quechua: Insights from an interactive task. In *Intonational Grammar in Ibero-Romance: Approaches across Linguistic Subfields*. Edited by Meghan Armstrong, Nicholas Henriksen and María del Mar Vanrell. Amsterdam: John Benjamins.
- Official Languages Act. 1988. Government of Canada. Available online: <https://laws-lois.justice.gc.ca/eng/acts/o-3.01/fulltext.html> (accessed on 22 February 2022).

- Ortega-Llebaria, Marta, and Laura Colantoni. 2014. L2 English intonation: Relations between form-meaning associations, access to meaning, and L1 transfer. *Studies in Second Language Acquisition* 36: 331–53. [CrossRef]
- O'Rourke, Erin. 2009. Phonetics and phonology of Cuzco Quechua declarative intonation: An instrumental analysis. *Journal of the International Phonetic Association* 39: 291–312. [CrossRef]
- O'Rourke, Erin. 2012. The realization of contrastive focus in Peruvian Spanish intonation. *Lingua* 122: 494–510. [CrossRef]
- Patience, Matthew, Olivia Marasco, Laura Colantoni, Gabrielle Klassen, Malina Radu, and Olga Tararova. 2018. Initial pitch cues in English sentence types. Paper presented at Speech Prosody 9, Poznań, Poland, June 13–16.
- Patience, Matthew, Laura Colantoni, Gabrielle Klassen, Malina Radu, and Olga Tararova. 2020. The perception and comprehension of L2 English sentence types: Cross-linguistic influence and task effects. *Gradus* 5: 71–98. [CrossRef]
- Pérez Leroux, Ana Teresa, Alejandro Cuza, and Danielle Thomas. 2011. From parental attitudes to input conditions: Spanish-English bilingual development in Toronto. In *Bilingual Youth: Spanish in English Speaking Societies*. Edited by Kim Potowski and Jason Rothman. Amsterdam: John Benjamins, pp. 149–76.
- Petrone, Caterina, and Oliver Niebuhr. 2014. On the Intonation of German Intonation Questions: The Role of the Prenuclear Region. *Language and Speech* 57: 108–46. [CrossRef] [PubMed]
- Peterson, Tyler. 2016. Mirativity as Surprise: Evidentiality, Information, and Deixis. *Journal of Psycholinguistic Research* 45: 1327–57. [CrossRef] [PubMed]
- Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, MIT, Wayne, PA, USA.
- Queen, Robin. 2001. Bilingual intonation patterns: Evidence of language change from Turkish-German bilingual children. *Language in Society* 30: 55–80. [CrossRef]
- Queen, Robin. 2012. Turkish-German bilinguals and their intonation: Triangulating evidence about contact-induced language change. *Language* 88: 791–816. [CrossRef]
- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Available online: <http://www.R-project.org/> (accessed on 12 September 2021).
- Radu, Malina, Gabrielle Klassen, Laura Colantoni, Matthew Patience, Ana Pérez Leroux, and Olga Tararova. 2018. The perception and interpretation of intonational contours among L2 English speakers. *Linguistic Approaches to Bilingualism*, 1–31. [CrossRef]
- Rakow, Martin, and Conxita Lleó. 2011. *Comparing Cues of Phrasing in German and Spanish Child Monolingual and Bilingual Acquisition*. Amsterdam: John Benjamins. [CrossRef]
- Saindon, Mathieu, Laura Cirelli, Glenn Schellenberg, Pascal van Lieshout, and Sandra Trehub. 2017a. Children's and adults' perception of questions and statements from terminal fundamental frequency contours. *Journal of the Acoustical Society of America* 141: 3123–31. [CrossRef]
- Saindon, Mathieu, Sandra Trehub, Glenn Schellenberg, and Pascal van Lieshout. 2017b. When is a question not a question for children and adults? *Language Learning and Development* 13: 274–85. [CrossRef]
- Sánchez, Liliana. 2008. Literacy and the expression of social identity in a dominant language: A description of mi familia by Quechua-Spanish bilingual children. In *Bilingualism and Identity*. Edited by Mercedes Niño-Murcia and Jason Rothman. Amsterdam and Philadelphia: John Benjamins, pp. 109–26.
- Santiago, Fabián, and Elisabeth Delais-Roussarie. 2012. Acquiring phrasing and intonation in French as a second Language: The case of Yes-No questions produced by Mexican Spanish Learners. Paper presented at Speech Prosody, Shanghai, China, May 22–25.
- Shokeir, Vanessa. 2009. *Intonation in Inuktitut*. Toronto: Ms. University of Toronto.
- Sicoli, Mark, Stivers Tanya, N. Enfield, and Stephen Levinson. 2015. Marked Initial Pitch in Questions Signals Marked Communicative Function. *Language and Speech* 58: 204–23. [CrossRef]
- Simonet, Miquel. 2011. Intonational convergence in language contact: Utterance-final F0 contours in Catalan-Spanish early bilinguals. *Journal of the International Phonetic Association* 41: 157–84. [CrossRef]
- Smith, Lawrence. 1977. *Some Aspects of Labrador Inuttut (Eskimo): A Survey of the Inflectional Paradigms of Nouns and Verbs*. Ottawa: University of Ottawa Press.
- Spalding, Alex. 1979. *Learning to Speak Inuktitut: A Grammar of North Baffin Dialects*. London: Centre for Research and Teaching of Canadian Native Languages, London: University of Western Ontario.
- Statistics Canada. 2019. Evolution of the Language Situation in Nunavut: 2001–2016. Release date: 9 July 2019. Available online: <https://www150.statcan.gc.ca/n1/pub/89-657-x/89-657-x2019010-eng.htm> (accessed on 13 January 2021).
- Thalbitzer, William. 1904. *A Phonetical Study of the Eskimo Language*. Copenhagen: MoG. København, vol. 31.
- Truckenbrodt, Hubert. 2011. On rises and falls in interrogatives. Paper presented at IDP 2009, Paris, France, Septembre 9–11; pp. 9–17.
- Tsimpli, Ianthi. 2014. Aspects of meaning in bilingual minds. Paper presented at the Jackman Humanities Institute Lecture Series, Toronto, ON, Canada, February 7.
- Zárate-Sánchez, Germán. 2015. Perception and Production of Intonation among English-Spanish Bilingual Speakers at Different Proficiency Levels. Ph.D. dissertation, Georgetown University, Washington, DC, USA.

Article

Prosodic and Segmental Aspects of Pronunciation Training and Their Effects on L2

Silvia Dahmen ^{1,*}, Martine Grice ² and Simon Roessig ³¹ Herder-Institut, Universität Leipzig, 4275 Leipzig, Germany² IfL-Phonetik, University of Cologne, Albertus-Magnus-Platz, 50923 Köln, Germany³ Department of Linguistics, Cornell University, Ithaca, NY 14850, USA* Correspondence: silvia.dahmen@uni-leipzig.de

Abstract: Some studies on training effects of pronunciation instruction have claimed that the training of prosodic features has effects at the segmental level and that the training of segmental features has effects at the prosodic level, with greater effects reported when prosody is the main focus of training. This paper revisits this claim by looking at the effects of pronunciation training on Italian learners of German. In a pre-post-test design, we investigate acoustic changes after training in learners' productions of two features regarded as prosodic and two features regarded as segmental. The prosodic features were the pitch excursion of final rises in yes–no questions and the reduction in schwa epenthesis in word-final closed syllables. The segmental features were final devoicing and voice onset time (VOT) in plosives. We discuss the results for three groups (with segmental training, with prosody training, and with no pronunciation training). Our results indicate that there are positive effects of prosody-oriented training on the production of segments, especially when training focuses on syllable structure and prosodic prominence (stress and accent). They also indicate that teaching segmental and prosodic aspects of pronunciation together is beneficial.

Keywords: second-language learning; second-language acquisition; second-language teaching; pronunciation instruction; prosodic training; production; intonation; syllable structure; final devoicing; epenthetic schwa

Citation: Dahmen, Silvia, Martine Grice, and Simon Roessig. 2023.

Prosodic and Segmental Aspects of Pronunciation Training and Their Effects on L2. *Languages* 8: 74.

<https://doi.org/10.3390/languages8010074>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 22 March 2022

Revised: 21 February 2023

Accepted: 21 February 2023

Published: 6 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Phonetic-phonological competence of L2 learners is commonly assessed by categories such as (foreign) accentedness, intelligibility and comprehensibility, for example in the Common European Framework of Reference for Languages (CEFR, but see also [Derwing and Munro 1997](#); [Thomson 2017](#)). The CEFR states that the goal of pronunciation instruction is not to achieve a native-like pronunciation but rather to speak in a way that does not impair communication ([Council of Europe 2020](#); [Chun and Levis 2020](#)). This implies that while a learner's utterance can be heavily influenced by their first language (foreign accent), it may still be easily understood by native speakers ([Derwing and Munro 2015](#), p. 5), so the more important aspects of pronunciation for successful communication are that the listener can identify what has been said and the message the speaker intends to communicate (intelligibility) without investing excessive effort into the process of understanding (comprehensibility). Studies on native-speaker perception of L2 speech have indicated since the 1980s that prosodic features play an important role in comprehensibility and intelligibility and that teaching prosodic aspects leads to improvements in both prosodic and segmental features of pronunciation, while the converse has not been shown for segmental training ([Anderson-Hsieh et al. 1992](#); [Munro and Derwing 1995](#); [Derwing et al. 1998](#); [Gordon and Darcy 2016](#)). Nonetheless, Derwing and Munro argue that the findings of such studies do not imply that only prosodic features should be taught ([Derwing and Munro 2015](#), p. 9), as segmental errors can also lead to misinterpretations of utterances and add to the perception of foreign accent. However, these claims are probably true only for target

languages such as English and German. Recent studies have shown different patterns for the effects of segmental and prosodic influence on the strength of a perceived foreign accent and comprehensibility when the target language is a tone language (Yang et al. 2021) or when the native language of the listeners/raters is different from the target language (Kaunzner 2015, 2018). Yang et al. (2021) examined the effects of prosodic and segmental deviations in L2 utterances in Mandarin Chinese and found that native Chinese listeners' ratings of foreign accent and comprehensibility were influenced by segmental rather than prosodic correctness. Kaunzner (2015, 2018) compared comprehensibility ratings for L2 German utterances of Italian learners for native German, Polish, and Italian listeners/raters and found that only the German listeners rated utterances with prosodic deviations as less comprehensible than utterances with segmental deviations, while the Polish and Italian listeners were instead influenced by segmental deviations. In addition, more-recent findings (e.g., Ulbrich and Mennen 2016; van Maastricht et al. 2021) have indicated that there is a strong interplay between segmental and prosodic features when native listeners rate speech for intelligibility, comprehensibility, and degree of perceived foreign accentedness, where some prosodic features affect native ratings more than others. Research involving English speech manipulated such that native prosody was mixed with non-native segments and vice versa revealed that native listeners' ratings of foreign accentedness depended on both segmental and prosodic deviances and that the impact of prosody depended on the nativeness of the segments: non-native prosody on native segments led to the perception of a weaker foreign accent than on non-native segments, and native prosody on non-native segments led to a stronger perception of foreign accent than on native segments (Ulbrich and Mennen 2016). In a study involving native listener judgements of Spanish learners' L2 Dutch utterances, speech data were manipulated such that a combination of rhythmic or intonational patterns or the speech rate of L1 Dutch speakers was transferred to original learners' utterances. The results showed a stronger influence of intonation on perceived foreign accentedness and comprehensibility when it was the only native feature transferred, while a syllable-timed rhythm (as in Spanish) and a slow speech rate had no such effects (van Maastricht et al. 2021). Thus, the question whether and to what extent it is prosodic or segmental features that mostly affect comprehensibility and perceived foreign accentedness is not as clear as previous research has indicated.

While there is a large number of publications on the general effectiveness of pronunciation instruction (see Saito and Plonsky (2019) for a discussion on intervention studies conducted until 2017), only a few studies have examined the effects of prosodic training on L2 production of segmental features and of segmental training on L2 production of prosodic features. Among these, Missaglia (1999a) found that Italian learners' production of German vowels improved more for a group that received training focused on prosody than for a group that received segmental instruction. While the segmental training consisted of a common set of discrimination and production tasks for German vowels, mixed with articulation exercises, she used the contrastive prosody method (CPM) for her prosodic training. In this method, learners are first made aware of their native language features, such as the rules for sentence-stress or word-stress placement, and of the phonetic features used to mark prominence. This awareness enables them to detect the differences between their L1 production and that of native speakers of the target language and to adapt their production accordingly. The basic assumption behind the method is that in order to know how to produce L2 features, learners need to know explicitly what the corresponding features are in their L1 and what they have to change to correctly produce an utterance in the target language. Learners are treated as bilinguals who are able to make use of their L1 competence in order to improve their L2 productions (Missaglia 1999b, 2007). Common tasks within the CPM are comparing utterances of native speakers to the same utterances produced by L2 speakers and describing the differences, or deliberately producing utterances in the target language with prosodic features of L2 speakers and then changing those features to approximate L1 production. Missaglia's CPM training included stress placement and intonation, including how to produce deaccentuation. Since the CPM

training also included the effects of deaccentuation on the phonetic realisation of vowels, it is unsurprising that vowel production improved for the group receiving this training. The distinction between prosody and segments is difficult to uphold here in that both stress and accentuation have cues that are linked to the production of segments.

Li et al. (2022) examined training effects of embodied prosodic training (involving hand gestures) on the pronunciation skills of Catalan learners of French. They found that embodied prosodic training has positive effects not only on perceived foreign accentedness ratings but also on F2 values of front rounded vowels.

In a larger-scale study on Italian learners of German, Dahmen (2013) compared the results of segmental training (including vowel length, VOT for plosives, and final obstruent devoicing) to those of prosodic training (including intonational focus marking, rhythmic syllable reduction, and syllable structure) for two training groups and a control group of L2 German learners from Northern Italy. Both trainings were based on a method described by Dieling and Hirschfeld (2000), which includes perception and production tasks. For the perception, learners are usually first introduced to a phonetic or phonological feature by listening to utterances that focus on the respective feature. An introductory task for the length contrast in German vowels, for example, could be listening to a story about animals at the zoo, where the teacher first names only those animals whose names contain stressed long vowels and then animals whose names contain stressed short vowels. The learners would not be expected to know all the words, but they should be able to say that there are differences in the vowels between the two sets of names. Further listening tasks include discrimination of contrasting sounds, stress patterns or intonation contours, using minimal pairs and identification exercises in which the learners are presented with speech stimuli and have to signal which of the stimuli contain a certain sound, stress pattern, or intonation contour. Other identification tasks involve detecting rules such as final obstruent devoicing.

For the production part, simple listen-and-repeat exercises are combined with articulation exercises and with tasks involving hand gestures or other visual support. Further production exercises progress from simple repetition to free production. The comparison of all three groups in the study showed that both training groups improved on both the segmental and the prosodic levels but that the group receiving the prosody training improved in more aspects than the group with segmental training. Training effects were assessed for VOT in alveolar plosives /t/ and /d/, final obstruent devoicing, and the quantity and quality of German long versus short vowels for the group that received training labelled as 'segmental', and for rhythmic reduction in unstressed syllables, syllable structure (the realisation of word-final codas and avoidance of epenthetic vowels), and prosodic marking of corrective focus for the group that received training labelled as 'prosodic'. During the study, other aspects of L2 German were also trained, namely the intonation of yes–no questions and answers, as well as stress and accent (word and sentence stress) for the so-called prosody group and the pronunciation of German r-sounds as well as /h/ versus the glottal stop in syllable onsets for the so-called segment group. The training effects in these areas were not assessed.

In this paper, we revisit some of the data collected during the training project that was the basis for Dahmen (2013), using state-of-the-art statistical analyses and making the results more accessible by presenting them in English. We also revisit the terms 'prosodic' and 'segmental', since many features are traditionally assigned to one of the two categories, although they have effects on both. We report in detail on two features that were assigned to the prosodic level and two that were assigned to the segmental level in (Dahmen 2013). The two 'prosodic' categories are the intonation of yes–no questions¹ (cf Section 3) and syllable structure, more specifically the production of epenthetic schwa after word-final consonants (cf Section 4). The 'segmental' categories are final obstruent devoicing (cf Section 5) and VOT in fortis plosives (cf Section 6).

These four features all contribute considerably to the intelligibility, and ultimately to the comprehensibility, of L2 speech. Intonation is crucial for signalling sentence modality because, even in German, questions can often be fragments that are not necessarily syn-

tactically marked as interrogative. The production of epenthetic schwa can lead to the perception of an extra syllable, which in turn can be interpreted as a suffix (such as the plural form in nouns), thus leading to problems at the grammatical level. Although the absence of final obstruent devoicing does not in itself create lexical confusions, the voiced consonant may be followed by epenthesis, leading to the same problem, that of being interpreted as an extra syllable. VOT, especially a lack of aspiration, can lead to lexical confusions, especially if these are in stressed syllables, where the aspiration in German is enhanced. Although language is highly redundant and minimal pairs can often be distinguished by virtue of the context in which they occur, intelligibility and comprehensibility are improved if the listener does not have to deal with conflicting information from the context and the pronunciation. These considerations were the motivation for investigating the effects of training on these four aspects of pronunciation.

These four features also provide clear evidence of the difficulty in upholding the prosodic–segmental dichotomy. For example, even in an aspect of pronunciation that could be regarded as clearly prosodic, i.e., the intonation of yes–no questions, a rise or complex pitch movement can lead to schwa epenthesis or the lengthening of a vowel, both of which are usually treated as segmental (Grice et al. 2015, see discussion in Section 4 below). This is referred to as tune–text interaction, indicating that the intonation and the segmental structure cannot be treated separately. A clearer case in our investigated features is the pronunciation of word-final consonants. This is not only segmental but also prosodic. This is because obstruent devoicing is related to syllable structure: an error in syllable structure, e.g., the epenthesis of schwa in *Rad* ‘bike’ [rad.də], leads to a possible resyllabification, in addition to other adjustments, such as the lengthening of the plosive (transcribed as a geminate) and possibly the shortening of the vowel. This resyllabification runs the risk of removing the (syllable final) context for the devoicing of <d> to apply. Voice onset time is not purely segmental either: it depends on the temporal coordination of laryngeal and supralaryngeal gestures, and it interacts with syllable prominence, such that the strength of plosive aspiration depends on whether the syllable is lexically stressed or accented (e.g., Lisker and Abramson 1967; Jessen and Ringen 2002; Savino et al. 2015; Lein et al. 2016).

Given these interactions, our research question is concerned with how far each of these features of L2 speech can improve with targeted explicit training. Specifically: (1) How successful is training in intonation and syllable structure (suppressing epenthesis) and does it affect the production of individual consonants? and (2) How successful is training in final devoicing and VOT of voiceless plosives and does this training affect the production of syllable structure and intonation?

2. Materials and Methods

2.1. Subjects and Recordings

The data were recorded during a training project in Germany, one day *before* and one day *after* each training phase. The recordings were conducted in a quiet room using a mobile DAT recorder and head-mounted microphones. The trainings took place in Bischofswerda (Saxonia) as part of a training camp for students from all over Italy who were preparing to take part in the German language diploma (*Deutsches Sprachdiplom der Kultusministerkonferenz*) for the level B2/C1 of the Common European Framework of Reference. The training camp consisted of two phases of 10 days each, in which different groups of students took part in the courses. In the following, we give details on the speakers in the groups.

In the first phase of the training project, students attended courses on reading and listening comprehension as well as on oral and written communication. During the first phase, 8 students (3 male, 5 female) from one school class in Turin were recorded. They were 17 or 18 years old at the time of the recordings and had learned German for 3.5 to 7 years. They reported no German relatives or friends and thus used German only in the classroom. They did not receive any pronunciation training during the duration of the project. Therefore, this group is the *control group* in the present study.

In the second phase of the training project, the reading and listening comprehension group was split in two subgroups, which took turns attending reading/listening comprehension and pronunciation training. Students recorded were from Montagnana and Turin. The groups undertook training in what was referred to as either segmental or prosodic aspects of pronunciation. The groups are heretofore referred to as the *segment group* and the *prosody group*, respectively. The segment group consisted of 13 subjects altogether, 7 from Turin (2 male, 5 female) and 6 from Montagnana (1 male, 5 female). The prosody group consisted of 12 subjects, 6 from Turin (all female) and 6 from Montagnana (1 male, 5 female). All subjects in the test groups were between 17 and 19 years old, had learned German for 4 to 5 years, and used German only in the classroom at the time of the recordings. More information about the training is given in the next section.

The students were randomly assigned to the training groups. The metadata of the students do not indicate any systematic differences in pronunciation competence between groups. Differences between the groups before training are most likely due to individual factors not controlled for in this study. The analysis presented here concentrates on differences between the time point before and the time point after training rather than absolute differences between groups.

2.2. Speech Materials

The speech materials presented in this article consist of read sentences as well as semi-spontaneous utterances. The semi-spontaneous utterances were yes–no questions (cf Section 3) elicited in specially designed card games. We first give an overview of the read sentences and explain the card games below. The following sentences were used in the study:

- (1) *Dina gab Elmar ein neues Rad.* ('Dina gave a new bike to Elmar')
- (2) *In der gelben Hütte lebte ein großer Hund.* ('In the yellow hut lived a big dog')
- (3) *Tina gab Hanna einen guten Rat.* ('Tina gave good advice to Hanna')
- (4) *Die billigen Hüte waren ganz schön bunt.* ('The cheap hats were pretty colourful')
- (5) *Helga spielte einmal Tennis.* ('Helga once played tennis')

The sentences were presented to the students in random order to reduce the chance of their identifying the minimal pairs. For the occurrence of word-final epenthetic vowels (cf Section 4), we examined the target words *Rad*, *Hund*, *Rat*, and *bunt* (sentences 1 to 4). *Rad* and *Rat* (sentences 1 and 3) were the target words for measuring final obstruent devoicing (cf Section 5). For VOT (cf Section 6), we looked at *Tina* and *Tennis* (sentences 1 and 5).

The card games were played in pairs. The cards in this game depicted day-to-day objects in different colours. The participants had the task of collecting cards with the same colour or the same object by exchanging cards with their fellow player. To initiate the exchange, participants formulated a yes–no question, e.g., *hast du einen gelben Teller?* (English: 'do you have a yellow plate?'). This question was followed by the answer, and if desired, the card was exchanged.

The materials used in the analyses of the different phenomena will be described in the respective subsections to make them more accessible to the reader for the interpretation of the results.

2.3. Training

During the training phases, the control group received 90 min of reading and listening comprehension training per day. This course was taught by the same teacher as the pronunciation training classes to rule out a teacher effect. The test groups received 45 min of pronunciation training per day. Each pronunciation training session contained perception and production exercises for the respective segmental or prosodic areas, usually with one or two new phenomena introduced in each session and then repeated in the following sessions. For instance, the segment group engaged in discrimination and production exercises for long versus short vowels and for aspirated versus unaspirated plosives in the first session, and then in the second session, they engaged in production exercises for both and for a

first introduction to final obstruent devoicing. The prosody group received training on sentence intonation, nuclear accent placement (sentence stress) and focus marking, word stress, rhythm (reduction in unstressed syllables), and syllable structure. The segment group received training in aspirated plosives, final obstruent devoicing, the long-short and tense/lax distinction in German vowels, consonantal and vocalised realisations of <r>, the fricative allophones [ç] and [x] of orthographic <ch>, word-initial /h/ versus glottal stop, and front rounded vowels. The students were asked not to exchange pronunciation exercises between the groups, and their teachers reported when they did. For that reason, two subjects that had originally been recorded had to be excluded from the study. These speakers were not included in the study (they are thus not part of the speaker sample described in the previous subsection). The training sessions for the areas relevant to the present study are briefly described below.

2.3.1. Intonation of Yes–No Questions (for Results, cf Section 3)

Only the prosody group received training in the intonation of yes–no questions. To make the participants aware of the high final rise in German yes–no questions, the teacher wrote questions such as *ist das ein Tisch? hast du ein Buch? kennst du München?* ('is this a table?', 'do you have a book?', 'do you know Munich?') on a board and drew lines over the sentences to indicate at which point and to which extent the intonation contour rose while the participants listened to the questions and identified the rise in pitch and in the line drawn over the sentence. Next, other questions of the same type were presented in oral and written form, and the students drew their own lines to represent the intonation contours they perceived. The point at which the contour starts rising in German (i.e., the accented syllable) was identified by the group, and a rule was formulated. Again, yes–no questions were used to apply the rule (task: find the syllable where the rise starts). This task was combined with oral production exercises and with hand gestures that imitated the rising pitch contours. The use of hand gestures in combination with oral output has been found to enhance L2 production of both segmental and prosodic features (e.g., Baills et al. (2022); Li et al. (2020)). Other production tasks included dialogues of the form *hast du [Objekt]?* ('do you have [object]?')–*ja/nein* ('yes/no'), where each participant asked others for a matching object on a card, knowing that there were pairs of identical cards. Similar tasks had one participant at a time choose an object from a set of possible objects (e.g., an orange, a banana, a book, a newspaper etc.), the others asking questions such as *kann man es essen? ist es gelb?* ('can you eat it?', 'is it yellow?') to find out which object the candidate had chosen. Hand gestures were used during production throughout the training phase.

2.3.2. Avoiding Word-Final Epenthetic Vowels (for Results cf Section 4)

The first step in the training of participants of the prosody group was to make them aware that they had produced epenthetic vowels after words ending in consonants, e.g., *Tisch, Stuhl, Blatt* ('table, chair, leaf'). Recordings of participants were played, and all cases of epenthetic schwa were pointed out by the teacher. As word-final schwa is a very common grammatical marker in German (orthographically represented by <-e>), word pairs such as *Tisch–Tische* ('table–tables') were presented as auditive stimuli to make the participants aware that epenthetic schwa can lead to the perception of unintended grammatical forms by German native listeners. In order to avoid word-final schwa epenthesis, participants were asked to produce words ending in fricatives, e.g., *Tisch*, and lengthen the final consonant for as long as they could, in order to prevent the reflex of adding a vowel. Subsequently, the final consonant was shortened (where the teacher indicated via a hand gesture when to stop producing the consonant, thus indicating the duration of the sound) until a normal duration was reached. For word-final plosives, as in *Blatt*, participants were asked to lengthen the aspiration of the plosive, first driving small balls of paper over a table with the force of the aspiration and then shortening it until the appropriate duration was achieved. In following sessions, words with more-complex codas were used for similar tasks, e.g., *eins, einst, Herz, Herbst* ('one, once, heart, fall'). In these tasks, the participants had to 'build up'

the words sound by sound in order to carefully pronounce all consonants in the complex codas. Another productive exercise included the oral production of the above-named word pairs of the type *Tisch–Tische*, with a special focus on the different pronunciations of each member of a word pair.

2.3.3. Final Obstruent Devoicing (for Results of Section 5)

In order to be made aware of the rule of final obstruent devoicing in German, the segment group was first presented with orthographic stimuli, focusing on the graphemes <b, d, g>. For example, in the sentence *Sabine ist sehr hübsch und lieb* ('Sabine is very pretty and kind'), they were asked to first find all graphemes and then listen to a recording of the sentence and mark all instances of being pronounced as [p]. The same procedure was carried out for other sentences, including words with <b,d,g> in the onset and coda positions. After this identification exercise, the rule for final obstruent devoicing was formulated in written form and then applied to other words, e.g., *Korb* ('basket'), *Land* ('country'), and *Tag* ('day'). In the next step, the graphemes <s> and <v> were treated in the same fashion. As a productive exercise, singular and plural forms of nouns ending in <b, d, g, s, v> were pronounced by the participants, focusing on the change in pronunciation of these graphemes when they change their position within syllables. For instance, in *Tag*, <g> is pronounced [k], but in the plural *Tage*, it is pronounced [g]. For word-final plosives <b, d, g>, participants held a sheet of paper before their mouths and produced aspiration strong enough to move the paper. For word-final fricatives <s, v>, they put a finger on their larynxes to feel whether their vocal folds were vibrating for words such as *Haus* ('house'), where there should be no vibration during the final consonant, versus *Häuser* ('houses'), where there should be.

2.3.4. Voice Onset Time (for Results of Section 6)

The segment group was first presented with written words present in German and Italian (and English), namely *Pizza* and *Taxi*. Participants were asked to pronounce the words in their Italian form, then the teacher pronounced them in the German way, with aspirated plosives. After thus making the participants aware of the difference in the production of plosives in German and Italian, the next step was a discrimination task with minimal pairs, such as *Pass–Bass* ('passport–bass'), *Tank–Dank* ('tank–thanks'), or *Karten–Garten* ('cards–garden'), where they indicated which of the words of a word pair they had heard. The term 'aspiration' was introduced, and the different use of voicing versus aspiration in Italian and German was explained. The need for the aspiration of fortis plosives in German was explained by the fact that unaspirated [t], for example, can be perceived as [d] by German listeners, which might result in misunderstandings. In order to obtain a strong aspiration, the participants were asked to hold a sheet of paper in front of their mouths and make it move by producing a puff of air after the release of the plosives. This was repeated for a great number of German words with initial [t^h, p^h, k^h]. Additionally, a card game was played during which the participants had to find words with matching initial sounds written on cards. For example, the words *Pass* and *Polizei* ('police') would be a match, but *Pass* and *Bass* would not be. In order to receive the cards of a matching pair, the participants had to pronounce the words loudly, and the other players decided whether aspiration was produced in the correct places.

2.4. Overview of Groups and Training

To provide a better overview of the methodology used in this study, Table 1 lists all speaker groups with their origin and a summary of the training they received.

Table 1. Overview of training groups.

Group	Number of Speakers	City/Region	Training
Control group	8 (3 male, 5 female)	Turin (all)	No pronunciation training
Segment group	13 (3 male, 10 female)	Montagnana (6 speakers) Turin (7 speakers)	Final obstruent devoicing; aspiration of fortis plosives
Prosody group	12 (1 male, 11 female)	Montagnana (6 speakers) Turin (6 speakers)	Intonation of yes–no questions; epenthetic vowels

2.5. Statistical Analyses

The data were statistically modelled with Bayesian mixed models. For tutorial introductions of Bayesian statistics with phonetic data, see [Vasishth et al. \(2018\)](#), [Roettger and Franke \(2019\)](#), and [Nalborczyk et al. \(2019\)](#). Bayesian statistics were carried out because they are known to provide reliable results, even for small samples ([van de Schoot et al. 2015](#)). The models were fit with `brms` 2.16.3 ([Bürkner 2018](#)) in R 4.1.2 ([R Core Team 2021](#)). The package `brms` ('Bayesian regression modelling with Stan') implements an interface to Stan to compute Bayesian models via Markov chain Monte Carlo (MCMC) sampling ([Carpenter et al. 2017](#)). All models were checked for convergence by ensuring that they did not exhibit *Rhat* values larger than 1.00. The model fit was visually inspected by using predictive posterior check plots. To assess the training effects, we examined the differences between the posterior distributions before and after training by employing the hypothesis function of the `brms` package. Throughout the analysis, we used `tidyverse` 1.3.1 for data processing ([Wickham et al. 2019](#)). For plotting, we used `ggplot2` 3.3.5 ([Wickham 2016](#)).

3. Training Effects on Magnitude of Question Rises

In this section, we examine the effects of training on the final rise in yes–no questions. Both German and Italian commonly have final rises in such questions. Refer to Appendix A for an overview of the native patterns in the two languages. In this comparison, it becomes evident that final yes–no question rises in Italian are *smaller* in magnitude than those in German.

Moreover, we ask whether the magnitude of the rise produced by Italian learners of German is similar to their L1, i.e., whether learners exhibit smaller rise magnitudes in their L2 because of influences from their L1 before training. We can investigate how this element of their L2 changes through training and whether the three training groups, namely control, segment, and prosody, show different training outcomes with respect to the question rise. The reader is reminded that only the prosody group received explicit training on question intonation (see Section 2.4).

3.1. Data

The data analysed here were elicited with a card game specifically designed for this task. The players ask for cards with specific colour-object combinations (*do you have a blue coffee pot?* German: *hast du eine blaue Kanne?*). Each player has a tableau in front of them depicting specific colour-object combinations in two rows of eight numbered positions. In addition, each player has a stack of cards designating positions 1 to 8. At the beginning of one move, a player draws a position card (e.g., position 3) and looks up the colour-object combination in this position in the upper row of the tableau (e.g., green plate). The player then formulates a question for this specific colour-object combination, e.g., 'in position 3, do you have a green plate?' (German: *in Position 3, hast du einen grünen Teller?*). The other player looks up the position in the lower row of their tableau and produces an answer. The answer can be 'yes' (German: *ja*) or 'no, I have <alternative>', where <alternative> stands for a different colour-object combination, e.g., *no, I have a green ball* (German: *nein, ich habe eine grüne Kugel*). The colour adjectives were *blaue/blauen* 'blue', *gelbe/gelben* 'yellow', *graue/grauen* 'grey' and *grüne/grünen* 'green'. The object nouns were *Kanne* 'coffee pot', *Teller* 'plate', *Gabel* 'fork', and *Kugel* 'ball'.

In total, 317 recordings entered the analysis. Of these recordings, eight were excluded because the questions lacked a final rising movement. As a result, the magnitude of 309 final question rises could be assessed. An example contour of one question is given in Figure 1C. This instance is taken from the recordings before training.

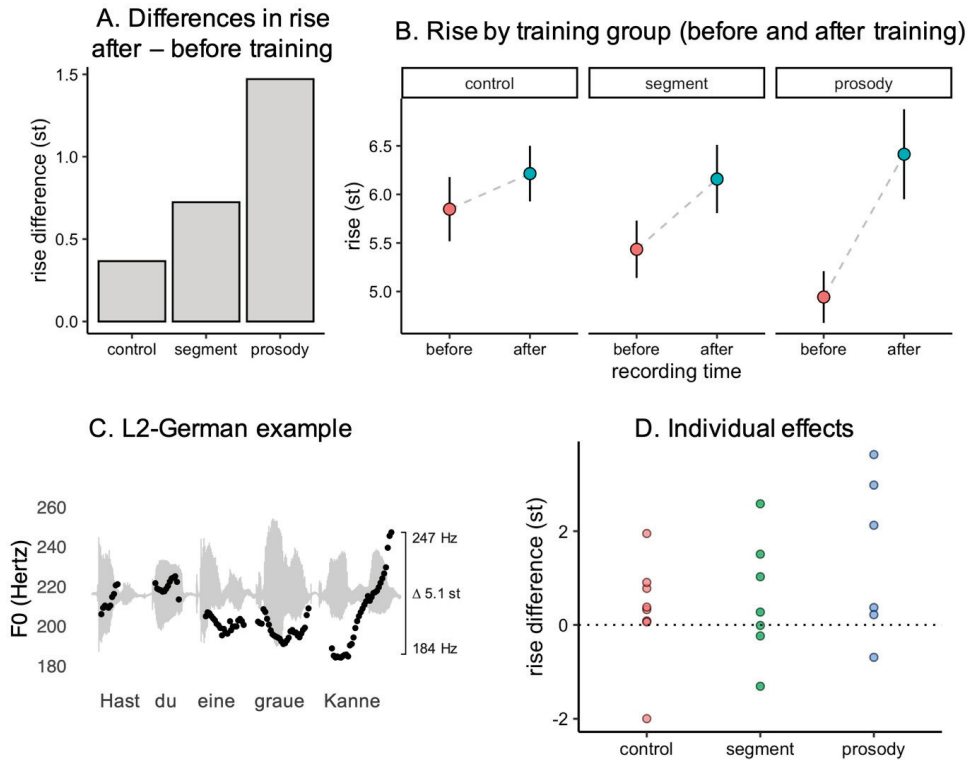


Figure 1. Differences in means before and after training (A), means and SE before and after training (B), example contour from one L2 speaker (C), separate differences in rise for speakers (D), where each dot corresponds to one speaker.

3.2. Analysis and Results

Table 2 gives the means and standard deviations of the final rise for the three training groups before and after training. In addition, the last column represents the difference between the mean before the training and the mean after the training.

Table 2. Results for the final rise of the three training groups in semitones (st).

Training	Time Point	Final Rise Mean (st)	Final Rise SD (st)	Difference (st) (Mean After–Mean Before)
Control	Before	5.85	2.17	0.37
	After	6.21	2.29	
Segment	Before	5.43	2.05	0.72
	After	6.16	2.63	
Prosody	Before	4.94	1.87	1.47
	After	6.41	3.25	

The results are illustrated in Figure 1A,B. Panel A shows the differences in means before and after training (mean before minus mean after). First, the differences in all groups are positive. This means that all groups adjust their final question rises to make them larger. The largest change is obtained by the prosody group, the smallest change by the control group. The segment group is situated in between these two poles. Panel B shows the means with standard errors before and after training. The slope of the dashed line illustrates the change within each of the groups between the two recording time points. In addition, it can be observed in this plot that the prosody group is not only the group with the largest improvement after training but also the group that exhibits the lowest values before training.

The statistical model used rise magnitude as the dependent variable. The fixed effects were time of recording (before or after training) and training type (control, segment, prosody), as well as the interaction between time of recording and training type. The model included random intercepts for speakers and by-speaker random slopes for the effect of recording time. In addition, the model used random intercepts for the nouns that the rise is realised on (e.g., *Teller, Kanne, . . .*).

We used a normally distributed prior probability distribution (prior) with a mean of 0.0 and a standard deviation of 1.0 for the regression coefficients. All the other priors were the default priors of brms. As priors for the intercept, we used a Student’s *t* distribution with degrees of freedom of 3.0, a median of the data as a mean of the distribution and a standard deviation of 2.5 ($\nu = 3.0, \mu = \text{median of the variable}, \sigma = 2.5$). As priors of the standard deviations of the random intercepts and slopes and as the residual standard deviation of the model, we used a Student’s *t* distribution ($\nu = 3.0, \mu = 0, \sigma = 2.5$). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions ($\eta = 1$). MCMC chains were run for 7000 iterations, with 3500 warmup iterations at four chains, resulting in a total of 14,000 posterior samples used for inference.

We are interested in the differences in posterior distributions between the recording time points (before vs. after) in each group to assess the evidence for an improvement in the groups. Therefore, we calculated the posterior distribution of the differences before and after training (after minus before). We report the estimated difference β , the standard error of the estimate (SE), the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is positive $Pr(\beta > 0)$. The parameter β indicates how large the model estimates the difference in rise magnitude between the two recording time points. $Pr(\beta > 0)$ indicates how certain we can be that the difference between before and after training is indeed positive, i.e., that the rise indeed became larger during training. Table 3 presents the results of the statistical model. The table shows that the estimate of the differences is largest in the prosody group (1.20). The 90% CI does not include zero, and $Pr(\beta > 0)$ is 0.99. Given the model and the data, we can conclude that this constitutes strong evidence for an increase in the rise magnitude from before training to after training. The other two groups also yield positive estimated differences, where the estimate for the segment group is larger. However, for both training groups, the 90% CI includes zero, and $Pr(\beta > 0)$ is only 0.84. Hence, given the model and the data, the evidence for a positive difference (or an increase in rise magnitude during training) is much weaker.

Table 3. Results of the Bayesian mixed model regarding the difference in rise magnitude between recording time points (after training minus before training) in the three groups.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	$Pr(\beta > 0)$
Control	0.41	0.41	−0.26	1.07	0.84
Segment	0.47	0.49	−0.33	1.27	0.84
Prosody	1.20	0.52	0.35	2.04	0.99

An interesting question in the context of the training effects on the final rise magnitude is whether all subjects behave in a uniform way. Figure 1D gives insights into the development of the individual language learners in the groups. In this plot, each dot corresponds to one subject. The *y*-axis shows the differences between the recording time points before and after training (after minus before), just like Figure 1A for the whole group. It can be observed that there is indeed a considerable amount of variation among the individuals. While most subjects show a positive difference, i.e., a larger rise after training, a minority of subjects exhibit the reverse pattern or a difference close to zero. This is particularly true for the segment training group. In addition, we can see that the training groups overlap to a certain extent: not all individuals in the prosody group yield larger rise differences than all individuals in the segment or control group. However, in the prosody group, there are some speakers who yield much larger differences, and the only speaker who reverses the pattern is close to zero.

In addition, there are differences between the groups before training. In Figure 1B, we observe smaller rise magnitudes for the prosody group and the segment group compared with the control group at the recording time point before training. As outlined in the methods section, however, the metadata of the students do not indicate any systematic differences between the groups. It is also beyond the scope of this paper to assess whether the magnitude of the improvement during the training is causally linked to the base level before training.

3.3. Interim Summary and Discussion

In this section, we analysed the rise magnitude of yes–no question rises, and how it develops in the three training groups under discussion. In all training groups, we see some kind of increase in rise magnitude after training. Our analysis has demonstrated that these differences are largest for the prosody group, and our statistical modelling has provided strong evidence for a positive change in only this group. We have also shown, in addition to the general trend of an increase in the rise magnitude and the group differences, that there is considerable individual variation.

In Appendix A, we compare similar questions produced by German and Italian native speakers in their L1s. These results show that Italian L1 yes–no questions exhibit considerably smaller rise magnitudes than their German counterparts. The learners' results presented in this section seem to range in between the two extremes, with a tendency towards the German realisation pattern after training in the prosody group.

An interesting point to consider is whether the observed rise magnitudes can be explained by a phonetic or phonological transfer effect from the L1 to the L2 (Mennen 2007). At first glance, it may appear to be a clear phonetic effect. Both languages have a rising question intonation that can be described as a combination of low accent L* followed by a high or rising boundary tone. The phonetic implementation of the height of the final tonal target appears to differ across the languages, and Italian learners of German may transfer their phonetic knowledge about the final rise to their L2 German. However, we gain a different perspective from a closer look at the phonological descriptions of intonation contours in both languages. In German, a typical nuclear yes–no question contour is one that is best described as L* H-ˆH%, with an H intermediate phrase boundary tone and an upstepped ˆH% intonation phrase boundary tone (Grice and Baumann 2002). This contour is characterised by the rise towards an extra-high final pitch. Given enough syllables between the L* and the end of the phrase, a plateau occurs. The L* H-ˆH% contour contrasts with L* L-H% which is said to be used to convey indignation or for answering the phone (Grice and Baumann 2002).

For Italian, as Savino (2012) points out, there is considerable variation in the realisation of question contours in the different varieties of the language, and each variety has multiple intonation patterns in its inventory. In Savino's study, the final rise is not predominant for the Turin region that the speakers of the present study were from, although it was found in around 15% of (information-seeking) polar questions. However, for other Northern Italian

varieties, such as Bergamo and Milano, she identifies a rising contour as predominant and describes it as H+L* L-H% (representing not only the final rise but also a preceding fall, which we are not concerned with here). Although Savino's intonation contours were obtained from task-oriented dialogues, the task (a map task) was different from the card game used in the current study and could have affected the distribution of different contours. What is important here is that both Savino's study and our results (from a considerably smaller sample) show that the final rise is available to the speakers as an option and is part of their intonational repertoire. Consequently, we may hypothesise that these speakers of Italian map their H+L* L-H% onto the German L* L-H% contour. In this light, the outcome observed in this study can be seen as the result of a phonological transfer of the boundary tone sequence, in which L-H% is used instead of the German native H-^H% with the higher final target. This hypothesis needs to be tested in future research. In doing so, it would be interesting to investigate how the final part of the contour is realised over different numbers of syllables in both languages and compare it with the learners' productions.

4. Training Effects on the Reduction in Epenthetic Vowels

A striking characteristic of the Italian pronunciation of words ending in a consonant is the epenthesis of a word-final vowel. As native Italian words usually end in vowels, epenthesis is usually found in loan words such as *tennis* [ten:is:ə] (Sluyters 1990). However, epenthesis is not present across the board. Inter alia, it appears to depend on factors such as the metrical structure of the word (more often if the final syllable is stressed), the voicing of the final consonant (more often when the final consonant is voiced), and the intonation contour (more often with rises and complex contours) (Grice et al. 2015). Unsurprisingly, epenthesis is also found in the pronunciation of Italian learners when they speak German. This subsection investigates the effects of explicit instruction in the prosody group in syllable structure, concentrating on words with a final consonant. This training aimed at both making the learners aware of their production of epenthetic vowels and reducing them by focusing on producing the word-final consonants without a following vowel. The segment group did not receive any information or instruction on word-final epenthetic vowels, but they did receive training on final obstruent devoicing (cf Section 5). As this training also focuses on the word-final consonant, it may have also had an effect on the production of epenthetic vowels, at least for words ending in consonants that undergo final devoicing in German.

4.1. Data

In order to assess the training effects of both the explicit syllable structure training that the prosody group received and the (implicit) segmental training of final obstruent devoicing, we separately focus on words ending in <t> and those ending in <d> because the word-final <d> is prone to be interpreted and produced as a voiced stop by Italians. Before the training, all groups produced epenthetic vowels in both conditions, but not consistently within groups and not to the same extent between groups. The control group produced the smallest number of epenthetic vowels before and after the training, followed by the segment and prosody groups (cf Table 4). With regard to the two conditions, epenthetic vowels were more often produced in words with a final <d> than those with a final <t> by both training groups, but not in the control group.

The data analysed here were recordings of two words ending in <d>, specifically *Rad* 'bike' and *Hund* 'dog', and two words ending in <t>, specifically *Rat* 'advice' and *bunt* 'colourful'. They were produced in sentences (1) to (4) below with the target words (here underlined) accented and in sentence-final position. This position leads to accentual and phrase-final lengthening, but this effect is constant across conditions. They were read aloud by all 33 subjects (target words are underlined, cf Section 2.1):

- (1) *Dina gab Elmar ein neues Rad.*
- (2) *In der gelben Hütte lebte ein großer Hund.*
- (3) *Tina gab Hanna einen guten Rat.*
- (4) *Die billigen Hütte waren ganz schön bunt.*

The sentences were interspersed with fillers during the recordings. Each sentence was produced three times by members of the control group and five times by members of the training groups. In total, 192 word realisations of the control group (8 speakers × 4 words × 3 repetitions × 2 recording times), 480 word realisations of the prosody group (12 speakers × 4 words × 5 repetitions × 2 recording times), and 520 word realisations of the segment group (13 speakers × 4 words × 5 repetitions × 2 recording times) entered the analysis. All word realisations were analysed in Praat to detect the presence of epenthetic vowels, and each occurrence was counted. The percentage of word realisations with epenthetic vowels of the data set was calculated and compared for all groups before and after the training sessions. The percentage values before and after training for both sets of words and all three groups are presented in Table 4, together with the differences between recordings made before and after training.

Table 4. Epenthetic vowel results.

Rad and Hund			
Training	Time Point	% Epenthetic Vowels	Difference (%) (% after–% before)
Control	Before	14.58	–2.08
	After	12.50	
Segment	Before	64.62	–25.38
	After	39.23	
Prosody	Before	84.17	–16.67
	After	67.50	
Rat and Bunt			
Training	Time Point	% Epenthetic Vowels	Difference (%) (% after–% before)
Control	Before	14.58	–10.42
	After	4.17	
Segment	Before	41.54	–7.69
	After	33.85	
Prosody	Before	67.50	–20.00
	After	47.50	

4.2. Analysis and Results

All groups show reductions after training in the percentage of epenthetic vowels for both sets of words, but to different extents when we compare groups and target words. The control group has the lowest values before and after training for both sets of target words. The reduction is larger for words ending in <t>. The segment group exhibits a massive reduction for words ending in <d> and only small improvements for words ending in <t>. The prosody group improves in both sets of words, slightly more for words ending in <t>. The improvements in the reduction in epenthetic vowels are illustrated in Figure 2.

For the statistical analysis, epenthetic vowel (yes/no) entered the model as a binary dependent variable for each set of words: (1) *Rad* and *Hund* and (2) *Rat* and *bunt*. The fixed effects were time of recording (before or after training) and training type (control, segment, or prosody), as well the as the interaction between the two variables. The model included random intercepts for speakers and by-speaker random slopes for the effect of time of recording.

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 10.0 for the regression coefficients. All the other priors were the default priors of brms. As priors for the intercept, a Student’s *t* distribution was used with degrees of freedom

of 3.0, a mean of 0.0, and a standard deviation of 2.5 ($\nu = 3.0, \mu = 0.0, \sigma = 2.5$). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student's t distribution ($\nu = 3.0, \mu = 0.0, \sigma = 2.5$). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions ($\eta = 1$). The model ran with four MCMC chains for 4000 iterations.

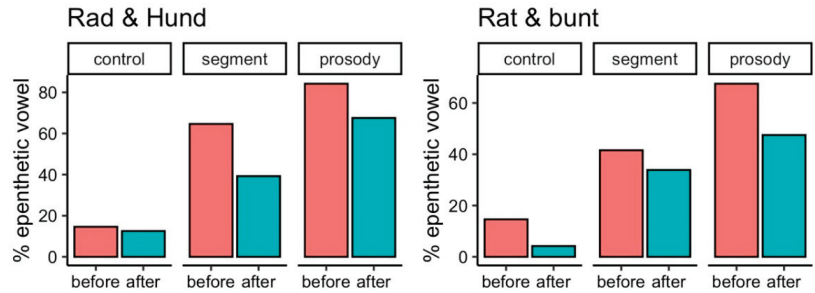


Figure 2. Percentages of epenthetic vowel occurrence for training groups before and after training.

We assess the training effects by looking at the posterior distributions for the differences between the recording time points (after minus before training) in terms of log odds. We report the model estimate of the difference β between the two time points, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\beta < 0)$. A negative estimate for the difference means that epenthetic vowels are reduced after training. $Pr(\beta < 0)$ gives an indication of how strong the evidence for a negative estimate is.

The results are presented in Table 5. For the set *Rad* and *Hund*, the results show a robust reduction in epenthetic vowels only in the segment group, with $Pr(\beta < 0) = 0.99$. For the set *Rat* and *bunt*, the results indicate a robust reduction in epenthetic vowels only in the prosody group, with $Pr(\beta < 0) = 0.98$.

Table 5. Statistical results for epenthetic vowels.

Rad and Hund					
Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	$Pr(\beta < 0)$
Control	-1.13	1.56	-3.80	1.28	0.77
Segment	-2.46	1.05	-4.28	-0.84	0.99
Prosody	-0.77	1.37	-2.87	1.54	0.73
Rat and Bunt					
Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	$Pr(\beta < 0)$
Control	-2.07	1.44	-4.53	0.20	0.93
Segment	-0.58	0.62	-1.62	0.41	0.84
Prosody	-1.33	0.64	-2.4	-0.31	0.98

4.3. Interim Summary and Discussion

In this section, we analysed the effects of two trainings on the realisation of word-final plosive codas with regard to the occurrence of epenthetic vowels. The results show that the prosody training was effective for both sets of words, but the effects are robust only for words ending in orthographic <t>, not for orthographic <d> (as in *Rad* and *Hund*). The voicing of final consonants plays an important role in the occurrence of final epenthetic vowels in Italian (Grice et al. 2015), which is reflected in our data set. Words ending in

voiced consonants (even if the voicing is the result of a spelling-based pronunciation) exhibit more cases of vowel epenthesis than do words ending in voiceless consonants. Consequently, the syllable structure training in the prosody group can result only in a reduction in epenthetic vowels in words ending in an orthographic voiced consonant when this consonant is interpreted as devoiced by the learners. This means that schwa epenthesis is best combined with an explicit training of final obstruent devoicing. The segment group that received explicit instruction in final obstruent devoicing shows a robust reduction in vowel epenthesis only for those words in which devoicing occurs, but not for others. This means that the segmental training had a positive effect for one set of words, probably due to the focus on the syllable coda and explicit instructions to produce final plosives with aspiration (precluding schwa epenthesis). However, the effects are not transferred to the other words with final consonants if these are voiced, so this does not constitute an improvement in the production of syllables in general. The results for all groups show that both trainings are effective but that they should be combined. We will next look at final devoicing in order to find out whether the syllable structure training of the prosody group had any effects on the production of (orthographically) voiced plosives.

5. Training Effects on Final Obstruent Devoicing

Final obstruent devoicing refers to a phonological phenomenon occurring in syllable codas in German words. Plosives and fricatives that are underlyingly voiced become voiceless in that position, so the word *Rad* ‘wheel’ is pronounced [ʁa:t], while the plosive is voiced when it is in syllable-initial position as in the plural form *Räder* [ʁɛ:ɖɐ]. German spelling does not reflect these differences, so learners interpret graphemes that usually represent voiced obstruents as such (Hayes-Harb et al. 2018). In Italian, obstruents usually occur in syllable codas when they are part of a geminate consonant, e.g., *fredda* [ˈfrɛd.da] ‘cold’, and there is a voicing distinction in that position (e.g., *fretta* [ˈfrɛt.ta] ‘hurry’). As a consequence, Italian learners tend to pronounce German *Rad* as [rad.də]. In this section, we investigate the effects of explicit training of final devoicing of plosives as conducted with the segment group (see Section 2.3). The other groups did not receive any information or instruction on final obstruent devoicing, but the prosody group received training focusing on word-final consonantal codas and the avoidance of an epenthetic vowels (cf Section 4). This may have led to more awareness of the syllable coda and even an improvement in final devoicing.

5.1. Data

Final devoicing is a neutralisation process, although many studies claim that this neutralisation is incomplete because German natives produce word pairs such as *Rad–Rat* (‘bike’–‘advice’) slightly differently (e.g., Roettger et al. 2014). However, the training was based on complete neutralisation, so this is what the learners aimed to achieve. The data analysed here are part of the data set described in Section 4. Here, we look at only the word pair *Rad–Rat* that we elicited as described above in the carrier sentences (cf Section 2.1):

- (1) *Dina gab Elmar ein neues Rad.*
- (3) *Tina gab Hanna einen guten Rat.*

In total, 48 repetitions for each target word were elicited from the control group (8 speakers × 3 repetitions × 2 recording times), 130 for the segment group (13 speakers × 5 repetitions × 2 recording times), and 120 for the prosody group (12 speakers × 5 repetitions × 2 recording times). All the word tokens were annotated in Praat, and the values were automatically extracted. The parameters examined here are the duration of the vowel and consonantal closure intervals and the duration of voicing during the closure interval for *Rad*. In order to achieve a neutralisation effect, these parameters should become more similar for the two target words after the training.

5.2. Analysis and Results for Neutralisation of Vowel Duration in L2 German

What we are interested in here is the absolute distance between the vowel /a:/ in *Rat* and the vowel /a:/ in *Rad* in terms of duration. We ask whether the distance becomes smaller, i.e., whether the vowels of *Rat* and *Rad* become more similar after the training. Because we are not dealing with a parameter on the level of one utterance but rather the relation between different productions, we first calculate the mean for each vowel for each speaker. That is, for each speaker, we calculate the mean duration of /a:/ from *Rat* and the mean duration of /a:/ from *Rad*. Next, we calculate the *absolute* distance between these durations. Table 6 presents the absolute distance between the vowels in both target words in milliseconds for each group before and after the training and the standard deviation as well as the changes in that distance after the training.

Table 6. Results for vowel duration distance.

Training	Time Point	Mean Vowel Duration Distance (ms)	SD Vowel Duration Distance (ms)	Difference (ms) (Mean After–Mean Before)
Control	Before	22.92	16.61	−4.50
	After	18.42	12.36	
Segment	Before	19.63	12.73	3.22
	After	22.85	17.97	
Prosody	Before	11.42	12.65	3.91
	After	15.34	11.36	

The statistical model used vowel duration distance as the dependent variable. The fixed effects were time of recording (before or after training) and training type (control, segment, or prosody), as well as the interaction between the two variables. The model included random intercepts for speakers (note that because we took the speaker means, there are only two observations per speaker—before training and after training).

We used a normally distributed prior with mean of 0.0 and standard deviation of 10.0 for the regression coefficients. All other priors were the default priors of brms. As priors for the intercept, a Student’s *t* distribution was used with degrees of freedom of 3.0, a mean of 15.3, and a standard deviation of 15.8 ($\nu = 3.0, \mu = 15.3, \sigma = 15.8$). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student’s *t* distribution ($\nu = 3.0, \mu = 0.0, \sigma = 15.8$). The model ran with four MCMC chains for 6000 iterations.

We assess the training effects by looking at the posterior distributions for the differences between the recording time points (after training minus before training). We report the model estimate of the difference β between the two time points, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\beta < 0)$. A negative estimate for the difference means that the distance between the vowel of *Rat* and the vowel of *Rad* was reduced during training, while a positive estimate indicates a growth of the distance between the two vowels and hence the opposite of neutralisation. $Pr(\beta < 0)$ gives an indication of how strong the evidence for a negative estimate is. The results are given in Table 7. They show no reliable effect in any group.

The results show that all groups produce different vowel durations for the two target words, so they distinguish between them by means of vowel duration. There are only very minor changes after training, however, and statistical analysis showed that none of the changes were robust. Thus, there are no training effects for this parameter.

Table 7. Statistical results for vowel duration distance.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	$Pr(\beta < 0)$
Control	0.51	3.98	−6.04	6.97	0.45
Segment	2.35	3.76	−3.9	8.43	0.26
Prosody	2.24	3.89	−4.04	8.66	0.28

5.3. Analysis and Results for Neutralisation of Closure Duration in L2 German

Another way of assessing neutralisation effects of the training is to look at the absolute distance between the closure duration of /t/ in *Rat* and *Rad*. We ask whether the distance becomes smaller, i.e., whether the consonants of *Rat* and *Rad* become more similar after the training. Because, as with vowel duration, we are dealing with the relation between different productions, we first calculate the mean for each closure duration for each speaker. That is, for each speaker, we calculate the mean closure duration of *Rat* and the mean closure duration of *Rad*. Next, we calculate the absolute distance between these durations.

Table 8 presents the distance between closure durations in both target words in milliseconds for each group before and after the training and the standard deviation as well as the changes in that distance after the training. Again, a negative value for the difference would indicate that the distance between the vowel of *Rat* and the vowel of *Rad* was reduced during training.

Table 8. Results for closure duration distance.

Training	Time Point	Mean Closure Duration Distance (ms)	SD Closure Duration Distance (ms)	Difference (ms) (Mean After–Mean Before)
Control	Before	22.05	16.59	8.79
	After	30.83	18.82	
Segment	Before	32.17	21.58	2.75
	After	34.92	33.02	
Prosody	Before	20.64	16.18	5.89
	After	26.52	14.38	

For the statistical analysis, we used a model with closure duration distance as the dependent variable. The fixed effects were time of recording (before training or after training) and training type (control, segment, or prosody), as well as the interaction between the two variables. The model included random intercepts for speakers (note that because we took speaker means, there are only two observations per speaker—before training and after training).

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 10.0 for the regression coefficients. All other priors were the default priors of brms. As priors for the intercept, a Student’s *t* distribution was used with degrees of freedom of 3.0, a mean of 23.1, and a standard deviation of 16.8 ($\nu = 3.0, \mu = 23.1, \sigma = 16.8$). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student’s *t* distribution ($\nu = 3.0, \mu = 0.0, \sigma = 16.8$). The model ran with four MCMC chains for 6000 iterations.

We assess the training effects by looking at the posterior distributions for the differences between the recording time points (after training minus before training). We report the model estimate of the difference β between the two time points, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\beta < 0)$. A negative estimate for the difference means that the distance between the closure duration of *Rat* and the closure duration of *Rad* was reduced during training.

$Pr(\beta < 0)$ gives an indication of how strong the evidence for a negative estimate is. The results, displayed in Table 9, show no reliable effect in any group.

Table 9. Statistical results for closure duration distance.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	$Pr(\beta < 0)$
Control	4.71	4.64	-3.11	12.27	0.15
Segment	2.41	5.09	-6.13	10.69	0.32
Prosody	5.54	4.92	-2.67	13.59	0.13

We can see that all groups distinguish between the two target words by means of closure duration at both time points. The changes after the training phase are minor, and according to our statistical analysis, none of them are robust, so no training effects are visible for this parameter either.

5.4. Analysis and Results for Reduction in Voicing during Closure in L2 German

As there are no measurable effects on vowel or consonant duration, we now look at vocal fold activity during the closure interval. Here, we look only at *Rad* because there was no voicing during closure in *Rat*. We measured the total duration of the consonant closure and that of the interval during which there was vocal fold vibration within the closure interval and calculated the percentage of voice during closure. Table 10 shows the mean percentages of voice during closure for all groups before and after the training as well as the standard deviation and the changes after the training. Negative values for the difference between the percentage of voicing during closure before and after training indicate an improvement. The results show improvements in all three groups but major changes only for the segment group.

Table 10. Voicing during closure results.

Training	Time Point	Voicing during Closure Mean (%)	Voicing during Closure SD (%)	Difference (%) (Mean After-Mean Before)
Control	Before	65.52	38.53	-12.60
	After	52.93	41.45	
Segment	Before	70.26	36.56	-50.93
	After	19.33	32.50	
Prosody	Before	82.02	34.66	-13.77
	After	68.25	38.74	

Figure 3 shows the raw data points as a jittered strip chart (grey dots) in addition to the means (coloured thick dots). It can be observed in the plot that the distributions of the voicing during closure data substantially deviate from a normal distribution. The points are half transparent, darker areas thus indicating the clustering of data points. There are many data points with values of 0% or 100%; i.e., there are a lot of closures that are either not voiced at all or fully voiced. Therefore, a model with a normal or skewed-normal distribution would produce a bad fit of the data. Instead, we transformed the data into the range of 0 to 1 (division by 100) and fitted a Bayesian zero/one inflated beta (ZOIB) model. The ZOIB model represents a mixture of a logistic and a beta regression. Therefore, the ZOIB model is able to estimate two interesting quantities in the context of this study. First, γ , the probability that an observation is 1. Second, μ , the mean of the continuous beta distribution in between 0 and 1. The two distributional parameters were estimated along with the precision of the beta distribution ϕ and the zero/one inflation α (the probability that an

observation is either 0 or 1), but we report only the results for γ and μ (for an introductory tutorial, see Vuorre 2021). The fixed effects were time of recording (before or after training) and training type (control, segment, or prosody), as well as the interaction between the two variables. The model included random intercepts for speakers and by-speaker random slopes for the effect of time of recording.

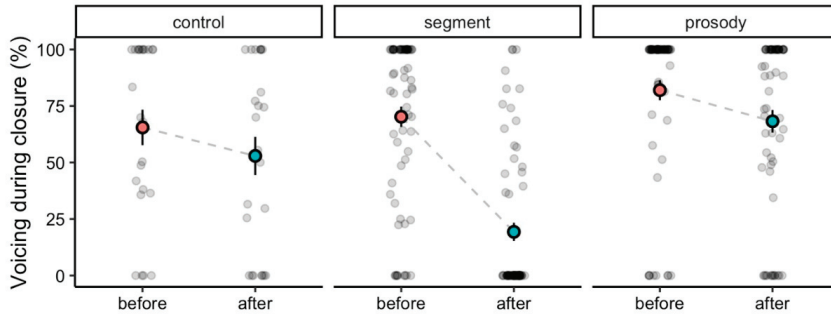


Figure 3. Voicing during closure (in %) for training groups before and after training (coloured thick points with bars: means and standard errors; grey points: raw measures).

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 1.0 for the regression coefficients. All other priors were the default priors of brms. As priors for the intercepts of μ and ϕ , a Student’s t distribution was used with degrees of freedom of 3.0, a mean of 0.0, and a standard deviation of 2.5 ($\nu = 3.0, \mu = 0.0, \sigma = 2.5$). As priors for the intercepts of γ and α , a logistic distribution was used ($\mu = 0, \sigma = 1$). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard deviation of the model, we used a Student’s t distribution ($\nu = 3.0, \mu = 0.0, \sigma = 2.5$). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions ($\eta = 1$). The model ran with four MCMC chains for 8000 iterations.

We assess the training effects by looking at the posterior distributions for the differences between recording time points (after training minus before training). We report the model estimate of the differences $\Delta\gamma$ and $\Delta\mu$, the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is negative $Pr(\Delta\gamma < 0)$ or $Pr(\Delta\mu < 0)$. A negative estimate for the differences means that the voicing during closure was reduced during training. A negative difference $\Delta\gamma$ indicates that the probability of 1, i.e., full voicing, is reduced. A negative difference $\Delta\mu$ indicates that the means of the beta distribution in between 0 and 1 decreases; i.e., the relative duration of partial voicing during the closure is reduced. The results are presented in Table 11 (all estimates are in logit). There is strong evidence for a reduction in full voicings in the segment group, but not for the other groups. No group reliably reduces the mean of the beta distribution, relating to the relative duration of the partial voicings.

Table 11. Statistical results for voicing during closure.

Training Group	γ				
	$\Delta\gamma$	SE	90% CI Low Boundary	90% CI High Boundary	$Pr(\Delta\gamma < 0)$
Control	-3.66	3.8	-10.12	1.59	0.89
Segment	-21.34	9.54	-38.63	-10.05	1.00
Prosody	-1.71	3.3	-6.05	2.91	0.80

Table 11. Cont.

Training Group	$\Delta\mu$	SE	μ		Pr($\Delta\mu < 0$)
			90% CI Low Boundary	90% CI High Boundary	
Control	−0.19	0.31	−0.68	0.34	0.74
Segment	−0.27	0.31	−0.77	0.24	0.82
Prosody	−0.14	0.37	−0.74	0.5	0.66

5.5. Interim Summary and Discussion

In this section, we analysed the effects of an explicit segmental training of final devoicing, compared with an (implicit) syllable structure training, by investigating whether the subjects learned to neutralise the distinction between the words *Rat* and *Rad* by producing more-similar duration values for vowels and consonants in both words after the training. The results showed that the segmental training was not effective in that respect, which could be because the focus of the exercises was not on these aspects but was rather on the mere voicing neutralisation, i.e., the avoidance of voicing during closure and final aspiration for words such as *Rad*. Moreover, Italian learners of German encounter additional challenges when learning to modulate vowel duration in closed syllables because in their L1, closed syllables can have only a short vowel (leading to a consonant cluster or a geminate word-medially, as in [ˈfrɛd.da] mentioned above).

Looking at voicing during closure, our results indicate that the segmental training was effective and led to a smaller number of word productions with fully voiced closures. In addition, as described in Section 4, there were positive effects with regard to the occurrence of epenthetic vowels in words with final (orthographically) voiced consonants. The control and prosody groups showed no reliable effects. Thus, the syllable structure training clearly had no effect on final devoicing. This once more supports the suggestion that final devoicing should be trained along with syllable structure, as syllable structure training helps to avoid epenthetic vowels, but only when the final consonant is voiceless; when the final consonant is interpreted as voiced on the basis of spelling, the training effects vanish. These results indicate that although training in final devoicing can support prosody training, the converse is not true: it is not implicitly acquired during prosody training, but it needs to be explicitly taught.

6. Voice Onset Time

German and Italian both have the plosives /p, t, k/ and /b, d, g/ in their consonant phoneme inventories, but they use different cues to distinguish between the two sets. Italian uses mainly voicing during closure (i.e., vocal fold activity during the consonant closure), whereas German uses mainly voice onset time, where /p, t, k/ is produced with a long voice lag (>30 ms) and /b, d, g/ with a short one (0–30 ms), while the vibration of the vocal folds during the consonant closure is not distinctive and generally only present when the plosive is surrounded by other voiced sounds (Jessen and Ringen 2002). The occurrence of aspirated plosives in Italian (i.e., with a positive VOT > 30 ms) is reported for some regions (Celata and Nagy 2022).

In this subsection, we examine the VOTs of all subjects from the three training groups for the word-initial plosive /t/ before and after the training phases to find out whether any changes towards longer positive VOTs are linked to the trainings that the test groups received. The segment group was explicitly made aware of the aspiration of plosives in German and of its significance for German natives to distinguish between words such as *'tennis'* and *Dennis* (a boy's name); see Section 2.3. The control and prosody groups received no explicit information or instruction on aspiration. However, the prosody group engaged in exercises for word stress, both on the phonological level (i.e., stress placement rules) and with regard to the phonetic features of word stress in German, which involve more articulatory effort and stronger air flow in stressed syllables, referred to as *Druckakzent*

(force accent). In order to generate the effort and pressure on stressed syllables, subjects were instructed to bang on the table with their fists when producing stressed syllables during the training sessions (not during the recordings). This may have had an effect on voiceless plosives in German stressed syllables, as the consonant release might have been stronger, resulting in a longer VOT. The influence of stress on VOT is reported in numerous studies (e.g., Lisker and Abramson 1967; Savino et al. 2015; Lein et al. 2016).

6.1. Data

The data analysed here were elicited in the reading tasks explained above. The target words were *Tina* and *Tennis* in the carrier sentences (cf Section 2.2):

- (3) *Tina* gab Hanna einen guten Rat.
- (5) *Helga* spielte einmal *Tennis*.

In total, 96 word realisations of the control group (8 speakers × 2 words × 3 repetitions × 2 recording times), 240 word realisations of the prosody group (12 speakers × 2 words × 5 repetitions × 2 recording times), and 260 word realisations of the segment group (13 speakers × 2 words × 5 repetitions × 2 recording times) entered the analysis.

6.2. Analysis and Results

Table 12 shows the mean VOTs for all groups before and after the training as well as the standard deviation and the difference between the mean values after the training and those before the training. All groups had already produced positive VOTs with mean values of over 30 milliseconds before the training, which shows that the subjects clearly pronounce voiceless plosives differently from their native productions, but with shorter VOTs than German natives speaking standard German (cf Kirby et al. 2020). The prosody group produced slightly shorter VOTs than the control and segment groups before the training. Positive values for the difference of mean VOTs before and after training indicate an improvement. Both test groups show longer VOTs after the training, with a slightly larger effect in the segment group. The control group exhibits a minor negative change. Figure 4 illustrates the changes in VOT for all groups.

Table 12. VOT results.

Training	Time Point	VOT Mean (ms)	VOT SD (ms)	Difference (ms) (Mean After–Mean Before)
Control	Before	33.54	21.66	–2.62
	After	30.92	14.10	
Segment	Before	34.02	17.99	9.51
	After	43.52	20.92	
Prosody	Before	28.77	16.06	6.68
	After	35.45	18.60	

For the statistical analyses, we used a mixed model with VOT as the dependent variable. The fixed effects were time of recording (before or after training), training type (control, segment, or prosody) and the interaction between the two variables. The model included random intercepts for speakers and target words, as well as by-speaker random slopes for the effect of time of recording. The model was fitted with a skewed-normal distribution to achieve a better model fit.

We used a normally distributed prior with a mean of 0.0 and a standard deviation of 10.0 for the regression coefficients. All the other priors were the default priors of brms. As priors for the intercept, a Student’s *t* distribution was used with degrees of freedom of 3.0, a mean of 14, and a standard deviation of 43 ($\nu = 3.0, \mu = 31, \sigma = 19.3$). As priors of the standard deviations of the random intercepts and slopes as well as the residual standard

deviation of the model, we used a Student’s *t* distribution ($\nu = 3.0, \mu = 0.0, \sigma = 19.3$). The priors of the Cholesky factors of the covariance matrix for random effects were Cholesky LKJ correlation distributions ($\eta = 1$). The prior for the skewness parameter α for the skewed-normal distribution was a normal distribution with a mean of 0.0 and a standard deviation of 4.0. The model ran with four MCMC chains for 4000 iterations.

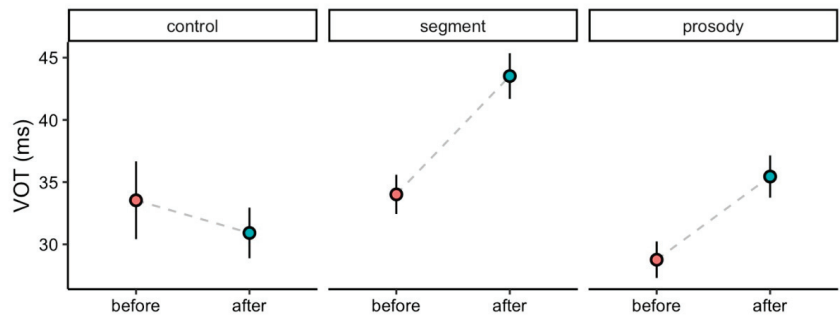


Figure 4. VOT for training groups before and after training (means and standard errors).

We assess the training effects by looking at the posterior distributions for the differences between recording time points (after training minus before training). We report the model estimate of the difference β between the two time points, the standard error of the estimate (SE) the lower and upper boundaries of the 90% credible interval (90% CI), and the probability that the estimate is positive $Pr(\beta > 0)$. A positive estimate for the difference means that the VOT became longer during training. $Pr(\beta > 0)$ gives an indication of how strong the evidence for a positive estimate is.

The results are presented in Table 13. The statistical estimates show that there is strong evidence for positive differences in the segment and prosody groups regarding the VOT with a $Pr(\beta > 0)$ of 1.0 in both cases, i.e., an increase in VOT during training. There is no reliable effect for the control group ($Pr(\beta > 0) = 0.59$). All in all, the statistical results show that the segment and prosody groups increase their VOTs for /t/ during training.

Table 13. Statistical results for VOT.

Training Group	β	SE	90% CI Low Boundary	90% CI High Boundary	$Pr(\beta > 0)$
Control	0.48	2.29	−3.30	4.24	0.59
Segment	7.08	1.74	4.24	9.93	1.00
Prosody	5.07	1.78	2.16	8.02	1.00

6.3. Interim Summary and Discussion

We analysed the effects of explicit segmental training and (implicit) prosodic training on the production of VOT in word-initial /t/. The results show positive effects for both trainings. Thus, training the phonetic features of word stress in German clearly improves learners’ VOT in fortis plosives similarly to purely segmental training. This does not mean that segmental training can be skipped for this aspect; after all, we examined only the plosives in stressed syllables here, and the effects of the prosody training might not be present in unstressed syllables. Again, a combination of both segmental training and prosodic training would be beneficial.

7. General Summary and Discussion

In this paper, we examined the effects of prosodic training in a prosodic feature (intonation) and of a prosody-oriented training in an area where prosody and segments

interact (word-final codas). The effects of segment-oriented training were assessed for final obstruent devoicing, which is linked to the syllable structure and is thus partly prosodic, and for VOT of voiceless plosives, which is regarded as a segmental feature, although the temporal coordination of laryngeal and supralaryngeal gestures is not typical of what is regarded as segmental in nature. Table 14 summarises our findings (✓ refers to a training improvement, X refers to no training improvement).

Table 14. Summary of results.

Area	Measure	Control Group	Segment Group	Prosody Group
Prosody	Final rise magnitude in yes–no questions	X	X	✓
Prosody	Epenthetic vowel	X	X (Rad/Hund)	✓ (Rat/bunt)
Segment	Final devoicing: voicing during closure	X	✓	X
Segment	Final devoicing: vowel duration distance	X	X	X
Segment	Final devoicing: closure duration distance	X	X	X
Segment	Voice onset time (VOT)	X	✓	✓

One result that is not at all surprising is that explicit segmental training improves the production of segments and that prosody training improves the production of prosodic features. The intonation training yielded reliable positive results for final rises in yes–no questions only for the prosody group, which is also not surprising given that there is no relation between question intonation and any of the segmental features examined here. VOT, which is dependent on the prominence of syllables (stress and accent), is a good example of a segmental area that can be influenced by prosody training. However, as noted above, we looked only at contexts in which the plosive was in a stressed (and accented) syllable, so we do not know whether the effects of the word-stress training will hold for unstressed syllables. This might be an interesting point to investigate in further research with Italian learners. Epenthetic vowels and final devoicing both focus on the word-final consonant in training. Our analyses showed that the segment-oriented training (final devoicing of /d/) had positive effects, both at the segmental level (the learners produced less voicing during closure) and at the prosodic level (they produced fewer epenthetic vowels after a word-final <d>). However, the effect does not hold for words ending in consonants that do not undergo final devoicing. For the prosody group, no reliable effects of the syllable structure training were found on final devoicing. The mere fact that the training focused on the syllable coda did not make the learners aware of final devoicing in German. The syllable coda training showed effects only for the target words ending in <t>. The voiced final consonants in the orthography of words such as *Rad* and *Hund* appeared to facilitate vowel epenthesis, analogous to the native Italian pronunciation. Thus, the lack of instruction on final devoicing prevented a positive effect for the <d>-words, at least for the small set of data examined here. More research in this area, specifically research that involves more final consonants, is needed to obtain a clearer picture.

In sum, this study provides (somewhat limited) confirmation for the previous claims, made by Anderson-Hsieh et al. (1992), Munro and Derwing (1995), Derwing et al. (1998) and Gordon and Darcy (2016), that there are positive effects of prosody-oriented training on the production of segments, but this crucially depends on the area of prosody that is being trained. In our study, the training of syllable structure and the production of prosodic prominence (lexical stress and the placement of pitch accents, which were part of the training but not of the testing) is likely to have had a greater effect on the segments than the training of intonation contours. Interestingly, also in line with the above-mentioned studies, there were no reliable positive effects of segment-oriented training on prosodic features. This was even the case when the training aimed at an area where segments and prosody interact, as is the case for final obstruent devoicing.

A limitation of this study is the small data set, making it difficult to generalise. Nonetheless, our results appear to indicate that prosodic training and segmental training are best treated in an integrated way. In particular, if aspiration is taught alongside stress and accent, aspiration can be learned in this hyperarticulated context, making the difference between L1 and L2 clearer. Moreover, there appear to be benefits in teaching final devoicing alongside syllable structure, including avoiding schwa epenthesis and thus restructuring of the word, making a “final” consonant in fact initial to a further syllable. Learning to devoice obstruents in syllable onsets instead of codas could otherwise lead to possible problems with learning to adequately produce the voicing distinction in onset position. Thus, our results support the conclusions drawn by [Derwing and Munro \(2015\)](#): if the segmental and prosodic levels are taught together, there is a greater likelihood of an overall beneficial outcome in pronunciation training.

Author Contributions: Conceptualization, S.D. and M.G.; methodology, S.D., M.G. and S.R.; software, S.R.; validation, S.D., M.G. and S.R.; formal analysis, S.D. and S.R.; investigation, S.D.; resources, S.D. and M.G.; data curation, S.D. and S.R.; writing—original draft preparation, S.D., M.G. and S.R.; writing—review and editing, S.D., M.G. and S.R.; visualization, S.R.; supervision, M.G.; project administration, S.D.; funding acquisition, M.G. and S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the collaborative research center SFB1252 Prominence in Language Project-ID 281511265) and project RO 6767/1-1 (Walter Benjamin program).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data and analysis scripts are available online on the Open Science Framework: <https://osf.io/mfbw3/> (accessed on 2 February 2023).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix complements the substudy on the yes–no question rise magnitudes in the paper. Because both German and Italian commonly have final rises in such questions (albeit with much regional variation), we can link the effects of training to patterns of realisation in the respective languages as produced by native speakers. In this comparison, it becomes evident that Italian yes–no question rises are smaller in magnitude compared with German rises.

The data analysed here are recordings of Italian and German native speakers. We collected two data sets in order to be able to compare Italian L1 and German L1 realisation patterns. The first data set consists of recordings of Italian native speakers. It was collected at a grammar school in Turin (Northern Italy). In total, eight students (three male, five female) were recorded while playing card games specifically designed to elicit yes–no questions. These students were also later part of the training groups described in the main text of the paper (three in the prosody group, five in the segment group). The card games were played in pairs. The recordings were conducted in a quiet room in a school in Turin with a mobile DAT recorder and head-mounted microphones. The speakers were 17 to 18 years old.

The second data set contains recordings of four native speakers of German, two of them authors of this paper (S.D. and S.R.). The recordings took place in a sound-attenuated recording booth at the University of Cologne, using head-mounted microphones (recorded directly on the hard disk of a computer through an external audio interface). The speakers were aged between 22 and 35 years; two of them identified as female, two as male.

The data were elicited in a card game setting in which the subjects played in pairs. The cards in this game depicted day-to-day objects in different colours. The participants’ task was to collect cards with the same colour or the same object by exchanging cards with their fellow player. To initiate the exchange, participants ask their fellow player whether

they are in possession of a specific card. For example, *do you have a green coffee pot/carafe?* (German: *hast du eine grüne Kanne?*, Italian: *hai una caraffa verde?*).

In each move, the colour or object that the player can use in their question is determined by a card from an additional stack. In one version of the game, it is the colour that is displayed by this card; in another version, it is the object. For example, if the card is green, the participant may ask *do you have a green coffee pot?* but not *do you have a blue coffee pot?* Similarly, when the card displays a coffee pot, the participant may ask *do you have a green coffee pot?* but not *do you have a green plate?* There was no visual contact between the participants; for communication, they relied solely on the auditory channel.

The German colour adjectives were *blaue/blauen* 'blue', *gelbe/gelben* 'yellow', *rote/roten* 'red', and *grüne/grünen* 'green'. In Italian, they were *azzurro/azzurra* 'blue', *giallo/gialla* 'yellow', *rosso/rossa* 'red', and *verde* 'green'. The German object nouns were *Kanne* 'coffee pot', *Teller* 'plate', *Gabel* 'fork', and *Kugel* 'ball'. In Italian, they were *caraffa* 'carafe', *piatto* 'plate', and *tazza* 'cup'. As exemplified above, the questions were of the form 'hast du eine(n) <colour> <object>?' for German and 'hai un(a) <object> <colour>?' for Italian.

From the 136 questions in the German L1 data, 13 were excluded because of hesitations, laughter, or mispronunciations; 13 were excluded because the speaker asked for two objects (in an alternative question such as *hast du eine grüne Kanne oder einen gelben Teller?* 'do you have a green carafe or a yellow plate?'). Moreover, 14 questions did not end in a simple rising intonation contour: seven of them were falling (H* L-%) and seven had a falling-rising nuclear contour (H* L-H%). Thus, for the investigation of the final rise magnitude, 96 German questions could be used. All these questions reflect the nuclear intonation pattern L* H-H% described in Grice and Baumann (2002) for neutral German yes-no questions. An example from the data set is given in Panel C of Figure A1.

For the Italian L1 data, 110 questions were recorded from a group of eight speakers. Here, 11 questions were excluded because of hesitations, laughter, or mispronunciations; four were excluded because they were alternative questions, as in the German data described above. Of the Italian questions, 24 ended in a falling boundary tone (see Panel B of Figure A1). These were mainly by two speakers who exclusively produced rising-falling contours (L+H* L-L%). For the analysis of the final rise, these speakers had to be excluded. Hence, 71 Italian questions with a final rise elicited from six speakers remained in the data set for this measurement. The nuclear intonation contour of these questions can be described as (H+)L* nuclear accent, followed by a rising boundary tone (see Panel A of Figure A1).

In both languages, the start and end points of the final rise were annotated. Start point "L" was placed on an F0 minimum in the vowel of the syllable with the nuclear L* accent, e.g., in [a] of *Kanne* in *hast du eine grüne Kanne?* and in [u] of *azzurro* in *hai un piatto azzurro?*² End point "H" was placed on the F0 maximum at the end of the utterance. The rise magnitude was calculated in semitones: $12 \log_2 \left(\frac{E}{B} \right)$, where B denotes the F0 in Hz at the beginning of the rise and E denotes the F0 in Hz at the end of the rise.

Panel D of Figure A1 shows the results obtained from the measure of the final rise. The violin plots show the distributions of the data. The thick black dots represent the respective means of these distributions. The graph presents a clear picture, where German exhibits substantially larger final rises (13.1 st) than does Italian (5.23 st).

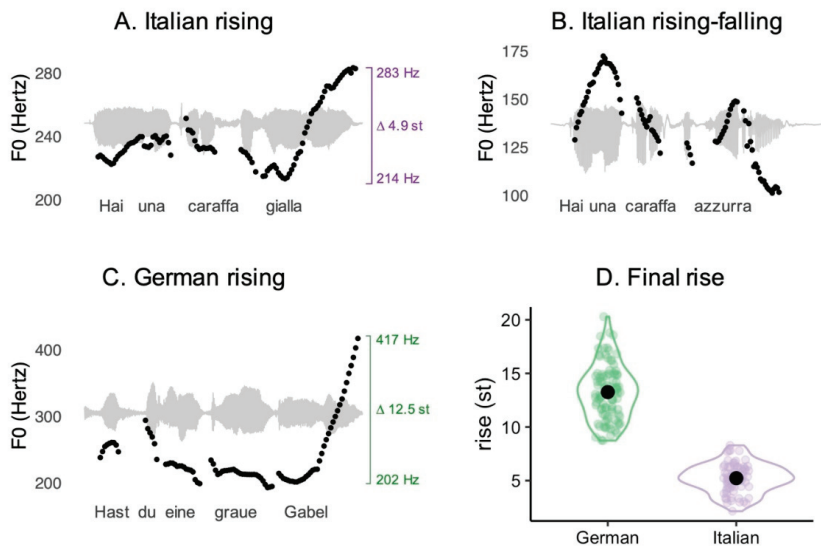


Figure A1. Example contours (A–C) and final rise magnitudes of German L1 and Italian L1 (D).

Notes

- For the intonation of questions, the data were collected during the project, but they have been analysed here for the first time.
- The labels deviated from this pattern only when there was no reliable F0 calculation in the vowel that was due to creaky voice.

References

- Anderson-Hsieh, Janet, Ruth Johnson, and Kenneth Koehler. 1992. The Relationship between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Language Learning* 42: 529–55. [CrossRef]
- Baills, Florence, Charlotte Alazard-Guiou, and Pilar Prieto. 2022. Embodied prosodic training enhances accentedness and general suprasegmental accuracy. *Applied Linguistics* 43: 776–804. [CrossRef]
- Bürkner, Paul-Christian. 2018. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* 10: 395–411. [CrossRef]
- Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76: 1–32. [CrossRef]
- Celata, Chiara, and Naomi Nagy. 2022. Sociophonetic Variation and Change in Heritage Languages: Lexical Effects in Heritage Italian Aspiration of Voiceless Stops. *Language and Speech. online first*. [CrossRef]
- Chun, Dorothy, and John Levis. 2020. Prosody in Second Language Teaching. In *The Oxford Handbook of Language Prosody*. Edited by Carlos Gussenhoven and Aojun Chen. Oxford: Oxford University Press, pp. 619–32.
- Council of Europe. 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment—Companion Volume*. Strasbourg: Council of Europe Publishing.
- Dahmen, Silvia. 2013. *Prosodie oder Segmente? Phonetische Untersuchungen zu Trainingseffekten bei Italienischen Deutschlernenden*. Cologne: Cologne Publication Service. Available online: <https://kups.uni-koeln.de/5368/> (accessed on 2 February 2023).
- Derwing, Tracey M., and Murray J. Munro. 1997. Accent, Intelligibility and Comprehensibility: Evidence from Four L1s. *Studies in Second Language Acquisition* 19: 1–16. [CrossRef]
- Derwing, Tracey M., and Murray J. Munro. 2015. *Pronunciation Fundamentals. Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Derwing, Tracey M., Murray J. Munro, and Grace Wiebe. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410. [CrossRef]
- Dieling, Helga, and Ursula Hirschfeld. 2000. Phonetik lehren und lernen. Fernstudieneinheit 21. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 5: 3.
- Gordon, Joshua, and Isabelle Darcy. 2016. The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation* 2: 56–92. [CrossRef]
- Grice, Martine, and Stefan Baumann. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191: 267–98.
- Grice, Martine, Michelina Savino, Alessandro Caffò, and Timo B. Roettger. 2015. The tune drives the text—Schwa in consonant-final loanwords in Italian. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14.

- Hayes-Harb, Rachel, Brown Kelsey, and Bruce L. Smith. 2018. Orthographic Input and the Acquisition of German Final Devoicing by Native Speakers of English. *Language and Speech* 61: 547–64. [CrossRef]
- Jessen, Michael, and Catherine Ringen. 2002. Laryngeal features in German. *Phonology* 19: 189–218. [CrossRef]
- Kaunzner, Ulrike. 2015. Die Verständlichkeit vorgelesener Texte durch nichtdeutsche Sprecher. Zur Sprechwirkung bei muttersprachlichen und nicht-muttersprachlichen Rezipienten. In *Aktuelle Forschungstendenzen in der Sprechwissenschaft: Normen, Werte, Anwendung*, hrsg. v. B. Teuchert (= *Sprache und Sprechen*). Baltmannsweiler: Schneider, pp. 65–81.
- Kaunzner, Ulrike. 2018. Das klingt sympathisch!. Selbst- und Fremdbild in der Sprechwirkung des italienischen Akzents. In *Gesprochene (Fremd-) Sprache als Forschungs- und Lehrgegenstand*. Trieste: EUT Edizioni Università di Trieste, pp. 139–55.
- Kirby, James, Felicitas Kleber, Jessica Siddins, and Jonathan Harrington. 2020. Effects of prosodic prominence on obstruent-intrinsic F0 and VOT in German. Paper presented at the 10th International Conference on Speech Prosody 2020, Tokyo, Japan, May 25–28; pp. 210–14.
- Lein, Tatjana, Tanja Kupisch, and Joost van de Weijer. 2016. Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism* 20: 732–49. [CrossRef]
- Li, Peng, Florence Baills, Lorraine Baqué, and Pilar Prieto. 2022. The effectiveness of embodied prosodic training in L2 accentedness and vowel accuracy. *Second Language Research*. online first. [CrossRef]
- Li, Peng, Xiaotong Xi, Florence Baills, and Pilar Prieto. 2020. Appropriately performing hand gestures cueing phonetic features facilitates simultaneous speech imitation in an L2. Paper presented at the 7th GeSpIn Conference, GeSpIn 2020, Virtual, September 7–9; Sweden: KTH Royal Institute of Technology.
- Lisker, Leigh, and Arthur S. Abramson. 1967. Some Effects of Context On Voice Onset Time in English Stops. *Language and Speech* 10: 1–28. [CrossRef] [PubMed]
- Mennen, Ineke. 2007. Phonological and phonetic influences in non-native intonation. In *Non-Native Prosody. Phonetic Description and Teaching Practice*. Edited by Jürgen Trouvain and Ulrike Gut. Berlin and New York: De Gruyter, pp. 53–76.
- Missaglia, Federica. 1999a. *Phonetische Aspekte des Erwerbs von Deutsch als Fremdsprache durch italienische Muttersprachler*. Frankfurt am Main: Hector.
- Missaglia, Federica. 1999b. Contrastive prosody in SLA—An empirical study with adult Italian learners of German. Paper presented at the ICPHS, San Francisco, CA, USA, August 1–7; pp. 551–54.
- Missaglia, Federica. 2007. Prosodic Training for Adult Italian Learners of German: The Contrastive Prosody Method. In *Non-Native Prosody. Phonetic Description and Teaching Practice*. Edited by Jürgen Trouvain and Ulrike Gut. Berlin and New York: De Gruyter, pp. 237–58.
- Munro, Tracey M., and Murray J. Derwing. 1995. Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning* 45: 73–97. [CrossRef]
- Nalborczyk, Ladislas, Cédric Batailler, Hélène Løevenbruck, Anne Vilain, and Paul-Christian Bürkner. 2019. An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research* 62: 1225–42. [CrossRef]
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online: <https://www.R-project.org/> (accessed on 20 February 2023).
- Roettger, Timo B., and Micheal Franke. 2019. Evidential Strength of Intonational Cues and Rational Adaptation to (Un-)Reliable Intonation. *Cognitive Science* 43: e12745. [CrossRef]
- Roettger, Timo B., Bodo Winter, Sven Grawunder, James Kirby, and Martine Grice. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43: 11–25. [CrossRef]
- Saito, Kazuya, and Luke Plonsky. 2019. Effects of Second Language Pronunciation Teaching Revisited: A Proposed Measurement Framework and Meta-Analysis. *Language Learning* 69: 652–708. [CrossRef]
- Savino, Michelina. 2012. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association* 42: 23–48. [CrossRef]
- Savino, Michelina, Martine Grice, and Alessandro Caffo. 2015. The Influence of Prominence on the Production of Plosives in Italian. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14.
- Sluyters, Willebrord. 1990. Length and Stress Revisited: A Metrical Account of Diphthongization, Vowel Lengthening, Consonant Gemination and Word-final Vowel Epenthesis in Modern Italian. *Probus* 2: 65–102. [CrossRef]
- Thomson, Ron. 2017. Measurement of accentedness, intelligibility, and comprehensibility. In *Assessment of Second Language Pronunciation*. Edited by Okim Kang and April Ginther. London: Routledge, pp. 11–29.
- Ulbrich, Christiane, and Ineke Mennen. 2016. When prosody kicks in: The intricate interplay between segments and prosody in perceptions of foreign accent. *International Journal of Bilingualism* 20: 522–49. [CrossRef]
- van de Schoot, Rens, Joris J. Broere, Koen H. Perryck, Mariëlle Zondervan-Zwijnenburg, and Nancy E. van Loey. 2015. Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology* 6: 25216. [CrossRef]
- van Maastricht, Lieke, Tim Zee, Emiel Kraemer, and Marc Swerts. 2021. The interplay of prosodic cues in the L2: How intonation, rhythm, and speech rate in speech by Spanish learners of Dutch contribute to L1 Dutch perceptions of accentedness and comprehensibility. *Speech Commun* 133: 81–90. [CrossRef]

- Vasishth, Shravan, Bruno Nicenboim, Mary E. Beckman, Fangfang Li, and Eun Jong Kong. 2018. Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71: 147–61. [CrossRef] [PubMed]
- Vuorre, Matti. 2021. Sometimes I R: How to Analyze Visual Analog (Slider) Scale Data? Available online: <https://web.archive.org/web/20210902234333/https://mvuorre.github.io/posts/2019-02-18-analyze-analog-scale-ratings-with-zero-one-inflated-beta-models/> (accessed on 2 February 2023).
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer. Available online: <https://ggplot2.tidyverse.org> (accessed on 2 February 2023).
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, and Jim Hester. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4: 1686. [CrossRef]
- Yang, Chunsheng, Jing Chu, Si Chen, and Yi Xu. 2021. Effects of Segments, Intonation and Rhythm on the Perception of L2 Accentedness and Comprehensibility. In *The Acquisition of Chinese as a Second Language Pronunciation. Prosody, Phonology and Phonetics*. Edited by Chunsheng Yang. Singapore: Springer. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Does Japanese/German L1 Metrical and Tonal Structure Constrain the Acquisition of French L2 Morphology?

Cyrille Granget^{1,*} and Elisabeth Delais-Roussarie²

¹ UR 4156-LNPL (Laboratoire de NeuroPsychoLinguistique), Université de Toulouse Jean Jaurès, CEDEX 9, 31058 Toulouse, France

² UMR 6310-LLING (Laboratoire de Linguistique de Nantes), Université de Nantes, CEDEX 3, 44312 Nantes, France

* Correspondence: cyrille.granget@univ-tlse2.fr

Abstract: In different studies dedicated to the acquisition of verbal morphology by bilingual children or by L2 learners, it has been noted that differences in the acquisition process cannot be accounted for by only considering the distance between L1 and L2 morphology. Some forms, such as auxiliaries, may occur in L2 productions without being motivated by L1 morphology. To account for this, the *prosodic transfer hypothesis*—according to which the acquisition of morphology in the non-dominant language is influenced by the prosody of the dominant language—has been formulated. That prosodic features may influence the acquisition of morphology is interesting as it shows that the acquisition process must be apprehended by considering interfaces and interrelations between the various levels of linguistic description. The aim of this contribution is thus twofold: (i) clarifying to which aspects of prosody *prosodic transfer hypothesis* refers (specifically, among tonal and metrical prosodic elements, which one comes into play to account for morphological development); and (ii) explaining the importance of considering grammatical interfaces in study on L2 development. To do so, an exploratory study, which relies on the analysis of L2 French narratives produced by two learners with L1 Japanese and two with L1 German, was achieved. This preliminary analysis of the data suggests that metrical structure—more precisely, the nature of the basic metrical unit—may constrain the occurrence of auxiliary and vowel-final forms in the productions of Japanese learners.

Keywords: L2 acquisition; cross-linguistic interferences; prosody-morphology interface; metrical structure

Citation: Granget, Cyrille, and Elisabeth Delais-Roussarie. 2022. Does Japanese/German L1 Metrical and Tonal Structure Constrain the Acquisition of French L2 Morphology? *Languages* 7: 305. <https://doi.org/10.3390/languages7040305>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 3 June 2022

Accepted: 23 November 2022

Published: 1 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

To account for the gradual process of acquisition of verbal morphology in inflectional second languages, it is common to show the evolution of verbal inflectional rate in an obligatory context. This rate of inflection allows us to compare different speakers and to analyse, for example, cross-linguistic influences in the acquisition of verbal morphology in a contrastive way. However, it may be insufficient to understand the specificity of certain morphological developmental aspects. Thus, recent studies have shown the limits of a binary approach to verbal morphology (inflected vs. uninflected verbs) and the interest of a qualitative analysis of verbal forms used at different stages (Benazzo and Starren 2007; Blom et al. 2013; Giuliano 2003). These studies have also highlighted the intermediate role of analytic verb forms—i.e., forms containing separate elements, such as auxiliaries as in (1)—in the acquisition of synthetic verb forms as in (2). Even if the constructions with auxiliaries do not fulfill all the functions (agreement, temporal, aspectual, agentive) assigned to them in the target languages, it is generally accepted that they are relevant clues to ongoing morphological development. The statement in (1), for example, taken from an account of family activities, is characteristic of the production of a late speaker of French.

- | | | | |
|-----|--|----------------------------|--|
| (1) | grand-mère
grand-mother
'the grand-mother is painting a boot from the Loch Ness' | [e] peint
[e] paint.3SG | un barque de la loch ness
a boot from the loch ness
(L36, Linguistic Development Corpus ¹) |
| (2) | la grand-mère
the grand-mother
'the grand-mother is painting the lake' | peint
paint.PRES.3SG | le lac
the lake
(L49, Linguistic Development Corpus) |

Only the verb forms *a peint* (has painted), with the preverbal element *a/has* or *peint* (paints/is painting), are attested in the speech of early French speakers, i.e., L1 French speakers. Studies dedicated to these constructions, also called dummy auxiliary constructions (Blom et al. 2013), have mainly focused on what motivate their occurrence and even on which functions they fulfill in the utterance, morphosyntactic vs. semantic. In (1), for example, there are at least two possible functions of the monophonemic element [e] located before the past participle form of the verb *peindre* (paint). A morphosyntactic analysis would emphasize the formal resemblance of this element with *est*, the inflected form of the auxiliary *be* in the third-person singular, and invite consideration that this element of morphological nature realizes in an economical way the agreement in person and number between the subject *grand-mère* and the verb *peindre*, as is the case for the auxiliary *do* in English (Parodi 2000). Alternatively, a semantic analysis would consider that an auxiliary form is motivated by an objective change in the temporal-aspectual context or a subjective change in the speaker's perspective on the event. Thus, the preverbal element could be interpreted as a progressive aspectual marker expressing the ongoing nature of the painting situation. This debate suggests that only one interpretation of the auxiliary would be valid at a given stage. Yet, a recent study shows that at the same stage of development in French as a second language (L2), the frequency of auxiliary verbal constructions in narrative discourse varies according to the first language (L1) of the narrators, German vs. Japanese (Granget 2018). This result suggests an influence of the first language and opens the way to other explanations than those previously outlined, which are mostly related to morpho-syntax and semantics. Phonological and prosodic features from the L1 may also come into play, as has been stated in the *prosodic transfer hypothesis* (Schlyter 1995; Goad and White 2004, among others). This is what we wish to investigate in this contribution by analysing productions in French from two learners with L1 Japanese and two with L1 German. Our aim is thus threefold: (i) to analyse the relevance of the hypothesis of a prosodic bias in the acquisition of L2 French verbal morphology by L1 Japanese and German speakers; (ii) to clarify the different prosodic domains and/or levels at which L1 could exert a constraint on the choice of verbal form in L2; and (iii) to present the results of a pilot analysis of four L2 French narratives, half of which were produced by L1 Japanese speakers and the other half by L1 German speakers.

In what follows, we first recall the main results of research on dummy auxiliary constructions and on the prosodic hypothesis. We then present the main morphological and prosodic features of the three languages under investigation—Japanese, German and French—by also clarifying the various domains/levels of prosodic analysis that could come into play. Then, we present the L2 French corpus used for the exploratory study presented in this contribution. In addition, the methodology used for annotating and analysing verbal phrases is explained. Finally, the results of this study are presented and discussed.

2. The Prosody/Morphology Interface in L2 Acquisition

2.1. The Use of Dummy Auxiliary Constructions

Research on the development of verbal morphology in inflectional second languages points to a proto-morphological stage in which dummy auxiliary verbal constructions, consisting of a lexical verb form preceded by a formally auxiliary-like but functionally restricted element, emerge (Blom et al. 2013; Noyau et al. 1995; Parodi 2000; Starren 2001). In the literature, this last element is variably called a “light or nuclear verb” (Viberg

2006), a “non-thematic verb” (Parodi 2000), a “proto-auxiliary” (Benazzo and Starren 2007; Starren 2001), or a “dummy-auxiliary” (Blom et al. 2013), the last of which is used in this article. Dummy auxiliary verbal constructions have been analysed in Dutch (Starren 2001; Jordens and Dimroth 2006; Verhagen 2011, 2013; Jordens 2012, 2013; Van de Craats and van Hout 2010), German (Schimke 2013), and French (Starren 2001; Giuliano 2003; Myles 2005; Benazzo and Starren 2007; Schimke 2013; Granget 2015, 2018). Although most studies focus on the auxiliaries HAVE (e.g., *hat*/has for German in (3) and *a*/has for French in (5)) and BE (e.g., *is* for Dutch in (4) or *est*/is for French in (6)), the dummy auxiliary class includes a variable set of elements from one study to another. For example, it also includes the semi-auxiliaries *go*, *gaat* (Dutch) or *aller* in French, as well as modals like *veut/veulent* (will) or *peut/peuvent* (can). Moreover, the copula *c’est*, */sE/* and *il y a* or the reduced form *ya/ja/* there is in French have often been considered as a preverbal central element in L2 acquisition as in (7a) and (7b) (see Noyau et al. 1995; Schimke 2013; Starren 2001; Véronique 2013).

- | | | | | |
|-----|---|---|--------------------------|---------------------------------------|
| (3) | <i>herr grün</i>
mister green | | <i>hat</i>
have.3SG | <i>schlafen</i>
sleep.INF/PRES.3PL |
| | ‘Mister Green is sleeping’ or ‘Mister Green (has) sleep(s)’ ² (Schimke 2013, p. 318) | | | |
| (4) | <i>de ouders</i>
the parents | <i>is</i>
be.3SG | <i>nog</i>
still | <i>slapen</i>
sleep.INF/PRES.3PL |
| | ‘The parents are still sleeping’ (Van de Craats and van Hout 2010, p. 491) | | | |
| (5) | <i>monsieur vert</i>
mister green | <i>il</i>
he | <i>est</i>
be.3SG | <i>dormir</i>
sleep.INF |
| | ‘Mister Green he is sleeping’ or ‘Mister Green (is) sleep(s)’ (Schimke 2013, p. 317) | | | |
| (6) | <i>il</i>
He | <i>a</i>
have.3SG | <i>dort</i>
sleep.3SG | |
| | ‘He has slept’ or ‘he (has) sleeps’ (Schimke 2013, p. 318) | | | |
| (7) | a. | <i>c’est prépar[e]</i>
it’s prepare.INF | | <i>le gâteau</i>
the cake |
| | | ‘il prepare le gâteau’ ou ‘he is preparing the cake’ (Schimke 2013, p. 318) | | |
| | b. | <i>il y a</i>
there is y.PRO | <i>have.3/2SG</i> | <i>dort</i>
sleep.PRES.SG |
| | | ‘he is sleeping’ or ‘he (has) sleep(s)’ (Schimke 2013, p. 319) | | |

Studies on dummy auxiliary constructions show also the relevance of complementing analyses of inflectional verbal morphology in terms of inflectional rate with analyses in terms of verbal forms.

Most studies have sought to determine the function of dummy auxiliaries, and most discussions have focused on the semantic vs. morphosyntactic properties of the preverbal auxiliary since there is a risk of over-interpretation if we infer the function of an element in learner varieties from the function it fulfills in the L1 (Jordens 2013; Myles 2004). In some cases, the dummy auxiliary is interpreted following Parodi (2000) as a proto-grammatical verbal element, which has no specific meaning but carries inflection (agreement), such as the auxiliary *do* in English. In other cases, it is considered as a verbal element expressing a semantic contrast, be it temporal or aspectual, depending on the first language at an intermediate stage (Benazzo and Starren 2007; Giuliano 2003; Starren 2001). Most studies on dummy auxiliary constructions have argued in favor of one interpretative hypothesis or the other. According to them, based on L2 data from speakers of the same first language who have reached different stages of development in the target language, semantic or morphosyntactic factors trigger the use of dummy auxiliary constructions.

The specificity of the singular auxiliaries *a/a* and *est/E* in French, in contrast to other inflected elements in flexional languages, is that they are monophonemic. In the French *input* they are often cliticized to the subject pronoun that precedes them, *il/il* or *elle/èl/*, and build a phonological block */ilE/*, */èlE/* and */ila/*, */èla/*. Studies of verbal morphology in French L2 have often pointed out the difficulty of perceiving the auxiliary as distinct

from the pronoun for learners who are only exposed to spoken French (Noyau et al. 1995; Benazzo and Starren 2007). In French L1 at early stages, there is no clear evidence for a phonological vs. morphological status of the preverbal element in the child's productions (Bassano 2000; Veneziano and Parisse 2010). We may expect that L2 learners exposed to written French and the visual chain of words get an early representation of pronouns and auxiliaries as free independent morphemes and do not go through this proto-morphological stage. But this is not the case since dummy auxiliaries are also used in L2 varieties produced by instructed learners of French (Granget 2015, 2018; Myles 2005). As the usual triggers of a dummy auxiliary are no longer relevant, this form may be a phonological element due to segmental or suprasegmental constraints.

In a contrastive study on the use of auxiliary constructions in the plural narrative utterances of Japanese and German learners of L2 French at an intermediate level, Granget (2018) observes that the rate of auxiliary verbal constructions depends on the first language: it is significantly higher in plural utterances produced by L1 Japanese speakers as in (8) than in those produced by L1 German speakers.

- (8) il(s) sont trainé de danse
 he/they be.3PL train.PAST.PART to/of dance
 'they are training to dance (L1 Japanese, Granget 2018)

Explaining this difference in terms of morphological development and the distance between L1 and French is rather difficult. Even if subject-verb agreement is not a relevant category in Japanese L1, the subject-verb agreement in French L2 (plural contexts) occurs more often in the production of the Japanese L1 group than in that of the German L1 group. Consequently, morphosyntactic transfer cannot explain the realizations obtained. An additional phenomenon catches our attention here: in the utterances where subject-verb agreement is realized with a plural subject, the preferred verbal form in the Japanese L1 group is the auxiliary verbal construction with a dummy auxiliary in 41.8% of the cases, compared to 4.7% of the cases in the German L1 group. The simple plural form is used in only 14% of cases compared to 35.8% in the German L1 group. This result suggests that the frequency of auxiliary constructions does not depend on morphosyntax; other properties of the learner's L1 must be involved.

As Japanese and German verbal morphology cannot allow explaining the forms observed in terms of morphological transfer (see Section 3), a prosodic transfer of the first language is likely to account for the preposed morphophonological element, as suggested by various studies that refer in such cases to a prosodic bias (Schlyter 1995) or a prosodic transfer (Goad and White 2004). Other studies also argue for the influence of L1 prosody in the realisation of epenthetic vowels (Yazawa et al. 2015; Sauzedde 2018). The aim of the following section is to review studies on prosodic bias that are relevant to the morphology/prosody interface.

2.2. Prosodic Bias, Transfer, or Epenthesis?

The occurrence of a form comparable to an auxiliary could also be considered as an epenthesis. As phonological epenthesis are often realised to satisfy prosodic constraints (well-formedness of syllable structures, stress-clash avoidance, etc.), several studies on L2 development have analysed these forms as resulting from a prosodic bias. Being influenced by the prosodic and phonological structures and patterns of their L1, learners may insert segments in the speech chain of the L2 in order to conform to their L1 prosodic patterns.

In a case study dealing with the use of different morphemes in English L2 produced by an advanced Turkish learner, Goad and White (2004) proposed to account for morphological variability and to examine the prosodic influence of the first language on the acquisition of inflected verb forms. In order to explain why postponed verbal suffixes (agreement and past inflection) and plural morphemes on nouns are significantly more frequent than preposed definite and indefinite articles, they refer to the *Prosodic Transfer Hypothesis* (PTH). Their claim is that the production of L2 inflectional morphology and function

words is constrained by the prosodic representations available in L1. When L1 prosodic representations are not identical to those required for L2, as it may appear in a contrastive analysis of the prosodic structure (syllable, foot, prosodic words, phonological phrases), they can be minimally adapted to represent the morphological material of L2. In this case, L2 speakers are predicted to build appropriate prosodic representations and produce functional morphology, as is the case for verbal suffixes in the English interlanguage of the Turkish learner. If prosodic representations are not adapted, learners are predicted to omit functional morphology, as for articles since the morphological material cannot be represented in prosodic structure. According to PTH, the distance between the prosodic structures of the verbal phrase in L1 and L2, especially at the low level, is a good predictor to account for the learnability of preposed or postposed morphemes. However, the authors do not address the particular case of our study, namely the frequency of dummy auxiliary construction due to the L1.

In a study dedicated to the acquisition of verbal morphology by children with early exposure to French and Swedish, Schlyter (1995) highlights a possible effect of prosodic dominance on verbal morphology that she calls the Prosodic Bias hypothesis. The study compares the morphological development of two bilingual children, one French dominant, Ann, and one Swedish dominant, Jean. According to Schlyter, children are sensitive to their dominant language in the way they construe verbal forms. This sensitivity is not only related to morphology (preposed auxiliary-stem as in (9) vs. postposed stem-suffix as in (10)), but also to the prosody of the language. The morphological analysis of the productions of the children consists of classifying verbal forms into two categories: preposed morphemes as in (9) and postposed morphemes as in (10). As for prosodic analysis, which is qualitative and quantitative, it consists of encoding the metrical patterns associated with small utterances. To assign a metrical form to each phrase, a distinction is made between FINAL stress patterns that are typical for French (iambic (ia), weak strong; anapestic (ana), weak- weak-strong) as in (9), where the stress syllable is in capital letters, and INITIAL stress patterns that are typical for Swedish (trochaic grave (gr), strong-weak, grave; trochaic acute (ac), strong-weak, acute; and dactylic (dac), strong-weak-weak) as in (10), as well as patterns attested in both languages (monosyllable (m), grave word preceded by a weak syllable (xgr)).

- (9) Final prosodic pattern, preposed verbal morpheme, French

ils	sont	veNUS
They	are	come.PAST.PART

 'They have come' or 'they come'
- (10) Initial prosodic pattern, postposed verbal morpheme, Swedish

KOmm-er	dom
Come.PRES	they

 'Will they come' or 'they come'

Since French spoken by adults has regular phrase-final stress, stress being culminative at the accentual phrase or clitic group (i.e., (*ils sont veNUS*) 'they have come', (*je VIENS*) 'I come', etc.), French-speaking children in a French environment easily pick up words and phrases with final stress. Those patterns facilitate the acquisition of preposed grammatical morphemes, i.e., prefixes, auxiliaries, and other preposed morphemes such as clitic pronouns in French: *il dort* 'he sleeps', *est cassé* 'is broken'. In contrast, as Swedish spoken by adults has a more variable stress pattern than French, Swedish-speaking children in a Swedish environment develop in the first stage initial stress patterns that favor the acquisition of postposed verbal morphemes.

The results of Schlyter's analysis show that the two children have different prosodic patterns due to their language dominance and different morphological preferences. The dominant prosodic pattern in the Swedish utterances of Jean (Swedish dominant due to his family situation) is a final one in the early stage, but an initial one typical for Swedish people in the late stage (age 2;2). During that period, most of the morphemes (verbal and

nominal) are post-posed (70%), as is the case in Swedish L1 acquisition (Table 1). The dominant prosodic pattern in the Swedish utterances of Ann (French dominant) is different, it is a final one during all stages, from age 2;6 to 2;10. During that period, most of the morphemes (verbal and nominal) and even all morphemes at the last stage are pre-posed, as is the case in French L1 before age 3, but not in Swedish L1 (Table 2).

Table 1. Proportion of initial and final stress pattern (% of the total number of patterns) and of pre- and postposed morphemes in the Swedish utterances of Jean (Swedish dominant).

Stage (Age)	1 (1;8)		2 (2;0)		3 (2;2)	
Stress	Initial ¹	Final ²	Initial	Final	Initial	Final
	28	62	56	38	77	28
Morphemes position	Pre-	Post-	Pre-	Post-	Pre-	Post-
	42	58	50	50	30	70

¹ (gr, ac, dac) ² iambic (ia), anapestic (ana), m, (xgr).

Table 2. Proportion of initial and final stress pattern (% of the total number of patterns) and of pre- and postposed morphemes in the Swedish utterances of Ann (French dominant).

Stage (Age)	1 (2;6)		2 (2;8)		3 (2;10)	
Stress	Initial ¹	Final ²	Initial	Final	Initial	Final
	14	80	8	71	3	74
Morphemes position	Pre-	Post-	Pre-	Post-	Pre-	
	79	21	94	6	100	

¹ (gr, ac, dac) ² iambic (ia), anapestic (ana), m, (xgr).

This case study clearly shows a close relationship between prosodic and morphological patterns and a clear prosodic bias: the dominant prosodic pattern of bilingual children constrains the position of emerging morphemes at the proto-morphological stage. According to Schlyter (1995, p. 102), it is still to be studied if the prosodic bias hypothesis is transferable to L2 acquisition since “the late acquisition of morphology in L2 may be partly due to prosodic patterns in L1 and L2, with prosodic habits from L1 which do not fit the habits of the L2”. If the prosodic bias hypothesis is true for French L2 acquisition, the stress patterns or some other metrical features of the first language may constrain the position of verbal morphemes. We also may expect that L2 learners will be able to acquire the position of French morphemes if they acquire French metrical patterns.

In de Bot’s multilingual model of speech production inspired by Levelt et al. (1999), the way syllables are realised at the surface level in an L2 is influenced by the syllabic structure inventory of the L1 (De Bot 2004). Yazawa et al. (2015) and Sauzedde (2018) have documented such influence through the production of vowel epenthesis in the speech data produced by Japanese learners of L2 English and L2 French: the insertion of a vowel between two consonants is due to constraints on syllables well-formedness. In Japanese, the basic rhythmic unit is the mora, as a consequence syllables of the form CV that consist of a single mora are highly preferred, and consonant clusters or syllable-final consonants are not allowed except in a few instances. Consequently, a vowel is often inserted to break up consonant clusters and to avoid word-final consonants (e.g., *cross*/krɔs/, CCVC in English, may be realised as /kuRosu/, CVCVCV, with epenthetic vowels /o/ and /u/ to overcome the consonant cluster and the final consonant). Even if vowel epenthesis has mainly been documented in experimental data on words and nonwords (Detey and Nespoulous 2008), it may also occur in more ecological speech data. Therefore, the verbal phrase *il mange*/ilmãʒ/, VCCVC, may be resyllabified to CVCVCV, with two epentheses realised as in /ilumãʒu/. A plausible scenario for L2 French is that this verbal form is realised /ilamãʒe/ because this form is already available in French input and has a moraic structure.

According to this hypothesis, the use of a monophonemic /a/a/, /est/é/E/ or a biphonemic auxiliary like *sont*/sõ/, which satisfies Japanese monomoraic syllabic templates, has

a double function: marking subject-verb number agreement and satisfying L1 metrical structure. In the narratives analysed for this study, the preverbal vowels or syllables clearly have the phonological forms of free morphemes (singular auxiliaries /a/ or /E/, plural, *sont/are*, [z]ont/[plurality liaison z]are, *vont/go*), but it is not excluded that this phoneme can take other forms closer to the intervocalic vowel /u/ identified in the studies mentioned above. Indeed, learners may prefer solutions that have several advantages: the forms exist in the target language and they are morphosyntactically functional, prosodically congruent, and easily pronounced.

2.3. Issues Raised and Objectives of the Paper: From Theory to Data

Despite some differences in the way to apprehend the prosodic transfer hypothesis, Schlyter (1995) and Goad and White (2004) consider that L1 prosody, be it expressed in terms of metrical patterns or prosodic structure, plays a role in the way to encode verbal inflection in an L2. According to Schlyter (1995), metrical and prosodic structure account for the linear position of verbal markers, whereas Goad and White (2004) consider that the frequency of occurrence of a morpheme depends, among other things, on the distance of the L1 and L2 prosodic representations. In our case, we do not want to compare the realisation of an inflectional morpheme, but for the same functional category, subject-verb agreement, we wish to explore how it is encoded in the L2 French discourse of learners with different L1s. The dependent morphological variable of interest is not the rate of inflection but the type of inflection, i.e., a complex form with auxiliary (mostly observed in the Japanese-speaking group) or a simple form (observed in the German-speaking group). This variable corresponds to the morphological variable observed by Schlyter (1995).

The question raised by the comparison of the two explanatory models is, which prosodic properties of Japanese would favour auxiliary forms in the L2 French spoken by Japanese learners at A2 stage? Is the accentual pattern sufficient? Does the prosodic phrase in Japanese present an accentual pattern that is transferred to L2 French and, thus, constrains the development of preverbal morphemes? Do some other metrical features of the language come into play? In order to investigate these issues, let us first present the prosodic and morphological characteristics of the languages involved, i.e., French, Japanese, and German.

3. Morphological and Prosodic Features of the Languages in Contact

The aim of this section is to present (i) the verbal elements or morphemes used to encode tense, aspect, person, and number, with a special attention on auxiliaries; and (ii) some aspects of Japanese, German, and French prosody, in particular those regarding metrical units/patterns and accentuation. This presentation will then allow formulating cross-linguistic differences and their consequences.

3.1. Morphology and Verbal Markers

Japanese and German verbal morphology differ in some aspects from French and are similar in others. The three languages have in common that they have simple verb forms as in (11), consisting of a synthetic prefixed, suffixed, or infixal lexical base, and complex forms as in (12), consisting of a lexical verb form and a free morpheme or auxiliary. However, the three languages differ in the marking of agreement, the meanings of auxiliary constructions, and the position of the inflected components of the verb.

While subject-verb agreement is a central functional category in French and German, this is not the case in Japanese. The Japanese verb form in the present tense does not vary according to number as in (11a). In German, most of the lexical verbs have an inflected present third-person singular form ending in *-t*, [t] in the spoken form, as in (11b), and an inflected present third-person plural verb form ending in *-en*, /ən/ or /n/ in the spoken form, as in (11b'). In French, there are different types of inflection in the present tense depending on the verbal class of the verb. Some verbs keep the same form in the third-person singular and the third-person plural in the present tense in spoken French. In

written French, as in (11c) and (11c'), the singular form ends in <e> and the plural form in <ent>. Those suffixes, typical for written French, have been called "silent morphemes" (Ågren 2008). Most of these verbs have their infinitive form in <-er> but some verbs do not, see *découvrir*/dekuvʁ/ / discover (3SG) and *découvrent*/dekuvʁ/ (3PL) from the verb *découvrir*. That is why Michot (2014) considers them all together as a uniform verb class (Vuni). She also considers two other classes of verbs whose plural form is different from the singular one: the class of verbs making their plural form with an additional final consonant (Vcons) as in (11d'). The verb form *disent*/diz/ corresponds to the addition of the consonant /z/ to the singular form *dit*/di/. A third class of verbs includes those making their plural in the third-person present tense with a changing stem (Vste), keeping most of the time one or more element of the consonantal architecture of the stem, e.g., the verb *savoir* (know/can), whose singular and plural forms in the third person are *sait*/sE/ and *savent*/sav/, respectively.

(11)	a.	onnanohito ni		kiki-masu	(Japanese)
		woman	to	ask.NONPAST	
	b.	sie	frag-t	eine Frau	(German)
		she	ask.PRES.3SG	a woman	
	b'	sie	frag-en	eine Frau	
		they	ask.PRES.3PL	a woman	
	c.	elle	demande	à une femme	(French)
		she	ask.PRES.1/3SG	to a woman	
	c'.	elles	demandent	à une femme	
		they	ask. PRES.3PL	a woman	
	d.	il	dit	au revoir	
		he	say.PRES.3SG	goodbye	
	d'.	ils	disent	au revoir	
		they	say.PRES.3PL	goodbye	
	e.	il	sait	danser	
		he	can.PRES.3SG	dance	
	e'.	ils	savent	danser	
		they can.	PRES.3PL	dance	

Even if agreement is not a relevant category for Japanese, verbs are indeed inflected because their form varies in tense, namely past and non-past. In (11a), the verb form *kiki-masu* is 'non-past', and contrasts with the past verb form *kiki-mashta*. The same is true for auxiliary forms where the auxiliary can be non-past as in (12) or past tense as in (13).

(12)	onnanohito ni		kiki-te	i-masu	(Japanese)
	woman	to	ask.TE	Aux.NONPAST	
	'she is asking a woman'				
(13)	onnanohito ni		kiki-te	i-mashta	(Japanese)
	Woman	to	ask.TE	Aux.PAST	
	'she was asking a woman'				

The meaning of auxiliary forms varies according to the auxiliaries and the language. In French, the constructions *avoir*/have + V (14) express the past perfective or present perfect, while the constructions *est*/be + V express these temporal-aspectual categories with motion verbs, but also passive meaning with transitive verbs.

(14)	elle	a	demand-é	à une femme	(French)
	she	AuxHave.3SG	ask.PAST.PART	to a woman	
	'she asked a woman'				

In German, auxiliary constructions with the auxiliaries *haben*/have as in (15) and *sein*/be also have past meaning, but it has often been claimed that aspect is not a relevant category for German (Klein 1994; Lasser 1997). As for the auxiliary choice, *haben*/have is

more frequently used, *sein* /be being restricted to motion verbs. The auxiliary *werden*, by contrast, is used for future and passive constructions.

(15)	sie	hat	eine Frau	ge-frag-t	(German)
	she	AuxHave.3SG	a woman	ask.PAST.PART	
	'she asked a woman'				

The auxiliary form *V-te imasu* in Japanese (12)–(13) has a quite unusual double meaning, imperfective progressive or resultative, depending on the aspectual lexical class of the verb. Moreover, it may occur in the past or non-past tense (Shirai 1998).

Syntactically, the languages also differ: French is a VO language, German is a V2 (declarative sentences) and OV (subordinate sentences) language, and Japanese an OV language. It follows that the position of the lexical verb and the auxiliary differ in declarative utterances. In French, the auxiliary is after the subject and before the lexical verb as in (14), whereas in Japanese the auxiliary is in final position, preceded by the lexical verb as in (12) and (13). In German auxiliary forms, as any inflected component, are in the second position of the sentence, and the lexical component in final position (see, e.g., *hat* and *gefragt* respectively in (15)). It is important to mention these syntactic positions as their form may be realised differently depending on the prosodic and metrical structure of the language.

3.2. Prosodic Features

This section does not offer an exhaustive description of the prosodic features of French, German, and Japanese. We are only interested in the prosodic features that are essential to accounting for metrical and intonational patterns at the level of the prosodic word and the accentual phrase (which may also be called the clitic group or minor phrase in the literature), i.e., prosodic phrases in which verbal forms are wrapped. The prosodic representation and analysis presented here are developed within an adapted version of the AM model (see Ladd (2008) for a review). In this framework, the prosody associated with an utterance is represented by means of two distinct representations or structures, the metrical structure and the tonal profile. Metrical structures encode which mora, syllables, or other units (i.e., foot, prosodic words, etc.) are prominent and explain at which level of structuring stress is culminative (Lieberman and Prince 1977; Prince 1983, among others). The tonal pattern consists of a linear sequence of pitch accents (associated with stressed or metrically strong positions in metrical structures) and edge tones associated to the edges of prosodic phrases, especially intermediate and intonational phrases (Pierrehumbert 1980; Pierrehumbert and Beckman 1988; Ladd 2008, among others). The two types of representations are independently constructed, but language-specific association principles are necessary. In order to present and compare the prosodic features of the languages under investigation, we will first present the metrical features and then the tonal ones. But, as previously said, the descriptions will only focus on the metrical and tonal structure up to the level of the prosodic word and the accentual phrase (AP). Basic metrical units, i.e., units that can be prominent and associated with a pitch accent, are the mora and the syllable. At the phrasal level, we refer to prosodic or phonological words as the domain of primary stress assignment in many languages and accentual phrases.

In any language, metrical structure is construed from a basic metrical unit. In Japanese, this unit is the mora, whereas in French and German it is the syllable. Japanese is thus clearly different from German and French. Based on Japanese linguistic tradition, as well as on the study of versified poetry and on numerous language games, Labrune (2012) has shown that the syllable seems to have no cognitive reality in Japanese. In contrast, in German and French, the syllable is the basic metrical unit (Wiese 1996 among others), and it plays an important role in versification. As for syllable structure, there are differences between the three languages. In German, complex syllable structure with consonant clusters and/or coda (CVC, CCV, VCC, etc.), which also appear in French, are frequent. In Japanese, by contrast, syllables are usually of the form CV in order to coincide with a mora.

Concerning stress patterns, in German and Japanese, stress is culminative at the level of the prosodic word, i.e., among the basic metrical units that compose a lexical word, one is more prominent, i.e., considered as stressed or accented. By contrast, in French, stress is culminative postlexically at the level of the accentual phrase. As for the localization of the stressed/distinguished metrical unit, stressed syllables are usually in the rightmost trochaic foot of the word in German (Wiese 1996, among others); and, in terms of realisation, the stressed syllable does not necessarily receive a tonal marking. As for Japanese, the location of the strong mora cannot be derived straightforwardly, but it is given in the lexical representation. Moreover, the strong or accented mora is always realized by a melodic movement that corresponds to a melodic fall from the prominent mora, noted as H*+L (Venditti 2005). In addition, note that some words remain unaccented. As for French, it has no lexical stress, which could allow distinguishing lexical words having the same phonemic form. Nevertheless, the elaboration of French stress patterns derives from underlying metrical templates. Their construction is based on (i) word classification and (ii) the principle of bipolarity (Di Cristo 1999, among others). The different words of the language can be classified into two classes according to whether they are capable or not of receiving a final stress on their last syllable (unless their nucleus is a schwa). A [+stress] word is any word that can receive a final accent, and a [−stress] word is any word that never receives a final accent. Determiners, weak pronouns, the complementor *que* ‘that’, the negative prefix *ne*, and monosyllabic prepositions such as *à*, *de* or *en*, i.e., monosyllabic grammatical words, are generally [−stress]. Other words, such as nouns, verbs, adjectives, interrogative pronouns, and adverbs, are [+stress]. In terms of metrical patterns, [−stress] words are represented as a simple sequence of weak syllables, whereas the initial and final syllables of [+stress] words are strong (s), as shown in (16a) and (16b), respectively.

- (16) a. Metrical templates associated with [−stress] words
 ɔ̃ə (je) lE (les) ā (en)
 w w w
- b. Metrical templates associated with [+stress] words
 demain livre partira
 ‘tomorrow’ ‘book’ ‘will leave’
 də mē livR paʀ ti ʁa
 s s s s w s

Note that at the lexical-level final strong position does not dominate initial strong position in French. It results from the fact that stress is not culminative at this level. Thus, depending on the context, the noun *chaton* ‘kitten’ may be stressed on the initial syllable or on the final one, and the forms *chaton* /'ʃa.tɔ/ and *chaton* /ʃa.'tɔ/ always refer to the same word. The same occurs for *demain* (16b). In the prosodic phrase *demain soir* in the utterance in (17a), an initial accent is often realized on the syllable /də/ and the final one falls on /swaʀ/, the syllable /mē/ remaining unstressed. In contrast, in the prosodic phrase *demain* in (17b), the syllable /də/ is unstressed and /mē/ is. The representation in (16b) thus represents the fact that both positions can potentially be stressed at a higher level in conformity with the principle of bipolarity (Di Cristo 1999).

- (17) a. *Demain soir,* *Pierre viendra*
 Tomorrow evening Peter will come
 ('də.mē.'swaʀ)_{AP}
- b. *Demain* *Pierre viendra*
 (də.'mē)_{AP}

In French, German, and Japanese, utterances are segmented into larger prosodic phrases marked tonally (APs, intermediate phrases, and intonational phrases). Among them, focus is given here to any prosodic phrases containing minimally a pitch accent, and sometimes edge tones or accents. In French, the left edge of the AP is associated with a L tone, whereas its right edge receives a pitch accent ((L)H*) on the last metrically

strong position (see (17b)). Moreover, when the AP contains more than three syllables, an internal accent with the form of a rising melodic movement (LH) may be realized on a strong metrical position (initial or final prominent syllable in (16b)). By and large, the melodic profile associated with an AP is thus of the form [L (H*) L H*] (Post 2000; Jun and Fougeron 2000; Delais-Roussarie et al. 2015, among others). In Japanese, the melodic pattern associated with the accentual phrase is either (H- H*+L L-) or (H- L-), depending on whether the AP includes an accented or an unaccented lexical word (Pierrehumbert and Beckman 1988; Venditti 2005). As for German, tonal patterns are assigned at the level of the intermediate or intonational phrase (Wiese 1996; Truckenbrodt 2005, 2007). However, according to Uhmann (1991) and Truckenbrodt (2005, 2007), pitch accents are usually realized on primary stressed syllables in a prosodic word, and pitch accented prosodic words are predictable on the basis of information and syntactic structure. Moreover, this accent is realized by a L*+H pitch movement when it does not coincide with the IP final accent. As for the IP final accent, it is of the form H+L*. Following Truckenbrodt (2005, 2007), we may consider that any XP (NP subject, NP object, VP) receives a pitch accent as shown in (18).

- (18) Die Lena will dem Werner im Jänner ein Lama malen
 (die LEna) (will dem WERner) (im JÄnner) (ein LAmA malen)
 L*+H L*+H L*+H (H+)L* L%
 (Truckenbrodt 2005, example (4), p. 277)

In terms of phonetic implementation, accented syllables are realized by means of a pitch movement in German and Japanese, whereas they are also lengthened in French as they coincide to the right edge of a prosodic phrase.

3.3. Summary and Hypotheses

At the morphological level, inflections and, in particular, verbal auxiliaries are attested forms in all the languages studied. Hence, no difference in the rate of inflection and auxiliary forms should be observed in the French L2 speech spoken by Japanese or German learners. Nevertheless, the study of subject-verb agreement in the third-person plural in an oral narrative corpus in L2 French shows that agreement in number is more frequent in the narratives of L1 Japanese speakers than in those of L1 German speakers, but this agreement is achieved by means of an auxiliary in 42% of the cases in the productions of L1 Japanese speakers, compared to only 4% of the cases in the narratives of L1 German speakers (Granget 2018). Since the differences observed cannot be attributed to the verbal morphology of the learner L1, other explanatory paths need to be explored, such as the prosodic bias hypothesis as formulated by Schlyter (1995).

Prosodically, German, Japanese, and French differ along several dimensions. Metrically, the basic metrical unit is the syllable in German and French, whereas it is the mora in Japanese. This difference has implications for syllable structure: consonant clusters and coda are frequent in German, and to a lesser extent in French, whereas syllables of the form CV are usually observed in Japanese. As a consequence, Japanese L1 speakers may prefer CV syllables. As for stress, it is assigned and thus culminative at the lexical level in Japanese and German, in contradistinction to French. At the level of APs, unaccented APs, i.e., APs with no syllable receiving stress, are possible in Japanese, but not in French and German.

As for tonal patterns associated with prosodic phrases, in French and Japanese, tones are realized at the edge of accentual phrases, be it a pitch accent as in French or an edge tone as in Japanese. By contrast, in German, edge tones are only associated at the level of the intermediate or intonational phrase. Note, however, that a pitch accent may be associated with the primary stressed syllable of a lexical word. Consequently, the tonal marking of APs should not be a problem for learners with L1 Japanese, in contradistinction to those with L1 German. Note, however, that features concerning tonal marking at the level of the

AP should not have an impact on morphological development (see, among other, previous studies on prosodic bias, Schlyter 1995; and Goad and White 2004).

4. Data and Methods

4.1. Corpus and Materials

4.1.1. Corpus Recording Protocol

The data used for this exploratory study were collected in a European research program on crosslinguistic influences on subject-verb agreement³. In fact, they are extracted from a larger corpus that consists of oral productions of Japanese and German learners at A2 or B1 level according to the CEFR. The level was evaluated by mean of the DIALANG lexical test (Alderson 2005).

The participants were recorded in a quiet room in a university environment. They were asked to tell a story based on 30 images projected on a computer screen one by one. This story, entitled *Paul et Pauline font la fête* ‘Paul and Pauline are having a party’, adapted from Ågren and Van De Weijer (2013), has been designed to account for the acquisition of number agreement in L2 French verbal morphology (Ågren et al. 2021; Granget et al. 2021).

Before recording the narratives, the following instructions were given to the participants: (i.) they had to look at all 30 pictures to understand the narrative before starting to tell the story orally; (ii.) they were then asked to tell the story with the help of the visual support shown on the computer screen; and (iii.) they had to consider the events to be happening now and, thus, retell the story in the present tense. This last point aimed to encourage the use of present tense and simple verbal forms, and it was recalled by the interviewer before each recording session. Note also that lexical help was provided on request during the preparatory phase prior to the narration proper, and very rarely some lexical help was also given during the recording phase.

4.1.2. Participants

For the current study, four participants were selected on the basis of their linguistic profile: two participants in each L1 group (Japanese and German). To increase comparability, only female participants at A2 level were chosen; moreover, their amount and type of exposure to French, their age at the time of recording, and the duration of their learning were more or less the same.

At the time of recording, the speakers with Japanese L1 were living in France. Consequently, for the German speakers, priority was given to German L1 students who spent time in France or declared in the questionnaire that they had had contact with French people outside the classroom. The relevant information is summarized in Table 3.

Table 3. Participants chosen for the pilot study.

Speaker ID and L1	Level	Age at the Time of Recording (First Exposure)	L2 “Spoken” in Addition to English	Level in English (Self-Assessment)
JAPA201 (J01)	A2	43 (37)	Korean	1
JAPA202 (J02)	A2	26 (18)		2
GERA207 (G07)	A2	43 (13)	Latin, Italian	4
GERA205 (G05)	A2	26 (12)	Spanish	6

As Table 3 shows, the learners’ multilingual repertoires are not strictly comparable in terms of previously learned L2s and the self-reported level of English: the 43-year-old German-speaking learner reports having learned Latin and Italian and having a self-assessed level of 5 in English on a scale ranging from 1 to 6, whereas the Japanese-speaking learner of the same age reports having learned Korean and having a level of 1 in English. According to multilingual models of speech production (De Bot 2004), L2 English might have an influence on L3 French verb forms, which is greater in (German) advanced learners

than in (Japanese) intermediate learners. However, it is very difficult to achieve such a high level of comparability in any research on L2 acquisition; thus, this issue should certainly be investigated in further studies.

Because of the limited number of participants, we are also conscious that the study cannot produce generalizable results. Our first intention here is rather to show the interest of looking at metrical and prosodic features in a cross-linguistic manner in order to account for morphological development. More widely, it is important to consider the interrelation that could exist between the various grammatical levels of linguistic description (phonology, morpho-syntax, semantics, prosody, etc.) when studying L2 acquisition.

4.1.3. Data Used

The data recorded with the protocol mentioned above are gathered in a corpus designed to carry out research on cross-linguistic influences in the acquisition of Subject-Verb agreement in French L2. The narratives were produced by speakers with L1 having different degrees of morphological richness, e.g., Dutch, Italian, French, German, and Swedish (Ågren et al. 2021; Granget et al. 2021). A sub-corpus was constructed from the productions of 14 speakers with German (7 speakers) and Japanese (7 speakers) as a L1, which were divided into two subgroups according to their lexical level: one subgroup estimated at A level and another subgroup estimated at B level (Granget 2018). The present case study uses data from this sub-corpus: the 4 narratives contain 1940 words, among which 1205 were produced by the Japanese learners and 735 by the German learners.

4.2. Methods and Annotation Protocol

The narratives were transcribed orthographically according to the protocol described in Section 4.2.1. As for the morphological annotation, we used the morphological protocol described in Section 4.2.2. and designed to account for subject-verb number agreement in third-person context, as well as an ad hoc protocol to account for dummy auxiliary constructions (Granget 2018). As for the prosodic annotation, we used the protocol presented in Section 4.2.3. Morphological and prosodic information was encoded from a careful listening of the audio forms and not from the transcriptions.

4.2.1. Orthographic Transcription

When the content and the form of the words used in the narratives were clearly identifiable, they were transcribed orthographically, sometimes with a simplified and adapted orthography, using the CHAT format in CLAN (MacWhinney 2000). Silent morphemes or letters were used with verbs in ambiguous contexts since they are the only cues indicating if one or more protagonists are involved in the situation. The silent plural morphemes <-s> and <-nt> are written in *ils préparant* (19) because the plural and singular forms of the pronoun *il/ils* and the verb *prépare/préparent* are homophonous. However, the silent letter <-t> is not written in *boi* and *von* in (20) and (21) because *boi/bwa/is* is a non-ambiguous singular form, in opposition to the plural form */bwav/*, here in a plural position, whereas *von/vō/is* is a non-ambiguous plural form (opposed to singular */va/*) in a plural position as well.

Audio and text were also aligned at roughly the level of the inter-pausal unit. The choice of orthography for oral data is motivated by the fact that this level of annotation allows (i) readability of the data and (ii) immediate access to the meaning of (frequent) homophones, as in French (see also on this issue Delais-Roussarie and Post 2014, among others). In addition, in all cases of hesitation or unexpected forms, the International Phonetic Alphabet (IPA) was used as in (22).

- (19) &euhm et ils préparent un cadeau pour ## un ami je crois
&hm and they prepare a gift for ## a friend I think (G05)
- (20) et boi [*] de [/] de limonade [/] de la limonade
and drinks [*] limonade limonade (G07)
- (21) ils von plépalé de de cadeau pour son ami
they will prepare some some gift for his friend (J01)
- (22) et paul &èh sectionne [/] choisit les pantalons et [/] et /kɛvɛt/[/] /kɛvɛt/oui [=! rires]
and paul sect [/] chooses the pants and and tie tie yes [=! laughs] (J02)

The spelling has been enriched by diacritics to encode pauses (#), repetition ([/]), reformulation ([//]), hesitations (&), para-verbal information (!=), and morpho-phonological errors on the verb [*], as shown in the above transcribed utterances (19, 20, and 22). All of these symbols are requested by CLAN to allow a proper use of the automatic analysis module. Pauses are encoded by # from an attentive listening of the audio data, and the longer the pause is perceived, the more important the number of #s. The orthographic representation sometimes differs from standard orthography, despite EAGLES and TEI's recommendations on this issue (see, for instance, (21)), and includes symbols to allow CLAN to perform an automatic morphological annotation with the use of tools, such as the morphological tagger (%mor). Note, however, that we did not use the tools that were mostly designed for research in L1 acquisition. In (21), the verbal form derived from the verb *préparer* 'to prepare' is transcribed *plépalé* to best reflect the learners' pronunciation [plepale].

As for simplified spelling, it is mostly used in cases where verbal agreement markers, which are usually encoded in French orthography, are not audible (the so-called silent morphemes), such as <-nt> in *arrivent* (arrive.3PL), pronounced [aʁiv], identical to *arrive* (arrive.3SG). Chosen an orthography over the other in cases of silent morphemes would reflect the representations of the transcriber and not necessarily those of the learners. That is why a simplified orthographic form is used to account for the indeterminacy of the morphological representations. For all of these reasons, we decided to use an ad hoc protocol for morphophonological annotation that does not presuppose the existence of grammatical or semantic functions attached to verb forms.

4.2.2. Morphological Annotation

The morphological annotation protocol used in this cross-linguistic project on the acquisition of subject-verb agreement in French L2 was designed to account for subject-verb agreement in number in third-person contexts. Consequently, only utterances with a third-person subject were coded morphologically using a special tier (%over) for annotating the following information:

- the number of protagonists involved in the situation described according to the picture (Context): one participant (ContSing for singular contexts), more than one (ContPlur for plural contexts), or ambiguous (ContAmb for ambiguous or indeterminate contexts);
- the type of verb used by the learner (Vuni, Vcons, Vste as described in 3.1 and Vont for verbs whose form in the third-person plural ends in /ō/like *von/vō/*);
- if the lexical verb was preceded by a classical auxiliary form, *avoir*/have or *être*/be as is the case in (23);
- the type of subject (nominal, pronominal, both, relative, etc.);
- the form of the verb (accurate morphophonological form, which is encoded Facc and includes the existing form, e.g., singular instead of plural, and inaccurate and nonexistent form)

- (23) *STU: il a brossé ses chaussures
%ver: abrosse&ContSing&Vuni&aux&Facc
he has brushed his shoes

According to this annotation protocol, utterances with a simple invariable verb form (Vuni) in singular and plural contexts were excluded (e.g., *préparent* (3PL)/*prépare* (3SG), both pronounced [pʁɛpaʁ] as in (19)), whereas complex (auxiliary) verb forms, such as *a brossé*/brushed, were not excluded because the auxiliary is not an invariable form. When the verbal form is reformulated as in (22), only the last form is considered, i.e., *choisit*. For the purpose of the present analysis, we analysed the global rate of agreement in plural and singular contexts, not only in plural contexts as in Granget (2018). The rate of agreement was calculated on the basis of the selected verb forms by means of the following formula:

$$\begin{array}{c}
 \text{number of plural verb forms (simple } \textit{boivent/bwav/} \text{ and complex } \textit{ont bu/\delta} \text{ by/)} \\
 \text{in plural contexts} \\
 + \text{ number of singular verb forms (simple } \textit{boit/bwa/} \text{ and complex } \textit{a bu/a by/)} \\
 \text{in singular contexts} \\
 = \text{ number of accurate verb forms} \\
 \text{————— (divided by)} \\
 \text{total number of verb forms in plural and singular contexts} \\
 (\textit{boivent, ont bu, boit, a bu, boivu, sont boi, etc.})
 \end{array}$$

The principal limit of this annotation system is that it is based on a restricted definition of auxiliary in which only *avoir*/have and *être*/be are considered. In order to account for the broad class of dummy auxiliaries in the sense of Blom et al. (2013) or Starren (2001), we used an ad hoc protocol for the present study that considered as dummy auxiliary constructions (encoded as DAC in the annotation tier) all verbal constructions containing a lexical verb—be it an accurate, an infinitival, or a past participle form- or not (e.g., *plépalé, arrivé, sorti, mette*)—preceded by a form of the following verbs: *avoir/to have* (*a/has, ont/have*, including forms preceded by the plural consonant [z] as in *ils ont* [ilzɔ̃]/they have); *être/to be* (*est/is, sont/are*); *aller/to go* (*vont/go.3PL*); *c’est/it is*; and *pouvoir/can* (*peut/can.3SG*).

- (24) &euhm et ils préparent un cadeau pour ## un ami je crois
&hm and they prepare a gift for ## a friend I think
%ver: préparent&ContPl&Vuni&Facc (G05)
- (25) pauline collé [//] pauline a collé &euh le [lə] [ləbato:]
pauline pasted [//] pauline has pasted &euh the the boat
%ver: acollé&ContSg&Vuni&DAC&Facc (J02)
- (26) il von fai de danser
they go do to dance
%ver: vonfai&ContPl&DAC&Vont&Facc (J01)

The above examples from (24) to (26) illustrate the descriptive categories we used in the ad hoc annotation protocol (%ver) to account for the rate of dummy auxiliary constructions (DAC) among all verb forms. Among these examples, DAC are observed in (25) and (26).

4.2.3. Prosodic Annotation and Analysis

Prosodic annotation has been done using the prosodic annotation system IV, which derives from IViE (Grabe et al. 2001; Delais-Roussarie and Post 2014). On the basis of a careful listening of each utterance or inter-pausal chunk, prominent syllables are encoded and then categorized as corresponding to initial or final stressed syllables depending on their position within words or phrases. As for the tonal patterns, the tonal movement associated with prominent syllables and edges was determined on a perceptive basis by listening to the implementation domain. Roughly, such a domain corresponds to one of the following chunks:

- from a pause to a prominent syllable and the following non-prominent one;
- from an already tonally encoded prominent syllable to the next prominent syllable and the following non-prominent one;
- from an already tonally encoded prominent syllable to the next prominent syllable when followed by a pause.

The tonal movements were encoded by assigning a relative tone to the prominent syllable, as well as to what precedes and follows, with the tone associated to the prominent syllable being in a capital letter (e.g., *LMh* for a rising movement that continues after the prominent syllable, *hL* for a fall on the accented syllable, etc.). These movements were then translated into pitch accents of the form LH*, H*, and L*, or into edge movements of the form L-/L% and H-/H%.

Apart from this annotation, a prosodic analysis was achieved for all third-person verbal forms, be they realized on one or more prosodic phrases or APs. Take for instance (26): the sequence *ils vont fai* is selected for the analysis, and it could have been realized in a single AP (*ils vont fai*) or in two (*ils vont*) (*fai*). Note, however, that sequences with hesitations or repetitions that made prosodic analysis difficult or even impossible were discarded. Moreover, when the verbal form was fully repeated as in (25), only the repetition was considered. Because of the mentioned restrictions, only 106 forms (50 produced by the two German learners and 56 by the Japanese learners) were retained for the prosodic analysis. The prosodic annotation achieved with the IV transcription system was then used to determine how verbal forms were wrapped into accentual phrases (AP). In addition, the tonal pattern associated with each phrase was noted, and we calculated the number of syllables within each AP and examined the structure of each syllable (CV, CVV, CCV, etc.).

5. Results and Discussion

The results presented here were obtained from the analysis of the forms according to the method exposed in Section 4. As already mentioned, the morphological analysis aimed to calculate both (i) the rate of agreement in number and (ii) the use of preverbal auxiliaries in the 138 forms under investigation. As for the prosodic analysis, it aimed to evaluate the characteristics of the APs produced by the four learners, as well as the syllabic structures usually observed. Japanese L1 speakers produced 77 forms (J01, 28 and J02, 49) compared to 61 for German L1 speakers (G05, 23 and G07, 38). One third of these statements are singular (48 statements) and two thirds are plural (90 statements).

5.1. Auxiliary Constructions

The morphological analysis of the 138 verb forms confirms the previous analyses carried out on a larger corpus. In order to analyse the frequency of agreement, only 95 verbal forms were taken into consideration, with 43 forms of the present tense of the third person having been excluded from the analysis because of their invariability in number in the spoken form ([mãʒ] eat, [uvɛ] open, [dãs] dance, etc.). In these data, no a priori L1-related differences were observed: 63% of the 95 verb forms agree in number (i.e., 60 forms), with 57% in the two L1 German and 67% in the two L1 Japanese. But there are important inter-individual differences (J01, 41%; G05, 53%; G07, 60%; J02, 88%).

Among the 60 agreed verb forms, we distinguish simple synthetic forms from complex forms with an auxiliary. The analysis shows that in the two French L2/Japanese L1 narratives, dummy auxiliary constructions are much more frequent than in the two French L2/German L1 narratives. Indeed, they represent 5% of the verb forms that agree in number in the German L1 narratives, but 72.5% in the Japanese L1 narratives, despite the fact that present tense, i.e., a simple verb form, was required. Here there are interindividual differences, but the likelihood of dummy auxiliary constructions clearly depends on the L1: G05, 12.5%; G07, 0%; J01, 81%; J02, 68%. This analysis of singular and plural contexts confirms that L1 Japanese speakers favor the use of a dummy auxiliary construction.

5.2. Prosodic Analysis

The analysis of the prosodic phrases in which the verbal forms were wrapped focused on their size and internal composition. There are two reasons for this. First, learners' productions, whatever their L1 (Japanese or German), lack fluency; thus, each phrase is often separated from the next by pauses and hesitations. As a result, phrases can be analysed in a way as an intermediate or intonational phrase (I-Phrase). Second, the

intonational contours associated with I-phrases and the form of the pitch accents associated with metrically strong syllables often present characteristics from the speaker’s L1. Thus, it is relatively frequent that in the productions of German speakers, prosodic phrases are longer than in French and contain pitch accents on stressed syllables. The tonal realization of these accents is often similar to what is observed in German. In (27), the accent on the syllable /*rɑ̃ʒ*/ (from the verb *rangent*) is of the form L*+H as in German, and a rising edge tone H- appears at the end of the I-phrase. This implementation, shown in Figure 1, is close to the German one, in which non-final pitch accents are of the form L*+H.

- (27) [ils rangent les affaires]_I
 They clean up the things
 ‘They clean up the things’ (G05)
 il rɑ̃ʒ le zafɛʁ
 L*+H L+H* H-

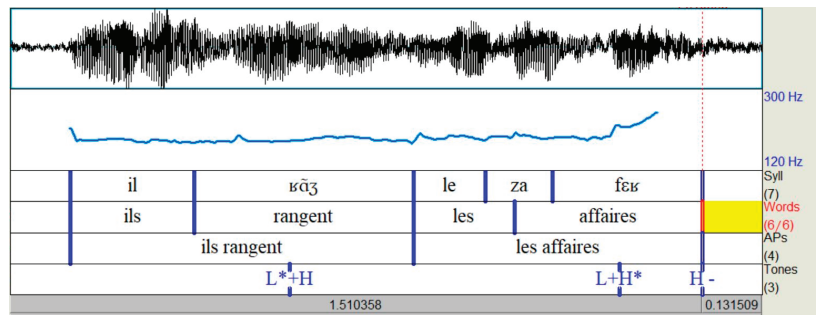


Figure 1. Waveform and pitch track associated with (27).

As for the Japanese learners’ productions, they are characterized by short APs of two syllables as in (28). In addition, APs often receive edge tone of the form H- at the left edge and L- at the right edge, whether they contain a pitch accent or not. In (28), the AP (paul) receive a H* pitch accent on the syllable /*po*/ and a L-edge tone, whereas the APs (ne peut pas) and (danser) just receive an initial edge tone H- and a final one L- as the final syllables of these APs are realized with a flat or slightly falling pitch contour and a significant syllabic lengthening, as shown in Figure 2.

- (28) [(Paul)_{AP} (ne peut pas)_{AP} (danser)_{AP} (bien)_{AP}]_I
 H* L- H- L- H- L- H* H% (J01)
 Paul cannot dance well

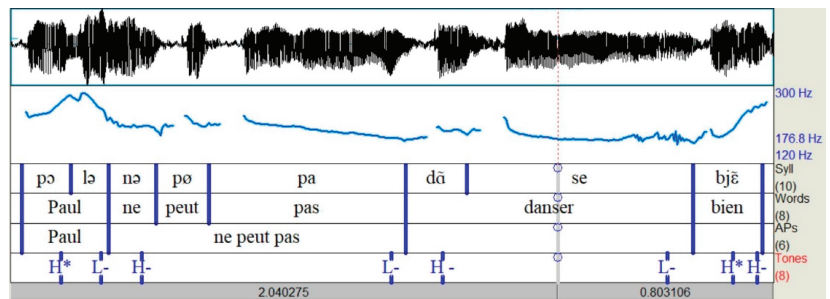


Figure 2. Waveform and pitch track associated with (28).

In summary, the rhythmic patterns and intonational contours observed in the different narratives are highly influenced by those of the speaker’s L1; thus, they cannot explain the

differences observed in the morphological development between German and Japanese learners.

Therefore, we decided to focus on verbal forms in order to evaluate how they were prosodically realized. This showed that the 106 verbal forms studied were wrapped in 121 accentual phrases (APs), a verb form sometimes being realized under two APs (e.g., *ils vont plépalé*/they will prepare is phrased as (il vont)_{AP}(plépalé)_{AP}; the same for example (28)). This segmentation in AP occurred especially for utterances in which verb forms were composed of an auxiliary. Moreover, it was much more frequently observed in the productions of the Japanese learners. The 50 verbs uttered by German L1 learners are realised on 51 APs, whereas the 56 verb forms produced by Japanese L1 speakers are realised on 70 APs. In addition, a closer analysis of the segmentation in APs showed that 12.6% of the Japanese speakers' APs are composed solely of elements that are unaccented in French (pronouns and auxiliaries of some sort), whereas the proportion is 1.6% for German speakers. This could result from the fact that APs in Japanese can be composed of unaccented words (see Section 3.2 and also (ne peut pas)_{AP} in (28)).

As for AP size, the APs realised by Japanese learners are slightly shorter (2.91 syllables/AP on average) than those by German speakers (3.25 syllables/AP on average). This may result from the minimality constraint at work in Japanese. Indeed, prosodic words, and consequently APs, often contain only two morae.

As far as syllabic form is concerned, the analysis indicates that Japanese speakers clearly prefer verbal forms ending with a CV syllable. Of all the final syllables of the verbal nuclei they produced, 80.21% are of CV form and 19.8% of CVC form, whereas for L1 German learners, the proportion is 49% and 51%, respectively. Auxiliary and schwa insertion in verb-final position thus correspond to a more general modification of verbal forms to allow resyllabification in accordance with Japanese CV moraic templates. Indeed, in introducing an epenthetic vowel (/a/or/E/) or a syllable (/zō/,/sō/), auxiliary use allows the emergence of monomoraic CV syllables (e.g.,:/i.zō.ɔ̃.sy/'ils ont reçu'/they have received vs /il.ɔ̃.swav/'ils reçoivent'/they receive). Similarly, the insertion of schwas in the case of consonant cluster or a coda leads to the production of CV syllables. In the case of monosyllabic verbs, it also contributes to satisfying the minimality constraint (two morae for a prosodic word or APs).

5.3. Discussion

The analysis of the data shows that, morphologically, Japanese learners use more dummy auxiliary constructions than German learners, regardless of the auxiliary used. As neither the morphological differences between the two L1s and French, nor the levels of the learners (that is equivalent) can explain this, the validity of the prosodic hypothesis as formulated by Schlyter (1995) was explored.

Among the prosodic features, special attention was given to metrical and tonal patterns up to the level of the accentual phrase as verbal forms are usually wrapped within APs in French. Concerning the rhythmic and tonal patterns observed at the level of the AP and higher, they are influenced by the learner's L1 in all cases. Indeed, the tonal patterns associated with APs shows that, even when verb forms are accurate (for instance, for the German L1 speakers), some prosodic features of the L1 are present. In fact, the APs produced by German L1s are longer, and above all, frequently realised with a melodic contour similar to what is often observed in German. The prosody of the L1 thus always plays a role independently of the morphological development. Thus, the overuse of dummy auxiliaries in the productions of the Japanese learners cannot be explained from prosodic constraints applying at a higher level in the prosodic structure (accentual, intermediate, and intonational phrases).

As Japanese differs metrically from German and French by having as basic metrical unit the mora and not the syllable, this difference was investigated more thoroughly. It appears that in Japanese, the use of auxiliaries and the past participle, as in "*pauline a collé e: le bateau*", allows the transformation of the verb form *colle*, metrically CVC (i.e., an

impossible form in the Japanese moraic-driven syllabic inventory), into a V.CV.CV form. In the same way, the use of the verbal form *vont fai* in “*ils vont fai de danser*”, realised [i.vɔ̃.fɛ] with the omission of the /l/ and the /ʁ/ in respectively *ils* and *faire*, is again of the form V.CV.CV, more compatible with the moraic-based metrical structure of Japanese. The presence of the auxiliary form could thus be motivated by metrical phonological constraints related to the nature of the basic metrical unit and its manifestation in the syllabic inventory.

The function of auxiliaries in L2s has been the subject of much discussion and analyses proposing a syntactic or semantic interpretation of these preverbal morphological elements, but this is not satisfactory to explain what really occurs. The analysis proposed here considers, following Granget (2018), that the sound dimension should also be considered. It appeared from the preliminary study that metrical principles related to basic units (i.e., syllables vs. morae) exert constraints on the segmental realisation of the verbal forms. The use of auxiliaries and past-participles as well as the insertion of a schwa are thus motivated to guarantee the well-formedness of the metrical templates associated with each word. In other words, at the observed stage, verbal forms are not only morphosemantic or morphosyntactic elements, but also morpho-prosodic elements. The vowel between the subject and the verb can be analysed as a prosodically constrained morpheme, or just as an epenthetic vowel constrained by Japanese metrics as it also occurs in loanwords or in lexical acquisition by Japanese-speaking learners of L2 French at a low intermediate stage (Sauzedde 2018).

6. Conclusions and Perspectives

The aim of this contribution was twofold: (i) analysing French verbal forms produced by learners with L1 German and Japanese in order to evaluate whether the overuse of auxiliaries in the productions of the Japanese learners could be explained by means of the *prosodic transfer hypothesis*; and (ii) explaining the importance of considering grammatical interfaces in study of L2 development. Concerning the first point, the analysis of the data showed that the presence of dummy auxiliaries in the forms uttered by the learners with L1 Japanese is probably due to the moraic structure of this language, which leads to preference for a syllable of the form CV. As for the second point, the results obtained clearly show that one cannot analyse language development in a modular manner, without considering other levels of linguistic description. Indeed, the occurrence of dummy auxiliaries in the data investigated appear to be motivated by metrical and phonological constraints.

In order to evaluate more precisely the weight of prosodic-phonological constraints on morphology, further research is necessary. As vowel epenthesis appears elsewhere in the production of the Japanese learners, can we really account for the occurrence of these vowels or forms in terms of dummy auxiliaries and not just in phonological terms (e.g., vowel epenthesis, metrical filler, etc.)? Would the forms analysed as dummy auxiliaries disappear as soon as learners modify the surface forms of subject pronouns, as is often done in standard spoken French (*ils ont* uttered as [izɔ̃] instead of [ilzɔ̃])? In order to investigate the nature of the correlation between morphology, phonology, be it segmental or metrical, in the acquisition process, we may also wonder whether a better control of morphology implies a suppression of auxiliary forms and epenthesis. In order to pursue this exploratory work, we will continue the analyses on a larger sample and in more developed narratives to observe the extent to which epenthesis occurs. This should allow understanding its relation to morphology. This can also be done by comparing it with written productions to verify whether auxiliaries would be as frequent.

Author Contributions: Conceptualization, C.G.; methodology, C.G. and E.D.-R.; collection and transcription, C.G. among other previous collaborators; new annotation C.G. and E.D.-R.; formal analysis, C.G. and E.D.-R.; writing—original draft preparation, E.D.-R.; writing—review and editing, C.G. and E.D.-R. All authors recognize that the present analysis has benefited from previous collaborative research mentioned in the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding authors. The data are not publicly available due to restrictions in data collection policy.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ French Oral Learner Corpora (FLOC), <http://www.floc.soton.ac.uk>. (accessed on 15 November 2022).
- ² For examples (3)–(7), we have proposed several translations as these are the results of our interpretation. Moreover, the authors of the articles from which these examples were taken did not always provide a translation.
- ³ This research program is not funded by a third party and not limited in time. It involves Marie-Eve Michot (Vrije Universiteit Brussels) who initiates the research group, Malin Ågren (University of Lund), Pascale Hadermann (University of Ghent), Sonia Gerolimich (University of Udine), Isabelle Stabarin (University of Trieste) and Cyrille Granget (University of Toulouse).

References

- Ågren, Malin, and Joost Van De Weijer. 2013. Input frequency and the acquisition of subject-verb agreement in number in spoken and written French. *Journal of French Language Studies* 23: 311–33. [\[CrossRef\]](#)
- Ågren, Malin, Marie-Eve Michot, Cyrille Granget, Sonia Gerolimich, Pascale Hadermann, and Isabelle Stabarin. 2021. “Les copains *dit au revoir”: On subject-verb agreement in L2 French and cross-linguistic influence. *Languages* 6: 7. [\[CrossRef\]](#)
- Ågren, Malin. 2008. À la Recherche de la Morphologie Silencieuse: Sur Le Développement du Pluriel en Français L2 Écrit. Ph.D. dissertation, University of Lund, Lund, Sweden.
- Alderson, Charles J. 2005. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Bloomsbury Academic. [\[CrossRef\]](#)
- Bassano, Dominique. 2000. Early Development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language* 27: 521–59. [\[CrossRef\]](#)
- Benazzo, Sandra, and Marianne Starren. 2007. L'émergence de moyens grammaticaux pour exprimer les relations temporelles en L2. *Acquisition et Interaction en Langue Étrangère* 25: 129–57. [\[CrossRef\]](#)
- Blom, Elma, Ineke Van de Craats, and Josje Verhagen. 2013. *Dummy Auxiliaries in First and Second Language Acquisition*. Berlin and Amsterdam: Mouton de Gruyter.
- De Bot, Kes. 2004. The Multilingual Lexicon: Modelling Selection and Control. *International Journal of Multilingualism* 1: 17–32. [\[CrossRef\]](#)
- Delais-Roussarie, Elisabeth, and Brechtje Post. 2014. Corpus annotation and transcription systems. In *Handbook of Corpus Phonology*. Edited by Jacques Durand, Ulrike Gut and Gjert Kristoffersen. Oxford: Oxford University Press, pp. 46–88.
- Delais-Roussarie, Elisabeth, Brechtje Post, Mathieu Avanzi, Carolin Buthke, Albert Di Cristo, Sun-Ah Jun, Philippe Martin, Trudel Meisenburg, Annie Riiland, Rafeu Sichel-Bazin, and et al. 2015. Developing a ToBI system for French. In *Intonational Variation in Romance*. Edited by Frota Sònia and Pilar Prieto. Oxford: Oxford University Press, pp. 63–100. [\[CrossRef\]](#)
- Detey, Sylvain, and Jean-Luc Nespoulous. 2008. Can orthography influence L2 syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua* 118: 66–81. [\[CrossRef\]](#)
- Di Cristo, Albert. 1999. Vers une modelisation de l'accentuation du français (première partie). *Journal of French Language Studies* 9: 143–80. [\[CrossRef\]](#)
- Giuliano, Patrizia. 2003. Negation and relational predicates in French and English as second languages. In *Information Structure and the Dynamics of Language Acquisition*. Edited by Christine Dimroth and Marianne Starren. Amsterdam: John Benjamins Publishing Company, pp. 119–57.
- Goad, Heather, and Lydia White. 2004. Ultimate attainment of L2 inflection: Effects of L1 prosodic structure. *EuroSLA Yearbook* 4: 119–45. [\[CrossRef\]](#)
- Grabe, Esther, Brechtje Post, and Francis Nolan. 2001. Modelling intonational variation in English: The IViE system. In *Prosody 2000*. Edited by S. Puppel and G. Demenko. Poznan: Adam Mickiewicz University, pp. 51–58.
- Granget, Cyrille, Marie-Eve Michot, Malin Ågren, Sonia Gerolimich, Pascale Hadermann, and Isabelle Stabarin. 2021. Acquisition des formes verbales alternantes en français L2: La L1 des apprenants fait-elle la différence? In *Réflexions sur les Usages et les Utilisateurs du Français: Aspects Acquisitionnels et Didactiques*. Edited by Inès Saddour and Cecilia Gunnarsson. Bruxelles: Peter Lang, pp. 27–62.
- Granget, Cyrille. 2015. Pourquoi le présent n'est-il pas si simple d'un point de vue psycholinguistique?: Une perspective anglophone sur le français L2. *Langage, Interaction, Acquisition* 6: 107–48. [\[CrossRef\]](#)
- Granget, Cyrille. 2018. Analyse de l'effet L1 dans l'émergence de l'auxiliation en français L2. *Discours* 23: 3–42. [\[CrossRef\]](#)

- Jordens, Peter, and Christine Dimroth. 2006. Finiteness in children and adults learning Dutch. In *The Acquisition of Verbs and Their Grammar: The Effect of Particular Languages. Studies in Theoretical Psycholinguistics*. Edited by Natalia Gagarina and Insa Gulzow. Dordrecht: Springer, pp. 173–200.
- Jordens, Peter. 2012. *Language Acquisition and the Functional Category System*. Berlin: Mouton de Gruyter.
- Jordens, Peter. 2013. Dummies and auxiliaries in the acquisition of L1 and L2 Dutch. In *Dummy Auxiliaries in First and Second Language Acquisition*. Edited by Elma Blom, Ineke Van de Craats and Josje Verhagen. Amsterdam and Berlin: Mouton de Gruyter, pp. 341–68.
- Jun, Sun-Ah, and Cecile Fougeron. 2000. A Phonological model of French intonation. In *Intonation: Analysis, Modeling and Technology*. Edited by Antonis Botinis. Dordrecht: Springer, pp. 209–42. [CrossRef]
- Klein, Wolfgang. 1994. *Time in Language*. London: Routledge.
- Labrune, Laurence. 2012. Questioning the universality of the syllable: Evidence from Japanese. *Phonology* 29: 113–52. Available online: <http://www.jstor.org/stable/41475385> (accessed on 13 January 2022). [CrossRef]
- Ladd, D. Robert. 2008. *Intonational Phonology*, 2nd ed. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press. [CrossRef]
- Lasser, Ingeborg. 1997. Finiteness in adult and child German. Ph.D. dissertation, City University of New York, New York, NY, USA.
- Levelt, Pim, Ardi Roelofs, and Antje S. Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1–75. [CrossRef]
- Liberman, A. Mark, and Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk*, 3rd ed. Mahwah: Lawrence Erlbaum Associates.
- Michot, Marie-Eve. 2014. L'acquisition de l'accord sujet-verbe en FLE: L'impact de la classe verbale. *Cahiers AFLS* 19: 57–75.
- Myles, Florence. 2004. From Data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the philological Society* 102: 139–68. [CrossRef]
- Myles, Florence. 2005. The emergence of Morpho-syntactic Structure in French L2. In *Focus on French as a Foreign Language: Multidisciplinary Approaches*. Edited by Jean-Marc Dewaele. Bristol: Multilingual Matters, pp. 88–113.
- Noyau, Colette, Et-Tayeb Houdaifa, Marité Vasseur, and Georges Daniel Véronique. 1995. The acquisition of French. In *The Acquisition of Temporality in a Second Language*. Edited by Rainer Dietrich, Wolfgang Klein and Colette Noyau. Amsterdam: John Benjamins Company, pp. 145–210. [CrossRef]
- Parodi, Teresa. 2000. Finiteness and verb placement in second language acquisition. *Second Language Research* 16: 355–81. [CrossRef]
- Pierrehumbert, Janet P., and Mary E. Beckman. 1988. *Japanese Tone Structure*. Cambridge: MIT Press.
- Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, MIT, Cambridge, MA, USA.
- Post, Brechtje. 2000. Tonal and Phrasal Structures in French Intonation. Ph.D. dissertation, Academic Graphics, The Hague, Holland.
- Prince, Alan. 1983. Relating to the grid. *Linguistic Inquiry* 14: 19–100.
- Sauzedde, Bertrand. 2018. L'acquisition de la Structure Syllabique par des Apprenants Japonais. Ph.D. dissertation, University of Nantes, Nantes, France.
- Schimke, Sarah. 2013. Dummy verbs and the acquisition of verb raising in L2 German and French. In *Dummy Auxiliaries in First and Second Language Acquisition*. Edited by Elma Blom, Ineke Van de Craats and Josje Verhagen. Amsterdam and Berlin: Mouton de Gruyter, pp. 307–38.
- Schlyter, Suzanne. 1995. Morphoprosodic Schemata in the acquisition of Swedish and French bilingual acquisition. In *The Development of Morphological Systematicity. A Cross-linguistic Perspective*. Edited by Hanna Pischwa and Karl Maroldt. Tübingen: Günter Narr Verlag, pp. 79–106.
- Shirai, Yasuhiro. 1998. The emergence of tense-aspect morphology in Japanese: Universal predisposition? *First Language* 18: 281–309. [CrossRef]
- Starren, Marianne. 2001. The second Time. The Acquisition of Temporality in Dutch and French as a Second Language. Ph.D. dissertation, LOT, Utrecht, The Netherlands.
- Truckenbrodt, Hubert. 2005. A short report on intonation phrase boundaries in German. *Linguistische Berichte* 203: 273–96.
- Truckenbrodt, Hubert. 2007. Upstep of edge tones and of nuclear accents. In *Tones and Tunes. Volume 2: Experimental Studies in Word and Sentence Prosody*. Edited by Carlos Gussenhoven and Tomas Riad. Berlin and New York: Mouton de Gruyter, pp. 349–86. [CrossRef]
- Uhmann, Susanne. 1991. *Fokusphonologie*. Tübingen: Niemeyer.
- Van de Craats, Ineke, and Roland van Hout. 2010. Dummy auxiliaries in the second language acquisition of Moroccan Learners of Dutch: Form and Function. *Second Language Research* 26: 473–500. [CrossRef]
- Venditti, Jennifer V. 2005. The J_ToBI Model of Japanese Intonation. In *Prosodic Typology: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford: Oxford University Press. [CrossRef]
- Veneziano, Edy, and Christophe Parisse. 2010. The acquisition of early verbs in French: Assessing the role of conversation and of child-directed speech. *First Language* 30: 287–311. [CrossRef]
- Verhagen, Josje. 2011. Verb placement in second language acquisition: Experimental Evidence for the different behavior of auxiliaries and lexical verbs. *Applied Psycholinguistics* 32: 821–58. [CrossRef]

- Verhagen, Josje. 2013. From dummy auxiliary to auxiliary in Moroccan adult learners' production and comprehension of Dutch. In *Dummy Auxiliaries in First and Second Language Acquisition*. Edited by Elma Blom, Ineke Van de Craats and Josje Verhagen. Amsterdam and Berlin: Mouton de Gruyter, pp. 281–306.
- Véronique, Georges D. 2013. La dislocation, l'emploi de *c'est* et *il y a* et l'organisation topique-focus en français parlé: Des variétés d'apprenants aux usages natifs. *TIPA (Travaux interdisciplinaires sur la parole et le langage)* 29 | 2013. Available online: <http://journals.openedition.org/tipa/1022> (accessed on 15 November 2022). [CrossRef]
- Viberg, Åke. 2006. Verbs. In *Encyclopedia of Language and Linguistics*, 2nd ed. Edited by Keith Brown. Oxford: Elsevier, pp. 408–11.
- Wiese, Richard. 1996. *The Phonology of German*. Oxford: Clarendon Press.
- Yazawa, Kakeru, Takayuki Konishi, Keiko Hanzawa, Greg Short, and Mariko Kondo. 2015. Vowel epenthesis in Japanese speakers' L2 English. Paper presented at the 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14; p. 969.

Article

Prosodic Transfer in Contact Varieties: Vocative Calls in Metropolitan and Basaá-Cameroonian French

Fatima Hamlaoui ^{1,*}, Marzena Żygis ^{2,†}, Jonas Engelmann ³ and Sergio I. Quiroz ²¹ Department of French, University of Toronto, Toronto, ON M5S 1J4, Canada² Leibniz-Zentrum Allgemeine Sprachwissenschaft, D-10117 Berlin, Germany³ Department of History and Philosophy of Science, Universiteit Utrecht, 3584 CS Utrecht, The Netherlands

* Correspondence: f.hamlaoui@utoronto.ca

† Current address: Department of French, University of Toronto, 50 St. Joseph Street, Toronto, ON M5S 1J4, Canada.

‡ These authors contributed equally to this work.

Abstract: This paper examines the production of vocative calls in (Northern) Metropolitan French (MF) and Cameroonian French (CF) as it is spoken by native speakers of a tone language, Basaá. While the results of our Discourse Completion Task confirm previous descriptions of MF, they also further our understanding of the relationship between pragmatics and prosody across different groups of French speakers. MF favors the vocative chant in routine contexts and a rising-falling contour in urgent contexts. In contrast, context has little influence on the choice of contour in CF. A melody consisting of the surface realization of lexical tones is produced in both contexts. Regarding acoustic parameters, context only exerts a significant effect on the loudness of vocative calls (RMS amplitude) and has little effect on their F0 height, F0 range and duration. A target-use of vocative calls in CF thus does not amount to target-like use of the original standard target language, MF. Our results provide novel evidence for the transfer of lexical tones onto the contact variety of an intonation language. They also corroborate previous studies involving the pragmatics-prosody interface: the more marked a prosodic pattern is (here, the vocative chant), the more difficult it is to acquire.

Keywords: vocative calls; intonation; lexical tones; contact variety; prosodic transfer; French; bilingualism; Cameroon; Basaá; Bantu

Citation: Hamlaoui, Fatima, Marzena Żygis, Jonas Engelmann, and Sergio I. Quiroz. 2022. Prosodic Transfer in Contact Varieties: Vocative Calls in Metropolitan and Basaá-Cameroonian French. *Languages* 7: 285. <https://doi.org/10.3390/languages7040285>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 8 June 2022

Accepted: 19 October 2022

Published: 7 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vocatives are generally understood as expressions, either simple (Marina! Misty!) or more complex (Professor Smith! Mrs. President!) whose aim is to attract one's attention ("calls" or "summonses") or help maintain and strengthen the relationship between interlocutors ("addresses") by spelling out the addressee (Di Cristo 2016; Hill 2014; Ritter and Wiltschko 2020; Schegloff 1968; Zwicky 1974). The expression of vocatives lies at the interface between different language components: phonology, syntax, morphology, semantics and pragmatics. It is also fundamentally related to how information is packaged to fit a context and this is what recent prosodic studies of vocatives have tried to establish across a variety of languages and prosodic types (Arvaniti et al. 2016; Borràs-Comes et al. 2015; Huttenlauch et al. 2018; Kubozono 2022; Kubozono and Mizoguchi 2019; Olawale 2021; Quiroz and Żygis 2017). A number of socio- and extralinguistic factors have long been known to affect how vocative calls are expressed: spatial distance, insistence, hierarchical relationship, politeness, intimacy (Brown and Levinson 1987; Brown and Gilman 1960; Hill 2014; Zwicky 1974). As stated by Di Cristo (2016), beyond the choice of nominal expression, the prosody of vocative forms also reflects the attitude adopted by the speaker: kindness, reprobation, etc. Many European languages associate vocative calls with a chanted tune consisting roughly of a rise followed by a sustained mid tone (e.g., English, German, Dutch, Polish). This contour, the 'vocative chant' or 'calling contour', is generally associated with

sweet and friendly contexts, and is known to display semantic and realizational differences across languages (Ladd 2008; citing Gibbon 1976). It has drawn considerable attention as part of the Auto-segmental Metrical (AM) framework, as there is no obvious answer to how to represent a phonetical mid tone using only phonological high and low tones (see Ladd 2008 and also Arvaniti et al. (2016) for relatively recent references of studies of vocative calls across a variety of languages). Other types of calling melodies have been associated with urgent or stern contexts, for instance a rising-falling contour in Polish (Arvaniti et al. 2016) and French (Delais-Roussarie et al. 2015; Di Cristo 2016).

How to use a language effectively in a particular context, or acquiring so-called *pragmatic competence*, is known to be particularly challenging for L2 learners (Kasper 2001; Leech 1983; Thomas 1983). Just as when it comes to grammatical knowledge, speakers tend to be influenced and transfer pragmatic knowledge from their L1 (Bardovi-Harlig 2002). Transfer from the L1 is also known to occur in the process of acquisition of different dimensions of the L2 prosody (Delais-Roussarie et al. 2015; Mennen 2004; Mennen and de Leeuw 2014; Trouvain and Braun 2020). A number of studies have shown that some aspects of sentence prosody, and in particular those that have to do with the discourse context (e.g., expressing focus and givenness), fail to be mastered even by advanced speakers. This is particularly the case when these aspects are absent from their L1 (Hamlaoui et al. 2021; Ortega-Llebara and Colantoni 2014; Rasier and Hilgsmann 2007; Trouvain and Braun 2020; Zerbian 2015). In addition to learning a form that may or may not exist in their native language, L2 speakers also have to master the pragmatic contexts in which this form can be used appropriately (Kang and Kermad 2019). This form-meaning association is also key in the appropriate expression of vocative calls, the focus of the present paper.

There are relatively few studies of the second language acquisition of vocatives. Pešková (2019) has recently investigated the acquisition of vocative calls in L2 Spanish and L2 Italian by L1 Czech speakers. The three languages have in common that they realize vocative calls with a chanting contour, and thus a rising pitch accent ($L^* + H, L + H^*$) and a following mid-tone analyzed as a downstepped high boundary tone ($!H\%$). Although her speakers tend to demonstrate a native-like production of vocative calls, they also show a pattern, in Italian L2, which is found neither in Italian L1 nor in Czech L1: $H^* + L L\%$. This is interpreted by Pešková as a case of prosodic overgeneralization (Brown 2000; Gabriel and Kireva 2014): the speakers use L2 tonal and durational patterns that otherwise exist in the target language but are appropriate in a different pragmatic context.

The present study concentrates on vocative calls in French. In Metropolitan French, vocative calls are also typically associated with a “chanting contour” (Ladd 2008). Previous descriptions of the language suggest that the chanting contour is favored in friendly contexts in which speaker and hearer have ‘a shared convention or agreement’ (Di Cristo 2016; Fagyal 1997). This contour is, however, not appropriate when an emergency call has to be made. Di Cristo (2016) and Delais-Roussarie et al. (2015) describe a distinct melody, that is, a rising-falling contour, which is felicitous in urgent calls. However, no study yet has systematically investigated the effect of context on the choice and realization of calling melodies in Metropolitan French. One of the aims of this paper is to fill this research gap.

As we are also interested in how bilingualism and contact conditions influence intonation systems, we also concentrate on vocative calls in French as it is spoken in Cameroon, and in particular by Basaá speakers.¹

There has been a growing interest in the topic of prosodic transfer in contact varieties of languages (Avanzi and Bordial Steien 2016; Bordial and Lyche 2012; Colantoni 2011; Colantoni and Gurlekian 2004; Delais-Roussarie et al. 2015; Gussenhoven and Udofot 2010; Gut 2005; Mamode 2015; Pešková et al. 2012). The idea of exploring these varieties from the perspective of Second Language Acquisition (SLA) has been put forward, for instance, by Williams (1989, p. 40). She argues that these varieties, which have developed over many generations of speakers and, in many cases, were once the result of individual language acquisition, crystallize inter-language features characteristic of the speech of second language learners in other settings (e.g., in the classroom). Their stability however

distinguishes them from learners' individual inter-languages and makes them another precious source of evidence regarding the processes involved in L2 acquisition. In this vein, our study investigates prosodic transfer through the study of the mature grammar of late bilingual speakers of a local variety of French, Cameroonian French. As is common in postcolonial contexts, for most speakers, Cameroonian French is primarily learned in school and sometimes outside, but almost always after at least one other language has been acquired (Onguene Essono 1999, p. 292). The transfer of characteristics from the L1 is widely acknowledged in the literature on Cameroonian French (Zang Zang 1999) and 'regional accents' have been described as 'strongly perceptible' (Tabi Menga 1999; Mendo Ze 1999). From a prosodic perspective, Cameroonian French has been described as being influenced by the tone languages with which it is in contact. This is no surprise when considering how early our processing system is tuned for our native language(s). According to studies such as Werker and Tees (1984) and Kuhl et al. (1992), there is evidence that native phonetic categories are acquired in infants between the age of 6 and 12 months and that they remain quite stable. What has been described in the case of Cameroonian French is reminiscent of what is observed in other contact varieties of European languages and, typically, varieties of New Englishes (Gut 2005; Gut and Pillai 2014; Lim 2009; Mesthrie 2008) and Latin American Spanish (Gabriel and Kireva 2014; O'Rourke 2005; Sosa 1999, and references therein). It is also consistent with the results of instrumental studies showing that bilinguals prosody differs from monolinguals' (Braunmüller and Gabriel 2012; Colantoni and Gurlekian 2004).

To date, and to the best of our knowledge, claims about Cameroonian French prosody are however mostly observational. The nature and scope of prosodic transfer from specific L1s remain little understood and for the time being we will focus on a particular L1, Basaá (Bantu A43). Although this language does not have any privileged status, it is one of the languages used for local inter-group communication in the South of the country and it is among a handful of local languages that have recently been selected to be used as a medium of education. As in the vast majority of Bantu languages, pitch in Basaá is phonemic. Makasso et al. (2016) found no evidence that post-lexical meanings such as focus and questions have an effect on Basaá tones, suggesting that the role of intonation is limited in this language. Vocative calls are expressed by means of a particle ((à-)) and the language distinguishes formal and informal vocative calls by means of morphology (Bitjaa Kody and Mutaka 1997; Makasso n.d.). There are however no available descriptions of the extent to which other variables such as kindness or reprobation can affect the realization of vocative calls. The question then arises as to the prosody of Cameroonian French as it is spoken by speakers who have Basaá as their L1 (or other tone languages with a limited role of intonation) and the influence of the context on the prosody of vocative calls. If the prosodic properties of Cameroonian French as spoken by Basaá speakers are at least in part the effect of transfer from the L1, we expect that syllables carry lexically specified tones and that the realization of lexical tones takes priority over the expression of post-lexical meanings which determine intonation in Metropolitan French. The fact that varieties of French in contact with African tone languages display lexical tones has been shown for instance by Bordal (2013, 2015) in relation to Central African French. Our study is particularly original in that it tests whether and how these tones can be influenced by the situational context and thus the extent to which a variety of French with lexical tone properties makes use of intonation.

The results of our Discourse Completion Task, adapted from Arvaniti et al. (2016) and Quiroz and Žygis (2017), tend to confirm previous descriptions of Metropolitan French regarding the preference for specific contours in particular pragmatic contexts. Whereas the chanting contour is most frequently realized in a routine (i.e., call for dinner) context, a rising-falling contour is favored in an urgent (i.e., call due to a broken vase) context. In contrast, in Basaá-Cameroonian French, the context has less influence on the choice of prosodic contour and a contour consisting of the surface realization of lexical tones is favored in both routine and urgent context. Target use of vocative calls in Basaá-Cameroonian French thus does not amount to target-like use of the original standard target language,

Metropolitan French (Williams 1989). Our results are consistent with previous studies showing a transfer of lexical tones onto the contact variety of an intonation language. In our perspective, the semantic deviance (Mennen 2015) from Metropolitan French observed in our Basaá-Cameroonian French speakers also corroborates what has been observed in other studies involving the pragmatics-prosody interface, that the more marked a prosodic pattern is (here, the vocative chant), the more difficult it is to acquire (Eckman 1987).

The paper is structured as follows. Section 2 provides some background on vocative calls in Basaá and Metropolitan French, as well as on the prosody of proper names in Basaá-Cameroonian French. Section 3 presents the material and methods of the present study. Section 4 lays out the results and Section 5 concludes the paper.

2. Some Background on the Formation of Vocative Calls

2.1. Basaá Vocative Calls

Basaá, a Northwest Bantu language (A43 in Guthrie's (1948) classification), is spoken by approximately 300,000 speakers as a native language in the Center and Littoral regions of Cameroon (Lewis et al. 2015). It is a relatively well-studied language. Several grammatical sketches were written by missionaries in the early twentieth century (Rosenhuber 1908; Scholaster 1914; Schürle 1912). Although a lot remains to be done, numerous studies have dealt with various aspects of the language's grammatical and speech properties (among others, Bitjaa Kody 1990; Bot Ba Njock 1964; Dimmendaal 1988; Hyman 2003; Lemb and de Gastines 1973; Makasso 2008; Makasso et al. 2016).

From the perspective of tone, the language underlyingly distinguishes high-toned (H), low-toned (L) and toneless moras. On the surface, a number of tonal processes apply that give rise to a five-way tonal contrast between high, low, downstepped high (!H), falling (HL) and rising (LH) tones (Dimmendaal 1988; Hamlaoui et al. 2014; Hyman 2003; Makasso et al. 2016).

According to Bitjaa Kody and Mutaka (1997), who offer a detailed description and analysis of the tonology and morphology of vocative calls in a set of Bantu languages from Southern Cameroon, Basaá has in common with a number of neighboring Bantu languages that it expresses vocative calls by means of the morphological marker {a-}. The language distinguishes a colloquial and a polite form, while the colloquial form is commonly used among children, peers and in informal contexts, the polite form is strongly preferred in formal contexts and to call a superior. These two forms are illustrated, respectively, in (1) and in (2) for the call "Mr. Bitjaa!" (Bitjaa Kody and Mutaka 1997, p. 56).

- (1) à tʃá⁺á!
 VOC Bitjaa
 'Bitjaa!' (Colloquial form)
- (2) à sáŋgó bítʃá⁺á
 VOC Mister Bitjaa
 'Mr. Bitjaa!' (Polite form)

Bantu languages are known for their complex nominal morphology and their use of noun class prefixes expressing morphological gender and number (Nurse and Philippson 2003). As in many cultures in this part of Africa and over the world, proper names often have a meaning and can refer to an artefact, action, activity or an individual who holds a significant meaning for the family. The name "Bitjaa" is thus a morphologically complex form, which consists of a noun class marker {bi-} and a lexical root. As visible in (1), in the presence of the vocative marker a-, the noun class prefix disappears.

In each of the five languages investigated by Bitjaa Kody and Mutaka (1997), that is, Akóósé (A15b), Basaá, Duala (A24), Ewondo (A72) and Mbòo (A10), the details of vocative call formation depend on the segmental and tonal makeup of the proper name onto which the vocative marker attaches. The marker itself can surface with different tones and sometimes even coalesce with the first vowel of a name, thus only surfacing as a tone on the initial of this name. As for the proper name, it can lose its class prefix, as seen in (1)

in Basaá, or lose its final vowel, depending on the name and the particular language. The tones of the lexical root however seem to remain relatively unaffected by the presence of the vocative marker in all five languages.

In the particular case of Basaá, the authors distinguish 4 groups of nouns:

- Group 1: nouns starting with a syllabic prefix and whose first vowel underlyingly carries a low tone → the vocative marker surfaces as {á} and the name loses its noun class prefix. No changes are observed in the tones of the lexical root.
- Group 2: nouns starting with a syllabic prefix and whose first vowel underlyingly carries a high tone → the vocative marker surfaces as {à-} and the name loses its noun class prefix. The tones of the lexical root remain unchanged.
- Group 3: nouns without a noun class prefix and whose first vowel underlyingly carries a high tone → the vocative marker surfaces as {à-} and the tones of the lexical root are unaffected.
- Group 4: nouns without a noun class prefix and whose first vowel underlyingly carries a low tone → the vocative marker surfaces as {à-} and high tone surfaces on the lexical root, either forming a falling tone on the first vowel of the noun or leading to a downstep on a following high tone.

Bitjaa Kody and Mutaka (1997) analyze the vocative marker as carrying a floating high tone, which surfaces on the vocative marker itself in Group 1, on the noun in Group 4 and fails to surface in the two other groups. From an acoustic perspective, Makasso (n.d.) argues that the tones of vocative calls are realized with a higher F0 than in citation forms. No systematic instrumental study as however been carried out yet.

2.2. Metropolitan French Vocative Calls

Metropolitan or Hexagonal Standard French is an intonation language: pitch is primarily used to express post-lexical meanings such as sentence modality (interrogation, exclamation etc.) and attitudes (surprise, irony etc.) (Gussenhoven 2004; Ladd 2008). Pitch is also used to indicate word groupings and dependency relations between them. At the word level, Metropolitan French is characterized by the absence of lexical stress. Stress in this variety of French is phrase-final and assigned by a combination of F0, intensity and duration cues (Delais-Roussarie et al. 2015; Delattre 1966; Di Cristo 2016; Di Cristo and Hirst 1999; Lacheret-Dujour and Beaugendre 1999; Martin 1981; Mertens 1993; Jun and Fougeron 1995, 2000).

Although French, and other Romance languages, tend to differ from other European languages (Germanic, Slavic) in how structural constraints on accentuation interact with pragmatic information (e.g., in the case of expressing focus/givenness), this is not as strikingly the case when it comes to vocative calls. As in a number of European languages (e.g., English, Dutch, German, Polish, Portuguese, varieties of Spanish and Italian), French vocative calls are typically associated with a “chanted intonation” or a stylized “calling contour” (Delais-Roussarie et al. 2015; Dell 1984; Di Cristo 2016; Fagyal 1997; Fónagy et al. 1983; Ladd 2008).

This type of call is illustrated in (3) with a scenario whose French version was used by Fagyal (1997) as part of an elicitation task.

(3) Chanting contour (Fagyal 1997, p. 81)

A, the aunt, is taking Joanna, her niece, out. She cannot see her, so she calls sweetly:

A: Joanna!

This melody, which is not exclusive to vocative calls, has been described as consisting of a penultimate high and a final lowered high or mid tone. The final tone is often carried by a lengthened vowel and in monosyllabic names, vowel doubling is sometimes observed, splitting the word into two syllables (Fagyal 1997, p. 78). This effect of the chanting contour on word length is illustrated in (4).

- (4) a. Marthe [maχt] → Ma-arthe! [ma.aχt]
 b. Yann [jan] → Ya-an! [ja.an]
 c. Louise [lwiz] → Lou-ise! [lu.wiz]

This calling contour has received different phonological representations in different frameworks. It has been represented as LMH (Dell 1984; Di Cristo and Hirst 1999) and lh\HH (Mertens 1987). In the Auto-Segmental Metrical framework (AM) (Pierrehumbert 1980), the French calling contour has been encoded as H* H-L% (Jun and Fougeron 1995) and more recently H + !H*!H% in the French ToBI conventions proposed by Delais-Roussarie et al. (2015).

The melody of vocative calls has been reported to vary depending on the context. A French more insistent or less friendly call, for instance, is described by Delais-Roussarie et al. (2015) as consisting of a rising-falling contour, represented H* L%. This description is similar to the one provided by Di Cristo (2016, p. 411) and acceptable, according to him, in a wider set of contexts (e.g., distant calls and reprimands). The H* L% contour of French urgent calls is similar to the contour observed in urgent contexts in Portuguese (Frota 2014; Frota et al. 2015) and has in common with the ones observed in languages such as Polish (H* L-L%, Arvaniti et al. 2016) and Catalan (L + H HL%, Prieto 2014) that it ends in a fall.

2.3. Basaá-Cameroonian French Vocative Calls

There is a long and well-established interest among Cameroonian scholars for the study of languages and a rich literature on both structural and sociolinguistic aspects of Cameroonian French, the variety of French resulting from the contact with local languages. In the historically complex and linguistically dense Cameroonian context, it is clear that there is not just one variety of Cameroonian French but many. The influence of tone on Cameroonian French has long been acknowledged. According to Fame Ndongo (1999, p. 198), each word rather than each phrase forms its own prosodic domain and pitch, rhythm and speech rate are all the effect of the speaker's L1. This seems consistent with the fact that, until recently, people used to learn to read and write French before they could speak it (Djoum Nkwescheu 2008).

To the best of our knowledge, no study has yet concentrated on the prosody of vocative calls in Basaá-Cameroonian French (or any other variety of Cameroonian French). What, from our perspective, is particularly interesting is that proper names provide evidence for prosodic transfer, as they are specified for lexical tones.² A few examples from the list of items we used, and that will be described in more detail in Section 3, are given in (5) to (8).

- (5) Mārínà (L H L)
 (6) Mâgdalénà (L L H L)
 (7) Yánn (HL)
 (8) Grégóry (H H HL)

The lexical tones of these proper names seem to reflect the intonation of their citation form, i.e., they end in a fall which either aligns with the last or the penultimate and last syllable. From this perspective, these words are not unlike Basaá borrowings from French and other European languages. Loanword phenomena have also been argued to involve a transfer from the L1 (Broselow 2000). According to Major (2008), loanword phonology can be considered a form of "forced transfer", by which foreign words are pronounced by speakers according to their L1 so as to avoid being perceived as 'too snobbish and affected'. Over time, and through contact with various European cultures and languages, in particular through colonization, a number of foreign words have been integrated into the lexicon of Basaá. In contrast to our proper names, these words have typically undergone morphological and/or (supra-)segmental adaptation. A few examples are given in (9) to (12) of loanwords originating from German, English and French (Emmanuel-Moselly Makasso, *p.c.*).

- (9) [káásà] < Kaiser (Ger), “emperor”
- (10) [màlêr] < Lehrer (Ger), “teacher”
- (11) [kìngê] < King (Eng)
- (12) [mésà] < Messe (Fr), “mass”

The words in (5) to (12) all have in common that they display a final H-L tonal pattern, suggesting that Basaá speakers follow the “stress-to-tone” principle (Silverman 1992), by which pitch accents and boundary tones in the intonation language are interpreted as lexical tone sequences. In relation to loanwords, this principle has been argued to find its source in the perceptual similarity between accentual and intonational phenomena on the one hand and lexically tonal phenomena on the other. It has been shown to apply in some Asian tone languages such as Cantonese (Chen 2000; Hao 2009; Kiu 1977; Silverman 1992) and Mandarin (Glewwe 2021). In African tone languages, it has been shown to apply to Hausa (Kenstowicz 2006; Leben 1996), as well Yoruba and Shona (Kenstowicz 2006), for instance. The same phenomenon seems to be at play in the French variety of our Basaá-Cameroonian French speakers, at least when it comes to proper names. As we know from previous studies on L2 prosody, speakers are greatly influenced, in their perception, by the prosody of their L1 (Mennen and de Leeuw 2014, p. 187). It is thus not surprising that the L2 prosody of our speakers would show influences from their L1 (Mennen 2004; Mennen and de Leeuw 2014; Pickering 2004). The question thus arises as to how the lexical tones of these proper names are realized as part of vocative calls in different situational contexts and this is what our study tries to establish.

3. Materials and Methods

3.1. Stimuli and Procedure

Following Arvaniti et al. (2016) and Quiroz and Žygis (2017), we used a Discourse Completion Task (DCT) in which 12 names were called by our participants, either under a routine or an urgent context. Names of various lengths were used (1 to 4 syllables) and appear in Table 1. We selected three names per syllable count. A phonetic transcription is given under each name. Dialectal variants are given in the order Metropolitan French/Basaá-Cameroonian French. The Cameroonian French phonetic transcription includes the surface realization of lexical tones.³

Table 1. The names used in this study presented by the number of syllables.

Syllables	Names		
1	<i>Yann</i> [jan]/[jân]	<i>Louise</i> [lwiz]/[lûiz]	<i>Marthe</i> [maʁt]/[mâ:t]
2	<i>Daniel</i> [daneɪ]/[dâɲɛɪ]	<i>Alice</i> [alis]/[âlîs]	<i>Patrick</i> [patʁik]/[pâtrîk]
3	<i>Natalia</i> [natalja]/[nâtáljà]	<i>Marina</i> [mavina]/[mârínà]	<i>Grégory</i> [gʁegovi]/[grégóʀi]
4	<i>Magdalena</i> [magdalena]/[mágdalénà]	<i>Alexandra</i> [aleksãdʁa]/[âlègzãdrà]	<i>Bénédictine</i> [benediktin]/[bènèdiktîn]

With respect to lexical tones, proper names can be grouped as in Table 2. All names have in common that they terminate in a fall. What distinguishes them is the tones that precede the fall (i.e., high or low) and whether the fall is anchored on the last (HL) or on the penultimate (H) and last (L) syllables.

Table 2. The names used in this study grouped by lexical tones

L (L) H L	L (L) HL	H H HL
Marina	Patrick	Grégory
Natalia	Daniel	
Alexandra	Alice	
Magdalena	Marthe	
	Louise	
	Yann	
	Bénédictine	

As in Quiroz and Žygis (2017), participants were asked to imagine a scenario in which they are inside a house and have to call a child who is playing outside. Under the routine context, the child is being called in for dinner, whereas under the urgent context, they are being called to be reprimanded for breaking a vase. All participants saw a prompt, illustrated in (13), consisting in a brief description of the scenario, and followed by the name they were asked to call.⁴

- (13) Vous entrez dans une pièce et vous voyez que votre enfant a cassé votre vase préféré. Vous l'appellez: ...
 'You enter a room and you see that your child has broken your favorite vase. You call them: ...'

Participants were asked to produce the name as naturally as possible in the given context. Names were presented in a semi-random order and all names were called by all participants. There was a total of 78 trial runs (12 items*3 repetitions*2 contexts). In total, 1008 items were produced by Metropolitan French speakers and 936 by Basaá-Cameroonian French speakers. Filler items were also included which consisted of basic questions to which the participants had to provide a short scripted answer.

For Metropolitan French, recordings were carried out in Berlin, in a sound-proof booth of the Leibniz Center for General Linguistics. Participants were sitting in front of a computer screen with a microphone placed approximately 20 cm away from their mouth. For Cameroonian French, recordings were made in a quiet room in Yaoundé. Participants were sitting in front of a laptop and recordings were made through a head-mounted microphone.

3.2. Participants

Data were elicited from a total of 27 speakers. 14 speakers (4 male speakers) originated mostly from Northern France (e.g., Lille, Metz, Paris) and were native speakers of Metropolitan French.⁵ Their age ranged from 19 to 31 years. They were all monolingual speakers (with various levels of proficiency in foreign languages such as German, English, Russian, Dutch, Italian and Portuguese; none with lexical tones) and had been living in Berlin from 3 months to 13 years.⁶ They all had completed secondary education and were either in the process of completing university or had a university degree.

13 bilingual speakers (9 male speakers) originated from Cameroon, had Basaá as their native language and were speakers of Cameroonian French. Their age ranged from 20 to 41 years. 9 speakers originated from the Centre region of Cameroon and 4 from the Littoral. Only three of them had lived a few years in another country than Cameroon (France, Germany). They all frequently spoke Basaá, in particular with other members of the Basaá community and at home, with their family members. None of our speakers declared speaking another Cameroonian tone language. All were schooled in French, had completed secondary education and were either in the process of completing university or had a university degree.

All speakers were naive as to the purpose of the experiment and were financially compensated for their participation.

3.3. Data Analysis

3.3.1. Data Annotation

The utterances were perceptually categorized by two native speakers of French independently (one of the authors and a naive speaker trained for the purpose). The data were manually annotated by 3 of the authors independently. Categorizations and annotations were compared and when the judges disagreed, the data was discussed until agreement was reached.

In Metropolitan French, three distinct contours were distinguished, namely, vocative chant, rising-falling and rising contours.⁷ See Figure 1. In both dialects of French, a number of utterances did not fit in any of these categories and are represented under the category “other”. These utterances were mostly produced in the urgent context. Typically, they failed to realize a call. Rather, they tended to express a form of (questioning) disapproval, sometimes mixed with what could be perceived as anger or exasperation, depending on speakers.

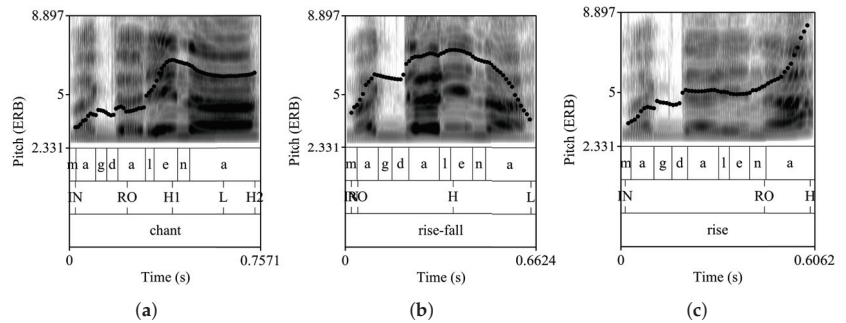


Figure 1. Illustration of the main three contours for the name ‘Magdalena’ in MF (male speaker). (a) Vocative chant; (b) Rising-falling contour; (c) Rising contour.

In Basaa-Cameroonian French, three main melodies were identified, namely, “default” (i.e., the surface realization of lexical tones, see Table 2), chanting and rising contour, which are illustrated in Figure 2.

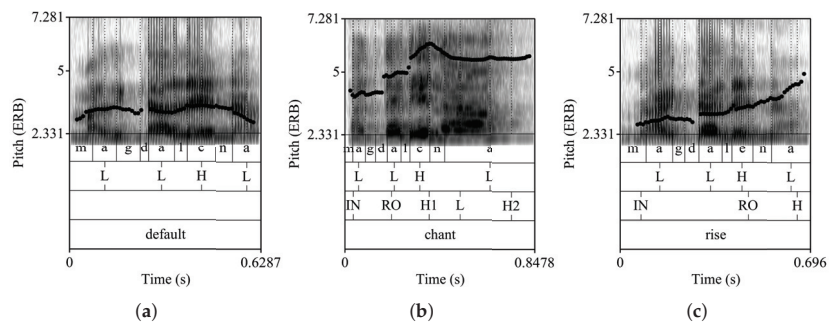


Figure 2. Illustration of the main three contours for the name ‘Magdalena’ in MF (male speaker). (a) Default contour; (b) Chanting contour; (c) Rising contour.

Again, some utterances could not be classified in any clear category based on perceptual and visual cues and were classified as “other”. In these utterances, our participants typically failed to realize a call.

3.3.2. Tonal Landmark Measures

Acoustic analysis of the data was carried out in PRAAT (Boersma and Weenink 2016). Again, following the methods used in Arvaniti et al. (2016) and Quiroz and Žygis (2017), we used the ERB scale to measure F0, that is, the perceptual equivalent bandwidth calculated from acoustic frequency in Hz, using the formula provided by PRAAT $(11.171n((x + 312)/(x + 14,680))^{43})$. The use of this scale was meant to reduce differences between male and female speakers.

As visible in Figures 1 and 2, measurements were taken at specific points in the contour. Areas of interest were selected and F0 maxima and minima were located and annotated using, respectively, the functions ‘maximum pitch’ and ‘minimum pitch’. For both varieties of French, the following measurements were subsequently obtained semi-automatically by means of PRAAT scripts.

- Initial Tone (IN): the F0 at the onset of the contour;
- Rise Onset (RO): RO corresponds to the F0 at the onset of the rise to the first (and sometimes only) peak in the contour. It was only measured in three and four-syllabic words. RO and IN coincided in 1- and 2-syllabic words.
- High 1 (H1): In the chanting melody, H1 was the F0 maximum of the first peak in the contour, while in rising-falling contours it was the only peak. Similarly, in rising contours there was also only one peak (H) found at the end of the word.
- Low (L): in the chanting melody, L was the F0 minimum in the dip in the contour (between the two H tones). In the rising-falling melody, it was the lowest point reached at the end of the contour.
- High 2 (H2): in the chanting melody, the F0 maximum at the end of the contour.

Additionally, in Basaá-Cameroonian French, each vowel was annotated for lexical tone, according to the tonal patterns presented in Table 2.

3.3.3. Additional Measurements

We also measured other acoustic parameters of the items including:

- F0 range, measured here as the difference between the F0 maximum and F0 minimum for a given name (Cosmides 1983, we thus focus on “span” in the sense of Ladd 2008)
- Duration (in seconds) for the entire name,
- Root mean square amplitude of the whole name. Prior to analysis, the RMS was log transformed as it was positively skewed.

3.3.4. Statistical Analysis

The statistical analysis of the data was conducted in the R studio software (version 4.0.0, RStudio Team 2020) by using the *lmer4* (Bates et al. 2020) and the *emmeans* package (Lenth 2019).

For Metropolitan French data, linear mixed effects models were employed for assessing the influence of Context [routine, urgent], Shape [chant, rise-fall, rise, other] and Number of Syllables of a given name [1:4] and Sex [female, male] on F0 range, RMS amplitude of a given name and name duration. In addition, an interaction of Context and Shape was included to test whether the dependent variables are affected when different contours are compared across the routine and urgent context. If the interaction was not significant, it was removed from the final model. In addition, participants and names were included as random intercepts and Context and Number of Syllables of a given name were taken as by-participant and Context as by-name random slope. The same statistical modeling was applied for Basaá-Cameroonian French data with the only difference regarding Shape, consisting of the following levels [chant, default, rise, other]. For a comparison of F0 range, RMS amplitude and duration between languages, we added the factor Language [MF, CF].

Since the factors Shape and Number of syllables consisted of four levels, we performed pairwise comparisons of the data by using the *emmeans()* function from the *emmeans* package (Lenth 2019).

4. Results

4.1. Calling Melody Frequency

4.1.1. Metropolitan French

As seen in Figure 3, the routine context elicited a majority of vocative chants (64%). In the urgent context, speakers mainly produced rising-falling contours (83%). Alternative melodies were also produced in each of the two contexts. A minority of rising (22%) and rising-falling contours (13%) were found in the routine context and some vocative chants (7%) and rising contours (2.5%) in the urgent context. The category “other” constitutes 0.7% calls in the routine and 7% in the urgent context.

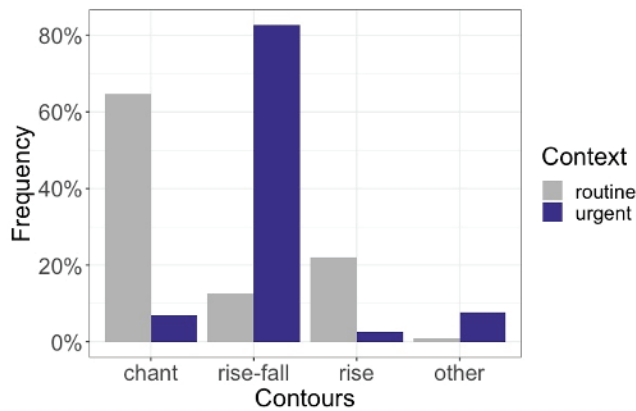


Figure 3. Frequency of occurrence of contours in routine and urgent context (Metropolitan French).

It is worth noticing that there was considerable variation among speakers as far as the choice of contour is concerned within a given context. Table 3 presents the frequency in percentages and counts appear in parentheses.

Table 3. Percentage of individual speakers’ calling melodies depending on context in MF.

	Chant		Rise-Fall		Rise		Other	
	Routine	Urgent	Routine	Urgent	Routine	Urgent	Routine	Urgent
S1	89 (32)			94.4 (34)	11 (4)	5.6 (2)		
S2	69.7 (23)		24 (8)	75.8 (25)	6 (2)			24 (8)
S3	83.3 (30)	2.8 (1)	8.3 (3)	80.6 (29)	8.3 (3)	13.9 (5)		2.8 (1)
S4	30.6 (11)		22.2 (8)	75 (27)	44.4 (16)	2.8 (1)	22.2 (8)	
S5	83.3 (30)		11.1 (4)	97.2 (35)	5.6 (2)	2.8 (1)		
S6	58.3 (21)			100 (36)	41.7 (15)			
S7	25 (9)		44.4 (16)	61.1 (22)	30.6 (11)			38.9 (14)
S8	83.3 (30)	22.2 (8)	2.8 (1)	75 (27)	11.1 (4)	2.8 (1)	2.8 (1)	
S9	91.7 (33)	58.3 (21)	2.8 (1)	38.9 (14)	5.6 (2)	2.8 (1)		
S10	5.6 (2)		11.1 (4)	97.2 (35)	83.3 (30)			2.8 (1)
S11	44.4 (16)	5.6 (2)	16.7 (6)	86.1 (31)	36.1 (13)	5.6 (2)	2.8 (1)	2.8 (1)
S12	72.2 (26)	2.8 (1)	25 (9)	97.2 (35)			2.8 (1)	
S13	77.8 (28)	2.8 (1)		100 (36)	19.4 (7)			
S14	91.7 (33)	5.6 (2)	5.6 (2)	77.8 (28)	2.8 (1)	16.7 (6)		

Some speakers (e.g., Speaker 4, Speaker 7, Speaker 10) only realized a minority of chanting contours in the routine context and either favored a rising or a rising-falling contour instead. In the urgent context, all speakers favored the rising-falling contour, except for Speaker 9, who realized a majority of chanting contours instead.

4.1.2. Basaá-Cameroonian French

As seen in Figure 4, Basaá-Cameroonian French participants produced a majority of default contours, that is, lexical tones, in both contexts. They constituted 57% of all calls in the routine context and 81% in the urgent context. The rising contour was the second most frequently appearing one, with 21% in the routine and 8% in the urgent context. The chant contour was also produced but to a very limited extent: 12% in the routine and 3% in the urgent context. Finally, speakers also produced other patterns: 10% in the routine and 9% in the urgent context.

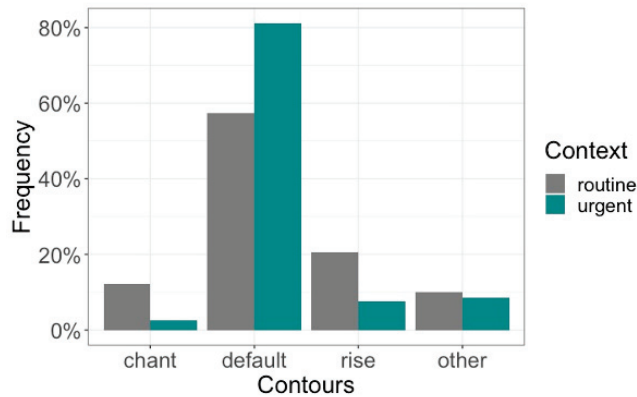


Figure 4. Frequency of occurrence of contours in routine and urgent context (CF).

Variation in the choice of contour based on context was also observed among Cameroonian-French speakers. See Table 4. Counts are given in parentheses.

Table 4. Percentage of individual speakers’ calling melodies depending on context in CF.

	Chant		Default		Rise		Other	
	Routine	Urgent	Routine	Urgent	Routine	Urgent	Routine	Urgent
S1			100 (36)	100 (36)				
S2	3 (1)		14 (5)	75 (27)	75 (27)	19 (7)	8 (3)	6 (2)
S3	36 (13)	6 (2)		94 (34)	64 (23)			
S4			9 (3)	6 (2)	79 (27)	76 (25)	12 (4)	18 (6)
S5	6 (2)	6 (2)	91 (32)	85 (28)	3 (1)			9 (3)
S6			97 (35)	100 (36)			3 (1)	
S7	6 (2)		82 (27)	87 (27)			12 (4)	13 (4)
S8		3 (1)	94 (34)	97 (35)			6 (2)	
S9			97 (35)	97 (34)		3 (1)	3 (1)	
S10	81(29)		3 (1)	97 (35)	14 (5)	3 (1)	3 (1)	
S11			31 (11)	53 (19)	6 (2)		64 (23)	47 (17)
S12	22 (8)	8 (3)	67 (24)	86 (31)	6 (2)	3 (1)	6 (2)	3 (1)
S13	3 (1)	11 (4)	61 (22)	72 (26)	22 (8)		14 (5)	17 (6)

Some speakers (Speakers 1, 6, 8 and 9) almost exclusively produced default contours, independently of the context. Some speakers realized contours that were akin to the vocative chant, with a sustained final mid tone (Speakers 3, 10 and 12). Note that although Metropolitan French is not the local variety of French of our Cameroonian speakers, speakers still have access to it through the media. Some of our speakers have also spent a few years in Europe, which might explain why they use this contour while other speakers do not. Speaker 10, in particular, produced a higher percentage of chanting contours in the routine context. Speaker 4 favored rising contours in both the routine and the urgent context, while Speakers 2 and 3 favored rising contours in the routine context only.

4.2. F0 Scaling of Tonal Landmarks

4.2.1. Metropolitan French

Beyond the choice of contour exerted in different contexts, we were also interested in the effect of context on the phonetic realization of contours. Starting with the chanting contour, our results show that the context did not have a significant effect on F0. Pairwise F0 comparisons within each tone (IN, RO, H1, L, H2, see Figure 5) across the two contexts did not reveal any significant differences either. Note that chanting contours only represent 7% of urgent calls whereas they represent 64% of routine calls.

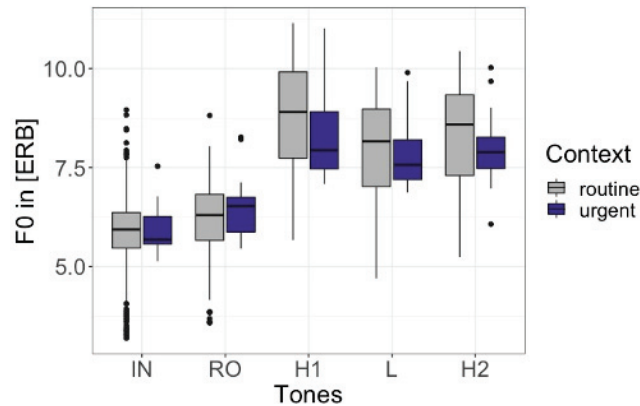


Figure 5. F0 scaling of the chanting contour in the routine and urgent context in MF.

Similarly, in the rising-falling contour, the context did not show a significant effect on F0. The pairwise comparisons did not reveal any significant differences in F0 between the routine and urgent context in IN, RO and H tones, see Figure 6. However, as far as the final L tone is concerned, it was lower in the urgent than in the routine context ($t = 6.38$, $p < 0.001$). Note again that the rising-falling contour represents 83% of all urgent calls and 13% of all routine calls.

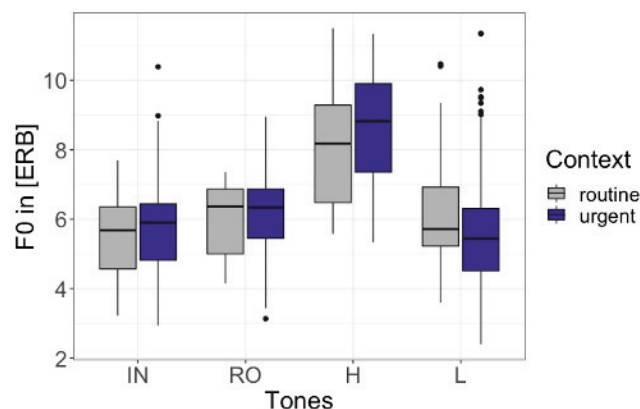


Figure 6. F0 scaling of the rising-falling contour in the routine and urgent context in MF.

Finally, in the rising contour, F0 was not significantly different in the routine and urgent context. None of the tones significantly differed across contexts, see Figure 7. Remember that rising contours count for 22% of routine calls and only 2.5% of urgent calls.

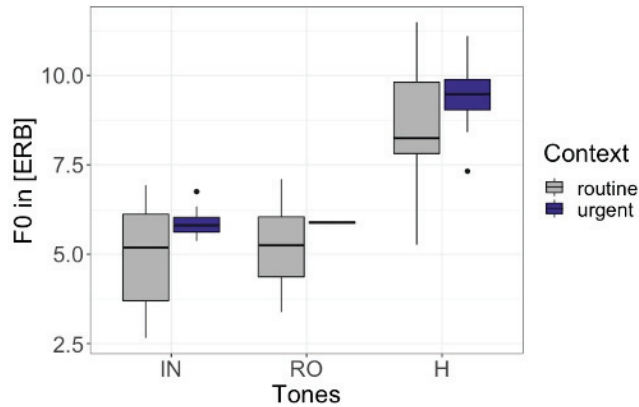


Figure 7. F0 scaling of the rising contour in the routine and urgent context in MF.

4.2.2. Basaá-Cameroonian French

Turning now to Basaá-Cameroonian French, and as context had little influence on the choice of contour, we were particularly interested in seeing whether context would significantly affect contour realization.

In contrast to what is observed in MF, the context exerted a significant effect on the realization of the chanting contour: F0 was higher in the urgent context as compared to the routine context ($t = 3.60, p < 0.01$). Remember that chanting contours represented 12% of all routine calls and 3% of urgent calls. Additionally, the interaction context*contour was at the level of statistical tendency (urgent*IN vs. routine L1; $t = -1.90, p = 0.057$), see Figure 8. Please note that our data set for the chanting contour in CF is extremely small (340 items), see also Figure 4.

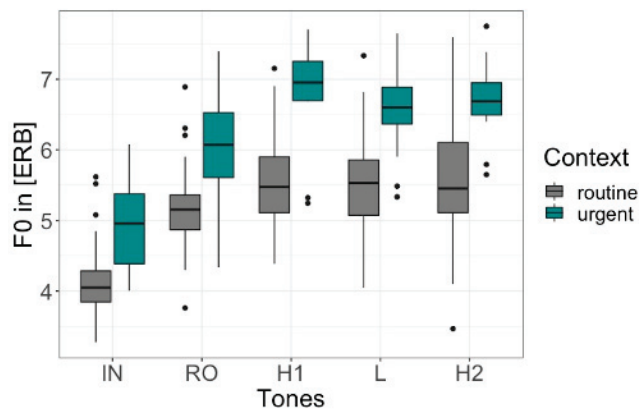


Figure 8. F0 scaling of the chanting contour in the routine and urgent context in CF.

As for the rising contour, the context appeared not to be significant, but the interaction context*contour turned out to be significant (urgent*IN vs. routine*H, $t = -2.88; p < 0.01$). We interpret it as a smaller difference between the routine context and the urgent context with regard to H than with regard to IN and RO. A pairwise comparison, however, revealed no significant effect of context for any of the three tones. Again, our data set is relatively

limited, as rising contours represented 21% of all routine calls and 8% of all urgent calls. See Figure 9.

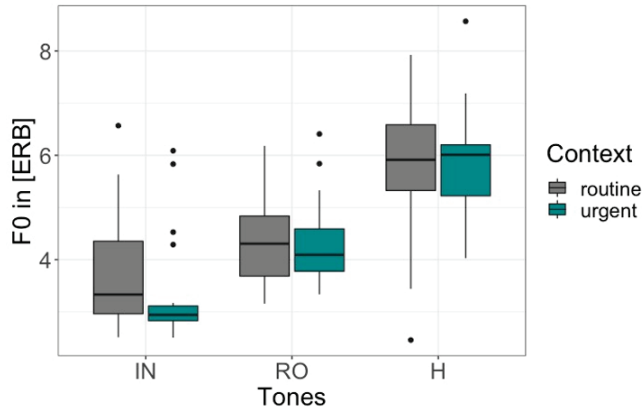


Figure 9. F0 scaling of the rising contour in the routine and urgent context in CF.

We also examined whether context exerted a significant effect on tones in the default contour, but the results did not reveal a significant difference. Remember that the default contour represented 57% of all routine calls and 81% of all urgent calls in CF. As the default contour is the surface realization of a sequence of lexical tones and different names consisted of different lexical tones (see Table 2), we also analyzed categories of names separately. Figure 10 shows F0 scaling for three syllabic words with L H L pattern (i.e., Marina, Natalia), and Figure 11, four syllabic names with the L H L pattern (i.e., Alexandra, Magdalena). The effect of context on tones did not reach significance.

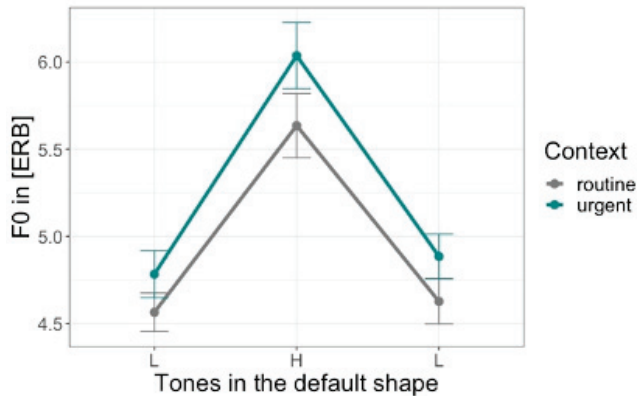


Figure 10. F0 scaling of the default contour in the routine and urgent context for three syllabic words with L H L lexical tones in CF.

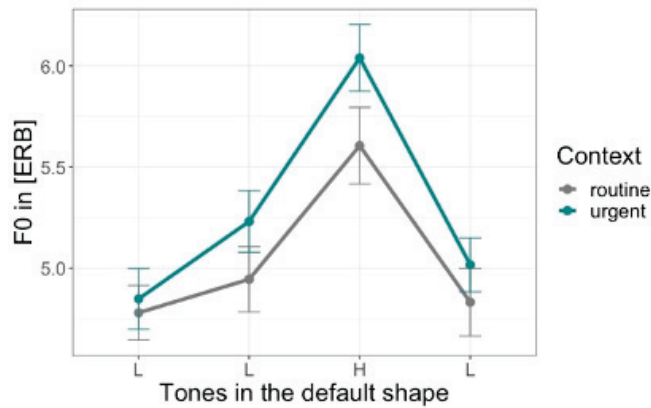


Figure 11. F0 scaling of the default contour in the routine and urgent context for four syllabic words with L L H L lexical tones in CF.

Let us now turn to F0 range.

4.3. F0 Range of Tonal Landmarks

4.3.1. Metropolitan French

F0 range for different calling melodies is represented in Figure 12 for MF. The interaction between shape and context was significant ($t = 3.83, p < 0.001$). Pairwise comparisons of similar contours across the routine and urgent context revealed no significant effect of context on F0 range, except for the category “other”, with lower F0 range values in the routine than in the urgent context ($t = 3.60, p < 0.05$). Our results also reveal, rather unsurprisingly, that F0 range values were lower in the chanting contour than in the rising-falling contour ($t = -6.85, p < 0.001$), and rising contours ($t = -8.23, p < 0.001$). No significant difference was found between the F0 range of rising-falling and rising contour. It should also be noted that longer words showed a larger F0 range (3 syllabic words vs. 1 syllabic words, $t = 3.73, p < 0.01$ and 4 syllabic words vs. 1 syllabic words, $t = 4.11, p < 0.01$).

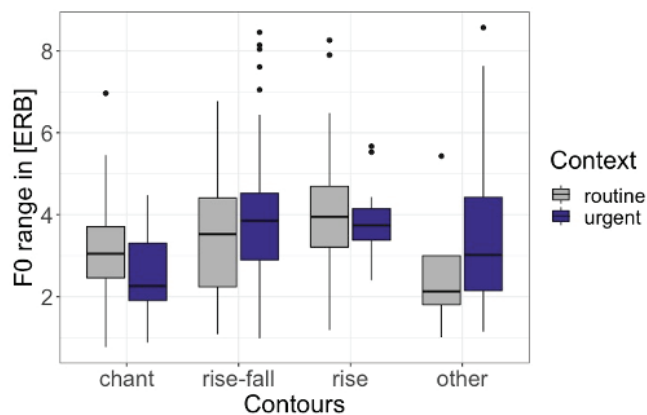


Figure 12. F0 range in the routine and urgent context across different calling melodies in MF.

4.3.2. Basaá-Cameroonian French

F0 range for different calling melodies is represented in Figure 13 for CF. The context did not exert a significant effect. When comparing F0 range across contours, our results reveal that F0 range values were significantly higher in the rising contour as compared to

the chanting contour ($t = 4.22, p < 0.001$), the default contour ($t = 3.85, p < 0.001$) and the category “other” ($t = 6.79, p < 0.001$). In contrast to MF, the number of syllables had no significant effect on F0 range.

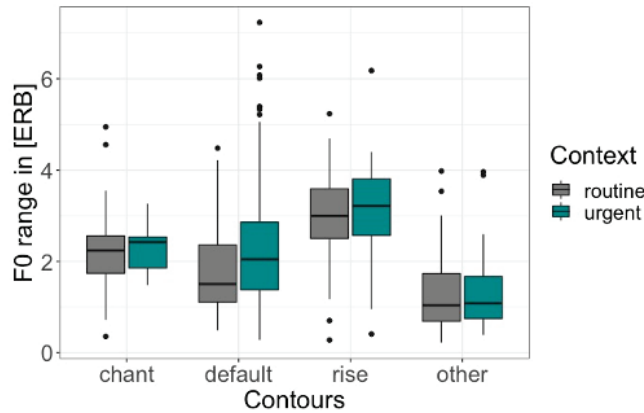


Figure 13. F0 range in the routine and urgent context across different calling melodies in CF.

Figure 14 shows F0 range in all CF speakers in both the routine and the urgent context. Overall, no uniform pattern emerges as to a possible consistent effect of context on F0 range in CF. The majority of speakers however realize the default contour with higher F0 range values in the urgent context (7 speakers out of 11). Other speakers seem to show the opposite tendency, i.e., a higher F0 range values in the routine context (Speakers 4, 7 and 12).

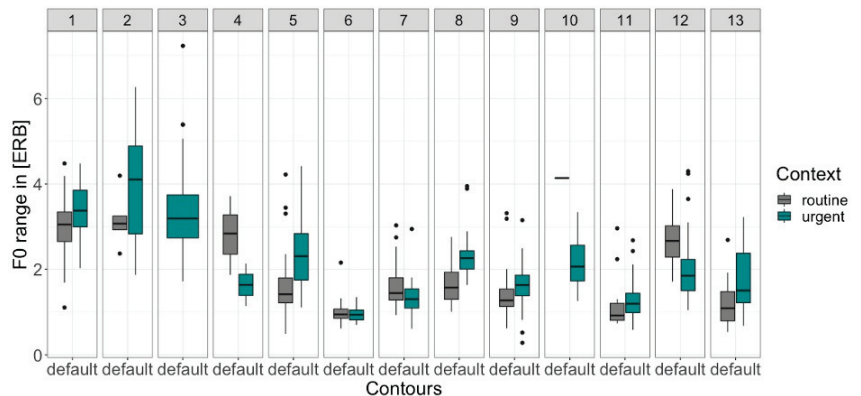


Figure 14. F0 range in the routine and urgent context for the default contour in CF speakers.

4.3.3. Cross-Dialectal Comparison

Languages are known to differ in their F0 range (Altenberg and Ferrand 2006; Eady 1982; Keating and Kuo 2012; Mennen et al. 2012; Nguyễn 2020). As in other varieties of European languages in contact with lexical tone, the intonational organization of Basaá-Cameroonian French looks different from the one of Metropolitan French. Observationally, some contours look like they result from a succession of tones carried by each syllable, corroborating previous literature on Cameroonian French. Pitch excursions, even in the rising and chanting contours, look much smaller than in Metropolitan French. If CF intonational contours are the implementation of successive lexical tones transferred from

the phonological/phonetic inventory available in Basaá, we expect CF to show a narrower F0 range than MF (Makasso et al. 2016). This would be reminiscent of what has been reported, for instance, in Nigerian English as compared to British English (Gut 2005, and references therein).

When comparing F0 range between the two varieties of French, it turns out that F0 range values are indeed significantly lower in Basaá-Cameroonian French than in Metropolitan French ($t = -9.06, p < 0.001$). Pairwise comparisons revealed that F0 range values in the chanting contour in MF were higher than in the same contour in CF ($t = 7.03, p < 0.001$). Similarly, the rising contour showed a higher F0 range in MF than in CF ($t = 6.89, p < 0.001$) and the “other” contours were also produced with significantly higher F0 range values in MF than CF ($t = 11.11, p < 0.001$). See Figure 15.

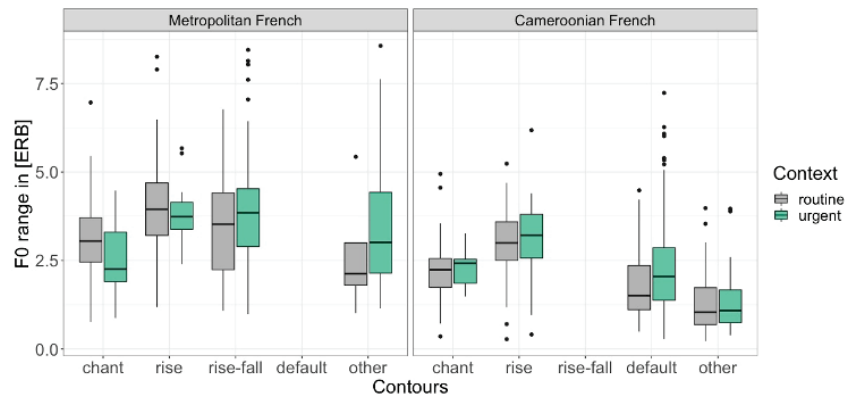


Figure 15. F0 range in the routine and urgent context across different calling melodies in MF and CF.

4.4. RMS Amplitude

4.4.1. Metropolitan French

Beyond the effect of context on F0, the question arises as to whether context influences the loudness of vocative calls, here measured in RMS amplitude. Our results show that calls produced in the urgent context were significantly louder than in the routine context ($t = 2.77, p < 0.01$). Pairwise comparisons also revealed that the amplitude was significantly higher in the rising-falling contour in the urgent than in the routine context ($p = 5.10, p < 0.001$). The same conclusions applied to the rising contour ($t = 4.15, p < 0.01$); see Figure 16.

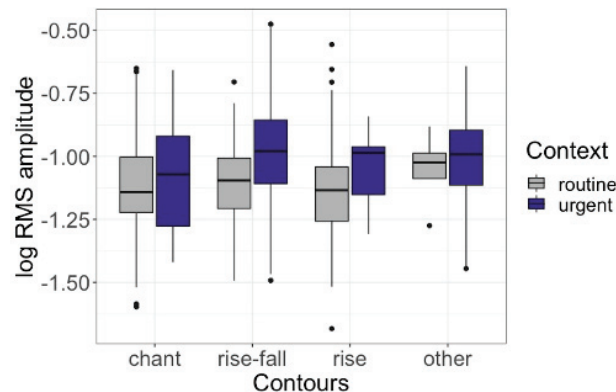


Figure 16. RMS amplitude in the routine and urgent context across different calling melodies in MF.

4.4.2. Basaá-Cameroonian French

In Cameroonian French, the items produced in the urgent context were significantly higher in comparison to those produced in the routine context ($t = 3.08, p < 0.05$); see Figure 17. Pairwise comparisons across contexts for individual contours did not reveal significant effects.

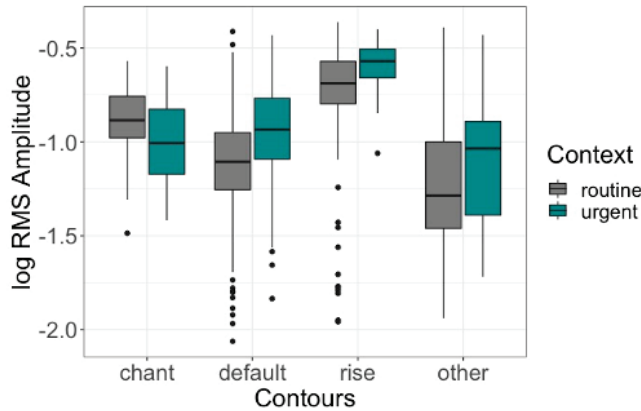


Figure 17. RMS amplitude in the routine and urgent context across different calling melodies in CF.

4.4.3. Cross-Dialectal Comparison

When comparing the RMS amplitude of calls in MF and CF, it turns out that the effect of Language was significant, i.e., names produced by the Basaá-Cameroonian French speakers were significantly louder in comparison to those produced by the Metropolitan French speakers ($t = 4.67, p < 0.001$). In particular, chants produced in CF were louder than chants produced in MF ($t = 7.03, p < 0.001$) and similarly, rising contours in CF were louder than rising contours in MF ($t = 6.89, p < 0.001$). Finally, a significant difference was also found in “other” contours, which were louder in MF than in CF ($t = 11.10, p < 0.001$). The results appear in Figure 18.

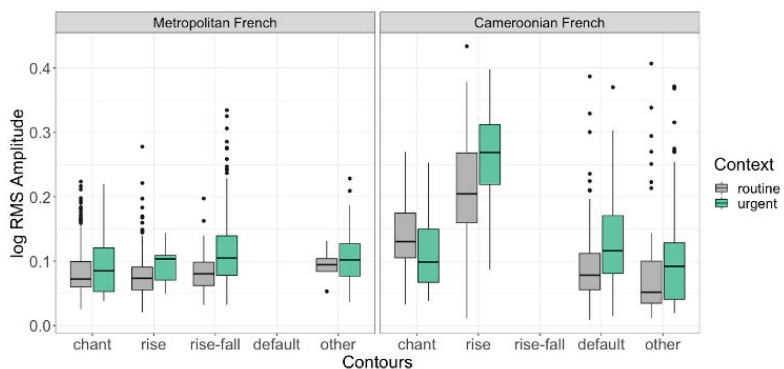


Figure 18. RMS amplitude in the routine and urgent context across different calling melodies in MF and CF.

Note, however, that the recordings were not made in the same conditions, as speakers were in two different countries, and that this acoustic parameter is particularly sensitive to the distance between speaker and microphone. These results should thus be interpreted with caution.

4.5. Word Duration

4.5.1. Metropolitan French

Finally, we were interested in determining whether context or contour would have a significant effect on word duration. As the syllable carrying a sustained downstepped high tone in the chanting contour has been reported to be lengthened, we expect names with a chanting tune to be longer than names carrying the other contours identified in our data (although see again Figure 1: the three contours are produced in the same context by the same speaker and they happen to all show a lengthened final vowel).

Word durations are visible in Figure 19 for Metropolitan French. The chanting contour was significantly longer than the rising-falling ($t = 6.03, p < 0.001$) and rising contour ($t = 6.71, p < 0.001$). No significant difference was found between the chanting contour and “other” shapes. The context did not however affect the word duration. Interestingly, overall, only 4 syllabic words were significantly longer from one-syllabic words ($t = 5.39, p < 0.001$).

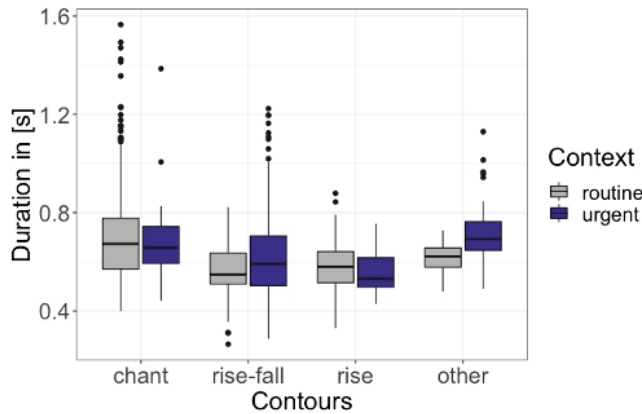


Figure 19. Word duration in the routine and urgent context across different calling melodies in MF.

In addition, we were also interested in whether one-syllabic words were longer in the chanting contour as opposed to the rising-falling and the rising contour, as they tend to become bi-syllabic words (Fagyal 1997; Ladd 2008). Results appear in Figure 20.

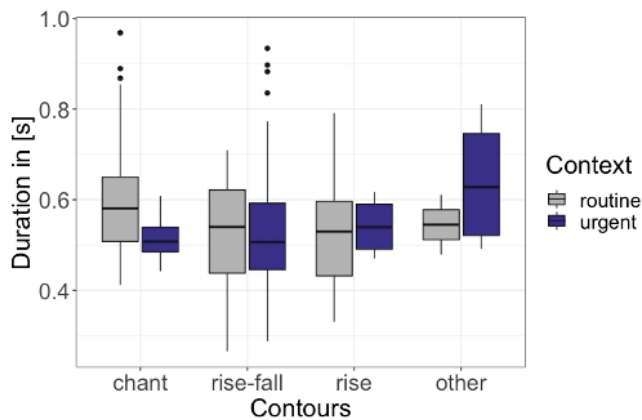


Figure 20. Word duration of one-syllabic words in the routine and urgent context across different calling melodies in MF.

The results show that one syllabic words were indeed longer when realizing the chanting contour than when realizing the rising-falling and the rising contour ($t = -5.14, p < 0.001$).

4.5.2. Basaá-Cameroonian French

The findings for word duration are shown in Figure 21. The context did not exert a significant effect on word duration. As in MF, the names produced with a chanting melody were significantly longer as compared to words with the default contour ($t = 4.14, p < 0.001$). Please note that there were only a few one syllabic words produced with the chanting contour so that a statistical comparison with respect to one-syllabic words with other contours was impossible.

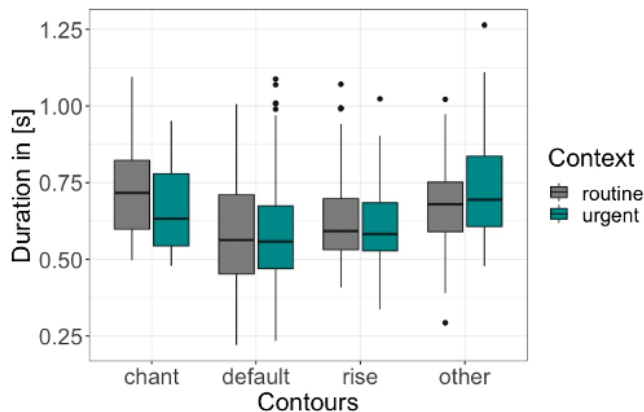


Figure 21. Word duration in the routine and urgent context across different calling melodies in CF.

4.5.3. Cross-Dialectal Comparison

Finally, for the sake of exhaustivity, we compared word duration in MF and CF and the effect of language turned out to be significant: words produced in MF were longer than those produced in CF ($t = -2.49, p < 0.001$). The only significant comparison across languages was found between the word duration in MF chants, that were longer in comparison to the default contours in CF ($t = 3.34, p < 0.05$). See Figure 22.

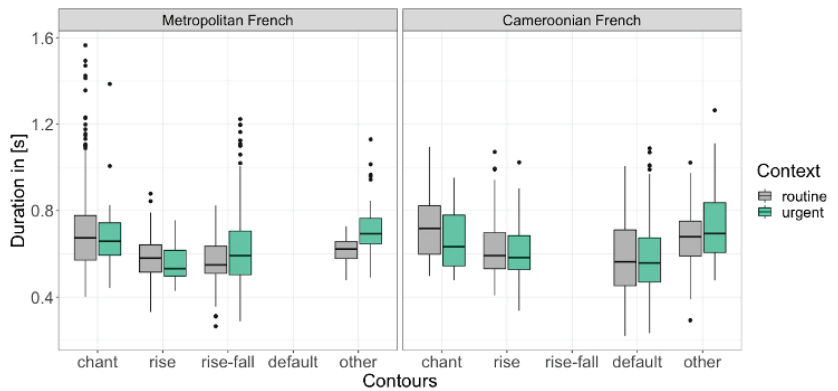


Figure 22. Word duration in the routine and urgent context across different calling melodies in MF and CF.

5. Discussion and Conclusions

As stated in the Introduction, the expression of vocatives lies at the interface between different language components: phonology, syntax, morphology, semantics and pragmatics. It provides us with a precious window onto the relationship between prosody and pragmatics in specific languages.

We have conducted a systematic study of the prosody of vocative calls in two different contexts (i.e., routine and urgent) among monolingual Metropolitan French speakers on the one hand and bilingual speakers of Basaá, a tone language, and Cameroonian French, on the other. The results of our DCT indicate that our two groups of speakers behave differently when it comes to the influence of context on the choice of contour. Context has a clear impact on the choice of contour in Metropolitan French, in which a chanting contour is favored in routine contexts and a rising-falling contour is preferred in urgent contexts. In Basaá-Cameroonian French, a contour consisting of a sequence of lexical tones is largely favored in both the routine and the urgent context. We have also observed a considerable amount of inter-speaker variation: in Cameroonian French, in particular, some speakers consistently produced the same contour independently of the context while others alternated between several contours including, for a minority, the chanting contour. As our Cameroonian speakers are also exposed to other varieties of French, for instance through the French media, we cannot exclude an influence from the exonormative standard, Metropolitan French.

Regarding acoustic parameters, context did not exert much effect on the F0 height of the different calling melodies in Metropolitan French. In contrast, we found an effect of context on the realization of vocative chants in Basaá-Cameroonian French, with a higher F0 in urgent contexts. This result is however to be interpreted with caution, as our speakers only produced a minority of chanting contours. We did not find an effect of context on F0 scaling for the two other contours (default and rising contour), indicating that our speakers did not realize the contours differently depending on the context.

When it comes to F0 range, neither of the varieties showed a significant effect of context. Interesting differences were observed across calling melodies in MF, with lower F0 range values in the chanting contour than in the rising-falling and rising contour. F0 range values were also partly dependent on word length (3 and 4 syllabic words showing significantly higher F0 range values than 1 syllabic words). In Basaá-Cameroonian French, the rising contour showed significantly higher F0 range values than the chanting and “other contours”. It did not however significantly differ from the default contour. No effect of word length was observed on F0 range, indicating that it remained stable across words of different lengths in Basaá-Cameroonian French. Interestingly, the two varieties of French significantly differ in their F0 range, with values being significantly lower in Basaá-Cameroonian French.

In both varieties, no effect of context on duration was found, but a significant effect of contour. Our results confirm that words carrying a chanting contour are longer than words with other contours in both varieties of French. For MF, in which our dataset was large enough for statistical analysis, this is also the case for 1-syllabic words, whose length only significantly differed from 4 syllabic words.

In sum, when it comes to the production of vocative calls, target-use of the contact variety, here Basaá-Cameroonian French, does not amount to target-like use of the original standard target language, here Metropolitan French (Williams 1989). We have seen that rather than using vocative chants in routine contexts, our speakers preferred a melody consisting of lexical tones reflecting the citation form of words in the standard target language (i.e., ending in a fall rather than a phonetic mid), and that this melody remained unchanged in urgent contexts. The very presence of lexical tones in the grammar of our speakers is consistent with what has been reported in other contact varieties of intonation languages in African as well as in Asian contexts. What is particularly interesting, and in our view further supports the idea of a prosodic transfer from the L1 (at an individual level in our present speakers but also maybe at a more collective level among Cameroonian French

speakers whose L1 is tonal), is the fact that the realization of these tones remains stable across pragmatic contexts. Remember that Basaá, the L1 of our speakers of Cameroonian French, is a language that has been described as showing little evidence of a use of prosody to encode postlexical meanings (Makasso et al. 2016), which is not uncommon among African tone languages (Downing and Rialland 2016). It is also a language in which vocative calls are expressed through morphological means (Bitjaa Kody and Mutaka 1997; Makasso n.d.).

To use Mennen's (2015) terminology, we observe a deviance in the semantic dimension from Metropolitan French. Such deviances have typically been reported in other areas involving information packaging: learners from different L1s and L2s have been shown not to mark prominence in focused elements and/or to fail deaccenting discourse-given ones (Gut and Pillai 2014; Hamlaoui et al. 2021; Ortega-Llebara and Colantoni 2014; Swerts and Zerbian 2010). This is particularly the case when their L1 does not use prosody to encode these information-structural categories (Rasier and Hiligsmann 2007; Zerbian 2015). The question arises as to the origin(s) of the present deviance. Historically, at the collective level, the present contact variety of French might have fossilized before vocative chants were acquired. This would be a natural result of the fact that very few speakers achieve native-like competence in an L2 (Edwards 1994). Although the vocative chant is common to a number of European languages and it was believed for a time that it might be a prosodic universal (Ladd 2008), it has not, to the best of our knowledge, been reported to exist in Basaá, Bantu or related languages. Its use within the grammar of Metropolitan French is actually limited to a small set of contexts, which might contribute to making it a marked prosodic pattern and thus particularly difficult to acquire for learners who do not already have it in their L1 (Eckman 1987). Studies of the prosody of vocative calls with other L1–L2 pairings might help shed further light on this issue, including among speakers of Cameroonian French with L1s other than Basaá. Further studies of Cameroonian French will also help determine whether other areas of the grammar of this contact variety support the idea of a prosodic transfer of lexical tones from local L1s and whether, as has been proposed for other contact varieties of European languages (Gussenhoven and Udofot 2010; Lim 2007), it should be classified as something else than an intonation language.

Author Contributions: Conceptualization, F.H. and M.Ž.; Methodology, M.Ž.; Validation, F.H. and M.Ž.; Formal analysis, J.E., F.H. and M.Ž.; Investigation, S.I.Q.; Data curation, J.E., F.H., S.I.Q. and M.Ž.; Writing—original draft preparation, F.H. and M.Ž.; Writing—review and editing, F.H. and M.Ž.; Visualization, F.H. and M.Ž.; Supervision, F.H. and M.Ž.; Project administration, F.H. and M.Ž. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the BMBF, grant number 01UG1411 to Fatima Hamlaoui and Marzena Żygis. Marzena Żygis was additionally supported by the European Research Council (PI: Manfred Krifka, Speech Acts in Grammar and Discourse (SPAGAD), Advanced Grant 787939, ERC Horizon 2020). This research was also partially supported by a Connaught grant from the University of Toronto to Fatima Hamlaoui.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are not available to the public. The participants did not consent to third party sharing of their audio data.

Acknowledgments: Heartfelt thanks go to Carole Ngo Sohna for recording Basaá-Cameroonian French participants, to Céline Bonnotte for assistance with the perceptual categorization of the data and Emmanuel-Moselly Makasso for discussions and insights on Basaá and Cameroonian French. We are also extremely grateful to the audiences at the Workshop on Intonation, Language Contact and Social Factors (ILCSF 20), the 3rd UoF L2 Intonation Workshop, the ICU Tokyo Linguistics Colloquium Series on Prosody and the Phonology Colloquium of the Goethe Universität Frankfurt, to the Editors, Laura Colantoni and Ineke Mennen, and two anonymous reviewers for their helpful feedback. Special thanks also go to all our participants. The usual disclaimers apply.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Notes

- ¹ With its highly heterogenous population and over 250 local languages from the 4 major African language families, Cameroon is often described as ‘Africa in miniature’. Although a few of the local languages serve as languages of major communication at the regional level, none of them dominates at the national level (Tabi Menga 1999). French was first introduced in 1916 and was imposed as the only language of education in the part of Cameroon under French rule. It has since acquired vehicular language status and is regularly used for inter-group communication (Mendo Ze 1999). For over 80% of the Cameroonian population, i.e., people living in present-day Francophone regions of Cameroon, it is still the main language of education and administration (Onguene Essono 1999, 2003). Together with English, which has been present in the country since around 1840 and is still dominant in the Western regions of Cameroon, French has remained one of the two official languages, even after the country gained its independence in 1960 (Mendo Ze 1999).
- ² We are extremely grateful to Emmanuel-Moselly Makasso (*p.c.*) for drawing our attention to this fact and for providing us with the tonal specifications for the proper names used in the experiment. Note that we cannot presently state whether speakers of Cameroonian French with other L1s will differ from our present speakers, as Cameroonian French is certainly the result of the contact of French with many typologically unrelated languages (tone languages and others) and not only Basaá. Further research will inform us of the specificities of Basaá-Cameroonian French speakers as compared to other speakers of Cameroonian French.
- ³ Several realizations of the phoneme /r/ were observed. For more detail on this point, the interested reader is referred to Hamlaoui et al. (2020).
- ⁴ A limit of the present design is that it does not distinguish between polite and informal calls, a distinction that has been described to be relevant in Basaá. Rather, the calls elicited within the present study all fall within the informal type. A design of the type found in Borrás-Comes et al. (2015) could be used in the future to determine whether this distinction is somehow transferred onto Basaá-Cameroonian French. This could also allow determining whether Basaá-Cameroonian French speakers have distinct means of expressing the type of speaker-hearer connivence that typically licenses the use of the vocative chant in Metropolitan French (Ladd 2008).
- ⁵ As pointed out by an anonymous reviewer, regional varieties of French also vary in pronunciation, making the label Northern Metropolitan French (NMF, Carignan 2013; Nicholas et al. 2019) more appropriate for the variety of European French considered here. On the segmental level, all speakers exhibited a Standard French pronunciation, consistent with NMF. Although there were audible differences in the way some of our speakers realized calls (e.g., in terms of the intensity of positive/negative emotion associated with our discourse contexts, or assumed distance between caller and callee), none of these differences could be attributed to a difference in geographic origin or the influence of another language. Varieties of Northern French have been described as undergoing a process of accent levelling or phonological uniformization, by which regionally localized features tend to disappear. This is particularly true of urban varieties, with differences across urban centers (Boughton 2005). The interested reader is also referred to Armstrong and Pooley (2010). A larger and more controlled sample of speakers from different regions of France would be needed to account for possible cross-dialectal differences in vocative calls within Metropolitan French.
- ⁶ As noted in previous studies (see for instance Mennen et al. 2014), in European countries it is generally difficult to find monolingual speakers in the strictest sense of the word, as people have usually received secondary instruction in at least one foreign language. As pointed out by an anonymous reviewer, it might be more accurate to consider our NMF speakers as having French as their native and dominant language, rather than as monolingual speakers. The same applies to our group of Cameroonian French speakers, who have also received secondary instruction in at least one foreign language, typically English. We expect the potential transfer effect from these other languages to be negligible.
- ⁷ The meaning associated with the rising contour was that of a question of the type “Can you hear me?”, “Are you there?”.

References

- Altenberg, Evelyn P., and Carole T. Ferrand. 2006. Fundamental frequency in monolingual English, bilingual English-Russian, and bilingual English-Cantonese young adult women. *Journal of Voice* 20: 89–96. [CrossRef]
- Armstrong, Nigel, and Tim Pooley. 2010. *Social and Linguistic Change in European French*. Chapter Accents and levelling in France. Belgium and Switzerland. Berlin: Springer, pp. 100–49.
- Arvaniti, Amalia, Marzena Żygis, and Marek Jaskuła. 2016. The phonetics and phonology of the Polish calling melodies. *Phonetica* 73: 338–61. [CrossRef]
- Avanzi, Mathieu, and Guri Bordal Steien. 2016. La prosodie du français en contact. *Languages* 202: 5–12. [CrossRef]
- Bardovi-Harlig, Kathleen. 2002. *Handbook of Applied Linguistics*. Chapter Pragmatics and Second Language Acquisition. Oxford: Oxford University Press, pp. 182–92.

- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, Bin Dai, Fabian Scheipl, Gabor Grothendieck, Peter Green, and et al. 2020. Lme4 Package Version 1.1-23. Available online: <https://cran.r-project.org/web/packages/lme4/index.html> (accessed on 10 October 2022)
- Bitjaa Kody, Zachée Denis. 1990. Le Système Verbal du Basaa. Ph. D. thesis, Université de Yaoundé, Yaoundé, Cameroon.
- Bitjaa Kody, Zachée Denis, and Nguessimo Mutaka. 1997. Le vocatif en bantu A. *Journal of West African Languages* 26: 55–69.
- Boersma, Paul, and David Weenink. 2016. Praat Doing Phonetics by Computer. Software. Available online: <https://www.fon.hum.uva.nl/praat/> (accessed on 10 October 2022)
- Bordal, Guri. 2013. Le français centrafricain: Un français à tons lexicaux. *Revue française de linguistique appliquée* XVIII: 91–102.
- Bordal, Guri. 2015. Traces of the lexical tone system of Sango in Central African French. In *Prosody and Language in Contact*. Berlin: Springer, pp. 29–49.
- Bordal, Guri, and Chantal Lyche. 2012. *La variation prosodique régionale en français*. Chapter Regard sur la prosodie du français d’Afrique à la lumière de la L1 des locuteurs. Bruxelles: DeBoeck.
- Borràs-Comes, Joan, Rafel Sichel-Bazin, and Pilar Prieto. 2015. Vocative intonation preferences are sensitive to politeness factors. *Language and Speech* 58: 68–83. [CrossRef]
- Bot Ba Njock, Henri-Marcel. 1964. Les tons en basaa. *Journal of African Languages* 3: 252–59.
- Boughton, Zoë. 2005. Accent levelling and accent localisation in French: Comparing Nancy and Rennes. *French Language Studies* 15: 235–56. [CrossRef]
- Braunmüller, Kurt, and Christoph Gabriel. 2012. *Multilingual Individuals and Multilingual Societies*. Amsterdam: John Benjamins.
- Broselow, Ellen. 2000. Stress, epenthesis, and segment transformation in Selayarese loans. In *Proceedings of the Twenty-Fifth Annual Meeting of the Berkeley Linguistics Society 1999: General Session and Parasession on Loan Word Phenomena*. Berkeley: Berkeley Linguistics Society, pp. 311–25. [CrossRef]
- Brown, H. Douglas. 2000. *Principles of Language Learning and Teaching*. New York: Addison Wesley Longman.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Brown, Roger, and Albert Gilman. 1960. The pronouns of power and solidarity. In *Style in Language*. Edited by Thomas. A. Sebeok Sebeok. Cambridge: MIT Press, pp. 253–76.
- Carignan, Christopher. 2013. When Nasal Is More Than Nasal: The Oral Articulation of Nasal Vowels in Two Dialects of French. Ph. D. thesis, University of Illinois at Urbana, Champaign, IL, USA.
- Chen, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.
- Colantoni, Laura. 2011. *Intonational Phrasing in Romance and Germanic: Cross-Linguistic and Bilingual Studies (Hamburg Studies on Multilingualism 10)*. Chapter Broad-Focus Declaratives in Argentine Spanish Contact and Non-Contact Varieties. Amsterdam: John Benjamins, pp. 183–212.
- Colantoni, Laura, and Jorge Gurlekian. 2004. Convergence and intonation: historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition* 7: 107–19. [CrossRef]
- Cosmides, Leda. 1983. Invariances in the acoustic expression of emotion in speech. *Journal of Experimental Psychology. Human Perception and Performance* 9: 864–81. [CrossRef]
- Delais-Roussarie, Elisabeth, Mathieu Avanzi, and Sophie Herment. 2015. *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*. Berlin: Springer.
- Delais-Roussarie, Elisabeth, Brechtje Post, Mathieu Avanzi, Carolin Buthke, Albert Di Cristo, Ingo Feldhausen, Sun-Ah Jun, Philippe Martin, Trudel Meisenburg, Annie Rialland, and et al. 2015. Intonational phonology of French: Developing a ToBI system for French. In *Intonation in Romance*. Edited by Sonia Frota and Pilar Prieto. Oxford: Oxford University Press.
- Delattre, Pierre. 1966. Les dix intonations de base du français. *The French Review* 40: 1–14.
- Dell, François. 1984. L’accentuation dans les phrases en français. In *La Forme Sonore du Langage: Structure des Représentations en Phonologie*. Edited by François Dell and Jean-Roger Vergnaud. Paris: Hermann.
- Di Cristo, Albert. 2016. *Les musiques du français parlé. Essais sur l’accentuation, la métrique, le rythme, le phrasé prosodique et l’intonation du français contemporain*. Berlin: De Gruyter.
- Di Cristo, Albert, and Daniel Hirst. 1999. *Intonation Systems: A Survey of Twenty Languages*. Cambridge: Cambridge University Press.
- Dimmendaal, Gerrit. 1988. *Aspects du basaa*. Translated by Luc Bouquiaux. Paris: Peeters/SELAF.
- Djourn Nkwescheu, Angéline. 2008. Les tendances fédératrices des déviations du français camerounais. De l’identité des processus linguistiques dans les changements diachroniques et géographiques. *Le français en Afrique* 23: 167–98.
- Downing, Laura Jo, and Annie Rialland. 2016. *Intonation in African Tone Languages*. Berlin: Mouton de Gruyter.
- Eady, Stephen J. 1982. Differences in the f0 patterns of speech: Tone language versus stress language. *Language and Speech* 25: 29–42. [CrossRef]
- Eckman, Fred R. 1987. *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Chapter Markedness and the Contrastive Analysis Hypothesis. Cambridge: Newbury House, pp. 55–69.
- Edwards, John. 1994. *Multilingualism*. London: Routledge.
- Fagyal, Zsuzsanna. 1997. Chanting Intonation in French. *University of Pennsylvania Working Papers in Linguistics* 4: 77–90.
- Fame Ndong, Jacques. 1999. L’enrichissement du français en milieu camerounais. In *Le Français Langue Africaine: Enjeux et Atouts pour la Francophonie*. Edited by Gervais Mendo Ze. Paris: Publisud, pp. 195–207.
- Fónagy, Iván, Éva Bérard, and Judit Fónagy. 1983. Clichés mélodiques. *Folia Linguistica* 17: 153–85. [CrossRef]

- Frota, Sonia. 2014. The intonational phonology of European Portuguese. In *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford: Oxford University Press.
- Frota, Sonia, Marisa Cruz, Flaviane Svartman, Gisela Collischonn, Aline Fonseca, Carolina Serra, Pedro Oliveira, and Marina Vigário. 2015. *Intonation in Romance*. Chapter Intonational Variation in Portuguese: European and Brazilian Varieties. Oxford: Oxford University Press, pp. 235–83.
- Gabriel, Christoph, and Elena Kireva. 2014. Prosodic transfer in learner and contact varieties. *Studies in Second Language Acquisition* 36: 257–81. [CrossRef]
- Gibbon, Dafydd. 1976. Calls. In *Perspectives of Intonation Analysis*. Bern: Lang, pp. 274–87.
- Glewwe, Eleanor. 2021. The phonological determinants of tone in English loanwords in Mandarin. *Phonology* 38: 203–39. [CrossRef]
- Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, Carlos, and Inyang Udofot. 2010. Word melodies vs. pitch accents: A perceptual evaluation of terracing contours in British and Nigerian English. *Proceedings of Speech Prosody* 100015: 1–4.
- Gut, Ulrike. 2005. Nigerian English prosody. *English World-Wide* 26: 153–77. [CrossRef]
- Gut, Ulrike, and Stefanie Pillai. 2014. Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition* 36: 283–302. [CrossRef]
- Guthrie, Malcolm. 1948. *The Classification of the Bantu Languages*. Oxford: Oxford University Press for the International African Institute.
- Hamlou, Fatima, Jonas Engelmann, and Kriszta Eszter Szendrői. 2021. Prosodic marking of focus and givenness in Kinyarwanda and Rwandan English. *Stellenbosch Papers in Linguistics Plus* 62: 135–60. [CrossRef]
- Hamlou, Fatima, Siri Gjerløe, and Emmanuel-Moselly Makasso. 2014. High tone spreading and phonological phrases in Basaa. Paper presented at 4th International Symposium on Tonal Aspects of Languages, Nijmegen, The Netherlands, May 13–16. pp. 27–31. ISCA Archive.
- Hamlou, Fatima, Milena Injac, and Emmanuel-Moselly Makasso. 2020. La réalisation du /r/ en Afrique Centrale: une première exploration du français des locuteurs basaaphones (Cameroun). In *Actes du 7e Congrès Mondial de Linguistique Française (CMLF)*. Les Ulis: EDP Sciences.
- Hao, Yen-Chen. 2009. Tonal adaptation of English loanwords in Cantonese. Paper presented at 21st North American Conference on Chinese Linguistics (NACCL-21), Smithfield, RI, USA, June 6–8. Edited by Yun Xiao. Smithfield: Bryant University.
- Hill, Virginia. 2014. *Vocatives: How Syntax Meets Pragmatics*. Leiden: Brill.
- Huttenlauch, Clara, Ingo Feldhausen, and Bettina Braun. 2018. The purpose shapes the vocative: prosodic realisation of Colombian Spanish vocatives. *Journal of the International Phonetic Association* 48: 33–56.
- Hyman, Larry. 2003. *The Bantu Languages*. Chapter Basaa (A43). London: Routledge, pp. 257–82.
- Jun, Sun-Ah, and Cécile Fougeron. 1995. The accentual phrase and the prosodic structure of French. Paper presented at 13th ICPhS in Stockholm, August 13–19. Edited by Kjell Elenius and Peter Branderud. pp. 722–25.
- Jun, Sun-Ah, and Cécile Fougeron. 2000. A phonological model of French intonation. In *Intonation: Analysis, Modeling and Technology*. Edited by Antonis Botinis. Dordrecht: Kluwer Academic Publishers, pp. 209–42.
- Kang, Okim, and Alyssa Kermad. 2019. *The Routledge Handbook of Second Language Acquisition and Pragmatics*. Chapter Prosody in L2 Pragmatics Research. London: Routledge, pp. 78–92.
- Kasper, Gabriele. 2001. *Pragmatics in Language Teaching*. Chapter Classroom Research on Interlanguage Pragmatics. Cambridge: Cambridge University Press, pp. 33–60.
- Keating, Patricia, and Grace Kuo. 2012. Comparison of speaking fundamental frequency in English and Mandarin. *Journal of the Acoustical Society of America* 132: 1050–60. [CrossRef] [PubMed]
- Kenstowicz, Michael. 2006. Tone loans: the adaptation of English loanwords into Yoruba. In *Selected Proceedings of the 35th Annual Conference on African Linguistics*. Edited by John Mugane, John P. Hutchinson and Dee. A. Worman. Somerville: Cascadia Proceedings Project. pp. 136–46.
- Kiu, K. L. 1977. Tonal rules for English loanwords in Cantonese. *Journal of the International Phonetic Association* 7: 17–22. [CrossRef]
- Kubozono, Haruo. 2022. Interactions between lexical and postlexical tones: Evidence from Japanese vocative prosody. In *Prosody and Prosodic Interfaces*. Edited by Haruo Kubozono, Junko Ito and Armin Mester. Oxford: Oxford University Press.
- Kubozono, Haruo, and Ai Mizoguchi. 2019. The phonetics and phonology of vocative intonation in Tokyo Japanese. Paper presented at 19th International Congress of Phonetic Sciences, Melbourne, Australia, August 5–9. pp. 497–501.
- Kuhl, Patricia, Björn Lindblom, and Francisco Lacerda. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606–08. [CrossRef]
- Lacheret-Dujour, Anne, and Frédéric Beaugendre. 1999. *La Prosodie du Français*. Paris: Editions du CNRS.
- Ladd, Robert. 2008. *Intonational Phonology*, 2nd ed. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.
- Leben, William. 1996. Tonal feet and the adaptation of English borrowings into Hausa. *Studies in African Linguistics* 25: 139–54. [CrossRef]
- Leech, Geoffrey N. 1983. *Principles of Pragmatics*. London: Longman.
- Lemb, Pierre, and François de Gastines. 1973. *Dictionnaire Basaa-français*. Douala: Collège Libermann.
- Lenth, R. 2019. Emmeans: Estimated Marginal Means, Aka Least-Squares Means. Available online: <https://cran.r-project.org/web/packages/emmeans/index.html> (accessed on 10 October 2022)

- Lewis, Paul M., Gary F. Simons, and Charles D. Fennig, eds. 2015. *Ethnologue: Languages of the world*, 18th ed. Dallas: SIL International.
- Lim, Lisa. 2007. Merger and acquisitions: On the ages and origins of Singapore English particles. *World Englishes* 26: 446–73. [CrossRef]
- Lim, Lisa. 2009. Revisiting English prosody. *English World-Wide* 30: 218–39. [CrossRef]
- Major, Roy C. 2008. Transfer in second language phonology: A review. In *Phonology and Second Language Acquisition*. Amsterdam: John Benjamins, pp. 63–94.
- Makasso, Emmanuel-Moselly. 2008. *Intonation et mélismes dans le discours oral spontané en basaa*. Ph. D. thesis, Université de Provence (Aix-Marseille 1), Marseille, France.
- Makasso, Emmanuel-Moselly. n.d. Morphology and tonology of the vocative in Basaa. *in prep.*
- Makasso, Emmanuel-Moselly, Fatima Hamlaoui, and Seunghun Julio Lee. 2016. Aspects of the intonational phonology of Basaa. In *Intonation in African Tone Languages*. Edited by Laura J. Downing and Annie Rialland. Berlin: Mouton de Gruyter.
- Mamode, Meï-Lan. 2015. Variation prosodique dialectale en français mauricien: étude de l'accentuation des syllabes proéminentes en parole spontanée. In *Actes du Congrès de l'Association Canadienne de Linguistique 2015*, pp. 1–14. Available online: <https://cla-acl.ca/actes/actes-2015-proceedings.html> (accessed on 10 October 2022).
- Martin, Philippe. 1981. Pour une théorie de l'intonation. In *L'intonation: De L'acoustique à la Sémantique*. Edited by Mario Rossi. Paris: Klincksieck, pp. 234–71.
- Mendo Ze, Gervais. 1999. *Le Français Langue Africaine. Enjeux et Atouts pour la Francophonie*. Paris: Publisud.
- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63. [CrossRef]
- Mennen, Ineke. 2015. Beyond segments: Towards a L2 intonation learning theory. In *Prosody and Language in Contact*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin: Springer, pp. 171–88.
- Mennen, Ineke, and Esther de Leeuw. 2014. Beyond segments: prosody in SLA. *Studies in Second Language Acquisition* 36: 183–94. [CrossRef]
- Mennen, Ineke, Felix Schaeffler, and Catherine Dickie. 2014. Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition* 36: 303–29. [CrossRef]
- Mennen, Ineke, Felix Schaeffler, and Gerard Docherty. 2012. Cross-language differences in fundamental frequency range: A comparison of English and German. *Journal of the Acoustical Society of America* 131: 2249–60. [CrossRef] [PubMed]
- Mertens, Piet. 1987. L'intonation du Français: De la Description à la Reconnaissance Automatique de la Parole. Ph. D. thesis, Catholic University of Leuven, Leuven, Belgium.
- Mertens, Piet. 1993. Accentuation, intonation et morphosyntaxe. *Travaux de linguistique* 26: 21–69.
- Mesthrie, Rajend. 2008. Introduction: Varieties of English in Africa and South and Southeast Asia. In *Varieties of English 4: Africa, South and Southeast Asia*. Berlin and New York: Mouton de Gruyter, pp. 23–31.
- Nguyễn, Anh-Thu'. 2020. F0 patterns of tone versus non-tone languages: The case of Vietnamese speakers of English. *Second Language Research* 36: 97–121. [CrossRef]
- Nicholas, Jessica A, Zsuzsanna Fagyal, and Christopher Carignan. 2019. What is with your nasals? Perception of nasal vowel contrasts in two dialects of French. Paper presented at 19th International Congress of Phonetic Sciences, Melbourne, Australia, August 5–9. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Canberra City: Australian Speech Science and Technology Association Inc., pp. 1208–12.
- Nurse, Derek, and Gérard Philippson. 2003. *The Bantu Languages*. London: Routledge.
- Olawale, Rukayat. 2021. La Relation Entre Tons et Intonation en Yoruba à travers L'étude du Vocatif. Master's thesis, University of Toronto, Toronto, ON, Canada.
- Onguene Essono, Louis Martin. 1999. Les statuts du français au Cameroun: essai de description des situations réelles du français au Cameroun. In *Le Français LANGUE Africaine: Enjeux et Atouts pour la Francophonie*. edited by Gervais Mendo Ze. Paris: Publisud, pp. 285–99.
- Onguene Essono, Louis Martin. 2003. L'écriture francophone: enrichissement ou appauvrissement du français? L'exemple camerounais. *Zeitschrift für französische Sprache und Literatur* 113: 225–38.
- O'Rourke, Erin. 2005. Intonation and Language Contact: A Case Study of Two Varieties of Peruvian Spanish. Ph. D. thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA.
- Ortega-Llebara, Marta, and Laura Colantoni. 2014. Relations between form-meaning associations, access to meaning, and L1 transfer. *Studies in Second Language Acquisition* 36: 331–53. [CrossRef]
- Pešková, Andrea. 2019. L2 Italian and L2 Spanish vocatives produced by L1 Czech learners: Transfer and prosodic overgeneralization. Paper presented at International Congress of Phonetic Sciences ICPhS, Melbourne, Australia, August 4–10. pp. 1932–36.
- Pešková, Andrea, Ingo Feldhausen, Elena Kireva, and Christoph Gabriel. 2012. *Multilingual Individuals and Multilingual Societies*. Chapter Diachronic Prosody of a Contact Variety. Analyzing Porteño Spanish Spontaneous Speech. Amsterdam: John Benjamins, pp. 365–89.
- Pickering, Lucy. 2004. The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes* 23: 19–43. [CrossRef]
- Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Prieto, Pilar. 2014. The intonational phonology of Catalan. In *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford: Oxford University Press, pp. 43–80.

- Quiroz, Sergio Israel, and Marzena Żygiś. 2017. The vocative chant and beyond: German calling melodies under routine and urgent contexts. Paper presented at Interspeech 2017, Stockholm, Sweden, August 20–24. pp. 1209–12.
- Rasier, Laurent, and Philippe Hiligsmann. 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française* 28: 41–66.
- Ritter, Elizabeth, and Martina Wiltschko. 2020. Interacting with vocatives! Paper presented at 2020 Annual Conference of the Canadian Linguistic Association.
- Rosenhuber, Simon. 1908. Die Basa-Sprache. *MSOS* 11: 219–306.
- RStudio Team. 2020. *Rstudio: Integrated development for R*. Rstudio, Inc., version 1.1.453. Boston: RStudio Team.
- Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist* 70: 1075–95. [[CrossRef](#)]
- Scholaster, Hermann. 1914. Basa Bakoko-Sprache. *Anthropos* 9: 740–59.
- Schürle, Georg. 1912. *Die Sprache der Basa in Kamerun: Grammatik und Wörterbuch*. Hamburg: Abhandlungen des Hamburgischen Kolonialinstituts.
- Silverman, Daniel. 1992. Multiple scansion in loanword phonology: evidence from Cantonese. *Phonology* 9: 289–328. [[CrossRef](#)]
- Sosa, Juan Manuel. 1999. *La entonación del español*. Madrid: Cátedra.
- Swerts, Marc, and Sabine Zerbian. 2010. Intonational differences between L1 and L2 English in South Africa. *Phonetica* 67: 127–46. [[CrossRef](#)] [[PubMed](#)]
- Tabi Menga, Jean. 1999. Proposition pour un aménagement du plurilinguisme en Afrique francophone. In *Le Français Langue Africaine. Enjeux et Atouts pour la Francophonie*. Edited by Gervais Mendo Ze. Paris: Publisud, pp. 31–44.
- Thomas, Jenny. 1983. Cross cultural pragmatic failure. *Applied Linguistics* 4: 91–112. [[CrossRef](#)]
- Trouvain, Jürgen, and Bettina Braun. 2020. *The Oxford Handbook of Language Prosody*. Chapter Sentence Prosody in a Second Language. Oxford: Oxford University Press.
- Werker, Janet, and Richard Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7: 49–63. [[CrossRef](#)]
- Williams, Jessica. 1989. Language acquisition, language contact and nativized varieties of English. *RELC* 20: 39–67. [[CrossRef](#)]
- Zang Zang, Paul. 1999. Le phonétisme du français camerounais. In *Le Français Langue Africaine. Enjeux et Atouts pour la Francophonie*. Edited by Gervais Mendo Ze. Paris: Publisud, pp. 112–29.
- Zerbian, Sabine. 2015. *Prosody and Language in Contact*. Chapter Markedness Considerations in L2 Prosodic Focus and Givenness Marking. Berlin: Springer.
- Zwicky, Arnold. 1974. Hey, whatsyourname! In *Papers from the Tenth Regional Meeting. ChicagoLinguistic Society* 10. Chicago: Chicago Linguistic Society, pp. 787–801.

Article

Sentence Prosody and Register Variation in Arabic

Sam Hellmuth

Department of Language and Linguistic Science, University of York, Heslington, York YO10 5DD, UK; sam.hellmuth@york.ac.uk

Abstract: Diglossia in Arabic differs from bilingualism in functional differentiation and mode of acquisition of the two registers used by all speakers raised in an Arabic-speaking environment. The ‘low’ (L) regional spoken dialect is acquired naturally and used in daily life, but the ‘high’ (H) variety, Modern Standard Arabic, is learned and used in formal settings. Register variation between the two ends of this H–L continuum is ubiquitous in everyday interaction, such that authors have proposed distinct intermediate register levels, despite evidence of mixing of H and L features, within and between utterances, at all linguistic levels. The role of sentence prosody in register variation in Arabic is uninvestigated to date. The present study examines three variables (F0 variation, intonational choices and post-lexical utterance-final laryngealization) in 400+ turns at talk produced by one speaker of San’ani Arabic in a 20 min sociolinguistic interview, coded for register on three levels: formal (*fusha*), ‘middle’ (*wusta*): and dialect (*ʕa:mijja*). The results reveal a picture of key shared features across all register levels, alongside distinct properties which serve to differentiate the registers at each end of the continuum, at least some of which appear to be under the speaker’s control.

Keywords: Modern Standard Arabic; San’ani Arabic; diglossia; multilingualism; prosody; F0

Citation: Hellmuth, Sam. 2022. Sentence Prosody and Register Variation in Arabic. *Languages* 7: 129. <https://doi.org/10.3390/languages7020129>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 3 March 2022

Accepted: 3 May 2022

Published: 24 May 2022

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Diglossia in Arabic

The Arabic language situation is a classic, and perhaps unique, example of diglossia, with speakers alternating between a ‘low’ spoken regional variety (L), acquired naturally, and a ‘high’ variety (H), Modern Standard Arabic (MSA), learned and used in formal settings (Ferguson 1959). Mastery of MSA as well as dialect is part of what it means to be a “socially competent” speaker of Arabic (Khamis-Dakwar and Froud 2019, p. 300). Acceptance of this stance is reflected in the increasing switch towards integrated approaches to teaching Arabic as a foreign language so that learners how to use both dialect and MSA, and in the process also learn when to use them (Younes 2014).

The classic characterization of diglossia in Ferguson (1959) distinguishes diglossia from both bilingualism and from a ‘standard-with-dialects’ model. In bilingualism, the learner acquires two languages which are structurally distinct but which can both be used in the same situations. In a standard-with-dialects context, a learner acquires two varieties of the same language which are used in different situations, but for some speakers the standard variety is their dialect. In diglossia, the learner acquires two varieties which are used in different situations, and the two varieties share enough linguistic features to be recognized as the ‘same’ language, despite differing in many ways; crucially, however, the standard variety is not the dialect of any speakers. Ferguson defined the H and L varieties in diglossia in terms of fundamental differences in their functional distribution, prestige, literary heritage, mode of acquisition and degree of standardization.

The practical reality is complex, however, with most speakers operating comfortably on a range of levels (Bassiouny 2009). A number of authors have therefore conceptualized the H–L distinction in terms of multiple levels, ranging from three (Mitchell 1984, 1986) to nine (Parkinson 1991), with five levels commonly proposed (e.g., Badawi 1973). For example, Mitchell (1984) identified three distinct levels by supplementing the basic

H (formal)–L (informal) divide, with a further subdivision of the informal register into ‘careful’ versus ‘casual’. This middle level, often referred to in Arabic as *wusta*: ‘middle’, is sometimes characterized as the form used in conversations between Arabs from different dialect backgrounds. In this communicative context, local variants which are unlikely to be accessible to those outside the relevant speech community are avoided, and replaced with words or linguistic features which are shared across spoken dialects, and may indeed also be found in MSA.

Despite the practical utility of conceiving of register variation in terms of levels, it is increasingly accepted that the formal and colloquial varieties do not form a dichotomy, but lie instead at opposite ends of a continuum of variation between MSA and spoken dialects (Mejdell 2019). Recent neurophysiological evidence also points to a complex interweaving of different levels of linguistic representation between MSA and dialect (Khamis-Dakwar and Froud 2014). An apparent middle variety thus arises as a result of mixing features from either end of the continuum within a single utterance or stretch of speech. The mixed or middle variety is not a separate attractor in its own right but rather a description of the range of possible points in the middle of the continuum, and the claim that register variation occurs both within and between linguistic levels predicts a potentially infinite number of such points along that continuum. This mixed production was in earlier literature identified as a distinct form (‘Educated Spoken Arabic’) but is now generally termed ‘diglossic mixing’ (Owens 2019). The expectation is that linguistic features of different registers will vary on all linguistic levels (i.e., lexicon, syntax, phonology, morphology). The present paper explores whether this is also true of sentence prosody, for the first time. Exploration of this prediction is relevant to the wider study of sentence prosody since Arabic diglossia presents a special case where we may see greater overlapping of prosodic features than seen in bilingual settings.

Another key point for our purposes here is that, although Ferguson argued that the H and L varieties are divergent, in that they have many different linguistic features, he did not claim or expect them to be *discrete*. Indeed, the overlap in features between H and L forms the common ground that underpins the recognition of the two varieties as related. Mejdell (2019) argues for greater attention to the *shared* features between MSA and dialects (cf. also Khamis-Dakwar and Froud 2019) and suggests that these shared features form the background which allows speakers to select distinctive features from either end of the continuum for stylistic purposes. In the present study, we are able to explore, for the first time, which features of sentence prosody, if any, are used this way.

Owens (2019) also notes that most (of the relatively few) prior studies of diglossic mixing in Arabic focus on linguistic features for which the differences between MSA and dialects are well-defined and clear-cut, which presupposes prior descriptions of those features in the literature on Arabic. As we will see in the next section, there are few comparative studies of the prosody of Arabic dialects and descriptions of prosodic differences between MSA and spoken dialects are even more scarce. As a result, it is not surprising that no prior studies of diglossic mixing in Arabic have included prosodic features in the list of variables investigated in their datasets. A search of two of the best quality recent studies, each based on a good volume of data, confirms that in both studies intonation was used solely as a diagnostic for identification of factors affecting other variables of interest; Mejdell (2006) uses intonation to determine whether relative clauses are restrictive or not, in her study of diglossic mixing in Egyptian panel show data, and Hallberg (2016) uses intonation solely to identify clauses as complete or incomplete.

A key aim of the present study is thus to provide the first investigation of register variation in Arabic in which the variables of interest are linguistic features at the sentence prosody level.

1.2. Sentence Prosody in Arabic

Work on Arabic sentence prosody has flourished in the last two decades, as evident from the expanding scope of literature summarized in two recent review chapters (Chahal

2006; El Zarka 2017). The majority of spoken Arabic dialects are stress accent languages in which pitch features have a post-lexical function in the form of intonation. The above review articles document a growing number of descriptions of the intonation patterns of individual Arabic dialects in the Autosegmental-Metrical framework (Ladd 2008), which are complemented by earlier descriptions in British School models (Alharbi 1991; Soraya 1966) or using acoustic analysis (Badawi 1965; Rosenhouse 2011). Few studies of intonational variation in spoken Arabic dialects are based on a direct comparison of parallel data. A contributing factor may be the perception of prosodic annotation, on which much analysis of sentence prosody relies, as ‘cumbersome’ (Watson and Wilson 2017).

Most studies of register variation in Arabic have addressed syntactic, morphological and lexical variation, with some work on phonological variation at the segmental level. It is typically assumed that the prosodic properties of an individual’s home dialect will transfer into their formal register (e.g., Benkirane 1998), and this has indeed been documented for some properties such as word-stress placement (Mitchell 1975). A laboratory study of intonational features in formal and spoken Cairene Arabic (El Zarka and Hellmuth 2008) found a greater incidence of secondary accents and shorter prosodic phrases in formal speech, but all other intonational parameters, such as peak alignment, were parallel across the two varieties. There has been no prior work specifically targeting register variation in suprasegmental features including intonation in non-laboratory speech data.

The present study examines register variation between MSA and dialectal San’aani Arabic (SA), spoken in and around the Old City of San’aa, in the capital of Yemen. SA has been described in detail on most linguistic levels except sentence prosody, including syntax, morphology and segmental phonology, by Watson (1993, 1996, 2002). A preliminary description of SA intonation is outlined in Hellmuth (2014). A distinctive feature of SA intonation observed in that study is the use of a rise–fall nuclear contour in information-seeking yes/no questions, in contrast to the rise contour typically observed in the same context in most other Arabic dialects outside North Africa (Hellmuth 2018). The use of a rise–fall contour in yes/no questions is also noted in a preliminary study with Yemeni speakers from the regional city of Taizz (Salem and Pillai 2020).

Another distinguishing feature of SA, shared as an areal feature with other dialects and languages of South Arabia, is utterance-final laryngealization. This term covers a set of related post-lexical phonological processes occurring utterance-finally, before a pause (Watson and Bellem 2011). The key generalizations are that in word- and utterance-final position: obstruents and long vowels are glottalized; oral stops are produced as ejectives; nasals are deleted; sonorants are glottalized, devoiced or deleted; vowels, fricatives and affricates are lengthened (Watson and Asiri 2008). The occurrence of this cluster of properties at the edges of prosodic domains makes laryngealization a potential variable of interest to investigate register variation in MSA–SA.

The choice to focus on register variation in MSA–SA, rather than another MSA–dialect pair, is also facilitated by the serendipitous (if unintentional) elicitation of a sociolinguistic interview recording in which register variation was displayed throughout, which is described in Section 2.1 below.

1.3. Sentence Prosody and Bilingualism

Sustained contact between languages in the context of community bilingualism has been shown to result in a range of different effects on the prosody of both first (L1) and second or additional language(s) (L2). The L2 may display prosodic features of a dominant L1 (Nance 2015; O’Rourke 2004), or the L2 may affect the prosody of the L1 (Colantoni and Gurlekian 2004; Fagyal 2005). There are also cases involving the creation of wholly new prosodic features which are properties of neither L1 nor L2 but instead a fusion of the two (Queen 2012), as well as a set of prosodic features which specifically characterize learner intonation (Mennen 2015). These diverse patterns have been argued to be a particular feature of prosody because all languages make use of the same phonetic exponents (pitch, duration and intensity) in some form or other (Bullock 2009). However, there is considerable

variation in the details of the mapping of prosodic form to meaning, both within and between languages, creating an ‘indeterminacy’ which Sorace (2004) argues is a context that fosters changes to bilingual grammars.

Exploration of sentence prosody and register variation in Arabic is relevant to the wider discussion of sentence prosody in the context of community bilingualism and/or second language (L2) acquisition because of past inference in the literature that MSA is an L2 for Arabic speakers (e.g., Kaye 1972). This assertion was typically based on the fact that MSA is learned in school, thus explicitly, and typically in the context of formal instruction.

Recent evidence suggests characterization of the dialect–MSA relationship as L1–L2 is an oversimplification. Albirini (2019) argues against this claim on the basis of emerging evidence that MSA is acquired implicitly to some extent, by Arabic children growing up in an Arabic-speaking environment, through exposure to media content which is aimed at children and produced in MSA such as cartoons (Albirini 2016). Khamis-Dakwar and Froud (2019) also question the tacit assumption that acquisition of MSA equates solely to literacy development since MSA differs from dialects on many levels of linguistic analysis (alongside many shared features, of course).

However, emerging evidence from neuroimaging studies suggests that dialect–MSA displays patterns of processing which also differ from those seen in balanced bilinguals. In a series of papers, Khamis-Dakwar and Froud (Froud and Khamis-Dakwar 2017, 2021; Khamis-Dakwar and Froud 2014, 2019) argue that L1 Arabic speakers who have grown up in an Arabic-speaking environment show the same type (if not magnitude) of brain response to stimuli in MSA and dialect, which also differs from the brain’s response to parallel stimuli in an L2 (such as Hebrew). They call for increased study of dialects and MSA in direct parallel to improve our understanding of the cognitive processing at work in diglossia.

A key point in Ferguson’s original proposals is that in diglossia we will observe a markedness relationship between the H and L varieties, in which the H features are a subset of the L features, in particular for phonology. Although this tendency is indeed commonly observed (e.g., an L affix can be added to an H stem, but not vice versa), Owens (2019) reports counterexamples, in the realm of phonology (e.g., dialectal Closed Syllable Shortening applying to an MSA stem); he suggests that future larger scale studies are likely to reveal bidirectional H–L mixing to be the general rule.

This study provides a first opportunity to explore whether there are any indications of a markedness relationship in suprasegmental properties between H and L in domains larger than the word. The primary hypothesis of the study, however, is that a complex interweaving of features, which is the hallmark of diglossic mixing on other levels of linguistic analysis, will be found also in sentence prosody.

1.4. The Present Study

The present study examines three variables operating at the level of sentence prosody: (i) F0 variation, within and between turns at talk; (ii) intonational choices, including the type and distribution of pitch accents and phrase boundaries; (iii) incidence of utterance-final laryngealization. These variables are investigated in data from a single speaker, whose utterances are first coded for register on three levels: formal or *fusha* (F), middle or *wuṣṭā*: (W) and dialect or *ṣa:mijja* (A). The coding is based on non-prosodic features to avoid circularity, generating 400+ turns for analysis. Owens (2019, p. 89) laments the lack of large-scale studies of diglossic mixing in Arabic but acknowledges the difficulty in eliciting or obtaining the data needed for larger studies. He further notes that the many existing small case studies, despite their limitations in size, are nonetheless valuable for generating hypotheses to explore in larger studies, and for identifying variables of interest to investigate further. To the best of my knowledge, this is the first study of diglossic mixing of sentence prosody in any MSA–dialect pair, but it is certainly the first to investigate sentence prosody in the context of MSA–SA mixing. This study also serves as a

potential model of methods for the investigation of sentence-level prosody across registers of Arabic.

2. Methods

2.1. Participants, Materials and Procedure

The data examined are from a single speaker (f2) in a 20 min sociolinguistic interview. Two female participants (f1/f2) took part, with the author as the interviewer. The participants are sisters, aged 20–25 years at the time of recording, recruited through personal contacts of the author. The participants are from the Al-Ga'a district, adjacent to the Old City of San'aa; their extended family originate from a village in Greater San'aa (name redacted for anonymity). The author/interviewer has British English as L1 and learned Arabic as an adult largely in formal educational settings. The interview is part of a small corpus of data collected in San'aa in 2008. Participants provided informed consent to record audio of their speech and to use the transcripts and data excerpts in research (but not to open access sharing of the audio recording, due to cultural sensitivities).

The sociolinguistic interview was conducted using the Sense Relation Network (SRN) tool (Llamas 2007). The SRN is designed to encourage participants to use their vernacular speech variety by inviting them to discuss dialect-specific lexical choices in a conversation with another member of the same speech community. A version of the SRN was created to target lexical items reported to vary between the variety spoken in the Old City of San'aa and other Yemeni dialects (Watson 1993, 1996, 2000).

The interview took place in a quiet office. The audio was recorded directly to wav format at 44.1 KHz 16 bit using a Marantz PMD660 recorder. Each participant was recorded to a separate channel in the stereo file, via a Shure SM10 headset microphone.

The SRN is an elicitation tool rather than an experimental 'word list' task. Target words and phrases are presented on a network diagram as prompts for spontaneous verbal discussion between a pair of participants about the lexical items they use in their variety. In Arabic, this involves the presentation of target words in written MSA, for the list of target meanings (i.e., 'senses') for which participants are invited to report their local variants. Of the two participants in this interview, only f2 was sufficiently confident in reading MSA to work directly from the text prompts. Speaker f2 thus took the lead in directing the conversation, liaising between her two interlocutors whose familiarity with different varieties of Arabic varied greatly: speaker f1 was an expert L1 speaker of SA whereas the author/interviewer was an L2 speaker of Arabic with relatively limited exposure to SA, but good fluency in other dialects/varieties of Arabic. Speaker f2's talented and sensitive navigation of this linguistic situation led to considerable intra-speaker register variation throughout the conversation, which forms the basis of this case study.

2.2. Analysis

Transcription: The interview data was manually segmented into turn-sized sections, typically mapping to one or two Intonational Phrases (IP). Each turn was orthographically transcribed by the author in ELAN (Sloetjes and Wittenburg 2008) using a phonetically transparent roman alphabet transliteration system for Arabic, devised by Hellmuth and Almbark (2019). The stereo wav file was split to extract the audio signal for each speaker, in Praat (Boersma and Weenink 1992–2018), and the text transcription was force-aligned to the mono audio file using Prosody Lab Aligner (Gorman et al. 2011) as an aid to later coding and annotation. The resulting mono sound file and aligned Praat TextGrid for speaker f2 (only) were then used for further analysis.

Coding: Each turn produced by f2 was coded for the register of Arabic on three levels: *fusha* (F), *wusṭa*: (W) and *ṣa:mijja* (A). These levels correspond to Mitchell's (1984, 1986) formal/careful/casual distinction but were defined and coded in the present study according to a specific set of criteria. The decision to code with only three levels was made for pragmatic reasons; the data displays consistent mixing of linguistic features from different registers within and between turns, and it would not have been possible to

determine, a priori, which constellations of features correspond to which level(s). Since the aim of the present study is to determine which features of sentence prosody participate in diglossic mixing, the coding in this study was performed with reference to lexical choices, morphology and segmental phonology only; some examples are shown in (1–2).

- | | | | | |
|----|----|---|-----------------|---------|
| 1. | a. | F | na'qu:lu | 'la-ha: |
| | | | say.1PL | to-it |
| | b. | W | na'gu:lu | 'la-hu |
| | | | say.1PL | to-it |
| | c. | A | nu'gulluh | |
| | | | say.1PL.to.it | |
| | | | 'we say for it' | |
| 2. | a. | F | la'daj-na: | |
| | | | to-us | |
| | b. | W | la'de-na | |
| | | | to-us | |
| | c. | A | 'iinda-na | |
| | | | to-us | |
| | | | 'we have' | |

In (1) we see the same lexical item in all three registers (the root <قال> [q-a-l] 'to say') but differences between F/W versus A in morphology, with the prepositional clitic affixed directly to the verb in A only; in contrast, we see a difference between F versus W/A in the segmental phonology in the realization of the target sound [q] <ق> 'qaf', with [q] in the F register but [g] in both W and A. In (2) we see different lexical choices in F/W versus A, with the distinction between F and W indicated through monophthongization of [aj] to [e:] in W only. Further codes were used to indicate turns produced in English (E) or where the content was uninterpreted, e.g., a hesitation marker (U).

All turns in the data were coded by the author at two time points more than one year apart, and by a second coder who is a first language speaker of Arabic. Inter-code agreement (between either of the two author codes and those of the independent coder) was initially 60% (327/548). The remaining data were discussed and the majority of differences arose from the treatment of mixed turns (where part of the turn was in one register and part in another). A 'whole clause' approach was thus applied: the register of the majority of information in a turn was applied to the whole turn, even if it contained an isolated word or phrase with features of a different register (an example will be seen in Figure 6 below). Any remaining discrepancies were resolved by discussion to reach a consensus. The final coding involved some adjustments to turn boundaries, yielding a final turn count of 469.

Annotation: The wav file and TextGrid were segmented into turn-sized short files, and then each turn was prosodically annotated by the author and labelled for the presence/absence of the post-lexical phonological process of turn-final laryngealization. Prosodic annotation was performed following the conventions of the Autosegmental-Metrical framework (Ladd 2008), using the putative 'language-neutral' tone label set proposed by Hualde and Prieto (2016). The use of this language-neutral annotation label tagset results in minor differences in the annotation here of tunes previously discussed in Hellmuth (2014) but none of these minor differences are at issue in the examples discussed below. The adopted inventory of tone labels assumes two levels of phrasing (intermediate and intonational phrases) and thus includes pitch accents (marked '*'), phrase accents ('-') and boundary tones ('%'). A stylized representation of the pitch contour for each pitch accent label is provided in Appendix A.

Categorical presence or absence of turn-final laryngealization was identified from auditory impression with reference to the spectrogram and waveform in Praat, with comparison to detailed descriptions of SA turn-final laryngealization in different phonological contexts (Watson and Asiri 2008). To control for phonological context in the analysis, the syllable type for each turn-final word was recorded during annotation e.g., CVVT [ʔari:g]

‘path’; CVVN [tama:m] ‘fine, okay’ (where T stands for ‘any obstruent’ and N stands for ‘any nasal’). A Praat script was used to extract annotation labels from each turn level TextGrid, along with a count of the number of words in each turn.

F0 measurement: A Praat Pitch object was created for each turn (using default settings). All turns coded as F/W/A were inspected and manually corrected for tracking errors. A Praat script was used to extract the following F0 measures from each corrected Pitch object, in Hz and semitones: minimum, maximum, mean, standard deviation (SD) and median; the maximum and minimum were then used to calculate the F0 range in octaves [$\log_2(\max F_0/\min F_0)$] for each turn.

Data visualization and statistical analysis: The descriptive results of each layer of analysis were visualized using ggplot2 (Wickham 2010) supported by further exploration of acoustic data using linear regression models run in R (R Core Team 2014); mixed models with random effects were not appropriate as the data do not involve repeated measures.

3. Results

3.1. Overview of the Data

The data comprise 469 turns: 35 were coded as uninterpretable (e.g., hesitation markers) and 5 were produced partially or in whole in English, and these were excluded leaving 429 for analysis. The split of codes for the remaining data was: F:N = 44 (10%); W:N = 200 (47%); A:N = 185 (43%). Figure 1 shows the number of turns by register (1a) alongside a count of the number of turns of each length (by word count) in each register (1b). Figure 2 visualizes the distribution of each register type along the timeline of the 20 min interview, generated using vistime (Raabe 2021).

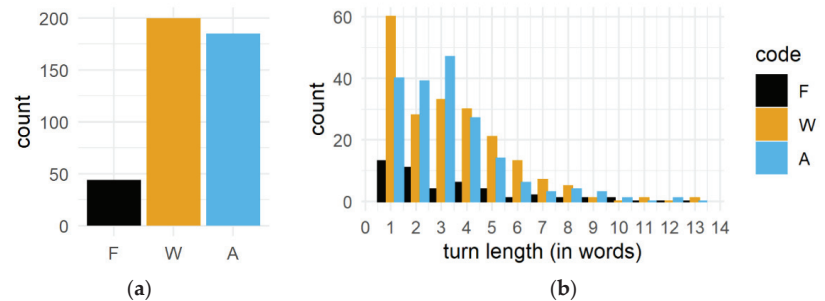


Figure 1. (a) Count of turns coded in each register type; (b) distribution of turn lengths in words, by register: *fusha* ‘formal’ (F), *wusta*: ‘careful’ (W) and *ša:mijja* ‘casual’ (A).

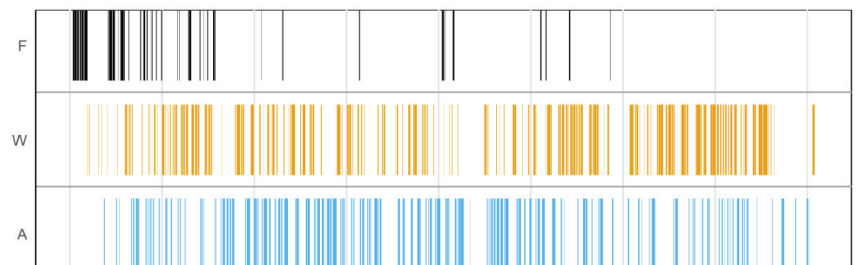


Figure 2. Distribution of turns in interview timeline by register: *fusha* (F)/*wusta*: (W)/*ša:mijja* (A).

Figures 1 and 2 show that the formal register was used in only a small proportion of the data and mostly at the beginning of the interview (in the first 4–5 min). Although the presentation of the target lexical items in written MSA initially elicited speech in formal register, the interactive nature of the SRN tool was successful in encouraging speaker f2

to gradually move towards the use of dialectal forms. In Figure 2 we can see that the careful register (W) is initially used to replace the formal (F) register, with the use of fully dialectal speech (A) following shortly afterwards; from about 5 minutes onwards, speaker f2 is largely using either W or A. Continued use of both registers was probably due to the presence of a non-vernacular speaker as the interviewer (favouring the use of the careful register, W), balanced against a shared focus on local lexis (favouring the use of the casual register, A).

Although fewer turns were produced in F than in W/A, the mean turn length in words is similar in all three registers (F = 3.14; W = 3.21; A = 3.11). The high number of single word turns coded as W is due to the decision to code all instances of the single word turn [tama:m] ‘okay’ (N = 41) as W (see discussion in Section 3.4). A data subset without these turns is used in relevant parts of the analysis (N = 388; F = 43 (11%); W = 160 (41%); A = 185 (48%)).

3.2. F0 Variation

Table 1 reports the mean and SD for the F0 measure by register code. The spread of values for these F0 measures across turns, by code, is illustrated in Figure 3.

Table 1. Mean (standard deviation) of measures of F0 variation across turns, by register code.

Register	Min (Hz)	Max (Hz)	Mean (Hz)	SD (Hz)	Median (Hz)	Range (Octaves)
F	215.58 (15.35)	329.68 (32.51)	258.97 (17.17)	29.34 (9.27)	251.84 (15.20)	0.61 (0.14)
W	212.41 (21.53)	318.89 (35.86)	256.64 (18.84)	27.09 (9.80)	252.00 (19.79)	0.58 (0.20)
A	209.11 (19.83)	325.53 (48.69)	261.74 (29.29)	29.77 (14.43)	259.46 (31.85)	0.63 (0.23)

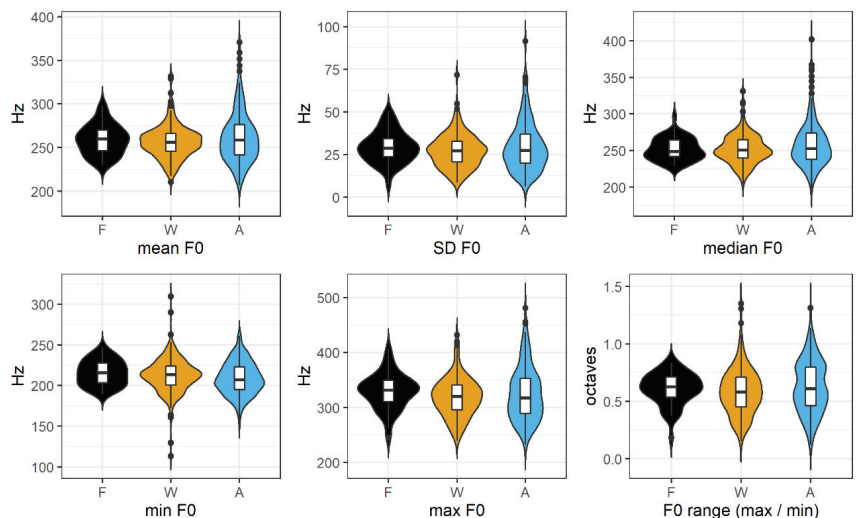


Figure 3. Median and interquartile range and frequency distribution of values across turns by register code for mean, SD, median, min and max values of F0, and F0 range (max/min) in octaves.

These measures reveal subtle differences only in the degree of F0 variation across registers. A wider range of variation is visible in turns labelled W or A, than F, but this is largely attributable to the larger number of tokens for those codes (90% of the data).

A series of linear regression models were run to predict each F0 measure in turn as the dependent variable, as a function of register *code* (e.g., minf0Hz~code) with treatment coding (i.e., with one level of the factor *code* as reference level); the model was re-run after re-leveling *code* to a different reference level to obtain pairwise comparisons. The only significant differences found in measures of F0 variation across registers were between W and A: median F0 is lower in W than A ($\beta = -7.458$; SE = 2.589; $t = -2.88$; $p = 0.0042$); mean F0 is lower in W than A ($\beta = -5.106$; SE = 2.427; $t = -2.104$; $p = 0.036$); SD of F0 is lower in W than A ($\beta = -2.674$; SE = 1.22; $t = -2.188$; $p = 0.029$); F0 range in octaves is narrower in W than A ($\beta = -0.0432$; SE = 0.021; $t = -2.034$; $p = 0.043$). There were no significant differences in measures of F0 variation between W and F.

The overall similar range of F0 variation across registers is perhaps to be expected as these are data from a single speaker and thus reflect her individual pitch range. The observed differences indicate greater use of higher and/or more expanded pitch by speaker f2 in A than W. The distribution of F0 range values is slightly bimodal in the A register, indicating a split which is also visible to a lesser extent in the distribution of values of max F0 in the A register. This split reflects the fact that f2 produced a subset of A-coded turns in a much wider pitch range, which have the auditory impression of being ‘performed’, as an example for the interlocutor of how an utterance would be produced naturally in context between SA speakers. Figure 4 shows an example in which f2 provides a sample of how a SA lexical item ([‘gawhaza’ ‘to sit’]) would be used; the reporting clause (‘And she says:’, coded W) is produced in a relatively narrow pitch span (0.5 octaves), but the reported clause (coded A) is produced in a very wide pitch span (1.3 octaves).

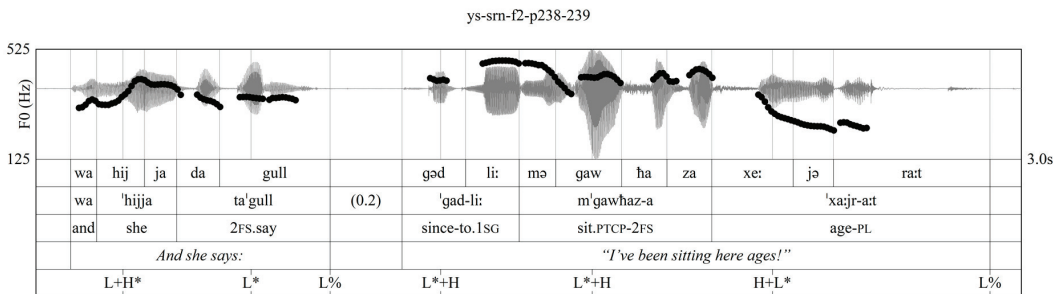


Figure 4. Sequence of turns (W then A) with narrow versus wide pitch span (0.5 versus 1.3 octaves).

In summary, then, the data reveal greater pitch variation in the casual (A) register than in the careful (W) and formal (F) registers. F0 variation is thus a linguistic feature relevant for the investigation of diglossic mixing in Arabic, as shown also for measures of F0 variation in formal versus informal speech in other languages such as Korean (Winter and Grawunder 2012). In the present data, however, the pattern observed in reported clauses suggests the variation here may be a by-product of differences in the semantic and/or pragmatic content expressed rather than an inherent property of any one register.

3.3. Intonational Phonology

Table 2 shows token counts for all pitch accent labels by register and Figure 5 illustrates the distribution of pitch accent types by register. The inventory of pitch accents used to label the F register data forms a subset of those needed to label the W/A registers. All registers share the property of using L* and H* as the most frequent pitch accents, with some use of bitonal rising pitch accents in all three also (L+H* and L*+H), but, bitonal falling pitch accents are used in W/A only, and the H+!H* pitch accent is more frequent in A than W.

Table 2. Token count of all pitch accent labels, by register code.

Register	L*	H*	!H*	L+H*	L*+H	H*+L	H+L*	H+!H*
F	66	28	5	14	3	0	0	0
W	246	144	56	33	15	11	3	7
A	170	112	58	32	37	28	3	36

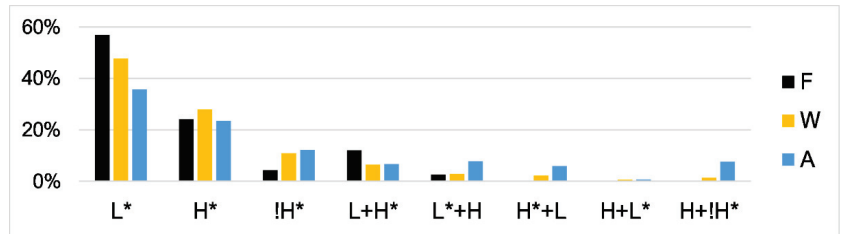


Figure 5. Observed pitch accent types as a percentage of total pitch accent tokens, by register.

Table 3 shows token counts for all edge tone labels, by register code. Most of the variation in the count of edge tones is due to the different volumes of data in each register. A count of the number of non-turn-final edge tones (phrase accents/boundary tones combined), as a proportion of the number of multi-word turns per register, in fact, reveals little difference in phrasing patterns between registers, as shown in Table 4. This is of note since differences in the distribution of phrase boundaries were reported as a feature of register variation for speakers from Egypt (El Zarka and Hellmuth 2008).

Table 3. Token count of all edge tone labels, by register code.

Register	L-	H-	!H-	Total	L%	H%	!H%	Total
F	1	19	0	20	13	31	6	50
W	25	48	0	73	108	114	0	222
A	25	47	3	75	144	51	12	207

Table 4. Incidence of turn-internal phrasing boundaries, by register.

Register	Turns with > 1 Phrase	Turns with > 1 Word	%
F	15	31	48%
W	74	140	52%
A	74	145	51%

Tables 5 and 6 show the distribution of ‘simple’ and ‘complex’ nuclear contours, respectively, by register, including all observed pitch accent boundary tone combinations. Cells of the table which account for 10% or more of the turns for that register are shaded in grey, in both tables. The ‘simple’ contours make up 98% and 91% of turns in the F/W registers respectively, but only 76% of turns in the A register.

Table 5. Observed ‘simple’ nuclear contours, as a percentage of all turns in that register.

Register	L* L%	L* H%	L* !H%	H* H%	H* L%	!H* L%
F	16%	56%	14%	0%	5%	7%
W	17%	43%	0%	5%	10%	18%
A	21%	15%	4%	3%	11%	21%

Table 6. Observed ‘complex’ nuclear contours, as a percentage of all in that register.

Register	L+H* L%	L+H* H%	L+H* !H*	L*+H H%	H*+L L%	H*+L H%	H+!H* L%	H+!H* H%	H+L* L%
F	0%	2%	0%	0%	0%	0%	0%	0%	0%
W	1%	1%	1%	0%	1%	2%	4%	0%	1%
A	0%	1%	1%	1%	5%	0%	13%	3%	2%

This apparent difference in the complexity of contours between the F/W versus A ends of the register continuum is largely driven by the high incidence of the H+!H* pitch accent in A-coded turns. Figure 6 shows an example of the distinctive H+!H* contour seen in many A-coded turns. Although the register coding was performed based on lexical and segmental features, the second transcriber remarked that many turns (which were later annotated with H+!H*) stood out as having ‘Yemeni intonation’.

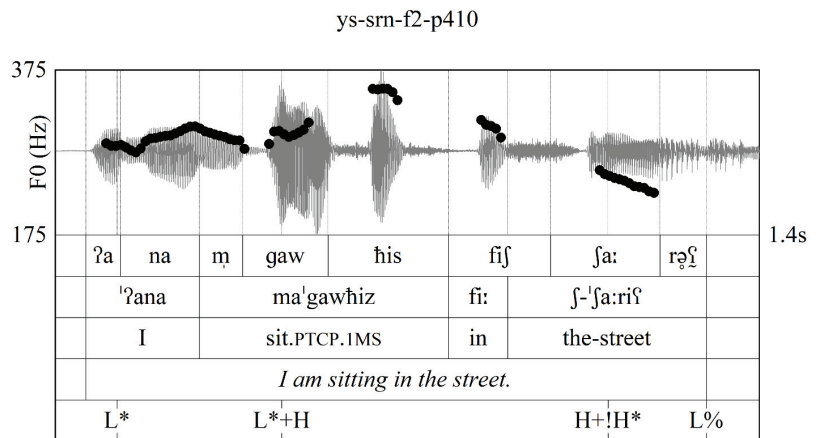


Figure 6. An A-coded turn realized with an H+!H* L% nuclear contour.

Another tendency in the data is a lower proportion of falling contours in the F register, in comparison to W/A, which may be due to speaker f2’s realization of many F turns as a sequence of short phrases, each of which bears a continuation rise, followed by a very short final phrase, in a pattern commonly heard in broadcast MSA speech; other patterns reported in broadcast MSA, such as sequences of early peak falls (Rastegar-El Zarka 1997), are not seen in the present data. Figure 7 shows an example of an F-coded turn realized with a series of continuation rises on short phrases, followed by a very short final phrase realized in a compressed pitch range.

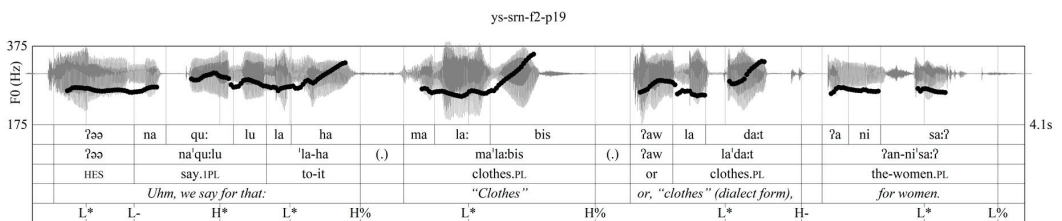


Figure 7. An F-coded turn with continuation rises and a broadcast MSA-style final short phrase.

One further feature that was shared across all three registers was the occasional use of secondary accents, whereby a word is realized with two pitch accents: one on the stressed syllable as expected, but another also on another syllable earlier in the word. The use of secondary accents in MSA but not dialectal speech was observed in a laboratory study of each register as produced by the same Egyptian speakers (El Zarka and Hellmuth 2008). In the present study, secondary accents are rare, but are more common in the F/W registers (three examples each): F: [ʕaːˈmijja] ‘dialect’ (turn 4); [ar-riˈzaːl] ‘the-man’ (turn 9); [ˌtalafazˈjoːn] ‘television’ (turn 324); and W: [al-ʔaˈðaːn] ‘the-ears’ (turn 84); [ˌtalafazˈjoːn] ‘television’ (turn 329); [al-ʕaˈʕiːd] ‘dumpling’ (turn 447)). There is just one example in an A-coded turn (A: [gusguˈsiː] ‘puppy’ (turn 354)). An example from a W-coded turn is shown in Figure 8.

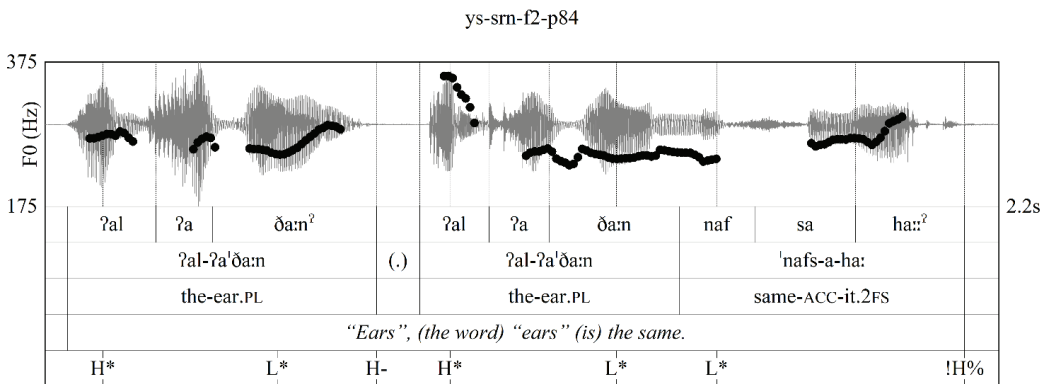


Figure 8. A W-coded turn showing secondary accents on the word [al-ʔaˈðaːn] ‘the-ears’.

In summary, the three registers share a core common inventory of pitch accents but falling bitonal pitch accents were only used in W-/A-coded turns, and more frequently so in the A register (particularly H+!H*). This difference contrasts with an F/W versus A distinction in the relative ‘complexity’ of nuclear contours. The incidence of turn-internal phrasing boundaries was similar across registers, but although all three registers contained examples of secondary accents they were more common in F/W than in A.

3.4. Post-Lexical Laryngealization

The proportion of turns in which laryngealization was identified in the final lexical item varied by register code: F had laryngealization in 18 out of 44 turns (41%); W in 98 out of 200 (49%), but A in 149 out of 185 (81%). Speaker f2 thus produces utterances with final laryngealization to an increasing extent as she moves from formal to dialectal speech. We might argue from this overall result that F/W pattern together in showing relatively low levels of laryngealization, in contrast to A where the rate is much higher. However, it is necessary to control for internal (linguistic) factors which also influence the incidence of laryngealization; the relevant factors in SA are the manner of articulation of the final consonant(s) and syllable structure (Watson and Asiri 2008).

Figure 9 shows the proportion of laryngealization for the most commonly observed syllable shapes (N = 400), by register code. The pattern in A-coded turns is of near categorical laryngealization of utterance-final obstruents and non-nasal sonorants, but slightly less of nasals; words ending in open syllables—which never attract stress—undergo laryngealization much less, consistent with Watson and Asiri’s (2008) observation that unstressed final syllables are less likely to be reduced. W-coded turns display a similar pattern of sensitivity to stress and final consonant (reduced incidence in CV open syllables and final nasals). The number of data points for F-coded turns is small, but we can see a

contrast in the treatment of CVV versus CVVN syllables, with the former much more likely to be laryngealized than the latter, matching the pattern in the A/W- coded turns.

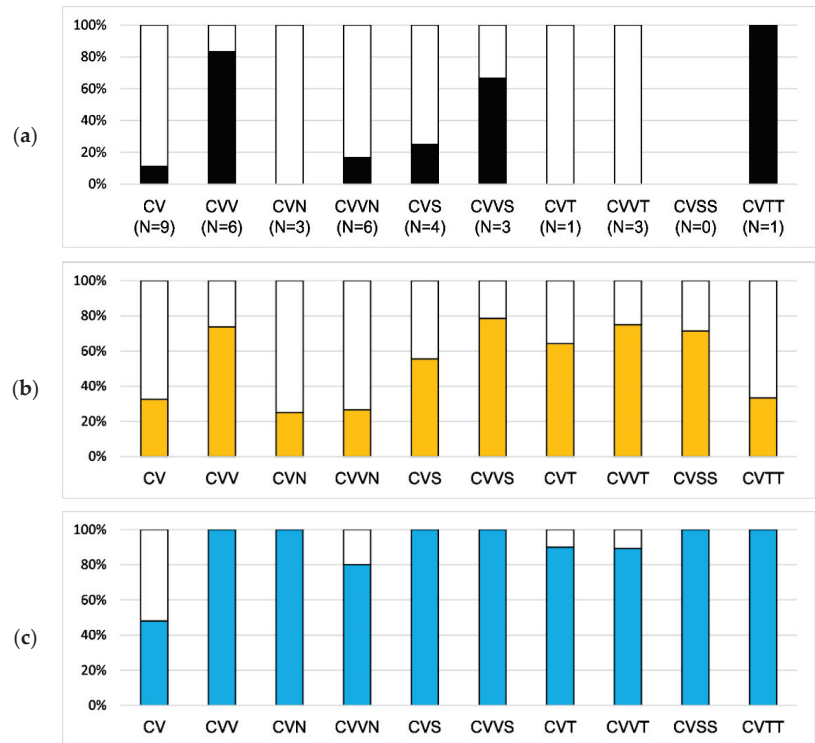


Figure 9. Proportion of turns displaying laryngealization by syllable type/shape and by register for (a): *fusha* ‘formal’ (F); (b): *wusta*: ‘careful’ (W); and (c): *sa.mijja* ‘casual’ (A). Key to syllable codes: C: any consonant; V: any vowel; N: any nasal; S: any non-nasal sonorant; T: any obstruent.

Overall then, the A-coded data show the expected patterns of laryngealization for a speaker of the SA dialect. The same speaker displays much less use of laryngealization in turns coded as F/W, but with some evidence of similar phonological conditioning to that observed in A-coded tokens.

There is an indication in the data that these patterns are under the speaker’s control. Figure 10 shows self-repair by speaker f2 of application of laryngealization on the word /jafʕal/ ‘do.IMPF.3MS’ (realized the first time as [jafʕa:lʔ]) at the potential completion point of a turn realized with F features (and where f2 is producing verbatim a text prompt written in MSA). She immediately produces an increment to the turn which is realized with the same prosodic contour as the host phrase which it repairs, except for suppression of laryngealization on the utterance-final word. Figure 11 illustrates the phonetic detail of the realization of the minimal pair realizations of the phrase-final word.

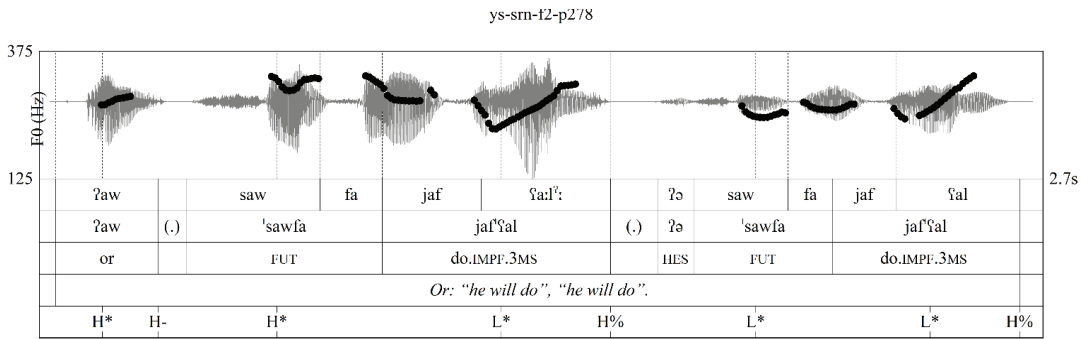


Figure 10. Sequence of F-coded turns produced with and without utterance-final laryngealization.

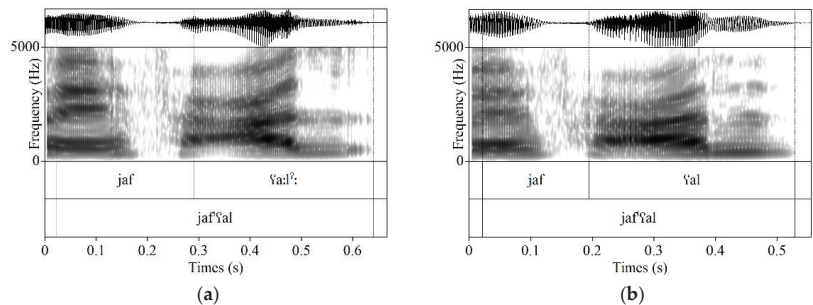


Figure 11. Two instances of [jaf.ʔal] (from Figure 10) with (a) and without (b) laryngealization.

Finally, there is some evidence also of co-variation between laryngealization and choice of prosodic contours. A large number of single word tokens of the discourse marker [tama:m] ‘okay/fine’ were produced by speaker f2 throughout the interview (N = 41). This word is a viable lexical item in both W or A registers, and does not display phonological features specific to W or A either; all tokens were coded as W, as in the interactional context of the sociolinguistic interviewer the intended audience was more likely to be the interviewer (for whom W is accessible) rather than the other participant. These 41 tokens vary in the incidence of laryngealization and also in the choice of nuclear contour, as set out in Table 7. The majority of [tama:m] turns are realized without laryngealization (78%), and the same proportion of turns are realized with a ‘simple’ rise contour (L* H%, also 78%). Although these choices do not strictly co-vary, the overall pattern is of a tendency to produce the discourse marker with prosodic features which fall in the common ground between F and A (since both registers use L* H%) but towards the formal end of the continuum of variation (and thus without SA dialectal laryngealization).

Table 7. Co-variation in laryngealization and choice of nuclear contour in tokens of [tama:m] ‘okay’.

Laryngealized?	L* H%	L* !H%	H* L%	L+H* L%	H*+L L%	Total
yes	5	0	2	2	0	9
no	27	1	3	0	1	32
Total	32	1	5	2	1	41

Interestingly, the only contour which does co-vary with the presence of laryngealization is L+H* L%, which is the contour typically observed in information-seeking yes/no-questions in SA (Hellmuth 2014). In these utterances we might conjecture that the intended audience of the turn was f2’s fellow participant (for whom A is accessible) rather than

the interviewer, leading to its realization in the A register and with SA prosodic features. Figure 12 shows an example of one of these [tama:m] turns, alongside a minimal pair realization of the same word with an MSA yes/no-question contour (L* H%).

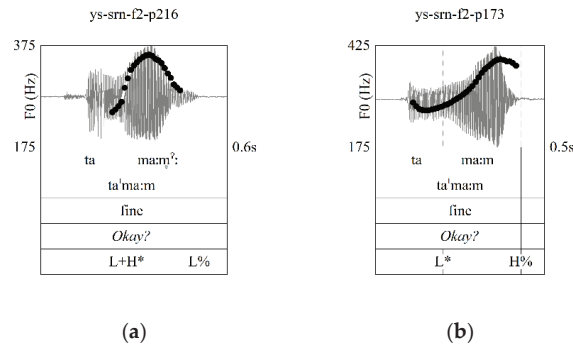


Figure 12. Single word W-coded turns of [tama:m] realized (a): with SA yes/no-question contour and with laryngealization; (b): with MSA yes/no-question contour and no laryngealization.

4. Discussion

Table 8 summarizes the observed differentiation of the identified registers of Arabic produced by speaker f2 in this case study data. The results show an interweaving of the use of different aspects of sentence prosody across the three registers of speech, with at least one feature serving as a cue to each of the possible ways of grouping the registers, though no feature fully differentiates F vs. W vs. A. All three registers shared the same density of phrasing boundaries, but the pattern of using a sequence of continuation rises on short phrases within a turn was a hallmark of F-coded turns only. A number of features distinguish A from W/F, and these features span all three of the investigated variables.

Table 8. Summary of observed variation in differentiation of registers, by prosodic feature type.

Pattern	F0 Variation	Intonation	Laryngealization
F&W vs. A	bimodal median F0 level of median F0	pitch accent inventory size use of secondary accents	% laryngealization
F vs. W&A	–	broadcast MSA-style contours	–
F vs. W vs. A	–	–	–
F&W&A	–	density of phrasing boundaries	–

The observed larger pitch accent inventory in A is a potential example of a markedness relationship between H and L predicted by Ferguson (1959). However, none of the features which distinguish A (in the top row of the table) are categorically exclusive to A. For example, W-coded turns also displayed a tendency towards bimodal median F0 and contained some tokens of falling bitonal pitch accents; also, some laryngealization was seen in all registers, and there was one example of a secondary accent in an A-coded turn.

This mixing of features across registers is consistent with the characterization of Arabic diglossic mixing as an interweaving of features of both A and F along a continuum of variation. The present dataset is small and limited in interactional scope, but there was some evidence here of the speaker displaying control over this variation, in the example of self-repair when A features were used in an otherwise F-framed turn. This self-repair was of a word produced with L features in H context—thus a counterexample of the type Owens (2019) cites as evidence of bidirectional mixing—but is repaired by speaker f2.

The present results for MSA–SA reveal a general picture of shared prosodic features across registers, alongside distinct features which serve to differentiate the registers at each end of the continuum, at least some of which appear to be under the speaker’s control. More work is needed to expand the volume of data, discourse context types and number of speakers investigated, but this study has identified variables and methods of analysis that can be used in future studies to further explore the role of sentence prosody in register variation in Arabic.

Funding: This research was funded by the University of York Research Priming Fund (Project Name: Prosodic Variation in Arabic). The APC was funded by the Department of Language and Linguistic Science, University of York.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the University of York Humanities and Social Sciences Ethics Committee (January 2008, Project Name: Prosodic Variation in Arabic).

Informed Consent Statement: Informed consent was obtained from all participants in the study.

Data Availability Statement: Derived data (such as measurements) are available from the author on request; the study participants did not consent to third party sharing of their audio data.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

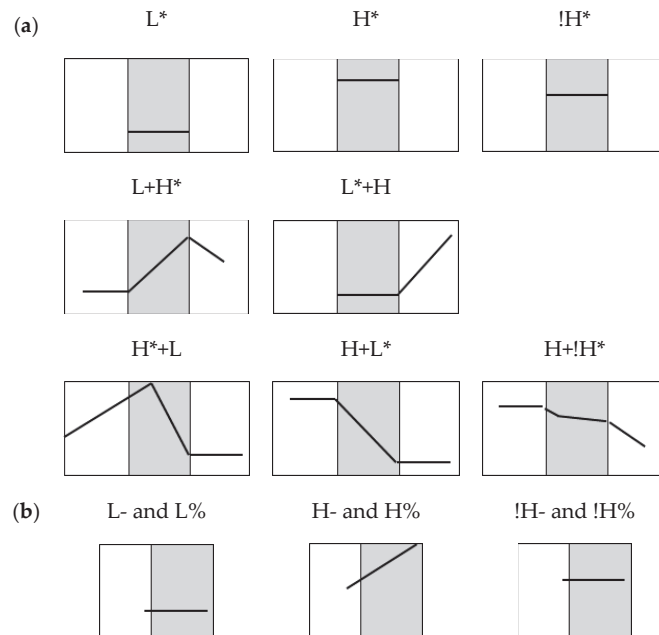


Figure A1. Schematized representation of a typical pitch contour labelled for (a) pitch accents and (b) edge tones. Boxes represent syllables; for pitch accents, the shaded box indicates the position of the accented syllable; for edge tones, the shaded box indicates the last syllable in the intermediate or intonational phrase.

References

- Alharbi, Lafi. 1991. Formal Analysis of Intonation: The Case of the Kuwaiti Dialect of Arabic. Unpublished Ph.D. thesis, Herriot-Watt University, Edinburgh, UK.
- Albirini, Abdulkafi. 2016. *Modern Arabic Sociolinguistics: Diglossia, Variation, Codeswitching, Attitudes and Identity*. London: Routledge.
- Albirini, Abdulkafi. 2019. The acquisition of Arabic as a first language. In *The Routledge Handbook of Arabic Linguistics*. Edited by Enam Al-Wer and Uri Horesh. London: Routledge, pp. 227–48.
- Badawi, El-Said M. 1965. An Intonational Study of Colloquial Riyadh Arabic. Unpublished Ph.D. thesis, SOAS University of London, London, UK.
- Badawi, El-Said M. 1973. *Mustawayaat al-'arabiyya al-mu'aasira fii miSr: BaHth 'ilaaqat al-lugha bi al-HaDaara*. Cairo: Daar al-Ma'aarif.
- Bassiouney, Reem. 2009. *Arabic Sociolinguistics*. Edinburgh: Edinburgh University Press.
- Benkirane, Thami. 1998. Intonation in Western Arabic (Morocco). In *Intonation Systems: A Survey of Twenty Languages*. Edited by Daniel Hirst and Albert Di Cristo. Cambridge: CUP, pp. 345–59.
- Boersma, Paul, and David Weenink. 1992–2018. *Praat: Doing Phonetics by Computer*, Version 6.0.40. [Computer Program]; Available online: www.praat.org (accessed on 2 March 2022).
- Bullock, Barbara E. 2009. Prosody in contact in French: A case study from a heritage variety in the USA. *International Journal of Bilingualism* 13: 165–94. [CrossRef]
- Chahal, Dana. 2006. Intonation. In *Encyclopedia of Arabic Language and Linguistics*. Edited by Kees Versteegh. Leiden: Brill Academic, vol. 2, pp. 395–401.
- Colantoni, Laura, and Jorge Gurlekian. 2004. Convergence and intonation: Historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition* 7: 107–19. [CrossRef]
- El Zarka, Dina. 2017. *Arabic Intonation*. Oxford Handbooks Online. Oxford: Oxford University Press. [CrossRef]
- El Zarka, Dina, and Sam Hellmuth. 2008. Variation in the intonation of Egyptian Formal and Colloquial Arabic. *Lingues et Linguistique: Revue Internationale de Linguistique* 22: 73–92.
- Fagyal, Zsuzsanna. 2005. Prosodic Consequences of Being a Beur: French in Contact with Immigrant Languages in Paris. *University of Pennsylvania Working Papers in Linguistics* 10: 91–104.
- Ferguson, Charles A. 1959. Diglossia. *Word* 15: 325–40. [CrossRef]
- Froud, Karen, and Reem Khamis-Dakwar. 2017. Neurophysiological investigations in studies of Arabic linguistics: The case of Arabic diglossia. In *The Routledge Handbook of Arabic Linguistics*. Edited by Elabbas Benmamoun and Reem Bassiouney. London: Routledge, pp. 249–62.
- Froud, Karen, and Reem Khamis-Dakwar. 2021. The Study of Arabic Language Acquisition: A Critical Review. In *The Cambridge Handbook of Arabic Linguistics*. Edited by Karin C. Ryding and David Wilmsen. Cambridge: Cambridge University Press, pp. 48–82.
- Gorman, Kyle, Jonathan Howell, and Michael Wagner. 2011. Prosodylab-Aligner: A tool for forced alignment of laboratory speech. *Journal of the Canadian Acoustical Association* 39: 192–93.
- Hallberg, Andreas. 2016. *Case Endings in Spoken Standard Arabic*. Lund: Studia Orientalia Lundensia.
- Hellmuth, Sam. 2014. Dialectal variation in Arabic intonation: Motivations for a multi-level corpus approach. In *Perspectives on Arabic Linguistics*. Edited by Samira Farwaneh and Hamid Ouali. Amsterdam: John Benjamins, pp. 63–89.
- Hellmuth, Sam. 2018. Variation in polar interrogative contours within and between Arabic dialects. Paper presented at the 9th International Conference on Speech Prosody 2018, Poznań, Poland, June 13–16.
- Hellmuth, Sam, and Rana Almbark. 2019. *Intonational Variation in Arabic Corpus 2011–2017*. Essex: UK Data Service.
- Hualde, José, and Pilar Prieto. 2016. Towards an International Prosodic Alphabet (IPRA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7: 25. [CrossRef]
- Kaye, Alan S. 1972. Remarks on diglossia in Arabic: Well-defined vs. ill-defined. *Linguistics* 10: 32–48. [CrossRef]
- Khamis-Dakwar, Reem, and Karen Froud. 2014. Neurocognitive modeling of the two language varieties in Arabic Diglossia. In *Perspectives on Arabic Linguistics XXVI*. Edited by Reem Khamis-Dakwar and Karen Froud. Amsterdam: John Benjamins, pp. 285–302.
- Khamis-Dakwar, Reem, and Karen Froud. 2019. Diglossia and language development. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London: Routledge, pp. 300–13.
- Ladd, D. Robert. 2008. *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press.
- Llamas, Carmen. 2007. A new methodology: Data elicitation for regional and social language variation studies. *York Papers in Linguistics Series* 8: 138–63.
- Mejdell, Gunvor. 2006. *Mixed Styles in Spoken Arabic in Egypt*. Leiden: Brill.
- Mejdell, Gunvor. 2019. Diglossia. In *The Routledge Handbook of Arabic Linguistics*. Edited by Enam Al-Wer and Uri Horesh. London: Routledge, pp. 332–44.
- Mennen, Ineke. 2015. Beyond segments: Towards a L2 intonation learning theory. In *Prosody and Language in Contact*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin: Springer, pp. 171–88.
- Mitchell, Terence Frederick. 1975. *Principles of Firthian Linguistics*. London: Longman.
- Mitchell, Terence Frederick. 1984. Soziolinguistische und stilistische Aspekte des gesprochenen Arabisch der Gebildeten in Ägypten und der Levante. *Sitzungsberichte der Sächsischen Akad. d. Wissenschaften in Leipzig, Philolog.-hist. Klasse* 123: 1–21.
- Mitchell, Terence Frederick. 1986. What is educated spoken Arabic? *International Journal of the Sociology of Language*, 7–32. [CrossRef]

- Nance, Claire. 2015. Intonational variation and change in Scottish Gaelic. *Lingua* 160: 1–19. [CrossRef]
- O'Rourke, Erin. 2004. Peak placement in two regional varieties of Peruvian Spanish intonation. In *Contemporary Approaches to Romance Linguistics. Selected Papers from the 33rd Linguistic Symposium on Romance Languages (LSRL)*. Edited by Julie Auger, J. Clancy Clements and Barbara Vance. Amsterdam: John Benjamins, pp. 321–42.
- Owens, Jonathan. 2019. Style and sociolinguistics. In *The Routledge Handbook of Arabic Sociolinguistics*. Edited by Enam Al-Wer and Uri Horesh. London: Routledge, pp. 81–92.
- Parkinson, Dilworth B. 1991. Searching for Modern Fusha: Real-life formal Arabic. *Al-Arabiyya* 24: 31–64.
- Queen, Robin. 2012. Turkish-German bilinguals and their intonation: Triangulating evidence about contact-induced language change. *Language* 88: 791–816. [CrossRef]
- R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.
- Raabe. 2021. *Vistime: Pretty Timelines in R*, Version 1.2.1. [Computer Program]; Available online: <https://cran.r-project.org/web/packages/vistime/> (accessed on 2 March 2022).
- Rastegar-El Zarka, Dina. 1997. Prosodische Phonologie des Arabischen. Unpublished Ph.D. thesis, Karl-Franzens-Universität, Graz, Austria.
- Rosenhouse, Judith. 2011. Intonation in the colloquial Arabic of Haifa: A syntactic-phonetic sociolinguistic study. *Israel Studies in Language and Society* 4: 117–42.
- Salem, Nada Mohammed, and Stefanie Pillai. 2020. An Acoustic Analysis of Intonation in the Taizzi variety of Yemeni Arabic. *Linguistics Journal* 14: 154–82.
- Sloetjes, Han, and Peter Wittenburg. 2008. Annotation by category: ELAN and ISO DCR. Paper presented at the 6th international Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco, May 28–30.
- Sorace, Antonella. 2004. Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition* 7: 143–45. [CrossRef]
- Soraya, Helmy A. I. 1966. An Intonational Study of Egyptian Colloquial Arabic. Unpublished Ph.D. thesis, SOAS University of London, London, UK.
- Watson, Janet C. E. 1993. *A Syntax of San'ani Arabic*. Wiesbaden: Harrassowitz Verlag.
- Watson, Janet C. E. 1996. *Sbahtu! A Course in San'ani Arabic*. Wiesbaden: Harrassowitz Verlag.
- Watson, Janet C. E. 2000. *Wasf San'a: Texts in San'ani Arabic*. Wiesbaden: Harrassowitz Verlag.
- Watson, Janet C. E. 2002. *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.
- Watson, Janet C. E., and Yahya Asiri. 2008. Pre-pausal devoicing and glottalisation in varieties of the south-western Arabian Peninsula. *Langues et Linguistique: Revue Internationale de Linguistique* 22: 17–38.
- Watson, Janet C. E., and Alex Bellem. 2011. Glottalisation and neutralisation in Yemeni Arabic and Mehri: An acoustic study. In *Instrumental Studies in Arabic Phonetics*. Edited by Zeki Majeed Hassan and Barry Heselwood. Amsterdam: John Benjamins, pp. 235–56.
- Watson, Janet C. E., and Jack Wilson. 2017. Gesture in Modern South Arabian Languages: Variation in multimodal constructions during task-based interaction. *Brill's Journal of Afroasiatic Languages and Linguistics* 9: 49–72. [CrossRef]
- Wickham, Hadley. 2010. ggplot2: Elegant graphics for data analysis. *Journal of Statistical Software* 35: 65–88.
- Winter, Bodo, and Sven Grawunder. 2012. The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics* 40: 808–15. [CrossRef]
- Younes, Munther. 2014. *The Integrated Approach to Arabic Instruction*. London: Routledge.

Article

Change across Time in L2 Intonation vs. Segments: A Longitudinal Study of the English of Ole Gunnar Solskjaer

Niamh Kelly

Department of Languages and Linguistics, University of Texas at El Paso, El Paso, TX 79902, USA; nekelly@utep.edu

Abstract: Research on L1 to L2 transfer has mainly focused on segments, while less work has examined transfer in intonation patterns. Particularly, little research has investigated transfer patterns when the L1 has a lexical pitch contrast, such as tone or lexical pitch accent, and the L2 does not. The current investigation is a longitudinal study of the L2 English of an L1 Norwegian speaker, comparing two timeframes. One suprasegmental feature and one segmental feature are examined: rise–fall pitch accents and /z/, because Norwegian and English have different patterns for these features. The results showed that the speaker actually produced more pitch movements in the later timeframe, contrary to the hypothesis, and suggesting that he was hypercorrecting in the earlier timeframe. In the early timeframe, virtually no /z/ was produced with voicing, while in the later timeframe, about 50% of /z/ segments were voiced. This suggests that the speaker had created a new category for this sound over time. Implications for theories of L2 learning are discussed.

Keywords: bilingualism; intonation; pitch accent; voicing contrast; longitudinal

Citation: Kelly, Niamh. 2022. Change across Time in L2 Intonation vs. Segments: A Longitudinal Study of the English of Ole Gunnar Solskjaer. *Languages* 7: 210. <https://doi.org/10.3390/languages7030210>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 28 February 2022

Accepted: 20 July 2022

Published: 9 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Among multilingual speakers, L1 to L2 transfer in segments has been well-described (e.g., [Flege and Port 1981](#); [Nagy 2015](#)), and recent research has also described transfer in suprasegmental aspects (e.g., [Mennen and de Leeuw 2014](#)). The goal of the current study is to examine, across time, the intonation patterns of a speaker whose L1 (Norwegian) uses lexical pitch accent but whose L2 (English) does not. I also examine a segmental pattern—voicing in English /z/—to see how that pattern develops across time in comparison with the intonation pattern.

1.1. Second Language Acquisition

Much work on L2 acquisition has focused on how segmental patterns transfer between a speaker's languages, whether at the individual level or at the community level. At the individual level, work on segments has found that the L1 and L2 can influence one another in either direction (e.g., [Flege and Eefting 1987](#); [Jarvis and Pavlenko 2009](#); [Lein et al. 2016](#); [Major 1992](#)). At the community level, when a whole group is bilingual, this can lead to larger scale transfer effects and the emergence of a contact variety ([Mayr and Siddika 2018](#); [McCarthy et al. 2013](#); [Nagy 2015](#); [Treffers-Daller and Mougéon 2005](#)).

One common measure in phonetic studies of bilingualism is voice onset time (VOT), a measurement of stop voicing which examines the time between the release of a stop closure and the onset of vocal fold vibration ([Lisker and Abramson 1964](#)). As languages can have different patterns for this measurement, it has proven to be a useful way of measuring transfer patterns in bilinguals' speech. For example, [Flege and Eefting \(1987\)](#) investigated stops in L1 Spanish speakers who were learning English, and found that in English, their voiceless stops had lower VOT values than monolingual English speakers, meaning that they were producing English voiceless stops with a more Spanish-like VOT pattern.

In terms of L2 acquisition theories, the Speech Learning Model (SLM) ([Flege 1987, 1995](#); [Flege and Bohn 2021](#); [Flege and Eefting 1987](#); [Flege et al. 2003](#)) states that L2 phonemes

are classified as new phonemes or as being similar to a phoneme in the L1 system. When the latter is the case, the L2 phoneme may be categorised as a phonetic realisation of an L1 phoneme, a process called equivalence classification. If the L2 phoneme is perceptually different from an L1 phoneme, it can be classified as a new sound and is thus easier to learn. The Perceptual Assimilation Model (PAM) (Best 1994; Best and Tyler 2007) also proposes that L2 sounds are perceived based on their similarity to L1 phonemes. As such, L2 phonemes may be assimilated in one of three ways: Two Category (TC) assimilation, where two L2 phonemes are categorised as two separate L1 phonemes (leading to accurate discrimination between them), Single Category (SC) assimilation, where two L2 phonemes are categorised as equally good (or poor) tokens of the same L1 phoneme, so discrimination between them is poor, and Category Goodness (CG), where two L2 phonemes are categorised as the same L1 phoneme, but they are not considered equally good examples of this phoneme.

1.1.1. Suprasegmentals

In terms of suprasegmental patterns in L2 learners, some general patterns have been found. For example, L1 Spanish speakers and L1 Dutch speakers learning English have a narrower pitch range in English than L1 English speakers (e.g., Backman 1979; Mennen 2008; Willems 1982). Similarly, Ordin and Mennen (2017) found that female bilingual speakers of English and Welsh used a wider pitch range in Welsh than in English. No such difference was found for male speakers. Intonation patterns have also been shown to transfer between a bilingual's languages, for example, bilinguals of Dutch and Greek were found to have transfer in alignment patterns in both directions (e.g., Mennen 2004).

In some instances, a speaker's languages may differ not just in alignment patterns or intonational pitch accent categories, but in whether pitch is used in lexical contrasts, as is the case in tonal languages. While intonation languages use different pitch accent categories for pragmatic purposes, such as focus, many languages use pitch lexically, including tonal languages and pitch accent languages. It is thought that over 40% of the world's spoken languages use lexical pitch (Maddieson 2011), with tonal languages being able to use pitch changes on every syllable, and pitch accent languages on every stressed syllable (Hayes 1995). In such languages, pitch accent is a tonal pattern found on stressed syllables (Beckman 1986; Hyman 2009), defined by Hualde (2012) as "a class of stress languages where words contrast in the tonal melody that is associated with the stressed syllable" (p. 1335). (More detail on the pitch accent system of Norwegian is provided in Section 1.2).

Some research has investigated transfer patterns when both of the languages use pitch lexically, but in different ways. One study on a speaker of L1 Swedish (a pitch accent language) and L2 Mandarin (a tonal language) found that the Swedish speaker could not produce a contour tone on a monosyllabic word, presumably since in Swedish, the pitch accent contrast only occurs on words of minimum two syllables (Tung 2006). As such, it appears that lexical pitch patterns in the L1 can interfere with the learning of tonal patterns in an L2. However, this study did not compare the Swedish speaker's productions to speakers of a language that does not use pitch lexically. A study comparing speakers of Swedish as an L2 who had L1s with or without tone found that those with a tonal L1 did not necessarily have an advantage in learning the Swedish pitch accents, but that it depended on what type of tonal language the L1 was (Tronnier and Zetterholm 2013).

Most relevant to the current study, research on L1 speakers of a tonal language who learn a non-tonal L2 has found that the L1 tonal patterns can transfer to the L2. For example, Cantonese English has been described as having tonal patterns (Gussenhoven 2012; Yiu 2014); similarly, Hong Kong English has been described as having high, mid, low, and falling tones (Wee 2016). French spoken by L1 Cantonese speakers has been analysed as having the Cantonese high tone on content words and the Cantonese low tone on function words (Lee and Matthews 2014). Japanese has a lexical pitch accent, and research on L1 Japanese speakers learning Spanish as an L2 found that they produce Japanese pitch accent patterns where none would occur in Spanish (Flores 2016). As such, lexical pitch patterns in

the L1 can be transferred to an L2 even if it does not use pitch lexically. However, beyond the studies just cited, little is known about the extent to which bilingual speakers transfer L1 tonal patterns to a non-tonal L2.

Acoustic measures related to intonation include f_0 level, f_0 range, and pitch dynamism quotient (PDQ). F_0 level is a measure of how high- or low-pitched a speaker's intonation sounds, while f_0 range measures the difference between a speaker's maximum and minimum f_0 . F_0 range is often measured as the middle 80% of a speaker's range (Busà and Urbani 2011; Meer and Fuchs 2021). PDQ is a measure of the overall variability of f_0 , meaning how much f_0 movement is produced, and this is measured as the standard deviation divided by its mean in Hertz (e.g., Hincks 2004; Meer and Fuchs 2021). These measures are relevant to the current study because they can provide a description of how much pitch movement a speaker is producing, which is relevant to rise–fall pitch accents.

1.1.2. Voicing Contrasts

The acquisition of voicing contrasts that differ between the L1 and L2 have been examined mainly in terms of stops (e.g., Bundgaard-Nielsen and Baker 2015; Flege and Eefting 1987). Mandarin has a stop contrast between voiceless unaspirated and voiceless aspirated stops, similar to English. In contrast, Russian has a contrast between voiceless unaspirated stops and voiced stops. Yang et al. (2022) examined the perception of Russian stops by L1 Mandarin speakers, and found that Russian voiced stops and voiceless unaspirated stops were perceived as being similar to Mandarin voiceless unaspirated stops, indicating that “when L2 sounds are perceptually similar to certain L1 sounds, their acquisition can be difficult even with an increase of L2 experience” (p. 20). If the voicing contrast exists at a different place of articulation, it may be possible for L2 learners to adapt their perception and production to a new place of articulation, as found by Flege and Port (1981) for L1 Arabic learners of English.

In English, the /s-z/ contrast also occurs as the allomorphic variants of the plural marker, with voicing assimilation occurring to that of the preceding obstruent; for example, in *cats* vs. *dogs*—in the former, the plural marker is [s] but in the latter it is produced as [z]. Recent research examined the acquisition of English sibilant voicing among speakers whose L1 varies in the use of /z/ (Contreras-Roa et al. 2020). French has /s/ and /z/ word-finally, Italian has word-final /s/ (although rarely) but not /z/, but has both word-medially, and Spanish does not have phonemic /z/ word-finally but has [z] allophonically (but non-obligatorily) in this position. Based on these differences, it was hypothesised that L1 French speakers would be able to produce English word-final /z/, L1 Italian speakers would be able to produce it by analogy with the word-medial contrast in their L1, and L1 Spanish speakers would have difficulty with it. The results supported the hypothesis regarding L1 French speakers, but for their measure of periodicity, L1 Spanish outperformed L1 Italian speakers in the production of English word-final morphemic /z/. The authors also note that L1 Italian speakers were able to produce English word-final /z/ when non-morphemic (that is, part of the stem, in a word such as *buzz*). They suggest that morphemic and non-morphemic /z/ in English are not treated the same by L2 learners.

1.2. Norwegian vs. English

Norwegian has a lexical pitch accent system (also called “tonal accent” (Kristoffersen 2000) or “word accent” (Bruce 1977)), whereby words carry either Accent I or Accent II. The phonetic implementation of the accent contrast varies by dialect, either in tonal makeup or tonal timing, but the pitch accent is generally a fall or rise–fall pattern (e.g., Almberg 2004; Fintoft 1970; Gårding 1973; Gussenhoven 2004). Figure 1 shows the accent contrast in disyllabic words in West Norwegian, which is relevant to the current study. In this figure, the thin vertical lines represent the beginning of the second syllable, showing that the high tone occurs earlier in Accent I than in Accent II.

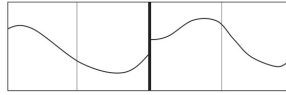


Figure 1. The West Norwegian lexical pitch accent contours (Accent I left, Accent II right), based on Gårding (1977).

In contrast, English uses (intonational) pitch accents for pragmatic reasons, such as on focused words (e.g., Pierrehumbert 1980). As such, L1 speakers of Norwegian learning English as an L2 are required to reduce the number of pitch accents in a sentence in order to approximate the intonational pattern of English. That is, they may be inclined to overuse pitch accents in English until they learn to suppress this. As with segmental features of L2 learning (e.g., Flege 1995; Flege et al. 2003; Heselwood and McChrystal 1999), it is likely that as the learner becomes more proficient in the L2, the intonation patterns become more L2-like. Mennen (2015) developed the L2 Intonation Learning theory (LILt) specifically about L2 learning of intonation patterns, and this may be more appropriate to describe what an L1 Norwegian speaker might do in the production of English intonation. Particularly, LILt includes the frequency dimension, which refers to the frequency with which a particular intonational element is used. This is relevant to the current study because Norwegian uses pitch accents more frequently than English. Previous studies have found that L2 learners tend to use pitch accent patterns from their L1 even when this is not the same pattern as the L2 (e.g., Jilka 2000).

It is not just the intonation patterns that differ between these two languages. One feature of Norwegian- or Swedish-accented English is the lack of voicing in the English /z/ sound (Hincks 2003). This pattern occurs due to transfer from the L1, since Norwegian and Swedish do not have /z/ in their phonological inventories and also do not have it allophonically (e.g., Engstrand 2004; Kristoffersen 2000). In fact, Norwegian has no voiced fricatives at any place of articulation. As such, it may be difficult for L1 Norwegian speakers to perceive voicing contrasts in fricatives. In the PAM model (Best 1994), this would mean that the English /z/ could be assimilated to the Norwegian /s/ phoneme category, in SC assimilation, or else CG may take place, where English /s/ may be a good example of the Norwegian /s/, but English /z/ may be a poor example of the same phoneme. In the SLM model (Flege 1995), the process of equivalence classification may take place, whereby due to acoustic similarity, the English /z/ may be categorised as a phonetic realisation of the Norwegian /s/. Similar to Norwegian, Danish has /s/ but not /z/, and research on L1 Danish speakers' perception of the /s/-/z/ contrast in English found that they had difficulty distinguishing them, and perceived /z/ as similar to /s/ (Bohn and Ellegaard 2019). This difficulty in perception may naturally correlate with a difficulty in production of the contrast.

1.3. Current Study

The current investigation is a longitudinal study of the L2 English of one L1 Norwegian speaker, Ole Gunnar Solskjær. Using interviews from two time periods, 1996–1998 and 2021, I examined one suprasegmental pattern and one segmental pattern of his English. The recordings were taken from YouTube in July 2021. In a similar longitudinal study, de Leeuw (2019) examined some German segments and average pitch of L1 German speaker Steffi Graf. Examining one speaker does not require controlling for interspeaker differences, and comparing an earlier timeframe with a later timeframe where the speaker has more exposure and practice with the L2 can provide insight into how speech patterns change over time. Using interviews available online allows for an examination of spontaneous speech. Solskjær was chosen as the subject of the current study because he moved to England in 1996 (aged 23) and has mostly lived in the UK (Manchester) ever since. He is from Kristiansund on the west coast of Norway, where a West Norwegian dialect is spoken. While he spoke English before moving to the UK, his English was audibly more

Norwegian-accented in the earlier timer period, and his English has been described in a number of YouTube comments as having features of the Mancunian accent. For example, “is it just me or does it sound like hes got abit of a manchester sort of accent he just pronounces his words in such away?” and “He sounds very Mancunian” (YouTube Channel 2011). The goal of the study was to examine how these two features (one suprasegmental, one segmental) changed over time as the speaker gained more experience with the L2.

2. Study 1: Intonation

2.1. Methods

2.1.1. Speaker

Ole Gunnar Solskjær is from Kristiansund, on the west coast of Norway. He moved to Manchester in 1996, aged 23, to play for Manchester United. He remained in the region ever since, except for a brief period where he managed a Norwegian team in 2011–2013.

2.1.2. Recordings

In total, ten English-language interviews were examined, five from each time period (1996–1998 and 2021). These were obtained from YouTube. The recordings were a combination of press conferences and interviews.

2.1.3. Labelling and Measurements

The interviews were divided into tokens that consisted of prosodic words, usually a 1–4 syllable span, totalling 1694 tokens. Each token was coded (by the author) for whether it exhibited a rise–fall pitch accent, based on auditory and visual examination of the spectrogram and pitch track. One of the clear acoustic correlates of a pitch accent is a rise–fall pattern, which entails a wider f_0 excursion than on words without a pitch accent, as shown in Figure 2. These tokens were also measured for two acoustic correlates of intonation patterns mentioned in Section 1.1.1: f_0 level (measured as the f_0 median per token in semitones) and f_0 range (measured as the middle 80% of the speaker’s range per token in semitones). Pitch dynamism quotient (PDQ) (the overall variability of f_0) was also measured, but not on the tokens as coded for the previous measures. Instead, the same interviews were broken up into longer phrases, usually constituting a sentence or phrase each, with pauses removed. This resulted in 271 tokens for PDQ.

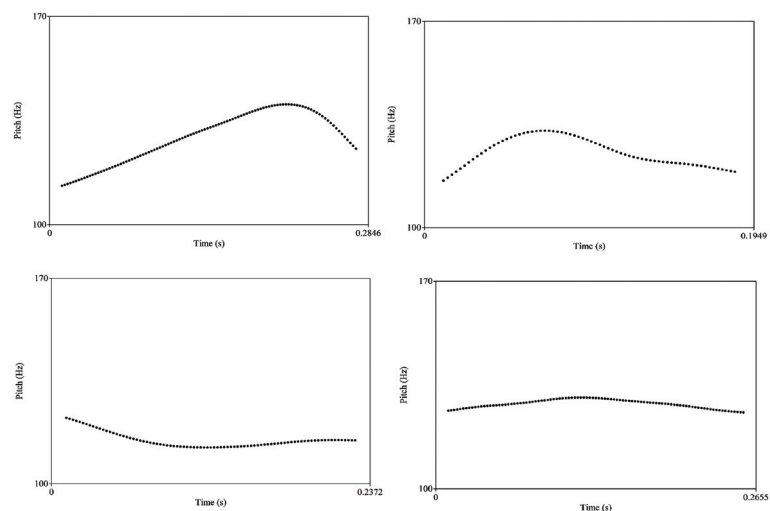


Figure 2. (Top): tokens labelled as pitch accents; (Bottom): tokens labelled as no pitch accent.

2.1.4. Statistical Analysis

A logistic regression was run to compare the proportion of speech that contained pitch accents in the Early vs. Late timeframes, while linear regressions were run on the acoustic measures. Pitch Accent was coded as Y (pitch accent present) or N (no pitch accent) and Timeframe was coded as Early or Late. All models had the random intercept of Interview. All tests were run using the *lmerTest* package in R (R Development Core Team 2008).

2.1.5. Hypotheses

It was hypothesised that there would be a higher proportion of pitch accents in the Early than Late timeframe, and relatedly, that there would be a higher pitch level, wider pitch range, and higher PDQ in the Early timeframe, due to a stronger influence of the L1 intonation pattern.

2.2. Results

The results for each measure will be discussed in turn.

2.2.1. Pitch Accents

The logistic regression comparing the proportion of speech that contained pitch accents was not found to differ between the two timeframes, with both having 17–18% of tokens carrying a pitch accent (Table 1, Figure 3).

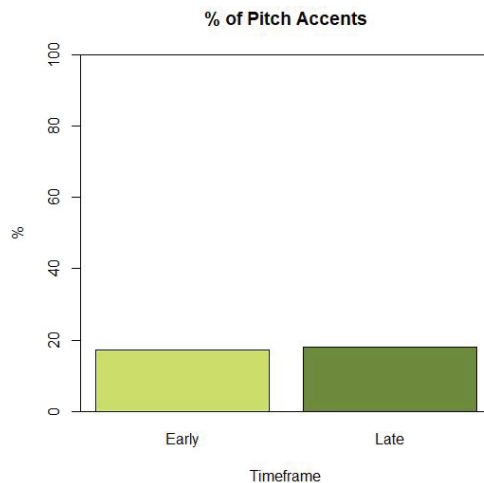


Figure 3. Proportion of tokens with a pitch accent in each timeframe.

Table 1. Statistical results for the logistic regression on the proportion of tokens with pitch accents. In all tables, * means the result is significant.

Measure	Coef.	SE	z	p
Intercept	−1.43	0.178	−8.1	<0.001 *
Timeframe.Late	−0.1	0.24	−0.41	0.681

2.2.2. F0 Level

In order to find the best model to predict the f0 level data, linear regression models were built up term by term and compared using the *anova* function in R (R Development Core Team 2008). In this approach, the goal is to find the model that best explains the data, meaning the fixed factors that most accurately predict the findings. The best model for f0 level was one with both Timeframe and Pitch Accent, with no interaction. There

was no significant effect of timeframe on f_0 level (Figure 4), but tokens containing a pitch accent had a significantly higher f_0 level (Table 2). These results mean that Early vs. Late Timeframe was not a significant predictor of the data, but the presence vs. absence of a pitch accent was, with the finding that when there was a pitch accent, the speaker's f_0 level was higher. Figure 5 shows this measure broken down by interview.

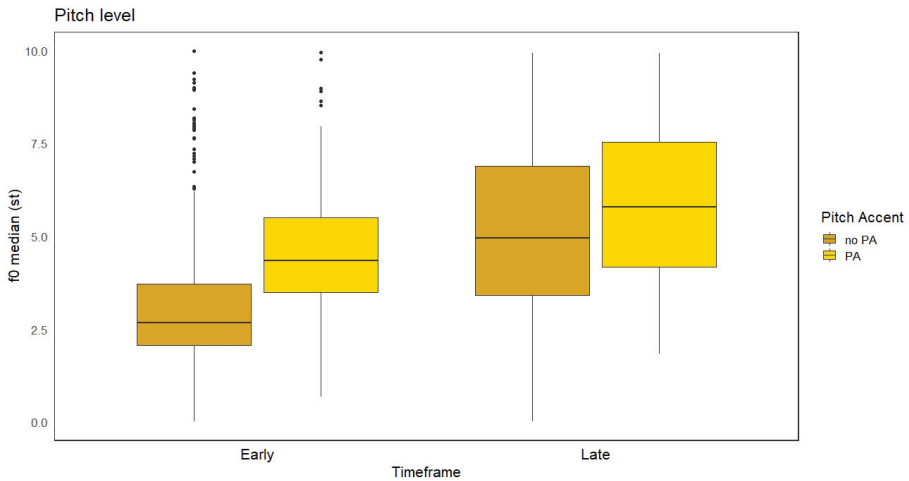


Figure 4. f_0 level for Early vs. Late timeframes and presence vs. absence of a Pitch Accent.

Table 2. Statistical results for the linear regression on f_0 level.

Measure	Coef.	SE	t	p
Intercept	3.87	0.86	4.5	<0.01 *
Timeframe.Late	1.02	1.3	0.81	0.44
PitchAcc.Y	1.36	0.11	12.7	<0.001 *

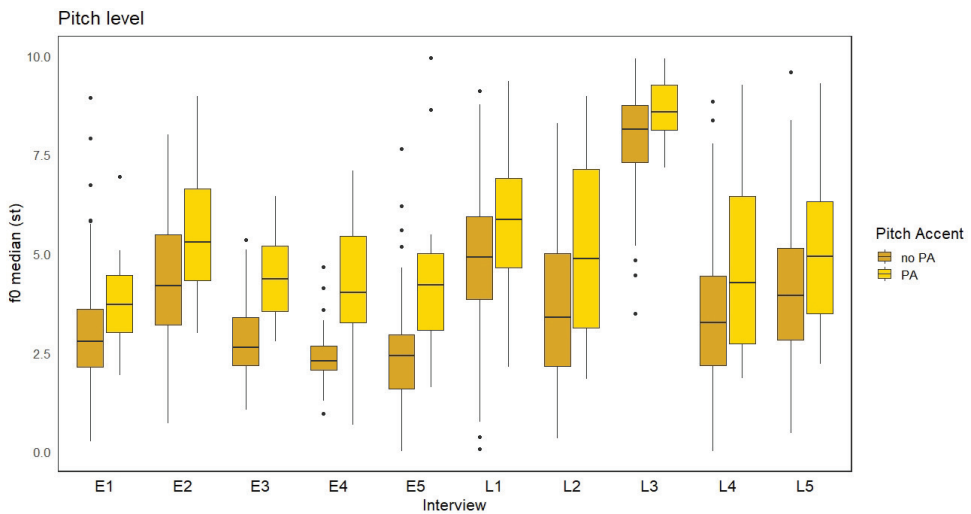


Figure 5. f_0 level by Interview (E = Early, L = Late) and presence vs. absence of a Pitch Accent.

2.2.3. F0 Range

The best model (the one that best explained the data) for f0 range was one with both Timeframe and PitchAccent and an interaction (Table 3). Since there was an interaction, a post-hoc pairwise test was conducted using the *emmeans* package in R. The results showed a significantly wider f0 range for the Late timeframe, the opposite of what was expected, as well as a wider f0 range for pitch-accented tokens (Figure 6). Figure 7 shows this measure broken down by interview.

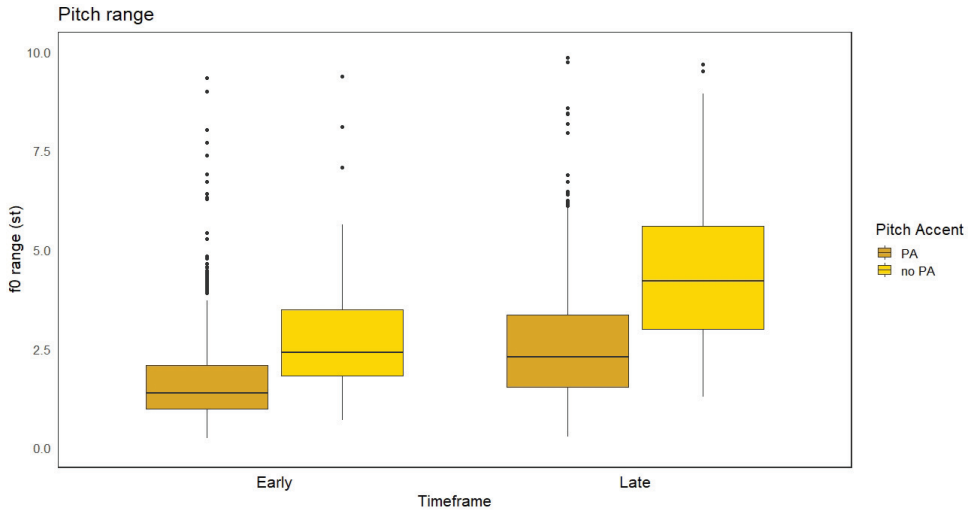


Figure 6. F0 range for Early vs. Late timeframes and presence vs. absence of a Pitch Accent.

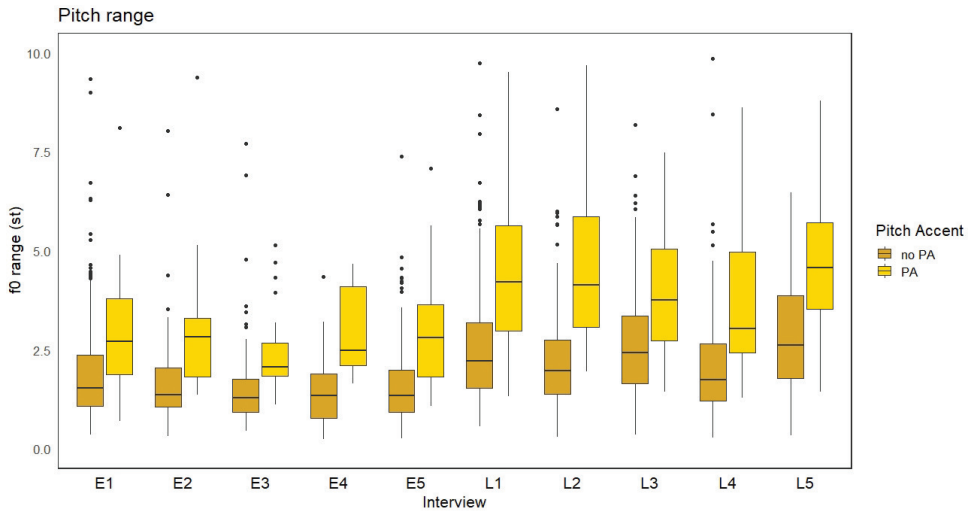


Figure 7. F0 range by Interview (E = Early, L = Late) and presence vs. absence of a Pitch Accent.

Table 3. Statistical results for the linear regression on f0 range. PA = Pitch Accent.

Measure	Coef.	SE	t	p
Intercept	1.89	0.1	18.9	<0.001 *
Timeframe.Late	0.75	0.13	5.7	<0.001 *
PA.Y	0.99	0.18	5.4	<0.001 *
Timeframe:PA	1.01	0.23	4.4	<0.001 *
Pairwise tests				
(Early, PA.N—Late, PA.N)	−0.75	0.14	−5.6	0.0013 *
(Early, PA.N—Early, PA.Y)	−0.99	0.19	−5.37	<0.001 *
(Late, PA.N—Late, PA.Y)	−2	0.14	−14.1	<0.001 *
(Early, PA.Y—Late, PA.Y)	−1.77	0.23	−7.74	<0.001 *

2.2.4. Pitch Dynamism Quotient

Since PDQ was measured over longer spans than those in the previous measures (which were not coded for Pitch Accent), only Timeframe was examined as a fixed factor. It was found to have a significant effect on the PDQ but in the opposite direction of what was hypothesised, that is, the speaker had a higher PDQ in the Late timeframe than in the Early timeframe (Table 4, Figure 8). Figure 9 shows this measure broken down by Interview.

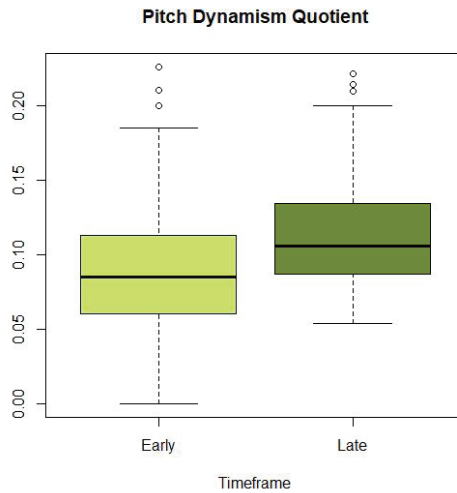


Figure 8. Pitch Dynamism Quotient for Early vs. Late timeframes. I used yellows when comparing Pitch Accent vs No Pitch Accent, and greens when comparing Early vs Late timeframes, but this is explained in every chart either by the legend or the labels on the axes.

Table 4. Statistical results for the linear regression on PDQ.

Measure	Coef.	SE	t	p
Intercept	0.091	0.006	16.22	<0.001 *
Timeframe.Late	0.024	0.008	3.1	0.02 *

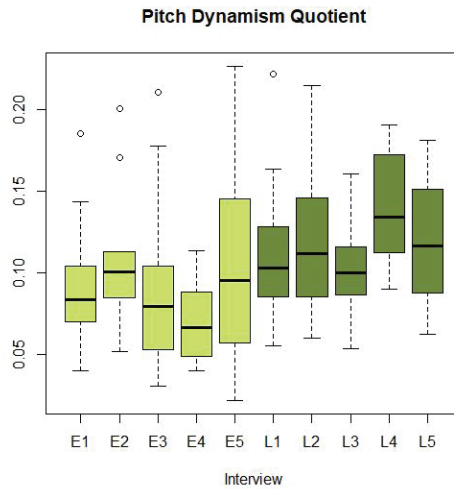


Figure 9. Pitch Dynamism Quotient by Interview (E = Early, L = Late).

2.3. Discussion

These findings indicate that, contrary to the hypothesis, the speaker did not seem to be transferring the L1 pitch accent system to the L2 even in the Early timeframe, at least in terms of the number of pitch accents, which did not differ between the two timeframes. The f_0 level was also (contrary to the hypothesis) not found to differ between the two timeframes, but as expected, it was higher for pitch accented-tokens in both timeframes. Additionally, contrary to the hypothesis, his f_0 range was wider and the PDQ was higher in the Late timeframe, which suggests more pitch movement in this timeframe. This may indicate that the speaker produced a more compressed f_0 range and movements when he was less fluent in the L2 (in the Early timeframe), and since becoming more comfortable speaking it, he has more dynamic f_0 patterns. This may suggest a type of hypercorrection in the earlier stages of learning the L2, resulting from an overreaction to the L1 influence (Eckman et al. 2013; Janda and Auger 1992; Odlin 1989). When a small number of tokens of the L1 were examined, the speaker's f_0 range and PDQ were higher than the Late L2 results (as expected for a pitch accent language), corroborating the idea that he was compressing his f_0 range and movements in the Early timeframe.

Figures 5 and 7 show that pitch-accented tokens have a consistently higher f_0 level and wider range than non-pitch accented tokens. While there is generally consistency in these patterns across interviews, it is possible that temporary factors such as emotional state could affect his intonation patterns. This may explain the higher average level in L3 (Figure 5).

These findings will be discussed further in Section 4.

3. Study 2: Segments

For this experiment, the speaker and recordings were the same as in Study 1.

3.1. Methods

3.1.1. Labelling and Measurements

For the segmental part, all words in which English has a /s/ or /z/ were coded for whether they were produced as voiced or voiceless (based on auditory analysis, similar to Dehé and Wochner 2022) for a total of 673 tokens. The English underlying /z/ tokens were also coded for their position in the word (medial or final) as well as their morphemic status,

that is, whether they were morphemic (for example, the plural marker) or part of a stem (e.g., the /z/ in *because*), based on results from Contreras-Roa et al. (2020).

Since the recordings came from interviews which were not of the highest sound quality, the types of acoustic measurements that could be made were limited. For this reason, duration was chosen as a useful variable, because it was easily measured and duration is also a cue to fricative voicing, with voiceless fricatives being longer than voiced ones (Contreras-Roa et al. 2020; Crystal and House 1988; Jongman et al. 2000).

3.1.2. Statistical Analysis

Two logistic regression tests were run on the auditory categorisation of /s/ and /z/ as voiced or voiceless.¹ The /z/ segments in the Late timeframe were also examined to determine whether morphemic status or word position had a significant effect on whether they were voiced.

A linear regression was run on the duration of /z/ phonemes, with possible independent variables of voicing and word position.

3.1.3. Hypotheses

It was hypothesised that the proportion of /s/ produced as voiced would not differ between timeframes but that /z/ would have a higher proportion produced as voiced in the Late timeframe. This is based on the fact that Norwegian does not have voiced fricatives, so it is likely difficult for the speaker to acquire the voiced /z/. Based on the literature previously cited, it is possible that over time and exposure to English, he has started to learn this pattern. In terms of duration, it was expected that voiceless fricatives would be longer than voiced ones, and that those in final position would be longer than those in medial position.

3.2. Results

3.2.1. Patterns of Voicing

The results showed a significant effect of timeframe only for the phoneme /z/, with more voiced productions of /z/ in the Late timeframe (Table 5). In the Early timeframe, 93% of /z/ were voiceless and in the Late one, 46% of /z/ were voiceless (Figure 10); in comparison, in the Early timeframe, 100% of /s/ were voiceless and in the Late one, 98.5% were voiceless. Both of these findings (the effect of Timeframe and the difference between the two phonemes) are in line with the hypotheses. Including morphemic status and word position did not improve the model, meaning that these factors do not significantly predict whether the segments were produced with voicing.

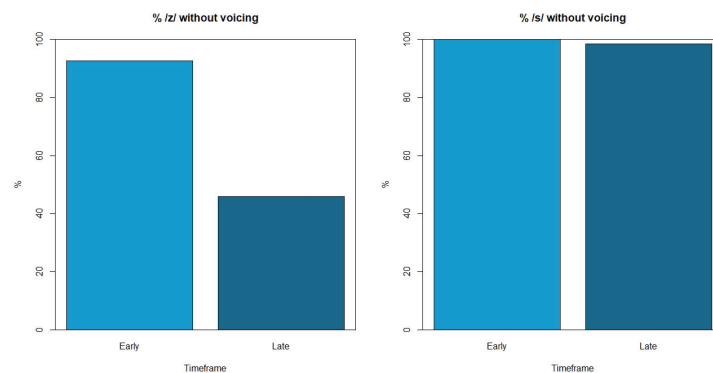


Figure 10. Proportion of /z/ (left) and /s/ (right) segments that were voiceless in Early vs. Late timeframes.

Table 5. Statistical results for the logistic regression on voicing.

Measure	Coef.	SE	z	p
Phoneme: /s/				
Intercept	−37.21	185.42	−0.2	0.841
Timeframe.Late	33.1	185.42	0.18	0.859
Phoneme: /z/				
Intercept	−2.6	0.51	−5.1	<0.001 *
Timeframe.Late	2.85	0.6	473	<0.001 *

As shown in Figure 11, there was no difference in voicing patterns for English /z/ based on whether it was a separate morpheme or part of a stem. Figure 12 shows that 61% of word-final /z/ were voiceless and 71% of word-medial /z/ were voiceless. That is, /z/ was more commonly voiced in final position than in medial position, although this effect was not found to be significant.

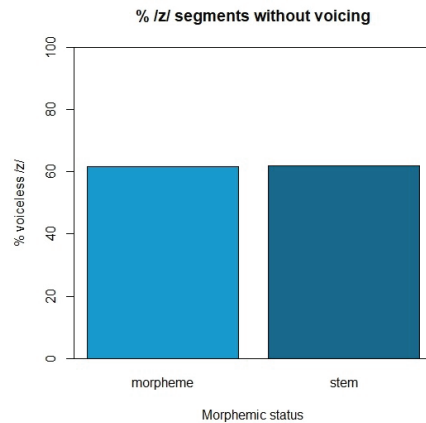


Figure 11. Proportion of voiceless /z/ in the Late timeframe, based on whether it was a morpheme or part of a stem.

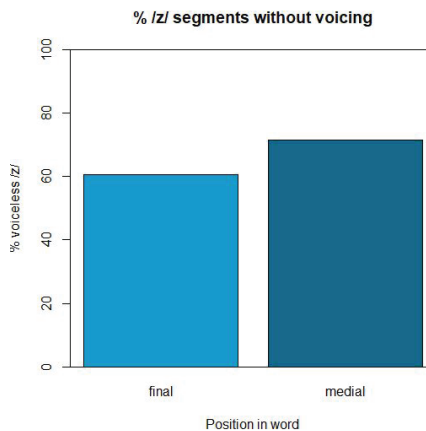


Figure 12. Proportion of voiceless /z/ in the Late timeframe, based on word position.

3.2.2. Duration

Figure 13 shows the duration of voiced and voiceless segments in word-medial and final position.

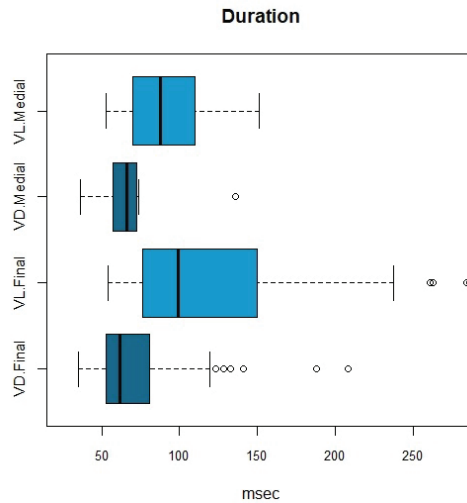


Figure 13. Duration of /z/ produced as voiced or voiceless, by word position.

The best model for the linear regression on duration included both voicing and position as independent variables, and showed a main effect of both. This means that, as shown in Table 6, duration was significantly longer in voiceless segments and in final position.

Table 6. Statistical results for the linear regression on duration.

Measure	Coef.	SE	t	p
Intercept	81.23	11.43	7.1	<0.05 *
Voic.Voiceless	37.66	6.88	5.48	<0.001 *
Position.Medial	−20.98	9.4	−2.32	<0.05 *

3.3. Discussion

These results suggest that more exposure to and practice with the L2 has led to an increase in L2-like voicing productions. This means the speaker is acquiring a voicing contrast that is not in the L1, but has not yet acquired it completely, since only just over half of /z/ productions were voiced in the Late timeframe. The results for morphemic status are different from those of Contreras-Roa et al. (2020), because the current study found the same pattern for both types of /z/. These results are discussed in terms of SLM and PAM in Section 4.

The duration results were expected in that the voiceless segments were longer than the voiced ones, similar to previous work (Crystal and House 1988; Jongman et al. 2000). (These results also add support to the auditory categorisation of the sounds as voiced or voiceless.) Additionally, as predicted, segments were longer in word-final than word-medial position.

4. General Discussion

For the intonation experiment, the results were contrary to the hypotheses, because it was found that the speaker did not appear to be transferring the lexical pitch accent system of Norwegian to his English even in the Early timeframe. He did not over-apply the L1 pitch accent pattern, contrary to what has been found in previous work on L2

intonation (Jilka 2000; Mennen 2015). Further, the wider f0 range and higher PDQ in the Late timeframe seem to suggest that he may have been compressing his pitch range and movements in the Early timeframe. It is also useful to note that since the proportion of speech that was counted as a pitch accent did not differ between the two timeframes, the difference found in f0 range and PDQ are not related to how many pitch accents occurred. This suggested the speaker is doing something qualitatively different between the two timeframes. If he has become more comfortable speaking the L2 over time, he may be allowing himself more pitch variation in the Late timeframe; that is, that lower fluency in the Early timeframe is connected to his compression, perhaps similar to previous work that found that L2 learners of English showed a narrower pitch range (Mennen 2008). Related to this, as noted in Section 2.3, is the idea that he was hypercorrecting (Eckman et al. 2013; Odlin 1989) in the Early timeframe, knowing that English has less pitch movement than Norwegian. Norwegian has substantial dialect variation in the lexical pitch accent system (Kristoffersen 2000), and Norwegian speakers are familiar with the different patterns. This may suggest that an L1 Norwegian speaker has an awareness of intonation patterns even in an L2. Previous work on hypercorrection in L2 learning has found that L1 French speakers learning English showed both h-deletion and h-insertion, although the latter at lower rates (Janda and Auger 1992). In the current study, it is possible that this awareness of the different patterns between Norwegian and English made the speaker hesitant in the Early timeframe, and he therefore hypercorrected by compressing his pitch range and movements. With increased comfort in the L2, he no longer does this. If he is aware that Norwegian and English have very different intonation patterns, the results fit in with the SLM (Flege 1995; Flege and Bohn 2021) prediction that patterns that differ substantially between languages are easier for learners to acquire.

For the segmental pattern, the results align with the SLM (Flege 1995; Flege and Bohn 2021) and PAM (Best 1994) insofar as the speaker did not voice the /z/ in the Early timeframe; that is, he used the closest L1 sound /s/ instead. The SLM (Flege 1995) of L2 learning would explain these results as the speaker categorising the English /z/ initially as /s/, which is in the L1—an example of equivalence classification. Over time, with more exposure to English, he has created a separate category for /z/ and has begun to distinguish the two phonemes. Particularly, though, the findings here may be best explained through the PAM CG analysis, where both /s/ and /z/ were originally considered to be the same /s/ phoneme, but /z/ was a less good example of it, so over time, the speaker has begun to learn that it is a separate phoneme in English. Future work could compare this speaker's productions to L1 English speakers' voicing patterns, because it has been reported that especially in word-final position, /z/ is not always fully voiced (Ogden 2009). However, in the current study, even in medial position, the speaker often does not produce English /z/ with voicing, indicating that he is not L1-like in this pattern. The results for morphemic status are different from those found by Contreras-Roa et al. (2020), since the morphemic status of /z/ was not found to have any effect on whether it was produced as voiced or voiceless. It may simply be that when no fricative voicing contrasts are in the L1, the first challenge is to perceive and produce the voiced fricative, and morphemic status is not (yet) relevant. However, the speaker is not voicing haphazardly, since he was not found to voice the phoneme /s/. In terms of duration, the results showed that the English /z/ phoneme was longer when in final position and when produced as voiceless.

It is also interesting to note that for the intonation pattern, the speaker's task is to suppress a feature from the L1, while for the segmental pattern, his task is to acquire a new contrast that does not exist in the L1. Perhaps it is easier to suppress a pattern than acquire a new one? Or perhaps prosodic systems that differ between languages are particularly salient to language learners.

This work provides insight into changes in the L2 over time by directly comparing changes in a segmental pattern with a suprasegmental pattern, in the same speaker. These findings contribute to descriptions of hypercorrection in L2 learning, specifically in the context of intonational features.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable since all recordings were publicly available.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

Note

- ¹ This analysis was used because including both fixed factors (Timeframe and Phoneme) as well as an interaction caused a scaling error, and an ANOVA could not be run because it is unsuitable for categorical dependent measures, so instead the two phonemes were examined separately.

References

- Almberg, Jørn. 2004. Tonal differences between four Norwegian dialect regions—Some acoustic findings. In *Nordic Prosody IX*. Edited by Gösta Bruce and Merle Horne. Lund: Peter Lang, pp. 19–28.
- Backman, Nancy Ellen. 1979. Intonation errors in second language pronunciation of eight Spanish speaking adults learning English. *Interlanguage Studies Bulletin* 4: 239–66.
- Beckman, Mary E. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.
- Best, Catherine T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge: MIT.
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins Publishing, pp. 13–34.
- Bohn, Ocke-Schwen, and Anne Aarhøj Ellegaard. 2019. Perceptual assimilation and graded discrimination as predictors of identification accuracy for learners differing in L2 experience: The case of Danish learners' perception of English initial fricatives. Paper presented at the 19th International Congress of Phonetic Sciences, Melbourne, Australia, August 5–9; pp. 2070–74.
- Bruce, Gösta. 1977. Swedish word accents in sentence perspective. *Travaux de L'Institut de Linguistique de Lund* 12.
- Bundgaard-Nielsen, Rikke Louise, and Brett Baker. 2015. Perception of voicing in the absence of native voicing experience. Paper presented at the INTERSPEECH, Sixteenth Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6–10; pp. 2352–56.
- Busà, Maria Grazia, and Martina Urbani. 2011. A cross linguistic analysis of pitch range in English L1 and L2. Paper presented at the 17th International Congress of Phonetic Sciences, Hong Kong, China, August 17–21; pp. 380–83.
- Contreras-Roa, Leonardo, Paolo Mairano, Marc Capliez, and Caroline Bouzon. 2020. Voice assimilation of morphemic -s in the L2 English of L1 French, L1 Italian and L1 Spanish learners. *Anglophonia—French Journal of English Linguistics* 30. [\[CrossRef\]](#)
- Crystal, Thomas H., and Arthur S. House. 1988. Segmental durations in connected speech signals: Current results. *Journal of the Acoustical Society of America* 83: 1553–73. [\[CrossRef\]](#)
- de Leeuw, Esther. 2019. Native speech plasticity in the German-English late bilingual Stefanie Graf: A longitudinal study over four decades. *Journal of Phonetics* 73: 24–39. [\[CrossRef\]](#)
- Dehé, Nicole, and Daniela Wochner. 2022. Voice quality and speaking rate in Icelandic rhetorical questions. *Nordic Journal of Linguistics* 1–10. [\[CrossRef\]](#)
- Eckman, Fred R., Gregory K. Iverson, and Jae Yung Song. 2013. The role of hypercorrection in the acquisition of L2 phonemic contrasts. *Second Language Research* 29: 257–83. [\[CrossRef\]](#)
- Engstrand, Olle. 2004. *Fonetikens Grunder*. Lund: Studenlitteratur.
- Fintoft, Knut. 1970. *Acoustical Analysis and Perception of Tonemes in Some Norwegian Dialects*. Oslo: Universitetsforlaget.
- Flege, James Emil. 1987. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65. [\[CrossRef\]](#)
- Flege, James Emil. 1995. Second language speech learning: Theory, findings and problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Baltimore: York Press, pp. 233–77.
- Flege, James Emil, and Ocke-Schwen Bohn. 2021. The revised speech learning model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Rtree Wayland. Cambridge: Cambridge University Press, pp. 3–83.
- Flege, James Emil, and Robert Port. 1981. Cross-language phonetic interference: Arabic to English. *Language and Speech* 24: 125–46. [\[CrossRef\]](#)
- Flege, James Emil, and Wieke Eefting. 1987. The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics* 15: 67–83. [\[CrossRef\]](#)
- Flege, James Emil, Carlo Schirru, and Ian R. A. MacKay. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40: 467–91. [\[CrossRef\]](#)

- Flores, Tanya. 2016. Declarative intonation in the Spanish of Japanese-Spanish bilinguals. *Proceedings of Meetings on Acoustics* 29: 060013. [CrossRef]
- Gårding, Eva. 1973. The Scandinavian word accents. In *Working Papers 8*. Lund: Phonetics Laboratory, Lund University.
- Gårding, Eva. 1977. *The Scandinavian Word Accents*. Lund: Gleerup.
- Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, Carlos. 2012. Tone and intonation in Cantonese English. Paper presented at the Third International Symposium on Tonal Aspects of Languages (TAL) 2012, Nanjing, China, May 26–29.
- Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Heselwood, Barry, and Louise McChrystal. 1999. The effect of age group and place of L1 acquisition on the realisation of Panjabi stop consonants in Bradford: An acoustic sociophonetic study. *Leeds Working Papers in Linguistics and Phonetics* 7: 49–68.
- Hincks, Rebecca. 2003. Pronouncing the Academic Word List: Features of L2 Student Presentations. Paper presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9; pp. 1545–48.
- Hincks, Rebecca. 2004. Processing the prosody of oral presentations. Paper presented at the Proceedings of InSTIL/ICALL, Venice, Italy, June 17–19.
- Hualde, José I. 2012. Two Basque accentual systems and the notion of pitch-accent language. *Lingua* 122: 1335–51. [CrossRef]
- Hyman, Larry. 2009. How (not) to do phonological typology: The case of pitch-accent. *Language Sciences* 31: 213–38. [CrossRef]
- Janda, Richard D., and Julie Auger. 1992. Quantitative evidence, qualitative hypercorrection, sociolinguistic variables—And French speakers' 'eadhaches with English h/Ø. *Language and Communication* 12: 195–236. [CrossRef]
- Jarvis, Scott, and Aneta Pavlenko. 2009. *Cross-Linguistic Influence in Language and Cognition*. London: Routledge.
- Jilka, Matthias. 2000. Testing the Contribution of Prosody to the Perception of Foreign Accent. In *Proceedings of New Sounds 2000*. Edited by Allan James and Jonathan Leather. Amsterdam: University of Amsterdam, pp. 199–207.
- Jongman, Allard, Ratree Wayland, and Serena Wong. 2000. Acoustic Characteristics of English Fricatives. *Journal of the Acoustical Society of America* 108: 1252–63. [CrossRef]
- Kristoffersen, Gjørt. 2000. *The Phonology of Norwegian*. Oxford: Oxford University Press.
- Lee, Jackson L., and Stephen Matthews. 2014. When French becomes tonal: Prosodic transfer from L1 Cantonese and L2 English. Paper presented at the 6th Annual Pronunciation in Second Language Learning and Teaching Conference, Santa Barbara, CA, USA, September 5–6.
- Lein, Tatjana, Tanja Kupisch, and Joost van de Weijer. 2016. Voice onset time and global foreign accent in German-French simultaneous bilinguals during adulthood. *International Journal of Bilingualism* 20: 732–49. [CrossRef]
- Lisker, Leigh, and Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20: 384–422. [CrossRef]
- Maddieson, Ian. 2011. Tone. In *The World Atlas of Language Structures Online*. Edited by Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library.
- Major, Roy C. 1992. Losing English as a first language. *The Modern Language Journal* 76: 190–208. [CrossRef]
- Mayr, Robert, and Aysha Siddika. 2018. Inter-generational transmission in a minority language setting: Stop consonant production by Bangladeshi heritage children and adults. *International Journal of Bilingualism* 22: 255–84. [CrossRef]
- McCarthy, Kathleen M., Bronwen G. Evans, and Merle Mahon. 2013. Acquiring a second language in an immigrant community: The production of Sylheti and English stops and vowels by London-Bengali speakers. *Journal of Phonetics* 41: 344–58. [CrossRef]
- Meer, Philipp, and Robert Fuchs. 2021. The Trini Sing-Song: Sociophonetic variation in Trinidadian English prosody and differences to other varieties. *Language and Speech*. [CrossRef] [PubMed]
- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63. [CrossRef]
- Mennen, Ineke. 2008. Phonological and phonetic influences in non-native intonation. In *Non-Native Prosody: Phonetic Description and Teaching Practice*. Edited by Juergen Trouvain and Ulrike Gut. Berlin and New York: De Gruyter Mouton, pp. 53–76.
- Mennen, Ineke. 2015. Beyond segments: Towards a L2 intonation learning theory. In *Prosody and Language in Contact*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin and Heidelberg: Springer, pp. 171–88.
- Mennen, Ineke, and Esther de Leeuw. 2014. Beyond Segments: Prosody in SLA. *Studies in Second Language Acquisition* 36: 183–94. [CrossRef]
- Nagy, Naomi. 2015. A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua* 164: 309–27. [CrossRef]
- Odling, Terence. 1989. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Ogden, Richard. 2009. *An Introduction to English Phonetics*. Edinburgh: Edinburgh University Press.
- Ordin, Mikhail, and Ineke Mennen. 2017. Cross-Linguistic Differences in Bilinguals' Fundamental Frequency Ranges. *Journal of Speech Language and Hearing Research* 60: 1493–506. [CrossRef]
- Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Ph.D. thesis, MIT, Cambridge, MA, USA.
- R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team, ISBN 3-900051-07-0.
- Treffers-Daller, Jeanine, and Raymond Mougeon. 2005. The role of transfer in language variation and change: Evidence from contact varieties of French. *Bilingualism: Language and Cognition* 8: 93–98. [CrossRef]
- Tronnier, Mechtild, and Elisabeth Zetterholm. 2013. Tendencies of Swedish word accent production by L2-learners with tonal and non-tonal L1. In *Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012*. Edited by Eva Liina Asu and Lippus Pärtel. Pieterlen and Bern: Peter Lang Publishing Group, pp. 391–400.

- Tung, Yi-Chen. 2006. The language interference of pitch accent language on tone language—A case study of Mandarin and Swedish. *The Journal of the Acoustical Society of America* 119: 3392–92. [[CrossRef](#)]
- Wee, Lian-Hee. 2016. Tone assignment in Hong Kong English. *English* 92: e67–e87. [[CrossRef](#)]
- Willems, Nico. 1982. *English Intonation from a Dutch Point of View*. Dordrecht: Foris.
- Yang, Yuxiao, Xiaoxiang Chen, and Qi Xiao. 2022. Cross-linguistic similarity in L2 speech learning: Evidence from the acquisition of Russian stop contrasts by Mandarin speakers. *Second Language Research* 38: 3–29. [[CrossRef](#)]
- Yiu, Suki S. Y. 2014. Tone spans of Cantonese English. Paper presented at the 4th International Symposium on Tonal Aspects of Languages (TAL) 2014, Nijmegen, The Netherlands, May 13–16.
- YouTube Channel. 2011. *Ole Gunnar Solskjaer FA Inquiry*. Fitchburg: Fitchburg Access Television.

Article

Spanish–English Cross-Linguistic Influence on Heritage Bilinguals’ Production of Uptalk

Ji Young Kim

Department of Spanish and Portuguese, University of California, Los Angeles, CA 90095, USA; jyoungkim@ucla.edu

Abstract: The present study examines the production of uptalk in Spanish and in English by Spanish heritage speakers in Southern California. Following the L2 Intonation Learning Theory, we propose that cross-linguistic influence in heritage bilinguals’ uptalk may occur along multiple dimensions of intonation. In this study, we examined the systemic dimension (i.e., presence of uptalk and presence of uptalk with IP-final deaccenting), the frequency dimension (i.e., frequency of uptalk and frequency of uptalk with IP-final deaccenting), and the realizational dimension (i.e., pitch excursion and rise duration) of heritage bilinguals’ uptalk. Our data showed that the three dimensions of intonation demonstrate varying degrees of cross-linguistic influence. The heritage bilinguals produced uptalk with IP-final deaccenting in both languages (i.e., systemic dimension), but produced it more in English than in Spanish (i.e., frequency dimension). That is, IP-final deaccenting emerges in heritage bilinguals’ uptalk in Spanish, but heritage bilinguals seem to recognize that this is an English feature that is not allowed in Spanish and try to suppress it as much as possible when producing uptalk in Spanish. However, in the realizational dimension, the heritage bilinguals demonstrated either phonetic assimilation to English (i.e., pitch excursion) or individual variability conditioned by language learning experience (i.e., rise duration). The asymmetry found across the dimensions suggests that, when bilinguals’ two languages are in competition for finite online resources, such as in the case of spontaneous speech production, phonological distinctions between L1 and L2 prosodic structures are kept, whereas phonetic differences that do not lead to any change in meaning are more prone to undergo cross-linguistic influence in order to reduce online processing cost. This study attempts to fill a gap in the literature on the cross-linguistic influence of intonation by bringing attention to heritage bilinguals. Heritage bilingualism introduces bilingual contexts that are often left unnoticed in traditional L2 acquisition scenarios (e.g., transfer from L2 to L1 intonation, asymmetry between order of acquisition and language dominance). Given that many aspects of cross-linguistic influence are shared across bilinguals, the investigation of heritage bilinguals’ intonation will contribute to building robust models of bilingual intonation.

Citation: Kim, Ji Young. 2023. Spanish–English Cross-Linguistic Influence on Heritage Bilinguals’ Production of Uptalk. *Languages* 8: 22. <https://doi.org/10.3390/languages8010022>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 25 September 2022

Revised: 3 January 2023

Accepted: 4 January 2023

Published: 9 January 2023

Keywords: cross-linguistic influence; heritage speakers; heritage language intonation; uptalk; L2 Intonation Learning Theory

1. Introduction

Bilingual speakers are not two monolinguals in one person (Grosjean 1989). They sometimes exhibit speech sounds that differ from the monolingual norm in one or both of their languages and, if noticeable enough to listeners’ ears, these differences could mark their speech as sounding non-native like. Thus, identifying areas of convergence and divergence between bilinguals’ L1 and L2 speech sounds sheds light on issues regarding the mechanism of cross-linguistic phonetic and phonological influence and the relative difficulty or ease when acquiring L2 phonetics and phonology. Based on comparisons of L1 and L2 segments, current models in L2 speech learning, such as the (Revised) Speech Learning Model (Flege 1995; Flege and Bohn 2021) and the Perceptual Assimilation Model (-L2) (Best 1995; Best and Tyler 2007), posit that L1 and L2 sound categories exist in a



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

common phonological space, leading to bidirectional cross-linguistic influence that can surface in various forms; depending on the perceptual similarity between L1 and L2 sound categories, a category in one language may approach a similar-sounding category in the other language or drift away from that category to maintain phonetic contrast (Flege and Bohn 2021; Flege et al. 2003). In some cases, variability is found in the presence, form, and direction of influence under the same linguistic contexts, conditioned by multiple factors associated with bilinguals' language learning experience (e.g., age of acquisition of the target language, speech community size, language proficiency, language use, and language attitude).

While bilinguals may differ from monolinguals in their production and perception of both segments and prosody, the majority of research on bilingual phonetics and phonology has focused on segments, whereas there is comparatively little work on prosody (Mennen 2015; Queen 2006). Studies in L2 prosody have shown that L2 learners demonstrate non-target-like patterns in various prosodic features that are conjectured as L1 transfer. Examples of such include the prosodic marking of information structure (Gut and Pillai 2014; Kim 2019; Nagano-Madsen 2015a; Nguyen et al. 2008; O'Brien and Gut 2010; Ortega-Llebaria and Colantoni 2014; Rasier and Hiligsmann 2007; Saito 2006; Swerts and Zerbian 2010; Turco et al. 2015; Ueyama and Jun 1998), prosodic phrasing (Horgues 2013; Nagano-Madsen 2015b; Nibert 2006; Santiago-Vargas and Delais-Roussarie 2012), and the types and phonetic implementation of pitch accents (Grabe 2004; Jilka 2000; Kim 2020; Mennen 2004; Mennen et al. 2014; Nagano-Madsen 2015a; O'Brien and Gut 2010; Trofimovich and Baker 2006) and boundary tones (Jilka 2000; Mennen et al. 2010). As in the case of L2 segments, depending on various linguistic and extralinguistic factors, deviations can appear in multiple forms, such as substitution (Jilka 2000; Mennen et al. 2010; O'Brien and Gut 2010), hybridization (de Leeuw et al. 2012; Mennen et al. 2014; Queen 2001; Rao 2016), phonetic assimilation (Colantoni et al. 2016; Kim 2020; Zuban et al. 2020), and phonetic dissimilation (de Leeuw et al. 2012).

Built on the types of intonational variability identified by Ladd (1996), the L2 Intonation Learning Theory (LILt) (Mennen 2015) recognizes four dimensions along which deviation from native norms may occur in L2 intonation. The systemic dimension refers to the inventory of structural phonological elements (e.g., tonal sequences and tune–text association). The realizational dimension refers to the phonetic implementation of the structural phonological elements (e.g., pitch scale, slope, and tonal alignment), while the semantic dimension is concerned with how such elements are used to convey meaning (e.g., information structure and question vs. statement). Lastly, the frequency dimension involves the frequency of use of the structural phonological elements. Consistent with the models in the acquisition of L2 segments (Best 1995; Best and Tyler 2007; Flege 1995; Flege and Bohn 2021), the LILt posits that L1 and L2 intonation categories exist in a common phonological space, leading to a cross-linguistic influence along any of the above-mentioned dimensions (Mennen 2015). Given the complexity and multidimensionality of intonation, the LILt's method of viewing L2 intonation through a multilayered lens allows us to answer questions, such as whether different dimensions of intonation are equally susceptible to native language influence and whether certain dimensions develop at a faster pace than others with more experience in the L2 (Mennen 2015). For instance, Mennen et al. (2010) found that, after 30 months of living in the UK, Punjabi and Italian L2 learners of English produced fewer rising pitch contours than when they first arrived to the UK and predominantly used the falling pitch contour, which is the most prevalent contour in British English (Grabe 2004) (i.e., frequency dimension). However, they did not use any complex contours (e.g., rise–fall and fall–rise) observed in British English, showing no improvement in the inventory of the tonal sequences of the target language (i.e., systemic dimension). Jun and Oh (2000) examined various aspects of the surface tone production in Korean accentual phrases (APs) by English L2 learners of Korean. They found that the learners were in general successful in using the high (H) tone in AP-final position (i.e., systemic dimension), but they failed to demonstrate f_0 differences between AP-initial tones which are realized as

the H tone when the AP begins with an aspirated or tense obstruent and as the low (L) tone in other contexts (i.e., realizational dimension). While surface AP tones in Korean do not change the meaning of an utterance, phrase boundaries do (Jun 2000). Unlike the AP-initial tones which are segmentally triggered, the AP-final H tone is a strong perceptual cue that marks the right edge of an AP. Thus, the better success observed in learners' production of the AP-final H tone suggests that L2 learners of Korean acquire the phonological properties earlier than the phonetic properties of intonation (Jun and Oh 2000).

Most evidence of deviation in L2 intonation has been found in the realizational dimension (Mennen 2015), particularly in tonal alignment (Atterer and Ladd 2004; Chen and Fon 2008; Graham and Post 2018; Kim 2019, 2020; Nagano-Madsen 2015a; Mennen 2004), pitch range (Aoyama and Guion 2007; Huang and Jun 2011; Jilka 2000; Kim 2019; Mennen et al. 2014; Willems 1982), and pitch scaling (Henriksen et al. 2010; Kelm 1987; McGory 1997). However, it is unclear whether the realizational dimension is the most susceptible among the four dimensions of intonation to cross-linguistic influence. Studies have demonstrated both improvement and deviation within the same dimension (Chen and Fon 2008; Huang and Jun 2011), as well as interactions between dimensions (Jun and Oh 2000; Kim 2019; Mennen 1999; Nagano-Madsen 2015a). That is, cross-linguistic influence of intonation is a complex process and it is sometimes difficult to identify the dimension of influence. More empirical studies should be carried out on a variety of L1-L2 pairings, prosodic aspects, and bilingual situations to better understand the interaction between bilinguals' two intonation systems.

While the LILt centers around L2 intonation, this model can be applied to the intonation of any bilinguals, including heritage speakers. Heritage speakers are a type of bilinguals who grew up speaking a home language (i.e., the heritage language) that is different from the majority language of the society. Heritage languages are minority languages acquired naturalistically in a bilingual or multilingual environment, such as diasporic languages spoken by immigrants and their children, aboriginal or indigenous languages whose linguistic status has been jeopardized by colonizing languages, and historical minority languages that have coexisted with other standard languages (Montrul and Polinsky 2021; Rothman 2009). Unlike L2 intonation, heritage language intonation is relatively underexplored. Given the minority status of heritage languages, many heritage speakers grow up becoming more dominant in the societally dominant language (Benmamoun et al. 2013; Polinsky and Kagan 2007). Thus, heritage language research has focused primarily on the cross-linguistic influence from the dominant language to the heritage language. Several studies have shown that heritage bilinguals exhibit intonational patterns that are present in both of their languages (Bullock 2009; Colantoni et al. 2016; Kim 2019; Queen 2006; Robles-Puente 2019; Zárate-Sánchez 2015). For instance, in the prenuclear position of declarative sentences, Spanish heritage speakers in the US demonstrate both the high-level tone, which is the most common tone in English (Jun 2014), and the rising tone with displaced f_0 peak, which is the most common tone in Spanish (Colantoni et al. 2016; Jun 2014; Robles-Puente 2019; Zárate-Sánchez 2015). Similarly, heritage speakers of French (Bullock 2009) and Spanish (Kim 2019) in the US prosodically mark focus by adopting the strategies used in both English (e.g., prominence in situ) and their heritage language (e.g., prosodic boundary after the focused constituent). Heritage bilinguals may also demonstrate mixed patterns in the phonetic implementation (Harris et al. 2014; Kim 2020; Mennen and Chousi 2018; Rao et al. 2022; Zárate-Sánchez 2015), the frequency of use (Dehé 2018; Rao 2016; Zuban et al. 2020), or the discourse functions of the prosodic categories of their heritage language (Alvord 2010; Queen 2001; Rao 2016). That is, deviation in heritage language intonation occurs along various dimensions of intonation, consistent with the LILt's claims on L2 intonation.

For many heritage bilinguals, although the heritage language is acquired earlier or simultaneously with the societally dominant language, it oftentimes becomes their less dominant language. Heritage language outcomes exhibit high interspeaker variability depending on the amount of heritage language use, proficiency, literacy, speech community

size, access to formal education, etc. (Amengual 2016, 2018, 2019; Chang et al. 2010, 2011; Kan 2021; Kissling 2018; Oh et al. 2003; Rao 2014, 2015; Repiso-Puigdellíura 2021; Robles-Puente 2014; Rodríguez 2021; Ronquest 2012; Saddah 2011; Yeni-Komshian et al. 2000). As heritage bilingualism introduces bilingual situations that are usually overlooked in traditional L2 acquisition scenarios (e.g., transfer from L2 to L1 intonation, asymmetry between order of acquisition and language dominance), the inclusion of heritage bilinguals in the discussion of the cross-linguistic influence of intonation will contribute to building robust models of bilingual intonation.

The present study examines the production of uptalk in Spanish and in English by Spanish heritage speakers in Southern California. In Section 2, we present an overview of previous work on uptalk in English and in Spanish, as well as studies on Spanish–English bilinguals’ uptalk, which motivated this study. Section 3 presents the research questions of this study and Section 4 provides details on the participants and the methods used to answer the research questions. Section 5 presents the statistical results and Section 6 discusses the findings in connection with the research questions, as well as directions for future research. Lastly, Section 6 concludes this paper.

2. Uptalk

According to Warren (2016, p. 2), uptalk is “a marked rising intonation pattern found at the ends of intonation units realized on declarative utterances, and which serves primarily to check comprehension or to seek feedback.” Here, the term “marked” is used to distinguish uptalk from other sentence types where rises are more expected, such as in declarative questions, which function as questions, but have the syntactic form of a declarative sentence (e.g., echo questions) (Warren 2016, p. 23), and continuation rises, which occur at the end of a set of listed items, except for the last one, or at the end of incomplete statements (e.g., subordinate clauses) (Warren 2016, p. 25). Uptalk is also distinguished from statements with the rise–plateau–slump contours often found in the Urban Northern British (UNB) varieties (Cruttenden 1995, 2007; Ladd 1996, pp. 125–26; Warren 2016, pp. 88–92) and from the circumflex contours in Chicano English (Fought 2003, p. 72; Santa Ana and Bayley 2008), originated from Mexican Spanish intonation (Kvavik 1979; Matluck 1952; Martín Butragueño 2004). Such rise–fall patterns differ from uptalk, not only systemically, but also functionally. The UNB pattern is usually found in affirmative statements and signals finality (Warren 2016, p. 88) and the Chicano English pattern has emphatic and assertive discourse functions (in Kvavik 1979, as cited in Martín Butragueño 2004). The commonality of the functions of the rise–fall patterns in these varieties is that they have “closed” meanings (e.g., finality and reinforcing), which, according to Cruttenden (1981), are generally associated with falling tones.

Uptalk, on the other hand, signals openness (Warren 2016, pp. 68, 169), which is commonly linked to rising tones (Cruttenden 1981). Some of the meanings of uptalk that frequently appear in the literature are uncertainty, politeness, deference, friendliness, open-ended, and checking (House 2006; Shokeir 2008; Warren 2016, pp. 47–68). While uptalk is stereotypically associated with uncertainty or lack of confidence, this interpretation is contentious, given that uptalk has multiple layers of meaning (e.g., indexical, linguistic, discourse, and attitudinal), which often simultaneously emerge in a single contour (House 2006; Warren 2016, p. 14). Moreover, uptalk may even signal conflicting meanings within the same layer (e.g., subjugation vs. socially ambitious), depending on the context and on the shared communicative conventions and norms (McLemore 1991; Warren 2016, p. 68). According to Warren (2016), despite the vast range of meanings, the main significance of uptalk is interactional; uptalk is primarily used to invite listeners to check for comprehension, to elicit feedback, and to signal information structure (Warren 2016, pp. 47–68).

2.1. Uptalk in English

Uptalk is widely used across English varieties and the forms of uptalk may vary from one variety of English to another, similar to how sound change in certain segments (e.g., back vowel merger) is present in some varieties, but not in others (Warren 2016, pp. 31, 42). In a review of the phonological description of uptalk across English varieties, Grice et al. (2020) stated that, in Autosegmental-Metrical (AM) terms, “uptalk has been labelled as L* H-H% and H* L-H% for Canadian English (Di Gioacchino and Jessop 2010; Shokeir 2008), L* L-H%, L* H-H%, and H* H-H% for American English (Hirschberg and Ward 1995; McLemore 1991; Ritchart and Arvaniti 2014), L* H-H%, H* H-H% and the longer sequence H* L* H-H% for Australian and New Zealand English (Fletcher 2005; Fletcher et al. 2005; Fletcher and Harrington 2001; McGregor and Paethorpe 2008), H* L-H% or H*+L H-H% for British English (Bradford 1997).”

The phonetic realization of uptalk in English also shows variability in whether and how uptalk differs from question rises (see Warren 2016, pp. 36–40 for a comprehensive review). For instance, uptalk in North American English (Di Gioacchino and Jessop 2010; Ritchart and Arvaniti 2014) demonstrates smaller pitch excursion, compared to question rises, while in Australian English the pitch level at rise onset appears to play a more important role in the distinction between uptalk and question rises (i.e., lower onset in uptalk) (Asano et al. 2020; Fletcher and Harrington 2001). With regard to the temporal aspect of the rise, uptalk is found to be produced with later rises than questions in Southern California English (Ritchart and Arvaniti 2014), in New Zealand English (Warren 2005), and in South African English (Dorrington 2010). This pattern may be found because question rises typically include the last stressed vowel, whereas rises in uptalk is generally aligned with “metrically strong post-nuclear syllables (MSPNS)” or with the final unstressed syllable at the periphery of the phrase (Dorrington 2010; Ritchart and Arvaniti 2014; Warren 2005). However, it is important to note that there is a wide range of areas that uptalk can cover in English; uptalk rises may occur in the final syllable of the intonational phrase (IP), over more than one syllable within the last word, or across multiple words (Britain and Newman 1992; Warren 2005; Warren 2016, p. 32). These findings suggest that, while there is some commonality in the forms of uptalk in English, considerable variation exists within and across varieties (Warren 2016, p. 45).

While the use of uptalk is not limited to a specific variety of English, in the context of the US, it is stereotypically associated with young female speakers from Southern California (Armstrong et al. 2015; Ritchart and Arvaniti 2014; Tyler 2015). It is deemed a typical trait of *Valley Girl* speech, which triggers images of “rich, white young females from the San Fernando Valley” in Los Angeles County (in Ritchart 2014, as cited in Tyler 2015). Such misconception of uptalk, which may have been popularized due to media exposure, has been overturned by empirical evidence. For instance, Armstrong et al. (2015) compared uptalk in Southern California English and Massachusetts English and found that the two varieties did not have any systematic gender or regional differences in the frequency of uptalk. Ritchart and Arvaniti (2014) also found that female and male speakers in Southern California used uptalk with similar frequency in non-floor-holding statements (17% and 16%, respectively), although the female speakers used uptalk more than twice as much as the male speakers for floor holding purposes (59% and 28%, respectively). With regard to the phonetic implementation of uptalk, Ritchart and Arvaniti (2014) found that the female speakers had greater pitch excursions and later rise onsets (i.e., steeper rises) than male speakers. Similarly, the female speakers in Armstrong et al. (2015) produced steeper rises than the male speakers, but they also had longer rises than the male speakers. That is, female speakers are likely to use greater “intonational gesture space” between short/steep and long/shallow rises (Armstrong et al. 2015, p. 5). According to Armstrong et al. (2015), the popular stereotype regarding the prevalence of uptalk in *Valley Girl* speech may have been formed because young female speakers in Southern California exploit the phonetic aspects of rises and/or use uptalk for more forward-looking purposes (e.g., directing attention to the upcoming utterance), which is associated with prolonged rising pitch

(Tomlinson and Tree 2011). In other words, rather than the use of uptalk per se, young female speakers' phonetic implementation of uptalk and/or the different pragmatic choices that they make may have led to the impression that their uptalk is more salient than others.

2.2. Uptalk in Spanish

Uptalk in Spanish has not been investigated as extensively as in English, but studies have reported that Spanish speakers commonly use uptalk (Henriksen 2017; Holguín Mendoza 2011; Kim and Repiso-Puigdelliura 2021; Martínez-Gómez 2018; Vergara 2015; Willis 2010). For instance, Willis (2010) reported that, similar to uptalk in English, statements in the Cibaño variety of Dominican Spanish were consistently produced with a final rise, which often involved a high boundary tone (H%), preceded by a falling nuclear pitch accent (H+L*). While these contours are mainly found in yes–no questions in this variety, the magnitude of rise was typically higher in statements than in questions (Willis 2010). Henriksen (2017), based on oral narratives collected from an on-going project with Armstrong-Abrami and García-Amaya, showed that non-question rises were much more common in Peninsular Spanish (57.8%) than in American English (20.4%). They found that male speakers produced more rises and that their rises were realized with greater pitch excursion and with longer duration than those of female speakers, contrary to the findings in English (Armstrong et al. 2015; Ritchart and Arvaniti 2014). According to Vergara (2015), Peninsular Spanish speakers mainly use the L* LH% melody when producing uptalk. According to the Spanish Tones and Break Indices (Sp_ToBI) annotation system (Beckman et al. 2002; Prieto and Roseano 2010), this contour is realized as a low plateau throughout the last accented syllable and a part of the subsequent syllable, followed by a rise to a high pitch level. Although less frequent than L* LH%, there were some instances of L+H* HH%. The L+H* HH% melody is the same contour used for counter-expectational questions in Peninsular Spanish, which is realized as a rise during the last accented syllable that continues into the following syllable(s), attaining a high pitch level (Estebas-Vilaplana and Prieto 2010).

In Mexican Spanish, uptalk is associated with *fresa* (Spanish word for “strawberry”), a word that is used in Mexico to call “a person, especially, women, who are or try to appear from the upper class by behaving, dressing, and speaking in a manner perceived as snobbish towards other people” (Holguín Mendoza 2011, p. 36). Holguín Mendoza (2011) showed that young women in Ciudad Juárez, Mexico, who demonstrate typical traits of *fresa* speech, produced many of their uptalk contours using the L* LH% melody, as in Peninsular Spanish (Vergara 2015). According to Holguín Mendoza (2011), this melody resembled the contours of information-seeking yes–no questions, echo yes–no questions, and imperative yes–no questions in Mexico City Spanish (De la Mota et al. 2010). Uptalk is also used among non-*fresas*. Martínez-Gómez (2018) argued that young speakers in the Guadalajara Metropolitan Area, frequently use uptalk, regardless of whether they are *fresas* or not. She found that the main difference between *fresa*-sounding and non-*fresa*-sounding speech derived from the phonetic realization of uptalk; *fresa*-sounding participants produced uptalk with greater pitch excursions and steeper rise slopes than non-*fresa*-sounding participants. In other words, uptalk in itself does not index a *fresa* persona, but rather the way it is realized (e.g., steep rises accompanied by other linguistic features) (Martínez-Gómez 2018, p. 92). The distinction between the uptalk in *fresa*- and non-*fresa* speech may also be characterized by their intonation contours. As in Holguín Mendoza (2011) and Vergara (2015), Kim and Repiso-Puigdelliura (2021) found that non-*fresa* speakers in Central Mexico used L* LH% when producing uptalk (11.4%), but the two most common melodies in their data were L+H* (H)H% (33%) and L* (H)H% (20.4%). Recall that the former contour was also found in Peninsular Spanish uptalk, but it was used with very low frequency (Vergara 2015). The L* H(H)% contour is generally used for invitation and confirmation yes–no questions in Mexico City Spanish and it is realized as a low plateau during the last accented syllable, followed by a rise to a (very) high pitch level (De la Mota et al. 2010). While there were very few instances of questions (3.4% of the entire data) to make any generalizations, the uptalk

rises in Kim and Repiso-Puigdelliura (2021) had greater pitch excursions and steeper rises (8.4 semitones, 32.3 semitones per second) than the questions (7.2 semitones, 23.3 semitones per second), which are consistent with the uptalk rises in Dominican Spanish (Willis 2010).

The findings of the above-mentioned studies demonstrate that the use of uptalk is widespread in Spanish, showing not only similarities, but also considerable differences within and across varieties in its intonation contour and phonetic implementation, which may index different linguistic and social meanings. Moreover, uptalk in Spanish seems to share the same intonation contours with yes–no question rises, with uptalk rises having a greater pitch excursion than yes–no question rises. Nonetheless, more research should be conducted to confirm that the distinction between the two sentence types in Spanish is truly phonetically based. The meanings and functions of uptalk have been even less investigated in Spanish. To the best of our knowledge, Vergara (2015) is the only study that examined various discourse functions of uptalk in Spanish. He found that Peninsular Spanish speakers use uptalk to hold the floor, to show camaraderie, to soften a command, and in the case of female speakers, to flirt (*coqueteo*). However, it is uncertain whether these functions are transferable to other Spanish varieties and whether they surface in different uptalk contours.

2.3. Uptalk of Spanish–English Bilingual Speakers

Due to the strong link between uptalk and English, uptalk observed in Spanish–English bilinguals' Spanish is often considered an indication of influence from English intonation (Buck 2016; Henriksen et al. 2010; Méndez Seijas 2019; Trimble 2013; Zárate-Sánchez 2018). Zárate-Sánchez (2018) examined the pitch values at the end of declarative sentences produced by six groups of speakers with varying degrees of language dominance, from Spanish dominant to English dominant: Spanish monolinguals, Spanish heritage speakers that are balanced bilinguals, three groups of English L2 learners of Spanish with different Spanish proficiency levels (i.e., very high, high, and intermediate), and English monolinguals. Results showed that speakers who are more dominant in English had higher final pitch values. Given the final rising intonation in uptalk, Zárate-Sánchez (2018) conjectured that the higher final pitch found in more English-dominant speakers suggests a more frequent use of uptalk by these speakers. However, given that high pitch at the end of an utterance alone does not signify a final rise, it is possible that uptalk does not explain the positive relationship found between English dominance and final pitch height.

While it is well accepted that uptalk is a widespread phenomenon in English, it is important to take into account that Spanish speakers frequently use uptalk as well (Henriksen 2017; Holguín Mendoza 2011; Kim and Repiso-Puigdelliura 2021; Martínez-Gómez 2018; Vergara 2015; Willis 2010). Thus, the presence of uptalk in bilinguals' Spanish in itself does not attest that uptalk has been transferred from English to Spanish. Kim and Repiso-Puigdelliura (2021) found that Mexican Spanish speakers did not differ in their uptalk frequencies, regardless of whether they are monolingual in Spanish or heritage bilinguals. Rather, the two groups differed in the forms of uptalk. Compared to the Spanish monolinguals in Mexico, the heritage bilinguals in Southern California produced uptalk with flatter, and to some extent, larger rises and with less dynamic intonation contours, similar to the low-rise pattern found in Southern California English uptalk (Ritchart and Arvaniti 2014). Moreover, in some cases, the heritage bilinguals produced uptalk over multiple words, which has been attested in English (Britain and Newman 1992; Warren 2005; Warren 2016, p. 32), whereas none of the Spanish monolinguals demonstrated rises beginning in a non-IP-final word.

According to Jun (2014), both Spanish and English are head-prominence languages, but the domain of the head (i.e., pitch accent) is approximately one content word in Spanish, whereas it is larger than one content word in English. Therefore, uptalk in English can begin in a non-IP-final word if the following words are deaccented, while rise onsets in Spanish should occur within IP-final words because, in Spanish, content words almost always carry a pitch accent. While deaccenting is also possible in Spanish, particularly in semantically

light words (e.g., high lexical frequency, given information, syntactic determiners, and copulas) in spontaneous speech (Face 2003; Rao 2009), it is not as common as in English (Face 2003); if it does occur, it is usually located in non-final phrase positions (Rao 2009). Thus, uptalk beginning in a non-IP-final word is an English feature associated with IP-final deaccenting. To the best of our knowledge, no study has reported such uptalk patterns in non-heritage Spanish varieties. Fought (2003, pp. 73, 76) reported that some of the heritage bilinguals she interviewed seemed to superimpose the uptalk contours of California English onto their Spanish. These findings suggest a potential influence from English to Spanish on how bilinguals produce uptalk in Spanish.

With regard to uptalk productions in English by Spanish–English bilinguals, studies have shown that heritage bilinguals demonstrate similar uptalk patterns as Anglo English speakers who grew up monolingually in English (Asch and Brogan 2022; Fought 2003; Santa Ana and Bayley 2008). Asch and Brogan (2022) found that heritage bilinguals in Southern California were very similar to Anglo English speakers in both the frequency and the phonetic implementation of uptalk (i.e., starting pitch, pitch scaling, rise alignment, and peak delay). These findings suggest that Spanish (i.e., a minority language) may not have a noticeable influence on heritage bilinguals' production of uptalk in English (i.e., the societally dominant language). However, in order to corroborate cross-linguistic influence of uptalk in heritage bilinguals, it is necessary to examine both languages because bilingual speakers are highly heterogeneous and transfer cannot occur if the target features are absent in one's grammar. To the best of our knowledge, no study has made a direct comparison between the uptalk in bilinguals' two languages.

3. Research Questions

The present study examines heritage bilinguals' uptalk in Spanish and in English to better understand the role of cross-linguistic influence in their uptalk production. We explore heritage bilinguals' uptalk patterns, focusing on those that have been found to differ from Spanish monolinguals (Kim and Repiso-Puigdelliura 2021): uptalk with IP-final deaccenting and uptalk realized with smaller and flatter rise. In Kim and Repiso-Puigdelliura (2021), the phonetic analysis of uptalk was conducted based on two interrelated properties, namely, pitch excursion and rise slope (i.e., the extent of pitch excursion per second). Thus, in this study, instead of rise slope, we analyzed the duration of uptalk rises.

Following the L2 Intonation Learning Theory (LILt) (Mennen 2015), we propose that cross-linguistic influence in heritage bilinguals' uptalk can occur along multiple dimensions of intonation. Here, we focus on the systemic, the frequency, and the realizational dimensions of heritage bilinguals' uptalk. The semantic dimension of uptalk was not considered in this study because, without a clear understanding of the meanings and functions of various uptalk contours in Spanish, it is premature to investigate the cross-linguistic influence in the semantic dimension. Moreover, the meanings of uptalk can be best understood through a perception task that tests how uptalk is interpreted by listeners of the target variety (e.g., Tomlinson and Tree 2011), which is outside the scope of the present study. Apart from the above-mentioned three dimensions of uptalk, we also examine how heritage bilinguals' uptalk is influenced by their language learning experience.

We aim to answer the following research questions.

- (1) Do heritage bilinguals produce uptalk in both Spanish and English (systemic dimension)? If so, do they produce it with similar frequency between the two languages (frequency dimension)?
- (2) Do heritage bilinguals produce uptalk with IP-final deaccenting in both Spanish and English (systemic dimension)? If so, do they produce it with similar frequency between the two languages (frequency dimension)?
- (3) Do heritage bilinguals produce uptalk with similar pitch excursion and rise duration between Spanish and English (realizational dimension)?
- (4) Do extralinguistic factors associated with heritage bilinguals' language learning experience have an effect on their production of uptalk?

4. Methods

4.1. Participants

Twenty-four Spanish–English bilingual Mexican Americans (18F, 6M) participated in the present study. Due to technical issues, the speech of one participant (HS3) was not recorded. In this paper, we report information regarding the remaining 23 participants. 16 of them (12F, 4M) were born and raised in Los Angeles County and their parents immigrated to the US from Mexico as adults. The other 7 speakers (5F, 2M) spent their childhood in Mexico and moved to Southern California during late childhood or adolescence (age range: 7–15 years). All of the 7 speakers were born in Mexico, except for one speaker (HS11), who was born in Los Angeles, moved to Mexico with her family soon afterwards, and lived there until 8 years of age.

Among the 23 participants, only 2 speakers learned Spanish and English at the same time (i.e., simultaneous bilinguals). The other 21 speakers learned Spanish first and English after that (i.e., sequential bilinguals); 14 of them learned English before entering elementary school (age range: 1–5 years), while 7 speakers reported that they learned English at school (age range: 7–15 years). All the participants acquired Spanish at home since birth and were fluent enough in both Spanish and English to carry on a conversation in the two languages. Table 1 summarizes participants’ language profile.

Table 1. Descriptive statistics of participants’ language profile.

	Mean	Standard Deviation	Min	25%	Median	75%	Max
Age (years)	21.87	3.02	18	19.5	21	24	28
Age of arrival (years)	3.43	5.57	0	0	0	7.5	15
English AOA ¹ (years)	5.61	3.95	0	4	4	8	15
Language dominance ² (−218–218)	6.54	50.76	−118.78	−11.26	15.98	43.23	77.64
Language use ³ (0–50)							
Spanish	20.7	8.29	9	13	20	26	37
English	29.3	8.29	13	24	30	37	41
Proficiency ³ (0–24)							
Spanish	19.87	3.93	12	17.5	21	23.5	24
English	21.61	2.76	14	20	22	24	24
Education ³ (years)							
Spanish	4.04	2.88	0	2	4	6	10
English	14.35	3.81	4	13	15	16.5	20
Picture naming (0–60)	54.3	3.38	48	52	55	57	59

¹ Age of acquisition, ² Bilingual Language Profile score, and ³ Sub-Component of the Bilingual Language Profile.

The age of the participants ranged between 18 to late 20s. All of the participants were either college students or recently graduated from college. Information regarding participants’ language dominance was obtained from their responses in Birdsong et al.’s (2012) Bilingual Language Profile (BLP). The BLP is a questionnaire that evaluates the overall language dominance of bilingual speakers based on self-reports on language history, language use, language proficiency, and language attitude in their two languages. It generates a continuous score from −218 to 218. A positive score indicates English dominant and a negative score indicates Spanish dominant. A score of or close to zero indicates balanced bilingualism. The BLP scores in our data ranged between −118.78 (Spanish dominant) and 77.64 (English dominant); 6 speakers were Spanish dominant (M = −51.51, SD = 47.95), 4 speakers were balanced bilinguals (M = −6.27, SD = 2.34), and 13 speakers were English dominant (M = 40.68, SD = 19.98).

One of the advantages of using the BLP in bilingualism research is that it not only evaluates the overall language dominance through a composite score, but also allows separate analysis of the amount of use and the proficiency of bilinguals’ two languages. Language dominance is a multidimensional construct that is relativistic in nature (i.e., Language A compared to Language B) (Birdsong 2016; Montrul 2016; Silva-Corvalán and Treffers-Daller

2016). In other words, language use and language proficiency are sub-constructs of language dominance (Birdsong 2016; Montrul 2016). Among the 23 participants, 6 speakers use Spanish more frequently than English and 17 speakers use English more frequently than Spanish. Regarding language proficiency, 4 speakers rated their Spanish higher than their English, 7 speakers rated the two languages equally, and 12 speakers rated their English higher than their Spanish. The BLP also provides information regarding bilingual's classroom experience in their two languages, which is an important factor of heritage bilinguals' language learning experience because exposure to formal speech has shown to influence heritage bilinguals' sound system (Rao et al. 2020). Since our participants spent all or many of their school years in the US, they had classes in English for a longer period of time than classes in Spanish; only one speaker who moved to the US at age 15 reported that she spent more time taking classes in Spanish than in English.

In this study, we also conducted a picture-naming task in Spanish (Kim 2016) to measure participants' lexical proficiency. According to Polinsky and Kagan (2007), lexical proficiency is a powerful diagnostic of heritage language proficiency. The picture-naming task includes black-and-white images of 60 Spanish object nouns across five frequency levels based on Davies' (2006) Spanish frequency dictionary. The images were selected from the International Picture-Naming Project (IPNP) database (Szekely et al. 2004) and were individually presented in PowerPoint slides. The participants were asked to say the word out loud in Spanish as quickly as possible. For a detailed explanation of the task design and the complete list of items used in the picture-naming task, refer to Kim (2016, pp. 54–55, 162). Out of 60, the participants scored between 48 (80%) and 59 (98.33%), suggesting that they had good lexical knowledge in Spanish.

4.2. Procedures

The participants were assigned to pairs that had similar backgrounds (e.g., country of birth, age, and gender) and based on their time availability. All pairs matched in their age of arrival to the US, except for one pair (HS11 and HS12); both participants were born in Los Angeles, but HS12 spent all her life in the US, whereas HS11 moved to Mexico soon after she was born and lived there until she came back to Los Angeles when she was 8 years old (see Section 4.1). After reading and signing a written informed consent form, the participants completed two production tasks (one in each language), the picture-naming task, and the Bilingual Language Profile (Birdsong et al. 2012). Recall that uptalk most likely occurs in interactional contexts (see Section 2), which indicates that it is unlikely to emerge in tasks where "the listener can be assumed to already know the general content of what the speaker is saying" (Warren 2016, p. 176). Thus, we conducted production tasks that involve conversations between two people.

For the Spanish conversation task, a dyadic interaction task was conducted in a sound-attenuated room, where the participants discussed topics related to Los Angeles in pairs. A list of topics was provided at the onset of the task (e.g., racism, undocumented immigrants, safety of women, maintenance of Spanish language, and housing) and each pair chose between two and four topics of interest to discuss. The instructions were provided in Spanish by a Spanish–English bilingual Mexican American research assistant. With regard to the English conversation task, the investigator, a second language (L2) speaker of Spanish and non-Latinx, asked questions regarding participants' experience interacting with their partners during the dyadic interaction task (e.g., Does your partner share similar backgrounds with you? Did you agree on the topics you discussed? If you were to choose a different topic, would you have similar perspectives as your partner?). While the English conversation task took place in the form of an interview, the investigator encouraged the participants to elaborate their responses and talk about any other topics of their interest. The conversations oftentimes diverged from the interview topics, which the investigator did not deter, given that the purpose of this study was to elicit spontaneous conversational speech.

After the Spanish conversation task, the participants took a short break of 5–10 min. After the break, one of the partners moved to a quiet furnished office room to complete the English conversation task. In the meantime, the other partner stayed in the lab to complete the picture-naming task and the BLP questionnaire described in Section 4.1. Then, the partners switched turns¹. Participants' spontaneous speech during the conversation tasks and their responses during the picture-naming task were recorded using an AKG C520 head-mounted microphone and a Zoom H4n digital recorder with a sampling rate of 44.1 kHz and a sample size of 16 bits.

4.3. Coding and Analysis

Uptalk was identified as rising contours at the end of non-question intonational phrases (IPs). In this study, we used pauses as the main cue to IP-final boundary, which surface as silence, glottalization, or final lengthening in the speech signal. Instances of pausing due to disfluency, such as stutters, self-repairs, fillers (e.g., like, you know, so, and uh), backchannel responses (e.g., yeah and uh-huh), and utterances interrupted by the interlocutor, were excluded from the analysis. We also excluded IP-final boundaries overlapped with laughter or background noises, in which the intonation patterns are unclear. Moreover, any English expressions at IP-final boundaries in the Spanish data or Spanish expressions at IP-final boundaries in the English data were excluded from the analysis, given that it is uncertain whether the language of the uptalk in such cases should be categorized as Spanish or English. We also did not consider the circumflex contours attested in Chicano English (Asch and Brogan 2022; Fought 2003; Santa Ana and Bayley 2008) and in Mexican Spanish (Kvavik 1979; Matluck 1952; Martín Butragueño 2004), which are clearly distinct from uptalk (Fought 2003; Santa Ana and Bayley 2008).

A trained Spanish–English bilingual research assistant identified non-question IPs in the Spanish and the English data and annotated whether they were produced as uptalk based on both auditory and visual inspections of the pitch contours (i.e., rising contours). The annotation and the visualization of the pitch contours were carried out in Praat (Boersma and Weenink 2021). By default, the pitch settings were set to 75–300 Hz for male speakers and 100–600 Hz for female speakers, but adjustments were made if an individual speaker's pitch ranged outside the default settings. For non-question IPs produced as uptalk, the investigator further annotated whether the last accented syllable occurred in a non-IP-final word (i.e., uptalk with IP-final deaccenting), based on auditory inspection and visual inspection of the pitch contour (i.e., no prominent f_0 movement between rise onset and offset) and the spectrogram (i.e., no clear distinction in the darkness of syllables between rise onset and offset). After extracting the labels of all non-question IPs using a Praat script (adapted from a custom script by Christopher Carignan), the investigator re-coded them in Excel spreadsheet based on the presence of uptalk (1 = uptalk, 0 = not uptalk). For instances of uptalk, further coding was carried out based on the presence of uptalk with IP-final deaccenting (1 = IP-final deaccenting, 0 = no IP-final deaccenting). The relative frequency of uptalk was calculated as the number of uptalk divided by the total number of non-question IPs. The relative frequency of uptalk with IP-final deaccenting was calculated as the number of uptalk with IP-final deaccenting divided by the total number of uptalk instances.

With regard to the phonetic realization of uptalk, the investigator extracted the pitch excursion and the rise duration using Praat (Boersma and Weenink 2021). For each rise, we first selected the regions in which the highest (i.e., f_0 maximum) and the lowest points (i.e., f_0 minimum) of the rise were identified and automatically extracted the f_0 (Hz) and time (seconds) of these points using a Praat script (adapted from a custom script by Christopher Carignan). To calculate pitch excursion, we converted the f_0 difference between these two points into semitones (st) ($=12 * \log_2[f_0 \text{ maximum} / f_0 \text{ minimum}]$), which is a logarithmic scale that best reflects listeners' intuitions about intonational equivalence (Nolan 2003; Pépiot 2014; Simpson 2009). Tokens in which the f_0 maxima and minima could not be measured (e.g., within voiceless segments and creak) were excluded from the analysis.

Rise duration was calculated as the distance between the time of the f0 maxima and the time of the f0 minima (seconds). Rise duration of uptalk with IP-final deaccenting was excluded from the analysis, given that uptalk realized across multiple words inevitably leads to longer rise duration than those realized at IP-final words. Statistical analyses and data visualization were performed using R (R Core Team 2021). More information of the packages and the statistical models used in this study is presented in Section 5.

With regard to the effect of extralinguistic factors associated with language learning experience, we did not include language dominance in the analysis, given that language dominance is a multidimensional construct that embraces most of the other variables. Moreover, since none of the participants spoke languages other than Spanish and English, their Spanish use was inversely related to their English use. Thus, we only included participants' Spanish use and not their English use in the analysis. This resulted in a total of 8 extralinguistic variables (i.e., age of arrival to US, age of acquisition of English, Spanish use, Spanish self-rated proficiency, English self-rated proficiency, education in Spanish, education in English, and picture-naming task score). The correlation matrix of these variables² revealed that, among the 28 pairs of variables analyzed, 10 pairs exhibited absolute correlation coefficients higher than 0.5, suggesting that many of these variables are correlated with each other (e.g., age of arrival to the US and age of acquisition of English: $r = 0.89$, Spanish use and Spanish self-rated proficiency: $r = 0.7$, and age of arrival to the US and education in English: $r = -0.66$). Therefore, we decided to conduct principal component analysis to reduce dimensionality.

Prior to performing the principal component analysis, we ran the Kaiser-Meyer-Olkin (KMO) test (Kaiser 1970) and Bartlett's Test of Sphericity for data screening. The KMO test gauges sampling adequacy. While values higher than 0.7 are considered adequate (Kaiser 1974), given the small number of participants in the present study ($n = 23$), we considered 0.5 as the acceptable lower limit, following Field's (2009) recommendation. Bartlett's test checks whether the correlations among the variables are large enough to be analyzed. A significant Bartlett's test indicates interrelationship among the variables. The KMO test and Bartlett's test were performed using the `KMO()` function and the `cortest.bartlett()` function, respectively, in the *psych* package (Revelle 2022).

The principal component analysis was performed using the `principal()` function in the *psych* package (Revelle 2022). We determined the number of components to extract based on the components' eigenvalues (Kaiser 1960) and their cumulative percentage of total variation (Jolliffe 2002). Following Kaiser's (1960) criterion, eigenvalues above 1 were extracted, given that an eigenvalue lower than 1 indicates that the component accounts for less variance than the original variable. However, if the combination of the extracted components based on this criterion does not sufficiently explain the total variation of the dataset, we extracted additional components. Jolliffe (2002, p. 113) suggested the cut-off point to be somewhere between 70% and 90%. In this study, we set 70% as the cut-off point.

5. Results

5.1. Principal Component Analysis of Extralinguistic Factors Associated with Language Learning Experience

We conducted principal component analysis on the eight extralinguistic factors associated with language learning experience (i.e., age of arrival to US, age of acquisition of English, Spanish use, Spanish self-rated proficiency, English self-rated proficiency, education in Spanish, education in English, and picture-naming task score) to reduce dimensionality. In our dataset, the KMO value was 0.66 and Bartlett's test was significant ($\chi^2(28) = 105.47$, $p < 0.001$), indicating acceptable sampling adequacy and interrelationship among variables. Based on these results, we concluded that the variables are suitable for principal component analysis. We initially extracted two principal components (PCs) based on their eigenvalues (PC1: 4.12, PC2: 1.33). However, given that their cumulative percentage of total variation was lower than the 70% cut-off point, we additionally extracted PC3 whose eigenvalue was 0.95. The three components in combination accounted for 80% of the variance (PC1: 38%,

PC2: 28%, PC3: 14%). Since the components were not correlated with each other (i.e., correlation coefficients close to zero), we chose varimax rotation, which is an orthogonal rotation. Table 2 contains the component loadings of the eight variables after rotation. Only variables with absolute loadings greater than 0.4 are shown.

Table 2. Component loadings.

Variables	PC1 (English Experience)	PC2 (Spanish Proficiency and Use)	PC3 (Education in Spanish)
Age of arrival	0.9		
English AOA	0.9		
Proficiency in English	−0.9		
Education in English	−0.7		
Spanish use		0.8	
Proficiency in Spanish		0.9	
Picture naming		0.6	
Education in Spanish			0.9

As demonstrated in Table 2, the rotated solution yielded three interpretable components. PC1 is strongly correlated with age of arrival to the US, age of acquisition of English, English self-rated proficiency, and education in English. We will interpret this component as “English experience.” PC2 is strongly correlated with Spanish use, Spanish self-rated proficiency, and the picture-naming task score (i.e., lexical proficiency) and PC3 is constructed mostly from education in Spanish. Thus, we will interpret PC2 and PC3 as “Spanish proficiency and use” and “education in Spanish”, respectively.

5.2. Systemic and Frequency Dimensions of Uptalk

In our data, uptalk was observed in heritage bilinguals’ Spanish and English. In total, 7431 instances of non-question intonational phrases (IPs) were produced. Among them, 2232 were IPs with uptalk contours (30.04%) and 5199 were non-uptalk IPs (69.96%). Average uptalk rates in English and in Spanish were 30.27% and 29.23%, respectively. Individual uptalk rates varied between 14.15% and 63.92% in English and between 14.29% and 58.06% in Spanish.

With regard to uptalk with IP-final deaccenting, we found that heritage bilinguals produce this pattern in both languages. Out of the 2232 instances of uptalk, 595 tokens (26.66%) were produced with IP-final deaccenting. Such rises occurred 31.66% of the time in English with individual rates ranging from 17.02% to 50.91%. In Spanish, they occurred 8.51% of the time and individual rates ranged from 0% to 30.77%. Figure 1 demonstrates an example of uptalk with IP-final deaccenting, in which the rise onset falls on the penultimate content word.

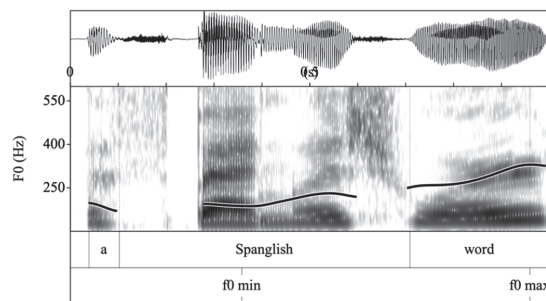


Figure 1. Example of uptalk with IP-final deaccenting in English produced by a female heritage bilingual (HS2). “A Spanglish word”.

We performed a mixed-effects logistic regression analysis using the *glmer()* function in the *lme4* package (Bates et al. 2015) to examine whether language and participants' language learning experience influence the presence of uptalk (1 = yes/0 = no). Moreover, we further analyzed the effects of the same factors on the presence of uptalk with IP-final deaccenting (1 = yes/0 = no). As fixed effects, we entered language (English/Spanish) and the three principal components (PCs) presented above (see Section 5.1). Recall that half of the participants completed the language survey first and the English conversation task after that, whereas the other half completed the English conversation task first and the language survey after that (see Section 4.2). Since the order of the language survey and the follow-up interview may have an impact on participants' performance, we included task order (language survey first/English conversation task first) as a covariate. We also included gender as a covariate, due to potential relationship between gender and uptalk use (Armstrong et al. 2015; Henriksen 2017; Ritchart and Arvaniti 2014; Tyler 2015). The categorical fixed effects were contrast-coded using simple coding, in which each level is compared to the reference level (language: English, task order: language survey first, gender: female) and the intercept is the grand mean. We entered participant as a random effect. For both the presence of uptalk and the presence of uptalk with IP-final deaccenting, the best fitting model selected through backward elimination included an intercept for participant with by-participant random slope for language. Statistical significance of the fixed effects was analyzed through likelihood ratio tests of the full model with all the predictor effects (i.e., language and the three PCs) against the model without the effect in question. Likelihood ratio tests were performed using the *anova()* function in the *car* package (Fox and Weisberg 2019) and visualization of the predictor effects was carried out using the *predictorEffect()* function in the *effects* package (Fox and Weisberg 2019).

With regard to the presence of uptalk, adding gender to the model significantly strengthened the model fit ($\chi^2(1) = 8.38, p < 0.01$), while task order did not show any improvement. Thus, we added gender to the full model. Results showed that none of the four predictor effects influenced the presence of uptalk ($ps > 0.05$). As in the case of the presence of uptalk, adding gender to the model significantly strengthened the model fit for the presence of uptalk with IP-final deaccenting ($\chi^2(1) = 6.26, p < 0.05$), while task order did not show any improvement. Thus, we added gender to the full model. Results showed that language ($\chi^2(1) = 31.5, p < 0.001$) affected the presence of uptalk with IP-final deaccenting, which indicates that heritage bilinguals produced this uptalk pattern more frequently in English than in Spanish (see Figure 2). None of the PCs had an effect on participants' production of uptalk with IP-final deaccenting.

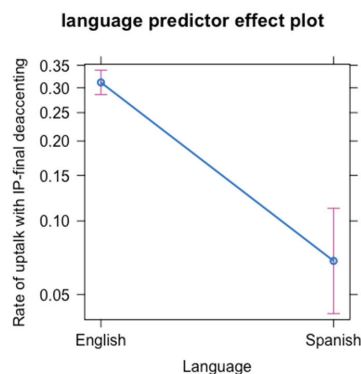


Figure 2. Rate of uptalk with IP-final deaccenting by language.

These findings suggest that heritage bilinguals use uptalk to varying degrees, regardless of language and their language learning experience. However, they make a distinction between the uptalk pattern of their two languages. In English, they sometimes produce

their uptalk with IP-final deaccenting, while such pattern was rarely found in their uptalk in Spanish.

5.3. Realizational Dimension of Uptalk

In this study, we examined two aspects of the phonetic realization of uptalk: pitch excursion (st) and rise duration (s). 488 tokens were excluded from the analysis of pitch excursion due to missing f_0 maxima and minima (see Section 4.3). For the analysis of rise duration, 464 tokens with IP-final deaccenting were additionally removed because uptalk realized across multiple words inevitably leads to longer rise duration than those realized at IP-final words (see Section 4.3). Furthermore, 22 proparoxytones tokens were removed, since this stress pattern was only found in the English data. Thus, the remaining data reflect a total number of 1744 tokens for the pitch excursion analysis and 1258 tokens in the case of the rise duration analysis.

We performed a mixed-effects linear regression analysis using the *lmer()* function in the *lme4* package (Bates et al. 2015) to examine whether language and participants' language learning experience affect the two phonetic properties of uptalk. As fixed effects, we entered language (English/Spanish) and the three principal components (PCs). Task order and gender were included as covariates due to the reasons mentioned above (see Section 5.2). For the analysis of rise duration, we additionally added stress pattern (oxytone/paroxytone/proparoxytone) as a covariate, since rise onset is expected to be influenced by stressed syllable location; rise onset is likely to occur earlier for words in which the stressed syllable is farther away from the right edge of the IP boundary. The categorical fixed effects were contrast-coded using simple coding, in which each level is compared to the reference level (language: English, task order: survey first, gender: female, and stress pattern: oxytone) and the intercept is the grand mean. We entered participant as a random effect. For the analysis of pitch excursion, the best fitting model selected through backward elimination included an intercept for participant with by-participant random slope for language. In the case of the analysis of rise duration, the best fitting model included an intercept for participant with no slope terms. Statistical significance of the fixed effects was analyzed through likelihood ratio tests of the full model with all the predictor effects (i.e., language and the three PCs) against the model without the effect in question. Likelihood ratio tests were performed using the *anova()* function in the *car* package (Fox and Weisberg 2019) and visualization of the predictor effects was carried out using the *predictorEffect()* function in the *effects* package (Fox and Weisberg 2019).

Regarding the analysis of pitch excursion, adding task order or gender to the full model did not have any effect on the model fit. Thus, we did not include these factors in the full model. Results showed that none of the four predictor effects influenced participants' pitch excursion ($ps > 0.06$). In the case of rise duration, the model fit improved by adding gender ($\chi^2(1) = 9.8, p < 0.01$) or stress pattern ($\chi^2(2) = 338.67, p < 0.001$) to the model, and adding both gender and stress pattern demonstrated a better fit than the models with one of the effects ($ps < 0.001$). Thus, we included both covariates in the full model. Results showed that language ($\chi^2(1) = 6.27, p < 0.05$), English experience (PC1) ($\chi^2(1) = 4.51, p < 0.05$), and Spanish proficiency and use (PC2) ($\chi^2(1) = 6.11, p < 0.05$) had an effect on participants' rise duration. In other words, the participants produced the uptalk with shorter rise duration in Spanish than in English (see Figure 3a). Moreover, participants with more English experience had longer rise duration (see Figure 3b), whereas those that had higher Spanish proficiency and use demonstrated shorter rise duration (see Figure 3c). Education in Spanish (PC3) did not have an effect on participants' rise duration.

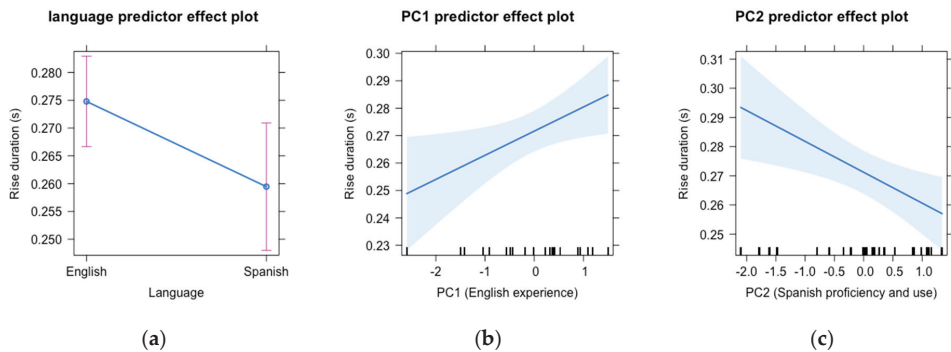


Figure 3. Rise duration (s) by (a) language; (b) English experience (PC1); (c) Spanish proficiency and use (PC2).

Overall, the findings of the phonetic properties of uptalk suggest that heritage bilinguals do not systematically distinguish the pitch excursion of uptalk in their two languages. Rather, the difference between their two languages is based on rise duration (i.e., longer rise in English uptalk than in Spanish uptalk). The rise duration of heritage bilinguals' uptalk was conditioned by their language learning experience; heritage bilinguals with more English experience and those with lower Spanish proficiency and use tend to produce uptalk with longer duration.

6. Discussion

The present study explores cross-linguistic influence of intonation, focusing on the production of uptalk by Spanish–English bilingual Mexican Americans in Southern California (i.e., heritage bilinguals). While uptalk is typically regarded as an intonational pattern of *Valley Girl* speech of Southern California English (Armstrong et al. 2015; Ritchart and Arvaniti 2014; Tyler 2015), there is a good deal of empirical evidence that uptalk is commonly used across English varieties (Asano et al. 2020; Bradford 1997; Di Gioacchino and Jessop 2010; Dorrington 2010; Fletcher 2005; Fletcher et al. 2005; Fletcher and Harrington 2001; Hirschberg and Ward 1995; McGregor and Paethorpe 2008; McLemore 1991; Shokeir 2008; Warren 2005), as well as in Spanish (Henriksen 2017; Holguin Mendoza 2011; Kim and Repiso-Puigdelliura 2021; Martínez-Gómez 2018; Vergara 2015; Willis 2010).

Given that uptalk occurs in both English and Spanish, the presence of uptalk in heritage bilinguals' Spanish does not provide enough support for transfer from English intonation. Rather, it would be informative to examine in what ways heritage bilinguals' uptalk differs from the uptalk of non-heritage Spanish varieties and whether the divergent patterns trace back to their own English. Previous studies have shown that heritage bilinguals use uptalk both in Spanish (Fought 2003; Kim and Repiso-Puigdelliura 2021; Zárate-Sánchez 2018) and in English (Asch and Brogan 2022; Fought 2003; Santa Ana and Bayley 2008). Moreover, while heritage bilinguals produce uptalk in Spanish with similar relative frequency as Spanish monolinguals, they exhibit English-like patterns that differ from those of Spanish monolinguals (Kim and Repiso-Puigdelliura 2021). On the other hand, uptalk in heritage bilinguals' English is comparable to that of Anglo English speakers in both frequency and phonetic implementation (Asch and Brogan 2022), suggesting that English-to-Spanish influence may be stronger than Spanish-to-English influence for heritage bilinguals.

This study attempts to fill a gap in the literature on cross-linguistic influence of intonation by investigating the uptalk patterns in the two languages of heritage bilinguals and by taking into account language learning experience to explain interspeaker variability. Following the L2 Intonation Learning Theory (LILt) (Mennen 2015), we propose that cross-linguistic influence in heritage bilinguals' uptalk can occur along multiple dimensions of intonation. Among the four dimensions recognized by the LILt (i.e., systemic,

frequency, realizational, and semantic), we focused on the systemic, the frequency, and the realizational dimensions of heritage bilinguals' uptalk. Below we summarize our findings (Section 6.1) and discuss cross-linguistic influence in heritage bilinguals' uptalk along the three dimensions (Section 6.2).

6.1. Summary of Findings

Our data showed that heritage bilinguals produced uptalk in both Spanish and English. The frequency of heritage bilinguals' uptalk did not systematically differ between their two languages. While individual uptalk rates varied across speakers (14.15–63.92% in English and 14.29–58.06% in Spanish), they were not conditioned by heritage bilinguals' language learning experience. That is, heritage bilinguals with more experience with English (i.e., earlier age of arrival to the US, earlier age of acquisition of English, higher self-rated English proficiency, and more education in English) were not the ones that produced more uptalk rises. These findings, in addition to the fact that heritage bilinguals use uptalk as frequently as monolingual speakers of Spanish (Kim and Repiso-Puigdelliura 2021) and English (Asch and Brogan 2022), confirm that uptalk is not an English-specific phenomenon. In other words, the presence and frequency of uptalk per se do not inform us about cross-linguistic influence of intonation between heritage bilinguals' Spanish and English.

We now turn to our findings regarding uptalk rises beginning in a non-final word (i.e., uptalk with IP-final deaccenting). Recall that, unlike Spanish monolinguals whose rise initiate within the last word of the intonational phrase (IP), heritage bilinguals' uptalk in some cases span multiple words (Kim and Repiso-Puigdelliura 2021), similar to what has been found in English (Britain and Newman 1992; Warren 2005; 2016, p. 32). To test whether this divergent pattern is transferred from English, we compared heritage bilinguals' production of uptalk with IP-final deaccenting in Spanish and in English. Our findings showed that the heritage bilinguals produced uptalk with IP-final deaccenting significantly more frequently in English (31.66%) than in Spanish (8.51%). This implies that heritage bilinguals are able to maintain the typological differences between Spanish and English prosodic structures. Language learning experience did not influence their use of uptalk with IP-final deaccenting; the interspeaker variability found in our data may be explained by factors not examined in the present study or it is simply idiosyncratic.

As for the phonetic realization of uptalk, we examined two acoustic properties of the rise: pitch excursion and rise duration. Our data showed that heritage bilinguals distinguished uptalk in their two languages based on rise duration, but not based on pitch excursion; their uptalk in English was produced with longer rise duration than their uptalk in Spanish. The individual variability in rise duration was conditioned by heritage bilinguals' experience with English (PC1) and by their Spanish proficiency and use (PC2). Education in Spanish (PC3) did not affect their rise duration. These findings suggest that heritage bilinguals make cross-linguistic distinction mainly in the duration of the rise, which is influenced by their experience with English and the local Spanish variety, not by their experience with standard Spanish.

6.2. Varying Degrees of Cross-Linguistic Influence along the Dimensions of Intonation

Our findings support the LILT's argument that cross-linguistic influence of intonation occurs along multiple dimensions (Mennen 2015). The heritage bilinguals in this study produced uptalk with IP-final deaccenting in both languages, but produced it more in English than in Spanish. That is, at least for this uptalk pattern, cross-linguistic influence from English to Spanish is likely to occur in the systemic dimension of heritage bilinguals' uptalk, but not so much in the frequency dimension. While IP-final deaccenting emerges in heritage bilinguals' uptalk in Spanish, heritage bilinguals seem to recognize that this is an English feature that is not allowed in Spanish and try to suppress it when producing uptalk in Spanish.

With respect to the realizational dimension of uptalk, the heritage bilinguals mainly used rise duration to distinguish the uptalk of their two languages (i.e., longer rise duration

in English than in Spanish), whereas they did not make any cross-linguistic distinction in pitch excursion. Long rise duration, especially in female speech, and small pitch excursion have been attested in Southern California English uptalk (Armstrong et al. 2015; Ritchart and Arvaniti 2014). While the phonetic properties of uptalk have not been investigated in Spanish as much as in English, what we can infer from our data is that heritage bilinguals do not associate the extent of pitch excursion (small or large) with any language. Studies have shown that heritage bilinguals produce a somewhat smaller pitch excursion in Spanish than Spanish monolinguals (Kim and Repiso-Puigdeliura 2021), whereas the pitch excursion of uptalk in heritage bilinguals' English does not systematically differ from that of Anglo English (Asch and Brogan 2022). This, together with the findings of our study, suggests that heritage bilinguals phonetically assimilate the pitch excursion of the uptalk in Spanish to the uptalk in English to the point that they no longer distinguish the two languages in this regard.

As for the rise duration of uptalk, our findings indicate that heritage bilinguals associate long rise duration with English uptalk and, importantly, long rise duration is more prone to emerge for individuals with more experience with English and less experience with the local Spanish variety. According to Putnam (2020), non-balanced bilinguals, similar to many heritage speakers, may fail to properly inhibit their more dominant language, especially in situations where the two languages are in competition for finite online resources, such as in the case of spontaneous speech production. In such situations, they would be pressured to select between representations that have similar and contrastive properties. Over time, properties that are shared between both languages can lead to restructuring in the grammar of the less dominant language to free up processing cost, whereas properties that contrast from the more dominant language has a better chance at survival (Putnam 2020). This line of reasoning is in accordance with current models of L2 speech learning (Best 1995; Best and Tyler 2007; Flege 1995; Flege and Bohn 2021; Mennen 2015), as presented in Section 1.

The asymmetry found between the frequency (i.e., uptalk with IP-final deaccenting) and the realizational dimensions (i.e., pitch excursion and rise duration) is noteworthy. While cross-linguistic distinction was observed in both the frequency of uptalk with IP-final deaccenting and the rise duration of uptalk, only in the latter, interspeaker variation was conditioned by individuals' language learning experience. In the case of pitch excursion, the heritage bilinguals did not make any cross-linguistic distinction; the uptalk in both languages resembled the English low rises characteristic of their region (i.e., Southern California). These findings imply that phonetic aspects are more prone to cross-linguistic influence at the individual level than the phonological aspects of intonation, consistent with the argument of Jun and Oh (2000). Perhaps for this reason, most support for cross-linguistic influence of intonation has been found in the realizational dimension (Mennen 2015). Bilinguals have only one vocal tract to produce an extensive set of speech sounds in their two languages (de Bot 1992) and, as a consequence, may experience difficulties in articulating sub-phonemic differences between the two languages. The separation of cross-linguistic differences is especially taxing for bilinguals when the same articulator(s) are used in the two languages. For instance, the distinction in pitch excursion is achieved primarily through fine adjustments of pitch, which involve laryngeal muscle activation that controls the stiffness and the tension of the vocal folds (Zhang 2016). Thus, for the sake of economy of online resources (Polinsky and Scontras 2020), it is likely that heritage bilinguals avoid making cross-linguistic phonetic differences as much as possible during spontaneous speech production if these differences do not lead to any change in meaning. Future research should conduct cross-linguistic analysis on the link between phonetic properties and meanings of uptalk in Spanish and in English. If the same meaning is realized differently between the two languages, it would be important to investigate whether phonetic assimilation is more prone to occur in such cases than in cases where cross-linguistic phonetic distinction leads to change in meaning.

Even if heritage bilinguals are found to be better at making cross-linguistic phonetic distinction that leads to different meanings, the phonetic properties of one language may still surface in the other. For instance, Queen (2006, 2012) demonstrated that Turkish heritage speakers in Germany produce a mix of Turkish and German phrase-final rises in both of their languages; apart from the normative low rise of German, the heritage bilinguals also employed rising pitch to indicate pragmatic prominence (e.g., emphasis and focus), which is generally used in Turkish. Queen (2006) interpreted this as a fusion of two intonational grammars into a single intonational grammar, within which the two rise patterns are contrasted. According to Queen (2012, p. 794), heritage bilinguals capitalize on the differences in their two intonational grammars, which “serve as conventionalized, strategic linguistic resources that speakers (and listeners) may use as cues to discourse structures and inference.” Although the heritage bilinguals in this study were able to distinguish the rise duration of uptalk in the two languages, they may employ a mix of Spanish-like short rises and English-like long rises for different purposes, which is shared across members of their speech community, similar to the case of Turkish heritage speakers in Germany (Queen 2006, 2012). To attest this, apart from understanding the form-function association of uptalk in Spanish and in English, it is important to demonstrate whether heritage bilinguals utilize the resources from both of their languages and whether such practice is recognizable to other heritage bilinguals with similar backgrounds.

7. Conclusions

Bilinguals’ two languages interact at multiple levels and intonation is no exception. Given the complexity and multidimensionality of intonation, cross-linguistic influence is expected to occur along different dimensions of intonation that interact with each other (Mennen 2015). While a good amount of work has been performed on L2 intonation, the intonation of heritage bilinguals has received relatively little attention. Heritage bilingualism offers bilingual contexts that are often left unnoticed in traditional L2 acquisition scenarios (e.g., transfer from L2 to L1 intonation, asymmetry between order of acquisition and language dominance). Given that many aspects of cross-linguistic influence are shared across bilinguals, the investigation of heritage bilinguals’ intonation will contribute to building robust models of bilingual intonation.

The present study attempts to fill a gap in the literature by comparing uptalk produced in Spanish and in English by Spanish heritage speakers in Southern California and by exploring whether individuals’ uptalk varies depending on their language learning experience. Our findings showed that heritage bilinguals produced uptalk with similar frequency between Spanish and English, confirming that uptalk is not English-specific. Consistent with the L2 Intonation Learning Theory (LILt) (Mennen 2015), the cross-linguistic influence of uptalk occurred along multiple dimensions of intonation. In the systemic dimension, the heritage bilinguals produced uptalk in Spanish with IP-final deaccenting, which is an English feature that has not been attested in non-heritage Spanish varieties. However, in the frequency dimension, they demonstrated significantly lower rates of uptalk with IP-final deaccenting in Spanish than in English. Heritage bilinguals’ overall success in separating their two languages in the frequency dimension implies that cross-linguistic influence occurs only to a small degree from English to Spanish in this dimension. With regard to the realizational dimension, the heritage bilinguals demonstrated either assimilation to English (i.e., pitch excursion) or individual variability conditioned by language learning experience (i.e., rise duration). In other words, English-to-Spanish influence appears to occur to a larger extent in the realizational dimension than in the frequency dimension of uptalk. The findings of this study suggest that different dimensions of intonation demonstrate varying degrees of cross-linguistic influence. Specifically, the phonetic aspects are more prone to change than the phonological aspects of intonation.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted according to the guidance of the Declaration of Helsinki, and approved by the Institutional Review Board of the University of California, Los Angeles (protocol code: 17-000417; approval date: 29 March 2017).

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ As a reviewer noted, proximity to the Spanish conversation task and/or heightened language awareness after the language survey may have an effect on participants' performance. We included this potential effect as a covariate in the statistical analysis.
- ² We used the `corr.test()` function in the `psych` package (Revelle 2022) in R (R Core Team 2021).

References

- Alvord, Scott. 2010. Miami Cuban Spanish declarative intonation. *Studies in Hispanic and Lusophone Linguistics* 3: 3–39. [CrossRef]
- Amengual, Mark. 2016. Acoustic correlates of the Spanish tap-trill contrast: Heritage and L2 Spanish speakers. *Heritage Language Journal* 13: 88–112. [CrossRef]
- Amengual, Mark. 2018. Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics* 69: 12–28. [CrossRef]
- Amengual, Mark. 2019. Type of early bilingualism and its effect on the acoustic realization of allophonic variants: Early sequential and simultaneous bilinguals. *International Journal of Bilingualism* 23: 954–70. [CrossRef]
- Aoyama, Katsura, and Susan G. Guion. 2007. Prosody in second language acquisition: Acoustic analyses of duration and F0 range. In *The Role of Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, pp. 281–97.
- Armstrong, Meghan, Page Piccinini, and Amanda Ritchart. 2015. The phonetics and distribution of non-question rises in two varieties of American English. Paper presented at 18th International Congress of Phonetic Sciences, Glasgow, Scotland, August 10–14; Edited by the Scottish Consortium for ICPHS 2015. Glasgow: The University of Glasgow, p. 0927.
- Asano, Yasuhiro, C. Yuan, Ann-Kathrin Grohe, Andrea Weber, Mark Antoniou, and Anne Cutler. 2020. Uptalk interpretation as a function of listening experience. Paper presented at Speech Prosody 2020 Conference, Tokyo, Japan, May 25–28; Edited by Nobuaki Minematsu. pp. 735–39.
- Asch, Madeline, and Franny D. Brogan. 2022. Uptalk in Chicano Southern California English. *American Speech*, 1–27. [CrossRef]
- Atterer, Michaela, and D. Robert Ladd. 2004. On the phonetics and phonology of 'segmental anchoring' of F0: Evidence from German. *Journal of Phonetics* 32: 177–97. [CrossRef]
- Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]
- Beckman, Mary, Manuel Díaz Campos, Julia T. McGory, and Terrell A. Morgan. 2002. Intonation across Spanish in the Tones and Break Indices Framework. *Probus* 14: 9–36. [CrossRef]
- Benmamoun, Elabbas, Silvina Montrul, and Maria Polinsky. 2013. Defining an "ideal" heritage speaker: Theoretical and methodological challenges. Reply to peer commentaries. *Theoretical Linguistics* 39: 259–94. [CrossRef]
- Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 107–26.
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception. In *Language Experience in Second Language Learning: In honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murry J. Munro. Amsterdam: John Benjamins, pp. 13–44.
- Birdsong, David, Libby M. Gertken, and Mark Amengual. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin: University of Texas at Austin. Available online: <https://sites.la.utexas.edu/bilingual/> (accessed on 1 November 2017).
- Birdsong, David. 2016. Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Edited by Carmen Silva-Corvalán and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, pp. 85–10.
- Boersma, Paul, and David Weenink. 2021. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.54. Available online: <http://www.praat.org/> (accessed on 9 October 2021).
- Bradford, Barbara. 1997. Upspeak in British English. *English Today* 13: 29–36. [CrossRef]
- Britain, David, and John Newman. 1992. High rising terminals in New Zealand English. *Journal of the International Phonetic Association* 22: 1–11. [CrossRef]

- Buck, Gavin. 2016. *Systematic Transfer of ‘Uptalk’ from English (L1) to Spanish (L2) for Native English Speakers: A Developmental Study*. Undergraduate. Senior thesis, University of Michigan, Ann Arbor, MI, USA.
- Bullock, Barbara E. 2009. Prosody in contact in French: A case study from a heritage variety in the USA. *International Journal of Bilingualism* 13: 165–94. [CrossRef]
- Chang, Charles B., Erin F. Haynes, Yao Yao, and Russell Rhodes. 2010. The phonetic space of phonological categories in heritage speakers of Mandarin. In *Proceedings from the 44th Annual Meeting of the Chicago Linguistic Society: The Main Session*. Edited by Max Bane, Juan Jose Bueno, Thomas Grano, April Grotberg and Yaron McNabb. Chicago: Chicago Linguistic Society, pp. 31–45.
- Chang, Charles B., Yao Yao, Erin F. Haynes, and Russell Rhodes. 2011. Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America* 129: 3964. [CrossRef] [PubMed]
- Chen, Sally, and Janice Fon. 2008. The peak alignment of prenuclear and nuclear accents among advanced L2 English learners. Paper presented at Speech Prosody 2008 Conference, Campinas, Brazil, May 6–9; Edited by Plínio A. Barbosa, Sandra Madureira and Cesar Reis. Campinas: State University of Campinas, pp. 643–46.
- Colantoni, Laura, Alejandro Cuza, and Natalia Mazzaro. 2016. Task related effects in the prosody of Spanish heritage speakers. In *Intonational Grammar in Ibero-Romance: Approaches Across Linguistic Subfields*. Edited by Meghan Armstrong, Nicholas Henriksen and Maria Vanrell. Amsterdam: John Benjamins, pp. 1–24.
- Cruttenden, Alan. 1981. Falls and rises: Meanings and universals. *Journal of Linguistics* 17: 77–91. [CrossRef]
- Cruttenden, Alan. 1995. Rises in English. In *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O’Connor*. Edited by Jack W. Lewis. London: Routledge, pp. 155–73.
- Cruttenden, Alan. 2007. Intonational diglossia: A case study of Glasgow. *Journal of the International Phonetic Association* 37: 257–74. [CrossRef]
- Davies, Mark. 2006. *A Frequency Dictionary of Spanish: Core Vocabulary for Learners*. New York and London: Routledge.
- de Bot, Kees. 1992. A bilingual production model: Levelt’s ‘speaking’ model adapted. *Applied Linguistics* 13: 1–24.
- De la Mota, Carme, Pedro Martín Butragueño, and Pilar Prieto. 2010. Mexican Spanish intonation. In *Transcription of Intonation of the Spanish Language*. Edited by Pilar Prieto and Paolo Roseano. München: Lincom, pp. 319–50.
- de Leeuw, Esther, Ineke Mennen, and James M. Scobbie. 2012. Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism* 16: 101–16. [CrossRef]
- Dehé, Nicole. 2018. The Intonation of Polar Questions in North America (“Heritage”) Icelandic. *Journal of Germanic Linguistics* 30: 213–59. [CrossRef]
- Di Gioacchino, Martina, and Lorena Crook Jessop. 2010. Uptalk-towards a quantitative analysis. *Toronto Working Papers in Linguistics* 33: 1–16.
- Dorrington, Nichola. 2010. ‘Speaking Up’: A Comparative Investigation into the Onset of Uptalk in General South African English. BA(Honours) thesis, Rhodes University, Grahamstown, South Africa.
- Estebas-Vilaplana, Eva, and Pilar Prieto. 2010. Castilian Spanish intonation. In *Transcription of Intonation of the Spanish Language*. Edited by Pilar Prieto and Paolo Roseano. München: Lincom, pp. 17–48.
- Face, Timothy. 2003. Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech. *Catalan Journal of Linguistics* 2: 115–31. [CrossRef]
- Field, Andy. 2009. *Discovering Statistics Using SPSS*, 3rd ed. London: Sage.
- Flege, James E. 1995. Second-language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issue in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 229–73.
- Flege, James E., and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Rtree Wayland. Cambridge: Cambridge University Press, pp. 3–83.
- Flege, James E., Carlo Schirru, and Ian R. A. MacKay. 2003. Interaction between the native and second language phonetic systems. *Speech Communication* 40: 467–91. [CrossRef]
- Fletcher, Janet, and Jonathan Harrington. 2001. High-rising terminals and fall-rise tunes in Australian English. *Phonetica* 58: 215–29. [CrossRef] [PubMed]
- Fletcher, Janet, Esther Grabe, and Paul Warren. 2005. Intonational variation in four dialects of English: The high rising tune. In *Prosodic Typology: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford: Oxford University Press, pp. 390–409.
- Fletcher, Janet. 2005. Compound rises and ‘uptalk’ in Spoken English. Paper presented at Interspeech 2005, Lisbon, Portugal, September 4–8; pp. 1381–84.
- Fought, Carmen. 2003. *Chicano English in Context*. Houndmills: Palgrave Macmillan.
- Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks: Sage.
- Grabe, Esther. 2004. Intonational variation in urban dialects of English spoken in the British Isles. In *Regional Variation in Intonation*. Edited by Peter Gilles and Jörg Peters. Tuebingen: Niemeyer, pp. 9–31.
- Graham, Calbert, and Brechtje Post. 2018. Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics* 66: 1–14. [CrossRef]
- Grice, Martine, James Sneed German, and Paul Warren. 2020. Intonation systems across varieties of English. In *The Oxford Handbook of Language Prosody*. Edited by Carlos Gussenhoven and Aoju Chen. Oxford: Oxford Handbooks, pp. 285–302.
- Grosjean, François. 1989. *Bilingual: Life and Reality*. Cambridge and London: Harvard University Press.

- Gut, Ulrike, and Stefanie Pillai. 2014. Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition* 36: 283–302. [CrossRef]
- Harris, Michael, Viola Miglio, and Stefan Th. Gries. 2014. Mexican and Chicano prosody: Differences related to information structure. Paper presented at 6th Annual Pronunciation in Second Language Learning and Teaching Conference, Santa Barbara, CA, USA, September 5–6; Edited by John Levis, Rania Mohammed, Manman Qian and Ziwei Zhou. Ames: Iowa State University, pp. 40–50.
- Henriksen, Nicholas, Kimberly L. Geeslin, and Erik W. Willis. 2010. The development of L2 Spanish intonation during a study abroad immersion program in León, Spain: Global contours and final boundary movements. *Studies in Hispanic and Lusophone Linguistics* 3: 113–62. [CrossRef]
- Henriksen, Nicholas. 2017. Non-question Rises in L2 Spanish: Data from the Study Abroad Context. Paper presented at the Hispanistentag 2017, Munich, Germany, March 31.
- Hirschberg, Julia, and Gregory Ward. 1995. The interpretation of the high-rise question contour in English. *Journal of Pragmatics* 24: 407–12. [CrossRef]
- Holguín Mendoza, Claudia. 2011. Language, Gender, and Identity Construction: Sociolinguistic Dynamics in the Borderlands. Ph.D. dissertation, University of Illinois at Urbana-Champaign, Champaign, IL, USA.
- Horgues, Celine. 2013. French learners of L2 English: Intonation boundaries and the marking of lexical stress. *Research in Language* 11: 41–56. [CrossRef]
- House, Jill. 2006. Constructing a context with intonation. *Journal of Pragmatics* 38: 1542–58. [CrossRef]
- Huang, Becky H., and Sun-Ah Jun. 2011. The effect of age on the acquisition of second language prosody. *Language and Speech* 54: 387–414. [CrossRef]
- Jilka, Matthias. 2000. The Contribution of Intonation to the Perception of Foreign Accent. Ph.D. dissertation, University of Stuttgart, Stuttgart, Germany.
- Jolliffe, Ian T. 2002. *Principal Component Analysis*, 2nd ed. New York: Springer.
- Jun, Sun-Ah, and Mira Oh. 2000. Acquisition of second language intonation. In *Proceedings of International Conference on Spoken Language Processing*. Beijing: University of Beijing, pp. 76–79. vol. 4.
- Jun, Sun-Ah. 2000. K-ToBI (Korean ToBI) labelling conventions: Version 3. *The Korean Journal of Speech Science* 7: 143–169.
- Jun, Sun-Ah. 2014. Prosodic typology: By prominence type, word prosody, and macro-rhythm. In *Prosodic Typology II*. Edited by Sun-Ah Jun. Oxford: Oxford University Press, pp. 520–39.
- Kaiser, Henry F. 1960. The application of electronic computers to factor analysis. *Educational and Psychological Measurement* 20: 141–151. [CrossRef]
- Kaiser, Henry F. 1970. A second generation little jiffy. *Psychometrika* 35: 401–15. [CrossRef]
- Kaiser, Henry F. 1974. An index of factorial simplicity. *Psychometrika* 39: 31–36. [CrossRef]
- Kan, Rachel T. Y. 2021. Phonological production in young speakers of Cantonese as a heritage language. *Language and Speech* 64: 73–97. [CrossRef] [PubMed]
- Kelm, Orlando R. 1987. An acoustic study on the differences of contrastive emphasis between native and nonnative Spanish. *Hispania* 70: 627–33. [CrossRef]
- Kim, Ji Young, and Gemma Repiso-Puigdelliuera. 2021. Keeping a critical eye on majority language influence: The case of uptalk in heritage Spanish. *Languages* 6: 13. [CrossRef]
- Kim, Ji Young. 2016. The Perception and Production of Prominence in Spanish by Heritage Speakers and L2 Learners. Ph.D. dissertation, University of Illinois at Urbana-Champaign, Champaign, IL, USA.
- Kim, Ji Young. 2019. Heritage speakers' use of prosodic strategies in focus marking in Spanish. *International Journal of Bilingualism* 23: 986–1004. [CrossRef]
- Kim, Ji Young. 2020. Discrepancy between heritage speakers' use of suprasegmental cues in the perception and production of Spanish lexical stress. *Bilingualism: Language and Cognition* 23: 233–50. [CrossRef]
- Kissling, Elizabeth M. 2018. An exploratory study of heritage Spanish rhotics: Addressing methodological challenges of heritage language phonetics research. *Heritage Language Journal* 15: 25–70. [CrossRef]
- Kvavik, Karen H. 1979. An interpretation of cadences in Mexican Spanish. In *Colloquium on Spanish and Luso-Brazilian Linguistics*. Edited by James P. Lantolf, Fancine Wattman Frank and Jorge M. Guitart. Washington: Georgetown University, pp. 37–47.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
- Martín Butragueño, Pedro. 2004. Circumflex configurations in the intonation of Mexican Spanish. *Revista de Filología Española* 84: 347–73. [CrossRef]
- Martínez-Gómez, Rebeca. 2018. Fresa Style in Mexico: Sociolinguistic Stereotypes and the Variability of Social Meanings. Ph.D. dissertation, University of New Mexico, Albuquerque, New Mexico.
- Matluck, Joseph H. 1952. La pronunciación del español en el Valle de México. *Nueva Revista de Filología Española* 6: 109–20. [CrossRef]
- McCory, Julia T. 1997. Acquisition of Intonational Prominence in English by Seoul Korean and Mandarin Chinese Speakers. Ph.D. dissertation, Ohio State University, Columbus, OH, USA.
- McGregor, Jeannette, and Sallyanne Palethorpe. 2008. High rising tunes in Australian English: The communicative function of L* and H* pitch accent onsets. *Australian Journal of Linguistics* 28: 171–93. [CrossRef]
- McLemore, Cynthia Ann. 1991. The Pragmatic Interpretation of English Intonation: Sorority Speech. Ph.D. dissertation, University of Texas at Austin, Austin, TX, USA.

- Méndez Seijas, Jorge A. 2019. Second Language Spanish Intonation: Systemic and Realizational Dimensions of Its Acquisition. Ph.D. dissertation, Georgetown University, Washington, DC, USA.
- Mennen, Ineke, and Denise Chousi. 2018. Prosody in first-generation adult immigrants and second-generation heritage-language users: The timing of prenuclear rising accents. Paper presented at 9th International Conference on Speech Prosody, Poznań, Poland, June 13–16; pp. 828–32.
- Mennen, Ineke, Aoju Chen, and Fredrik Karlsson. 2010. Characterising the internal structure of learner intonation and its development over time. Paper presented at New Sounds 2010 6th International Symposium on the Acquisition of Second Language Speech, Poznań, Poland, May 1–3; Edited by Katarzyna Dziubalska-Kolaczyk, Magdalena Wrembel and Malgorzata Kul. Poznan: Adam Mickiewicz University, pp. 319–24.
- Mennen, Ineke, Felix Schaeffler, and Catherine Dickie. 2014. Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition* 36: 303–29. [CrossRef]
- Mennen, Ineke. 1999. Second Language Acquisition of Intonation: The Case of Dutch Near-Native Speakers of Greek. Ph.D. dissertation, University of Edinburgh, Edinburgh, UK.
- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63. [CrossRef]
- Mennen, Ineke. 2015. Beyond segments: Towards a L2 intonation learning theory. In *Prosody and Languages in Contact*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin: Springer, pp. 171–88.
- Montrul, Silvina, and Maria Polinsky. 2021. Introduction: Heritage languages, heritage speakers, heritage linguistics. In *The Cambridge Handbook of Heritage Languages and Linguistics*. Edited by Silvina Montrul and Maria Polinsky. Cambridge: Cambridge University Press, pp. 1–10.
- Montrul, Silvina. 2016. Dominance and proficiency in early and late bilingualism. In *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Edited by Carmen Silva-Corvalán and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, pp. 15–35.
- Nagano-Madsen, Yasuko. 2015a. Acquisition process of L2 Japanese intonation by Swedish learners—Interlanguage or prosodic transfer? Paper presented at 18th International Congress of Phonetic Sciences, Glasgow, UK, August 10–14; Edited by The Scottish Consortium for ICPHS 2015. Glasgow: The University of Glasgow, pp. 1–4.
- Nagano-Madsen, Yasuko. 2015b. Prosodic phrasing unique to the acquisition of L2 intonation—An analysis of L2 Japanese intonation by L1 Swedish learners. Paper presented at Interspeech 2015, Dresden, Germany, September 6–10; pp. 820–23.
- Nguyen, T. Anh Thu, C. L. John Ingram, and J. Rob Pensalfini. 2008. Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics* 36: 158–90. [CrossRef]
- Nibert, Holly J. 2006. The Acquisition of the phrase accent by beginning adult Learners of Spanish as a second language. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*. Edited by Manuel Diaz-Campos. Somerville: Cascadilla Proceedings Project, pp. 131–48.
- Nolan, Francis. 2003. Intonational Equivalence: An Experimental Evaluation of Pitch Scales. Paper presented at 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 3–9; Edited by Josep M. Solé, Daniel Recasens and Joaquin Romero. London: International Phonetic Association, pp. 771–74.
- O'Brien, Mary, and Ulrike Gut. 2010. Phonological and phonetic realisation of different types of focus in L2 speech. In *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*. Edited by Katarzyna Dziubalska-Kolaczyk, Magdalena Wrembel and Malgorzata Kul. Frankfurt: Peter Lang, pp. 205–15.
- Oh, Janet S., Suh-Ah Jun, Leah M. Knightly, and Terry Kit-fong Au. 2003. Holding on to childhood language memory. *Cognition* 86: B53–B64. [CrossRef]
- Ortega-Llebaria, Marta, and Laura Colantoni. 2014. L2 English intonation. *Studies in Second Language Acquisition* 36: 331–53. [CrossRef]
- Pépiot, Erwan. 2014. Male and female speech: A study of mean f0, f0 range, phonation type and speech rate in Parisian French and American English speakers. Paper presented at 7th International Conference on Speech Prosody, Dublin, Ireland, May 20–23; Edited by Nick Campbell, Dafydd Gibbon and Daniel Hirst. Piscataway: International Speech Communication Association (ISCA) Online Archive, pp. 305–9.
- Polinsky, Maria, and Gregory Scontras. 2020. Understanding heritage languages. *Bilingualism: Language and Cognition* 23: 4–20. [CrossRef]
- Polinsky, Maria, and Olga Kagan. 2007. Heritage Languages: In The 'Wild' and in the Classroom. *Language and Linguistics Compass* 1: 368–95. [CrossRef]
- Prieto, Pilar, and Paolo Roseano. 2010. *Transcription of Intonation of the Spanish Language*. Munich: Lincom Europa.
- Putnam, Michael T. 2020. Separating vs. shrinking. *Bilingualism: Language and Cognition* 23: 41–42. [CrossRef]
- Queen, Robin M. 2001. Bilingual Intonation Patterns: Evidence of Language Change from Turkish-German. *Language in Society* 30: 55–80. [CrossRef]
- Queen, Robin M. 2006. Phrase-final intonation in narratives told by Turkish-German bilinguals. *International Journal of Bilingualism* 10: 153–78. [CrossRef]
- Queen, Robin M. 2012. Turkish-German bilinguals and their intonation: Triangulating evidence about contact-induced language change. *Language* 8: 791–816. [CrossRef]
- R Core Team. 2021. R: A Language and Environment for Statistical Computing. Available online: <https://www.r-project.org/> (accessed on 10 August 2021).

- Rao, Rajiv, Ting Ye, and Brianna Butera. 2022. The prosodic expression of sarcasm vs. sincerity by heritage speakers of Spanish. *Languages* 7: 17. [CrossRef]
- Rao, Rajiv, Zuzanna Fuchs, María Polinsky, and María Luisa Parra. 2020. The sound pattern of Heritage Spanish: An exploratory study on the effects of a classroom experience. *Languages* 5: 72. [CrossRef]
- Rao, Rajiv. 2009. Deaccenting in spontaneous speech in Barcelona Spanish. *Studies in Hispanic and Lusophone Linguistics* 2: 31–75. [CrossRef]
- Rao, Rajiv. 2014. On the status of the phoneme /b/ in heritage speakers of Spanish. *Sintagma* 26: 37–54.
- Rao, Rajiv. 2015. Manifestations of /bdg/ in heritage speakers of Spanish. *Heritage Language Journal* 12: 48–74. [CrossRef]
- Rao, Rajiv. 2016. On the nuclear intonational phonology of heritage speakers of Spanish. In *Advances in Spanish as a Heritage Language*. Edited by Diego Pascual y Cabo. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 51–80.
- Rasier, Laurent, and Philippe Hilgsmann. 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux Cahiers de Linguistique Française* 28: 41–66.
- Repiso-Puigdeliura, Gemma. 2021. Empty onset repairs in the semi-spontaneous speech of Spanish child and adult heritage speakers. *International Journal of Bilingualism* 25: 1311–26. [CrossRef]
- Revelle, William. 2022. psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. Version = 2.2.5. Available online: <https://CRAN.R-project.org/package=psych> (accessed on 10 November 2022).
- Ritchart, Amanda, and Amalia Arvaniti. 2014. The form and use of uptalk in Southern Californian English. Paper presented at the 7th International Conference on Speech Prosody, Dublin, Ireland, May 20–23; Edited by Nick Campbell, Dafydd Gibbon and Daniel Hirst. Piscataway: International Speech Communication Association (ISCA) Online Archive, pp. 20–23.
- Ritchart, Amanda. 2014. Uptalk in Southern California. *Huffington Post*. January 10. Available online: http://www.huffingtonpost.com/amanda-ritchart/uptalk-in-southern-california_b_4573004.html (accessed on 17 October 2014).
- Robles-Puente, Sergio. 2014. Prosody in Contact: Spanish in Los Angeles. Ph.D. dissertation, University of Southern California, Los Angeles, CA, USA.
- Robles-Puente, Sergio. 2019. Transferencia y erosión de patrones melódicos en bilingües de español e inglés en California. *Boletín de Filología* 54: 283–306. [CrossRef]
- Rodríguez, Nicole. 2021. The Perception and Production of Lexical Stress among Early Spanish-English Bilingual Children. Ph.D. dissertation, Rutgers University, New Brunswick, NJ, USA.
- Ronquest, Rebecca. 2012. An Acoustic Analysis of Heritage Spanish Vowels. Ph.D. dissertation, Indiana University, Bloomington, IL, USA.
- Rothman, Jason. 2009. Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13: 155–63. [CrossRef]
- Saddah, Eman. 2011. The Production of Arabic Vowels by English L2 Learners and Heritage Speakers of Arabic. Ph.D. dissertation, University of Illinois at Urbana-Champaign, Champaign, IL, USA.
- Saito, Hiroko. 2006. Nuclear-rising placement by Japanese learners of English: Transfer from Japanese. In *Prosody and Syntax*. Edited by Yuji Kawaguchi, Iván Fónagy and Tsunekazu Moriguchi. Amsterdam: John Benjamins, pp. 125–39.
- Santa Ana, Otto, and Robert Bayley. 2008. Chicano English: Phonology. In *A Handbook of World Varieties of English. A Multimedia Reference Tool. Volumen 1: Phonology*. Edited by Bernd Kortmann and Edgar W. Schneider. Berlin and Boston: De Gruyter Mouton, pp. 1566–82.
- Santiago-Vargas, Fabian, and Élisabeth Delais-Roussarie. 2012. La prosodie des énoncés interrogatifs en français L2. In *Actes des Journées d'études sur la Parole JEP/TALN 2012*. Edited by Laurent Besacier, Benjamin Lecouteux and Gilles Sérasset. Grenoble: AFCP/ATALA, pp. 265–72.
- Shokeir, Vanessa. 2008. Evidence for the stable use of uptalk in South Ontario English. *University of Pennsylvania Working Papers in Linguistics* 14: 16–24.
- Silva-Corvalán, Carmen, and Jeanine Treffers-Daller. 2016. Digging into dominance: A closer look at language dominance in bilinguals. In *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Edited by Carmen Silva-Corvalán and Jeanine Treffers-Daller. Cambridge: Cambridge University Press, pp. 1–14.
- Simpson, Adrian P. 2009. Phonetic differences between male and female speech. *Language and Linguistics Compass* 3: 621–40. [CrossRef]
- Swerts, Marc, and Sabine Zerbian. 2010. Intonational differences between L1 and L2 English in South Africa. *Phonetica* 67: 127–46. [CrossRef] [PubMed]
- Szekely, Anna, Thomas Jacobsen, Simona D'Amico, Antonella Devescovi, Elena Andonova, Daniel Herron, Ching Ching Lu, Thomas Pechmann, Csaba Pléh, Nicole Wicha, and et al. 2004. A new on-line resource for psycholinguistic studies. *Journal of Memory and Language* 51: 247–50. [CrossRef]
- Tomlinson, John M., and Jean E. Fox Tree. 2011. Listeners' comprehension of uptalk in spontaneous speech. *Cognition* 119: 58–69. [CrossRef]
- Trimble, John C. 2013. Acquiring Variable L2 Spanish Intonation in a Study Abroad Context. Ph.D. dissertation, University of Minnesota, Minneapolis, MN, USA.
- Trofimovich, Pavel, and Wendy Baker. 2006. Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30. [CrossRef]

- Turco, Giuseppina, Christine Dimroth, and Bettina Braun. 2015. Prosodic and lexical marking of contrast in L2 Italian. *Second Language Research* 31: 465–91. [CrossRef]
- Tyler, Joseph C. 2015. Expanding and mapping the indexical field: Rising pitch, the uptalk stereotype, and perceptual variation. *Journal of English Linguistics* 43: 284–310. [CrossRef]
- Ueyama, Motoko, and Sun-Ah. Jun. 1998. Focus realization in Japanese English and Korean English intonation. *Japanese-Korean Linguistics* 7: 629–45.
- Vergara, Daniel. 2015. Uptalk in Spanish dating shows. *University of Pennsylvania Working Papers in Linguistics* 21: 187–96.
- Warren, Paul. 2005. Patterns of late rising in New Zealand English. *Language Variation and Change* 17: 209–30. [CrossRef]
- Warren, Paul. 2016. *Uptalk: The Phenomenon of Rising Intonation*. Cambridge: Cambridge University Press.
- Willems, Nico. 1982. *English Intonation from a Dutch Point of View*. Dordrecht: Foris.
- Willis, Erik. 2010. Dominican Spanish intonation. In *Transcription of Intonation of the Spanish Language*. Edited by Pilar Prieto and Paolo Roseano. Munich: Lincom, pp. 123–54.
- Yeni-Komshian, Grace H., James E. Flege, and Serena Liu. 2000. Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition* 3: 131–49. [CrossRef]
- Zárate-Sánchez, G. A. 2015. Perception and Production of Intonation among English-Spanish Bilingual Speakers at Different Proficiency Levels. Ph.D. dissertation, Georgetown University, Washington, DC, USA.
- Zárate-Sánchez, Germán. 2018. Production of final boundary tones in declarative utterances by English-speaking learners of Spanish. Paper presented at 9th International Conference on Speech Prosody, Poznań, Poland, June 13–16; Piscataway: International Speech Communication Association (ISCA) Online Archive, pp. 927–31.
- Zhang, Zhaoyan. 2016. Mechanics of human voice production and control. *The Journal of the Acoustical Society of America* 140: 2614–35. [CrossRef]
- Zuban, Yulia, Tamara Rathcke, and Sabine Zerbian. 2020. Intonation of yes-no questions by heritage speakers of Russian. Paper presented at 10th International Conference on Speech Prosody, Tokyo, Japan, May 24–28; Piscataway: International Speech Communication Association (ISCA) Online Archive, pp. 96–100.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Article

Plasticity of Native Intonation in the L1 of English Migrants to Austria

Ineke Mennen^{1,*}, Ulrich Reubold¹, Kerstin Endes¹ and Robert Mayr²¹ Department of English, University of Graz, 8010 Graz, Austria² Centre for Speech and Language Therapy and Hearing Science, Cardiff School of Health Sciences, Cardiff Metropolitan University, Cardiff CF5 2YB, UK

* Correspondence: ineke.mennen@unigraz.at

Abstract: This study examines the plasticity of native language intonation in English-Austrian German sequential bilinguals who have migrated to Austria in adulthood by comparing it to that of monolingual English and monolingual Austrian control speakers. Intonation was analysed along four intonation dimensions proposed by the L2 Intonation Learning theory (LILt): the inventory of categorical phonological elements ('systemic' dimension), their phonetic implementation ('realizational'), the meaning associated with phonological elements ('semantic'), and their frequency of use ('frequency'). This allowed us to test whether each intonation dimension is equally permeable to L2-on-L1 influences. The results revealed L2-on-L1 effects on each dimension. These consistently took the form of assimilation. The extent of assimilation appeared to depend on whether the cross-language differences were gradient or categorical, with the former predominantly resulting in intermediate merging and the latter in a complete transfer. The results suggest that native intonation remains plastic in all its dimensions, resulting in pervasive modifications towards the L2. Finally, in this first application of the LILt to the context of L1 attrition, the study confirms the model's suitability not only to acquisition of L2 intonation but also for predicting where modifications of L1 intonation are likely to occur.

Citation: Mennen, Ineke, Ulrich Reubold, Kerstin Endes, and Robert Mayr. 2022. Plasticity of Native Intonation in the L1 of English Migrants to Austria. *Languages* 7: 241. <https://doi.org/10.3390/languages7030241>

Academic Editors: Juana M. Licerias and Raquel Fernández Fuertes

Received: 26 April 2022

Accepted: 1 September 2022

Published: 16 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: speech plasticity; malleability of speech; phonetic attrition; intonation; L2 Intonation Learning theory (LILt); cross-language influences; transfer; late bilingualism; English; Austrian German

1. Introduction

Bilinguals are in the unique situation of regularly having to use two languages, a situation which is known to lead to cross-language interaction (e.g., Green 1998; Van Hell and Dijkstra 2002). At the phonetic level, it is well established that such instances of interaction will often lead to transfer from the native (L1) to the second language (L2), such that traces of the L1 are almost inevitably present in the pronunciation of the L2, particularly when the L2 was acquired after the age of puberty¹. Far less research attention has been given to the effect the L2 can have on speech patterns in the L1, even though this influence is equally plausible. Indeed, studies show that the extent of L2 influences on L1 pronunciation can lead to individuals being perceived as non-native in their mother tongue (Bergmann et al. 2016; de Leeuw et al. 2010; Hopp and Schmid 2013). The latter type of influence, and the one we focus on in this paper, is usually referred to as *phonetic attrition* or *L1 attrition of speech*, the non-pathological and non-age-related pronunciation changes that late sequential bilinguals who are being immersed in an L2 environment may experience in their L1 (de Leeuw et al. 2013; de Leeuw 2019a; Major 2010).

While an increasing number of studies has evidenced changes to the L1 of late sequential bilinguals in segmental areas of speech production (see de Leeuw 2019a for an overview), only a handful of studies (de Leeuw et al. 2012; de Leeuw 2019b; Gargiulo and Tronnier 2020; Mennen 2004; Mennen and Chousi 2018) have examined the effect the

L2 may have on prosodic areas of the L1. These suggest that prosodic effects also occur and that listeners base their judgements of non-nativeness in part on perceived prosodic changes, in particular intonational ones (Mayr et al. 2020)². Given the above-described lack of studies on phonetic attrition of prosodic areas of the L1, the present study will focus on one such aspect, namely intonation. It will do so by examining the extent of L1 intonational changes of English-Austrian German late sequential bilinguals who grew up in the UK as L1 speakers of English and emigrated to Austria in adulthood where they acquired (Austrian) German as their L2. In particular, the present study will be the first to apply a model that was originally developed to account for the difficulties L2 learners may experience when acquiring L2 intonation—L2 Intonation Learning theory (LILt, Mennen 2015)—to the context of L1 attrition. As we will see later in this paper, this allows for a more comprehensive investigation of potential changes to L1 intonation as hitherto seen.

1.1. Plasticity of Speech in Bilingual Contexts

Early investigations of cross-linguistic influences between the two sound systems of late sequential bilinguals started from the assumption that an L1 sound system that has reached biological maturity, is unlikely to be susceptible to influences from the L2 (Lado 1957; Lenneberg 1967). This view of non-plasticity of an individual's native language sound system resulted from the prevailing influence of the critical period hypothesis (Lenneberg 1967; Penfield and Roberts 1959), which holds that while children will have no problem acquiring their L1 within the time period of brain maturation occurring during adolescence, these maturational processes constrain the ability to acquire an L2. While L1-to-L2 transfer was therefore expected to occur, the L1 was thought to be protected against any L2 influences once it had reached neural maturity. As a consequence, the focus of research was on unidirectional L1-to-L2 influences, and L2-to-L1 influences were largely ignored. More recent studies, however, show that an individual's native sound system is not as impermeable to L2 influences as previously assumed (see Flege 1995 for an overview, and below) and it is now widely acknowledged that bidirectional influences are not unusual but a logical consequence of the constant interaction or co-activation of a bilingual's two languages (Flege 1987; Odlin 1989, 2006; Sharwood Smith and Kellerman 1986). This view of the plasticity of both the L2 and the L1 is reflected in one of the most influential models on L2 acquisition of speech, the Speech Learning Model (SLM) (Flege 1995) and its recently revised version, SLM-r (Flege and Bohn 2021), which posit that bidirectional influences are expected to occur because the L1 and L2 share a common phonetic space³.

There is now an abundance of evidence for the plasticity of L1 speech in late sequential bilinguals. Such changes to the L1 have been observed in situations where a bilingual's two languages are active for a restricted period of time, leading to temporary drifts in L1 speech. Such short-term L2-induced influences on the L1 are typically referred to as instances of *gestural or phonetic drift*, rather than phonetic attrition (Chang 2012, 2013; Sancier and Fowler 1997). Examples are situations where (novice) foreign language learners receive, sometimes intensive, language instruction (Chang 2012, 2013, 2019; Dmitrieva et al. 2020; Kartushina et al. 2016; Osborne and Simonet 2021), where bilinguals regularly change their linguistic environment by moving between an L1 and L2-speaking country (Sancier and Fowler 1997; Tobin et al. 2017), or where intensive code-switching is observed (Reubold et al. 2021). These studies show a relatively subtle restructuring in segmental areas of L1 pronunciation, which is thought to be fully (Kartushina and Martin 2019) or partially (Chang 2019) reversible, and "may be a precursor to more persistent changes that may become apparent over time" (Reubold et al. 2021, p. 20).

Changes in L1 pronunciation have also been observed in situations where a bilingual's two languages are more permanently activated, i.e., in experienced L2 learners who have migrated to another country and have been long-term (or even permanently) immersed in an L2 environment. Phonetic attrition of this kind appears to be common in most (but not all) late sequential bilinguals, with documented changes in L1 pronunciation affecting a wide range of segmental areas of L1 production, at least in the L1-L2 combinations

investigated so far ((Alharbi et al. forthcoming) for L1 Arabic-L2 English and L1 English-L2 Arabic; (Bergmann et al. 2016) for L1 German-L2 North-American English; (de Leeuw et al. 2013) for L1 German-L2 North-American English; (de Leeuw et al. 2018a) for L1 Albanian-L2 British English; (de Leeuw 2019b) for L1 German-L2 American English; (Guion 2003) for L1 Quichua-L2 Spanish; (Flege 1987) for L1 French-L2 American English and L1 American English-L2 French; (Kornder and Mennen 2021) for L1 Austrian German-L2 American English; (Major 1992) for L1 American English-L2 Brazilian Portuguese; (Mayr et al. 2012) for L1 Dutch-L2 English; (Mayr et al. 2020) for L1 Spanish-L2 British English; (Stoehr et al. 2017) for L1 Dutch-L2 German and L1 German-L2 Dutch; (Ulbrich and Ordin 2014) for L1 German-L2 Belfast English). Modifications to prosody and intonation have also been observed, although they have received far less research attention. With the exception of Gargiulo and Tronnier (2020), who investigated the use of prosodic cues to pronominal anaphora resolution, all studies focused on L2-induced changes to L1 intonation ((de Leeuw et al. 2012) for L1 German-L2 English; (de Leeuw 2019b) for L1 German-L2 English; (Mennen 2004) for L1 Dutch-L2 Greek; and (Mennen and Chousi 2018) for L1 Greek-L2 Austrian German). These studies all investigated just one particular aspect of intonation, i.e., tonal alignment (i.e., how the start or end of pitch rises are coordinated in time with segments), and showed a change in tonal alignment patterns in the L1 under the influence of the L2. However, L2-induced changes in L1 intonation are unlikely to be restricted to just aspects of its phonetic realization. The current study will therefore investigate L2-induced modifications in L1 intonation along the four intonation dimensions proposed by Mennen's (2015) L2 Intonation Learning theory (LILt), as explained later in this paper. This approach will ensure that L1 attrition in intonation is investigated in a more comprehensive and theoretically motivated way.

1.2. Approaches to Intonational Description

1.2.1. The Autosegmental-Metrical Model of Intonation

Intonation is said to be particularly susceptible to cross-language influences (Mackey 2000), yet the focus of most research studies on L1 attrition of speech has been on segments rather than intonation. A likely reason for this lack of research may be that intonation poses a particular challenge for researchers given the fact that it interacts with other prosodic aspects, like for instance tempo, rhythm, and loudness (e.g., Nolan 2006) and it is difficult—more so than in segments—to separate influences that are categorical from those that are gradient (Ladd 1996). It has been argued (Mennen 2004, 2007, 2015), however, that this is an important distinction to make, as cross-language influences may differ depending on whether they concern categorical (phonological) aspects of intonation or whether the aspects are gradient (phonetic). The few studies on cross-language influences in intonation suggest that gradient aspects may be more vulnerable to cross-language influences than categorical elements (Graham and Post 2018; Jun and Oh 2000; Mennen et al. 2010; Sanchez 2020). Thanks to the advent of the Autosegmental Metrical (AM) framework it has become more feasible to consider both types of influences in intonation, as it provides the tools to separate categorical phonological elements of intonation from the phonetic nature of their implementation. While the AM theory originates from Pierrehumbert's (1980) intonational description of American English, a series of language-specific annotation systems for many other languages has been derived from it (see Jun 2005, 2014, for overviews), and it has now become the most dominant approach to intonational description (Ladd 2000).

In the AM approach, the intonation of an utterance is presented phonologically as a sequence of high (H) and low (L) tones which are internally structured into pitch accents (when they associate with metrically prominent syllables) or boundary tones (when they associate with the edges of phrases). Pitch accents, often referred to as 'starred' tones because of their notation with an asterisk (*), can be monotonal (L* or H*) or bitonal (e.g., LH*, L*H, or H*L, where the asterisk indicates the most prominent tone within the accented syllable). Boundary tones describe the L or H tones at the beginning or end of an intonational phrase (e.g., H%) or an intermediate phrase (e.g., H-). Phonetically,

intonation is represented by the phonetic shape of the phonological categories, i.e., how phonological categories are phonetically realized in terms of, for instance, their height or timing. That is, the same phonological category (e.g., L*H or H*L) may be realized differently in different languages and dialects. Similarly, languages and dialects also differ in the inventory, complexity, and distribution of categorical phonological elements (see Jun 2005, 2014, for an overview). A more detailed description of the AM notations used in our study is given in Section 2.3.

1.2.2. The L2 Intonation Learning theory (LILt)

It has long been established that late sequential bilinguals who are long-term immersed in an L2 environment experience difficulties with the acquisition of L2 intonation, and often transfer elements of L1 intonation to the L2. Mennen (2015) proposed a model—the L2 Intonation Learning theory (LILt)—with roots in the AM approach, in order to account for and predict the difficulties learners may have in producing L2 intonation. The model is based on the premise that cross-language influences in intonation may occur along four dimensions (modified from Ladd 1996)⁴. These are:

1. The *systemic* dimension (the inventory and distribution of categorical phonological elements of intonation, such as boundary tones and pitch accents);
2. The *realizational* dimension (the phonetic implementation of these intonational primitives);
3. The *semantic* dimension (the functionality of the categorical elements or tunes, i.e., how they are used to signal meaning);
4. The *frequency* dimension (the frequency of use of the categorical elements).

The *systemic* dimension comprises the categorical or phonological elements of intonation, i.e., the intonational primitives, which can differ between languages and be a source of cross-language influences. An example of a cross-language difference on this dimension is the so-called ‘early peak’ (H!H*L), which has been reported for nuclear accents in German German (Féry 1993; Peters 2018) and Austrian German (Schmid and Moosmüller 2013; Ulbrich 2005) but not in British English (Grabe 2004). The term early peak is used to describe an intonation contour where the pitch maximum is reached on a metrically weak syllable immediately *preceding* the accented syllable. The accented syllable itself is falling or low. Figure 1 gives a schematic representation of how an early peak (H!H*L) may look like and how it contrasts with a falling pitch accent (H*L) where the peak occurs *on* the accented syllable⁵. In Austrian German early peaks are said to occur in conditions of narrow contrastive focus (Schmid and Moosmüller 2013; Moosmüller et al. 2015), where females were observed to use it more often than males (Schmid and Moosmüller 2013). However, such a gender-related preference may be restricted to narrow contrastive focus only, as it has not been reported in larger studies on Austrian German examining other contexts (Moosmüller et al. 2015; Ulbrich 2005). Languages or language varieties can also differ in the boundary tones they use, with some languages using complex boundary tones (such as LH% or HL%), others using simple low or high boundary tones at the start or end of intonation phrases, and some languages such as Mandarin sometimes omitting final boundary tones (see Jun 2005, 2014, for overviews).

The *realizational* dimension comprises the gradient or phonetic elements of intonation, i.e., how the intonational primitives such as pitch accents and boundary tones are phonetically realized. Cross-language differences on this dimension typically involve how pitch accents are lined up (‘aligned’) with segments in time (i.e., whether they occur early or late in a prominent syllable), the extent to which pitch accents are truncated at the utterance end (i.e., whether they are fully realized or ‘cut off’ when there is little voiced material available to realize falling or rising pitch accents), or what their relative height (‘scaling’) is within an individual’s pitch range. For instance, the languages in the current study display differences in alignment patterns and overall pitch range but are relatively similar in the extent to which pitch accents are realized at the utterance end. That is, in pre-nuclear rising pitch accents in statements, speakers of Standard Southern British English (SSBE) typically

show a rise in pitch that begins close to the onset of the accented syllable (Ladd et al. 1999). In contrast, speakers of Austrian German begin prenuclear rises considerably later, i.e., well within the stressed vowel (Mennen and Chousi 2018). As for pitch range, speakers of SSBE typically deploy a wider pitch range than speakers of German German (Mennen et al. 2012) who, in turn, are found not to differ from speakers of Austrian German (Ulbrich 2005). Hence it can be concluded that speakers of Austrian German tend to use a narrower pitch range than speakers of SSBE. As for truncation patterns, both SSBE and Austrian German are found to compress rising and falling pitch patterns under time pressure (Siddins and Mennen 2019). Boundary tones may also differ in how they are cross-linguistically realized. For instance, Willems (1982) found that native speakers of British English realize the initial boundary tones at the start of their intonation phrases on a mid-level pitch, whereas native Dutch speakers start their intonation phrases on a low-level pitch. The *semantic* dimension is concerned with the use of categorical elements of intonation to convey meaning. For instance, languages may differ in how they mark informational and contrastive focus. In some languages (e.g., Germanic languages), focus is signalled by accenting new and contrastive information, while deaccenting given information (Nooteboom and Terken 1982). In other languages (e.g., Spanish), no intonational distinction is made between utterances with broad (where focus is on the whole phrase or sentence) and narrow focus (where the focus is on one part of the phrase or sentence), and the nuclear pitch accent⁶ is always placed at the end of the intonational phrase (Hualde 2005). Japanese and Korean, on the other hand, signal focus by placing a boundary tone before or after the word in focus and deaccenting everything that follows (cf. Jun 2014). With respect to the two languages in the current study, previous research suggests that English and German differ in how they signal sentence-internal continuation, with German, including Austrian German (Moosmüller et al. 2015), favouring a rising pitch accent (L*H) and English speakers, including speakers of SSBE, typically employing a falling (H*L) pitch accent (see Chen 2007, for an overview). Finally, the *frequency* dimension concerns the frequency with which a specific intonation category is used in a particular language or dialect. For instance, while English and German both have rising and falling pitch accents in their respective inventory, the latter is used considerably more frequently in English than in German (Mennen et al. 2012, p. 2258)⁷. This is also the case for the language varieties in the current study, with Austrian German speakers using rises more frequently, at least in statements (Moosmüller et al. 2015), than SSBE speakers, who favour the use of falls (Mennen et al. 2012). Cross-language differences have also been observed in the frequency of use of boundary tones. For instance, a higher frequency of high boundary tones (H%), used also in utterances that are not intended as questions, is found in some varieties of English, particularly in younger generations (including Australian English, like New Zealand English, Belfast English or Glaswegian English) than in other varieties of English (e.g., Cruttenden 1997).

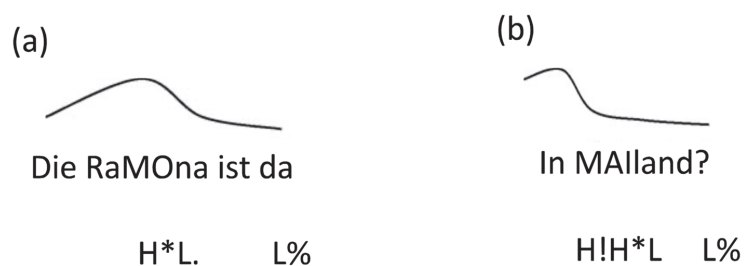


Figure 1. Schematic contour of the sentence (a) “Ramona is there” showing a falling (H*L) pitch accent with the peak on the accented syllable; and (b) “In Milan?” showing an early peak on the weak syllable ‘in’ before the accented syllable with low pitch. Capitals indicate accented syllables.

Drawing parallels to models of segmental learning, in particular the SLM (Flege 1995) and SLM-r (Flege and Bohn 2021), and based on previous findings (e.g., Atterer and Ladd 2004; de Leeuw et al. 2012; Mennen 2004, 2007; Mennen et al. 2010; Mennen et al. 2014), the LILt formulates a number of assumptions and hypotheses, which, in turn, generate testable predictions. While the LILt predominantly focuses on L1-to-L2 influences in intonation, it also allows for an explanation of L2-on-L1 influences. In particular, it assumes that a bilingual's L1 and L2 intonation systems are not entirely isolated but exist in a common space. This causes the intonation systems to interact with each other, which may result in bidirectional influences, such that L2-on-L1 effects are observed alongside L1-on-L2 effects (cf. Mennen 2015). Whether and where such influences are likely to occur depends to a large degree on the cross-language similarity in the various dimensions of intonation. According to the LILt, if an intonation category in the L2 is sufficiently different from any other L1 category already available in the L1, for instance when a pitch accent is part of the inventory of the L2 but not the L1, the chances that L2 learners will establish a new L2 category (i.e., chances of it being incorporated into their L2 inventory) are high. In such a case, the L2-on-L1 effect is likely to be completely absent. Alternatively, if a new L2 category is established, there may be a need for the new L2 category and already existing L1 categories to deflect away from each other in order to maintain contrast in a shared phonetic space. This could lead to an L2-on-L1 effect that is dissimilatory in nature. It is not entirely clear which factors guide the occurrence of the first or the latter scenario, i.e. which circumstances will lead to there being no effect on the L1 and which will lead to a shift of the L1 category to maintain contrast. The SLM offers "crowding" of the bilinguals' "combined L1-L2 phonetic space" when new L2 categories are added as a reason for the occurrence of the latter scenario (Flege 2002, p. 225). It is not specified though, neither by the SLM(-r) nor the LILt, at which point it becomes necessary "to augment inter-category distances in the common L1-L2 phonetic space of bilinguals" (Flege and Bohn 2021, p. 21), although one has to assume that when the L2 category is sufficiently different from any already existing categories in the shared phonetic space, there would be little need for dissimilation and thus the former scenario would be more likely.

If, on the other hand, the cross-language differences are gradient in nature, with differences in the phonetic implementation of the same intonational category, cross-language interaction is expected to occur and result in an assimilation or merging of L1 and L2 properties. This, in turn, may result in a shift of the L1 category towards the L2 category and the use of intermediate values somewhere between those found in the L1 and the L2. Cross-language influences may therefore not be equally pervasive on each dimension of intonation, with research suggesting that the realizational dimension may be more permeable to cross-language influences than the systemic dimension (Graham and Post 2018; Mennen 2007; Ueyama 1997). With regard to external factors, the LILt draws on models of segmental learning (Flege 1995; Flege and Bohn 2021) and suggests that factors such as—amongst others—the age of arrival (AoA) in an L2-speaking country, length of residence (LoR), or amount of L1 and L2 use, may play a role in the degree to which cross-language influences in intonation will be observed, although the evidence so far is extremely limited. The few studies that have explored the role of AoA in intonation suggest that there may be age effects on L2 intonation learning, with more successful acquisition of L2 intonation in learners who had arrived in the L2 environment at an earlier age (Chen and Fon 2008; Huang and Jun 2011; Mennen 2004). Similarly, studies suggest that experience with, and exposure to, the L2 may influence the degree of success, although acquisition of the various dimensions of intonation does not appear to proceed at the same rate (Graham and Post 2018; Jun and Oh 2000; Mennen et al. 2010, 2014; Trofimovich and Baker 2006).

Finally, although the LILt is a fairly recent working model that is subject to change when more data become available (Mennen 2015), a number of recent studies have shown its effectiveness in establishing cross-language similarity along the four dimensions of intonation, in predicting where cross-language influences are likely to occur, and whether such influences change under the influence of language experience and exposure (Albin 2015;

Busà and Stella 2015; Graham and Post 2018; Pešková 2020; Sanchez 2020; Schauffler 2021). However, as these studies all focused on the acquisition of L2 intonation, the effectiveness of examining these four dimensions of intonation for predicting where L2-on-L1 effects are likely to occur remains to be established.

1.3. Research Questions and Predictions

The main objective of this study is to arrive at a better understanding of the malleability of native language intonation in migrants who are being long-term immersed in an L2 environment. The first question posed in this study is whether L2-induced changes are observed in the intonation of late English-Austrian German sequential bilinguals by comparing their intonation patterns with those produced by monolingual SSBE speakers living in England and monolingual Austrian German speakers living in Austria. Based on the research studies reviewed above, we hypothesize that the late sequential bilinguals in our study will manifest L1 modifications of intonation due to L2 learning experience.

The second question posed is whether the L2-induced changes to L1 intonation are evidenced in each of the four dimensions of intonation, or whether some dimensions or sentence types are more permeable to L2 influences than others. As we saw in the research reviewed earlier, cross-language influences are only expected to occur when cross-linguistic differences exist between the L2 learners' two languages. This is the case for each dimension of intonation of the languages examined in our study. On the systemic dimension, SSBE and Austrian German are very similar in their respective inventories of pitch accent categories, differing only in the so-called 'early peak, which is present in Austrian German (Schmid and Moosmüller 2013) but not in SSBE (Grabe 2004). There is a suggestion in the literature that early peaks are predominantly used by female speakers, although this gender-preference may be restricted to contexts of narrow contrastive focus (Schmid and Moosmüller 2013). On the realizational dimension, the research reviewed earlier suggests that Austrian German and SSBE differ in the pitch range habitually used by its speakers (wider in SSBE than in Austrian German), and the alignment of prenuclear rising accents in statements (earlier in SSBE than in Austrian German). Cross-language differences are also found on the frequency dimension, with Austrian German speakers using rises more frequently than SSBE speakers. On the semantic dimension, there are cross-language differences in the type of pitch accent used to indicate sentence-internal continuation, with SSBE speakers using predominantly falling pitch accents and Austrian German speakers preferring the use of a rising pitch accent to signal continuation. If these cross-language differences are confirmed in our study, L2-induced influences on the L1 should—in principle—be evidenced on each dimension of intonation (but see below for our expectations for the systemic dimension), at least for the pitch accents. Given that there is no previous literature available on cross-language differences between SSBE and Austrian German in boundary tones, we are not able to make any specific predictions here.

The third question our study addressed is whether any observed L1 modifications in intonation will take the form of assimilation (with L1 values that have shifted towards the L2 when compared to monolingual SSBE speakers), or dissimilation (with L1 values shifted away from both monolingual SSBE and Austrian German groups). In light of the cross-language differences and the previously discussed literature, we hypothesize that gradient differences will cause the L1 and L2 intonation systems to interact, resulting in assimilation, which in turn will result in intermediate values between the L1 and the L2. We therefore predict that on the realizational dimension, the bilingual speakers in our study will produce values for pitch range and alignment that are intermediate between the two monolingual groups. On the frequency level, where the cross-language differences are also gradient, we also expect to find intermediate frequencies of use of rising (L*H) and falling (H*L) pitch accents between those of the two monolingual groups. Similarly, on the semantic dimension cross-language differences are gradient in nature, with SSBE speakers showing a preference for a falling pitch accent and Austrian German speakers favoring a rising pitch accent to indicate sentence-internal continuation. We therefore predict that

the bilingual speakers in our study will show evidence of assimilation and start using L*H more frequently to signal continuation than monolingual SSBE speakers, showing intermediate values for the frequency of use of L*H in sentence-internal continuations between those found for the two monolingual groups. In contrast, if the cross-language differences are categorical, and the L2 category is sufficiently different from any other L1 category available, there is likely to be no effect of the L2 on the L1. Based on the assumptions of the SLM(r) and the LILt, we therefore predict that there will be no L1 modification in the systemic dimension, as there is no reason why the acquisition of a new L2 pitch accent that does not exist in the L1 would influence any of the existing L1 categories, unless there is a need to maintain contrast between the new L2 category and already existing L1 categories in a shared phonetic space (see [Flege 2002](#)). In the latter case, we would expect the L2-induced influence on the L1 to be dissimulatory in nature.

2. Materials and Methods

2.1. Participants

Three groups of adults participated in this study: (i) late sequential English–Austrian German bilinguals (BIL, $N = 8$, 4 females, 4 males); monolingual speakers of SSBE residing in England (SSBE, $N = 8$, 4 females, 4 males); and monolingual speakers of Austrian German residing in Austria (AUT, $N = 8$, 4 females, 4 males). The participants in the BIL group were all raised as monolingual speakers of SSBE who moved to Austria in adulthood and now reside in Austria where they acquired Austrian German as an L2. Their average age of arrival (AoA) in Austria is 32.4 years (range: 19 to 59), their average length of residence (LoR) in Austria is 17.3 years (range: 3 to 38). The bilingual speakers reported that they did not speak any foreign languages other than Austrian German on a daily basis or above high-school level.

We also obtained global foreign accent ratings (FARs) of selected speech samples (comprising the same sentences used in the current study, cf. Section 2.2) produced by the participants in the BIL group mixed with 3 of the monolingual SSBE control speakers, by asking 25 monolingual SSBE listeners not familiar with any varieties of German in an online rating experiment to decide whether a speaker in a given sample sounded native or not (binary decision), followed by an indication of how confident they were of their choice on a 3-point scale: uncertain, semi-certain, or certain. Together, this resulted in a 6-point foreign accent scale, ranging from “1” = “certainly native”, to “6” = “certainly non-native”. This two-staged rating is a commonly used method in studies on L1 attrition of speech (e.g., [Bergmann et al. 2016](#); [de Leeuw et al. 2010](#); [Mayr et al. 2020](#)). The ratings showed that the group of BILs were perceived as sounding significantly less native than the SSBE controls (confirmed by a cumulative link model for ordinal regression: $\chi^2[1] = 389.3$, $p < 0.001$), receiving average FARs of 2.8 and 1.2, respectively. This shows that, on average, the group of BIL speakers is perceived as moderately accented in their L1.

Participants in the monolingual groups formed our control groups. They are monolingual speakers of either SSBE or Standard Austrian German, and have never lived outside England or Austria, respectively. While they all have some knowledge of other languages, none of them reported more than high school level knowledge, and therefore can be considered “functional monolinguals” with little active knowledge or use of foreign languages ([Best and Tyler 2007](#), p. 16).

2.2. Speech Materials and Recordings

There were two sets of speech materials, one for English and one for German. Each set consisted of twelve neutral sentences with various grammatical structures, including statements (e.g., There is phenomenal interest in the products.), wh-questions (e.g., Where is the manual?), yes/no questions (e.g., Do you live in Ealing?), declarative questions (questions without inversion, e.g., You live in Ealing?), and sentences containing sentence-internal continuation (e.g., Do you like Malaga or Malta best?). In order to ensure—as much as possible—a smooth fundamental frequency (f_0) contour, care was taken to have

sonorants, or in a few cases voiced obstruents, flanking the stressed vowels of the words we expected to bear the pitch accents. The sentences from the English set came from a study on alignment patterns in prenuclear rises (Atterer and Ladd 2004, see further Section 2.4) or from the Intonational Variation in English (IViE) corpus (cf. Grabe 2004). The German set was specifically designed to match the English set as much as possible in syntactic structure, length, number, and distribution of content words, and expected place of pitch accents.

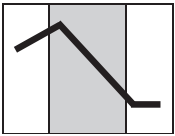
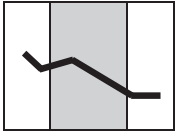
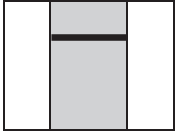
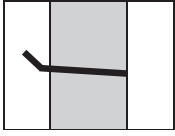
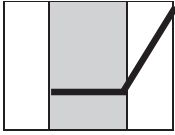
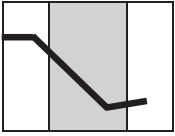
Participants were asked to read out two repetitions of each sentence in their respective L1s⁸. Due to contact restrictions during the COVID-19 pandemic, these recordings took place in the participants' own environment and using their own computer equipment. While this was not ideal (cf. Sanker et al. 2021), all recordings were carefully checked and, where necessary due to poor audio quality or misreading, participants were asked to re-record sentences. The latter was done just once, so as not to overburden the participants. In case the recordings still contained misreadings or were of poor audio quality, or the speaker failed to re-record the item, we discarded it, as happened in 4 cases. The two repetitions of each sentence were presented on the participants' computer monitor via WikiSpeech, an online tool designed to create web-based speech databases (Draxler and Jansch 2008). All sentences were presented in random order and interspersed with materials designed to test segmental changes to L1 speech, not reported here, with a 1.5 s pause between items. Thus, a total of 480 utterances (20 sentences \times 8 participants \times 3 groups) were elicited, of which we had to discard 4 (as described above). The remaining 476 utterances were annotated by hand using the same pool of tonal labels (as further explained in the following section), and generating a corpus of 2534 tonal labels, encompassing prenuclear and nuclear pitch accents and phrase-initial and phrase-final boundary tones, for subsequent analysis.

2.3. Intonational Description

Since our study compares intonation in different languages (Austrian German and SSBE) and in different groups (monolinguals versus bilinguals), it is essential to use the same system of intonational description in each comparison, as we may otherwise not be comparing like with like. As we have seen earlier, language-specific annotation systems have emerged that are grounded in the AM framework. These not only differ in their labelling conventions but are also often based on different underlying assumptions. The most crucial difference concerns assumptions about the left-headedness or right-headedness of bitonal pitch accents. Whereas left-headed systems see the pitch movement as starting on the accented syllable, and therefore account for the movement **from** the accented syllable onwards (e.g., Féry 1993 and Peters 2018, for German; Grabe 2004 and Grabe et al. 2000, for British English), right-headed systems see the movement **towards** an accented syllable as important (e.g., Baumann et al. 2000 for German; Beckman and Pierrehumbert 1986 and Beckman et al. 2005, for American English). These two approaches are sometimes respectively referred to as "off-ramp" versus "on-ramp" analyses (Gussenhoven 2004, pp. 127–28). Using two systems with different underlying assumptions in our study would unnecessarily complicate the comparison of the different languages and groups. We therefore decided to base the labels in our study largely on the tonal labels from Grabe's (2004) IViE system, which, in turn, is modified from Gussenhoven's (1983, 2004) left-headed approach to the description of English and Dutch intonation. This system was deemed particularly suitable because it has been extensively used in previous studies of intonational varieties of British English (Grabe 2004; Grabe et al. 2000) and German (Peters 2018) and has also successfully been used in a cross-language comparison of the two languages (Grabe 1998).

All data were thus transcribed using the same pool of labels, although not all labels were used for each language, sentence type, or participant group, as will become clear in the Section 3. These labels were found to suffice for a description of our data from the two languages and groups under investigation. A list of the labels that were used for pitch accents (panel a) and boundary tones (panel b) in our analysis is given in Table 1, along with a short description and schematic representation of their common shape.

Table 1. Labels used in our study, along with a description and schematic representation. Panel (a) lists the pitch accents and accent modifications. Panel (b) lists the boundary tones. The grey parts represent metrically strong (accented) syllables; the white parts represent unstressed syllables.

Panel (a): Pitch Accents		
Pitch Accents	Description of Commonly Observed Shape	Schematic Representation
H*L	High fall: a high tonal target on accented syllable followed by a low tonal target. The fall starts on the same or immediately following syllable	
!H*L	Downstepped fall: a high tonal target on accented syllable followed by a low tonal target, which is downstepped to a lower level compared to preceding high targets	
H*	High level: a high tonal target on accented syllable which remains high until the following high tonal target	
!H*	Downstepped high: a high tonal target on accented syllable which is downstepped to a lower level compared to preceding high-level targets. The contour remains at this level until the following high tonal target	
L*H	Low rise: a low tonal target on accented syllable followed by high tonal target. The rise takes place in the same or immediately following syllable.	
H!H*L	Early peak: a high tonal target that is associated with a metrically weak syllable immediately preceding the accented syllable. The accented syllable itself is falling or low.	
Panel (b): Boundary Tones		
Boundary Tones	Description	
%H or %L	High/low beginning of intonational phrase (IP)	
H% or L%	Rising/low ending of IP	

The intonation labelling was conducted by one main annotator, who is trained in IViE-style transcriptions. Annotations were inserted into Praat (Boersma and Weenink 2022) and were based on a combination of an auditory and visual inspection of the data, giving initial priority to auditory impressions. Intermediate phrase boundaries were determined on the existence of a pause, lengthening, or pitch reset, or a combination of these cues. After an annotation of the first set of repetitions of all sentences and speakers, a second annotator, also trained in intonation labelling, went through the annotated data, and identified possible disagreements. These were discussed and resolved, after which

the main annotator proceeded with annotating the second set of repetitions. In order to establish inter-annotator consistency, 35% of the second set of repetitions (given that the 2nd annotator had already seen the first set of repetitions) were annotated by a second annotator also trained in intonation labelling, after which inter-annotator agreement was calculated by means of Cohen’s κ (Cohen 1960). Agreement on the choice of tonal labels was 0.69, which corresponds according to Landis and Koch (1977) to a “substantial” agreement strength. In addition, the main annotator also re-labelled 12% of the data she had already annotated. Intra-rater agreement strength (Cohen’s κ : 0.92) on the choice of tonal events corresponds to an “almost perfect” agreement—again following Landis and Koch (1977). As these agreement levels are within the same order of magnitude as inter-rater and intra-rater agreement for other studies using AM based annotation systems (cf. Breen et al. 2012; Escudero et al. 2012; Yoon et al. 2004), we therefore proceeded with the labels provided by the main annotator. An example of our annotations is shown in Figure 2.

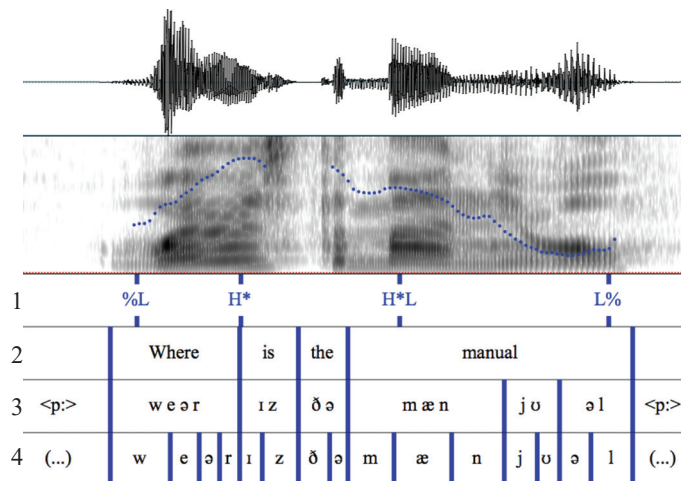


Figure 2. Example of a sentence produced by one of the SSBE speakers, annotated for intonation. Tier 1 shows the labels for boundary tones and pitch accents. Tier 2 shows the orthographic transcription. Tiers 3 and 4 show IPA transcriptions and delimitation of the syllables and segments, respectively.

2.4. Measures and Analysis

All recordings were digitized at 16 kHz. The audio recordings were automatically segmented and labelled, using the orthographic prompts used for the recordings, in Web-Maus, a web application that aligns recordings to their corresponding orthographic texts by means of text-to-phoneme conversion and forced-alignment algorithms (Kisler et al. 2017). The resulting phonetic segment boundaries were checked and hand-corrected where needed.

A number of measures was used to examine the production of the various dimensions of intonation. These measures were examined in the whole corpus, except for measures of alignment and sentence-internal continuation, which were examined only in the statements and sentences containing sentence-internal continuation, respectively (see below). To test the systemic and frequency dimensions, the labels for pitch accents and boundary tones were compared between the groups of speakers and sentence types. For the realizational dimension, we examined two aspects of phonetic implementation for which cross-language differences are reported between SSBE and Austrian German, namely pitch range and alignment (how pitch accents are lined up with segments in time). Therefore, these are likely candidates for L2-induced changes to L1 speech in the realizational dimension of intonation. We used Praat (Boersma and Weenink 2022) to calculate measures of pitch range

and alignment. For pitch range, we measured f0 in our corpus, with a pitch range setting of 50 to 400 Hz for males and 75 to 560 Hz for females, i.e., for both genders in a three-octave range, by means of the “To pitch (ac)...” routine in Praat. The parameter octave-jump cost was increased to 0.5 in order to penalize large frequency jumps; all other settings were left at default values for this routine. We then used a Praat script to obtain the speaker-specific 90% pitch range, i.e., the difference between the 95th and 5th percentile of the measured pitch range in semitones. As mentioned above, alignment was measured in the statements ($N = 96$) of our corpus. These statements, some of which were taken from Atterer and Ladd (2004), were designed to elicit a pitch rise on the first content word. In order to ensure that a prenuclear rise was elicited on the test word, care was taken to use “either an adjective followed by a noun, or a noun followed by a genitive construction” (Atterer and Ladd 2004, p. 182). In all cases, the stressed syllable of the test word was always preceded and followed by two or more unstressed syllables. While this construction generally attracted a prenuclear rise on the test word and a nuclear accent on the following noun, in some cases the following noun was deaccented. These cases ($N = 5$) were discarded. In the remaining 91 sentences, we measured the alignment of the start and end of the prenuclear rise. For the alignment of the start of the rise, the distance in milliseconds (ms) between the beginning of the initial consonant of the test word bearing the prenuclear accent (labelled as C0) and the start of the prenuclear rise was measured. For the alignment of the end of the rise, the distance between the end of the prenuclear rise and the start of the vowel of the post-accentual syllable (labelled as V1) was taken as our measure. Figure 3 shows an example of the alignment measures in one of the test words in our corpus. As the use of sonorants in the test syllables ensured a relatively smooth f0 trace, it was generally unproblematic to locate the local f0 peaks and valleys.

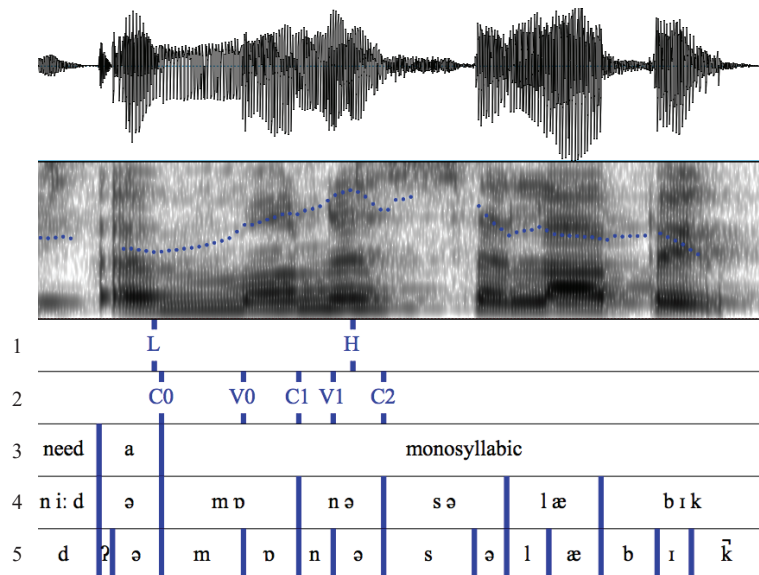


Figure 3. Example of the alignment measures in the test word *monosyllabic* in our corpus (extracted from the utterance “I need a monosyllabic word for my crossword puzzle”). Tier 1 shows the start (L) and end of the rise (H). Tier 2 shows the start of the initial consonant (C0) and vowel (V0) of the test word bearing the prenuclear accent, the start of the consonant (C1) and vowel (V1) of the post-accentual syllable, and the end of the post-accentual vowel (C2). Tier 3 shows the orthographic transcription. Tiers 4 and 5 show IPA transcriptions and delimitation of the syllables and segments, respectively.

Finally, in order to test the semantic dimension, we examined how sentence-internal continuation is signalled in our groups of participants. Previous studies have argued that English and German differ in how they signal sentence-internal continuation, with German, including Austrian German (Moosmüller et al. 2015), favouring a rising pitch accent (L*H) and English speakers typically employing a falling (H*L) pitch accent (see Chen 2007, for an overview). Therefore, we used the labels for nuclear pitch accents and the frequency with which they occur at the end of the first intonational phrase (e.g., in the sentence ‘Do you like Malaga or Malta best?’ we investigated the nuclear pitch accents occurring in the intonational phrase ‘Do you like Malaga’) in all the sentences with sentence-internal continuation in our corpus ($N = 94$) as our measure for examining differences between the groups in the semantic dimension.

3. Results

3.1. Systemic Dimension

Based on the existing literature, we expected that the inventory of pitch accents and boundary tones in SSBE and Austrian German would be very similar (Grabe 1998), with the exception of the early peak which is reported to be present in Austrian German (Schmid and Moosmüller 2013) but not in SSBE (Grabe 2004). This is indeed what we found when we compared all prenuclear and nuclear pitch accents and boundary tones used in our corpus. Both SSBE and AUT groups used the pitch accents H*L, !H*L, H*, !H* and L*H, as well as high and low initial (%H, %L) and final (H%, L%) boundary tones. The only cross-language difference was found in the use of H!H*L (early peak), a pitch accent which was present in the AUT speakers’ inventory but not in that of the SSBE monolinguals. In terms of their distribution across the different sentences types, we found that all pitch accents occurred in each sentence type (albeit to a different extent, as will be reported in Section 3.2), except for H!H*L which was only used in questions. As both monolingual groups used H*L, !H*L, H*, !H*, L*H, and initial and final high and low boundary tones, it is no surprise that these pitch accents and boundary tones are also used by the BIL group and, just as in the two monolingual groups, also occurred in each sentence type. However, the BIL group’s L1 inventory was found to also contain the early peak (H!H*L), a pitch accent which is not used by the SSBE monolingual group. Similar to the monolingual AUT speakers, the early peak was only used in questions.

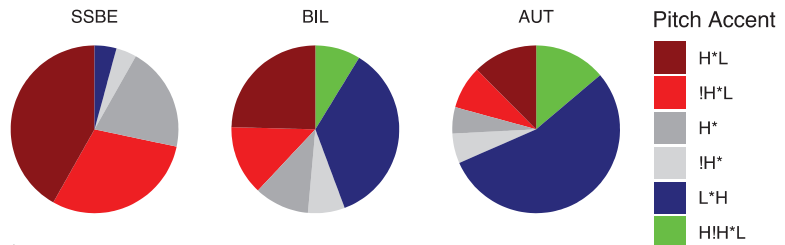
As there is a suggestion in the literature that there may be a gender-specific distribution in the use of early peaks (H!H*L) in Austrian German, we checked whether this was the case in our data. As mentioned above, early peaks were only used in questions. We therefore ran Chi-Square tests in the question data only with *percentage of occurrences of H!H*L* (i.e., H!H*L as opposed to non-H!H*L) as dependent variable and *gender* as independent variable, separately for the AUT and BIL groups. This showed no effect of gender, neither for the AUT group ($\chi^2[1] = 0.19$, n.s.) nor the BIL group ($\chi^2[1] = 0.29$, n.s.). While the bilingual speakers used early peaks to a lesser extent than the Austrian speakers did (as will be discussed in more detail in the next section), its use was not restricted to just a few bilinguals but used across all speakers.

3.2. Frequency Dimension

We first established whether the overall number of pitch accents was the same across the speaker groups: we ran an ANOVA⁹ with the *number of pitch accents per speaker and group* as dependent variable and *speaker group* (levels: SSBE, BIL, and AUT) as independent variable. This revealed that the groups differ in the overall number of pitch accents used ($F[2,21] = 7.9$, $p < 0.01$). Post-hoc pairwise *t*-tests with Bonferroni correction revealed, however, that AUT speakers have a lower number of pitch accents than the SSBE ($p < 0.05$) and BIL ($p < 0.05$) speakers, but that there is no significant difference in the number of pitch accents between SSBE and BIL speakers. We then proceeded with examining the frequency with which the pitch accents and boundary tones were used by the three groups of speakers. Figure 4 shows the overall frequency of use of the pitch accents and boundary tones in the

whole corpus by the three groups of speakers. It can be seen that the overall frequency of use of some pitch accents and boundary tones differs between the groups. In particular, SSBE speakers produced more falling (H*L and !H*L taken together) than rising (L*H) pitch accents (71.7 % vs. 4.2%), whereas the reverse was true for AUT-controls (20.7% versus 54.6%). This was found to be significant in a Chi-Square test with *percentage of pitch accent* (i.e., percentages of falling and rising pitch accents) as dependent variable, and *speaker group* (as above) as independent variable ($\chi^2[2] = 71.4, p < 0.001$). Post-hoc pairwise Chi-Square tests with Bonferroni-correction (correction factor 3 for the three tests) showed that falling and rising pitch accents were used in different amounts in SSBE vs. AUT ($p < 0.001$), SSBE vs. BIL ($p < 0.001$), and in AUT vs. BIL ($p < 0.01$). The level pitch accents (H* and !H* taken together) were produced to a greater extent by the SSBE (24.1%) than by the AUT speakers (10.8%), and BIL speakers' frequency of use is intermediate between that of the SSBE and the AUT speakers (at 17.7%). The early peak (H!H*L) was used by the AUT speakers in 13.8% of cases, whereas it did not occur in the SSBE speakers, and the BIL speakers were found to produce it in 8.7% of their utterances. A Chi-Square test with *percentage of pitch accent* (i.e., percentages of falling (H*L and !H*L), rising (L*H), level (H* and !H*), and early peak (H!H*L) pitch accents) as dependent variable and *speaker group* as independent variable confirmed that the frequency of use for the four pitch accent categories were generally very different for the three speaker groups ($\chi^2[6] = 90.2, p < 0.001$). Post-hoc tests with pair-wise comparisons (with Bonferroni-correction, i.e., with a Bonferroni factor of 3, due to the three pair-wise comparisons) of the speaker groups showed highly significant differences between BIL and SSBE and AUT and SSBE speaker groups ($p < 0.001$ each), and significant differences between BIL and AUT speakers ($p < 0.05$).

(a) Pitch Accents



(b) IP Boundaries

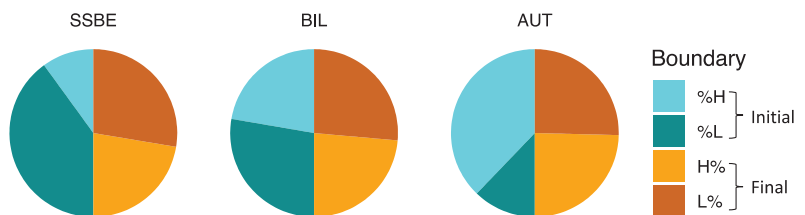


Figure 4. Overall frequency of use of pitch accents (panel a) and boundary tones (panel b) by the three groups of speakers.

As for the boundary tones, we found that SSBE speakers typically started (in 80.0% of the cases) their intonation phrases with a low boundary tone at the start of their intonation phrases (%L), whereas the speakers in the AUT group mostly used a high boundary tone (75.6%) at the start of their intonation phrases (%H). The BIL speakers' frequency of use of initial boundaries was found to be in between those for the two monolingual groups, with 44.6% use of a high boundary tone (%H) and 55.4% use of a low boundary tone (%L) at the start of their intonation phrases. These differences were found to be significant in a

Chi-Square test with *percentage of high and low initial boundary tones* as dependent variable and *speaker group* ($\chi^2[2] = 62.4, p < 0.001$). Post-hoc tests with Bonferroni correction revealed that all three pairwise speaker group comparisons showed significant differences (BIL vs. SSBE: $p < 0.01$, BIL vs. AUT and SSBE vs. AUT: $p < 0.001$). A Chi-Square test with *percentage of high and low final boundary tones* as dependent variable and *speaker group* showed no significant differences in the frequency of use of the boundary tones at the end of intonation phrases ($\chi^2[2] = 0.4, n.s.$), with a nearly 50/50 split for all three groups (SSBE: 44.7% H% vs. 55.3% L%; BIL: 47.2% H% vs. 52.8% L%; AUT: 49.2% H% vs. 50.8% L%).

We also observed differences in the frequency of use of the intonational primitives across sentence types. Due to the limits on article length, we restricted our analysis of the intonational primitives in the different sentence types to an analysis of nuclear accents. Table 2 shows the nuclear accents in each sentence type and participant group. In statements, the most frequently used nuclear pitch accent by all three groups is that of H*L (SSBE 96.9%, AUT 82.9%, BIL 90.9). AUT speakers additionally use L*H (in 14.3% of statements), whereas this nuclear pitch accent does not occur in the statements produced by SSBE speakers. The BIL speakers, on the other hand, use the nuclear accent L*H in their English nearly as often (in 9.1% of statements) as AUT speakers use it in their German statements. A Chi-Square test with *percentage of nuclear pitch accents* (i.e., with the percentages of H*L, L*H, and H*) as the dependent variable and *speaker group* showed a significant effect of speaker group ($\chi^2[4] = 20.7, p < 0.001$). Pairwise comparisons showed that BILs and AUTs did not show significantly different use of tonal categories (the other two pair-wise comparisons, i.e., SSBE vs. BIL and SSBE vs. AUT, resulted in (Bonferroni-corrected) $p < 0.001$).

Table 2. Nuclear accents and their usage in % by SSBE, BIL, and AUT speakers in statements (ST), wh-questions (WHQ), yes/no questions (YNQ), declarative questions (DQ), and sentence-internal continuation (CONT).

NUCLEAR ACCENT	ST			WHQ			YNQ			DQ			CONT		
	SSBE	BIL	AUT	SSBE	BIL	AUT	SSBE	BIL	AUT	SSBE	BIL	AUT	SSBE	BIL	AUT
H*L	96.9	90.9	82.9	96.7	78.1	9.4	96.8	29.4	0	84.4	15.7	0	46.9	12.5	0
H!H*L	0	0	0	0	12.5	31.3	0	38.2	62.5	0	78.1	71.9	0	0	0
H*	3.1	0	0	0	0	0	0	0	0	15.6	3.1	0	0	0	0
L*H	0	9.1	14.3	3.2	9.4	59.3	3.2	32.4	37.5	0	3.1	28.1	53.1	87.5	100

In wh-questions, both monolingual groups use the falling nuclear accent H*L, but to a very different degree. While SSBE speakers use it in nearly all of the wh-questions that were produced (96.7%), AUT speakers use it in just 9.4% of cases. The BIL speakers use considerably more falling nuclear accents than the AUT speakers, but less than the SSBE speakers (78.1%). Early peaks (i.e., H!H*L) were not observed in the wh-questions produced by the SSBE speakers, but are used in nearly a third of the cases (31.3%) by the AUT speakers. Despite this nuclear accent not being part of the L1 English inventory, the BIL speakers used it in 12.5% of cases, albeit to a lesser extent than the AUT speakers. Again, a Chi-Square test with dependent variable *percentage of nuclear pitch accents* (i.e., with the percentages of H*L, L*H, and H!H*L) and independent variable *speaker group* showed a significant effect of *speaker group* ($\chi^2[4] = 181.8, p < 0.001$); all three post-hoc pairwise comparisons with Bonferroni correction resulted also in $p < 0.001$.

In yes/no-questions, we again see clear and—as shown by a Chi-Square test—significant differences in *percentage of nuclear pitch accents* (the dependent variable with, in this case, the following three categories: H*L, L*H, and H!H*L) ($\chi^2[4] = 204.4, p < 0.001$) between the three *speaker groups*. Where the most frequent nuclear accent is H*L in the SSBE speakers (with a frequency of use of 96.8%), it is not used at all by AUT speakers, and the BIL speakers are, again, in-between (with a frequency of use of 29.4%). The reverse is true for the early peaks H!H*L, which is the most frequently used nuclear accent in yes/no questions by the AUT speakers (62.5%), but not used at all by the SSBE speakers, with the BIL speakers in-between the two monolingual groups (38.2%). In addition to their use

of H!H*L, the AUT speakers also use L*H, although with 37.5% it is used less often than H!H*L. SSBE speakers use L*H in yes/no questions on occasion (3.2%), whereas the BIL speakers use it almost as often as the AUT speakers (32.4%). Post-hoc tests with Bonferroni corrections, comparing speaker groups in pairs (i.e., SSBE vs. AUT, SSBE vs. BIL, and BIL vs. AUT), showed highly significant differences between all three pairwise tests ($p < 0.001$ each). In order to test as to whether there was a difference between the use of L*H vs. all other tonal categories in this context (H*L and H!H*L) combined, we conducted another Chi-Square test, with the dependent variable *percentage of nuclear pitch accents* (L*H vs. non-L*H) and independent variable *speaker group*, which showed an effect of speaker group ($\chi^2[4] = 204.4, p < 0.001$). Pairwise post-hoc comparisons showed significant differences for SSBE vs. AUT and SSBE vs. BIL, but not for AUT and BIL speakers ($\chi^2[1] = 0.4, n.s.$).

In declarative questions we again find a clear effect of *speaker group* in a Chi-Square test with dependent variable *percentage of nuclear pitch accents* (categories: H*L, H!H*L, L*H, and H*) ($\chi^2[6] = 263.7, p < 0.001$). Pairwise comparisons for SSBE vs. BIL, SSBE vs. AUT, and BIL vs. AUT showed highly significant differences for each ($p < 0.001$ each). In 28.1% of the cases, AUT speakers use a L*H nuclear accent in declarative questions, whereas L*H is not present in this type of questions of SSBE speakers. The BIL speakers show minimal use of L*H in declarative questions (3.1%). While H*L is the most frequently used nuclear accent by SSBE speakers (84.4%), it is not used at all by the AUT speakers, and the BIL speakers have a frequency of use that is in-between the two monolingual groups (15.7%). Instead of a H*L, AUT speakers' most frequently used nuclear accent is H!H*L (71.9%), a nuclear accent which is not used at all by the SSBE speakers. The H!H*L is also the most frequently used nuclear accent by the BIL speakers, with frequency of use patterns that are even slightly higher (78.1%) than those of the monolingual AUT speakers, i.e., overshooting the AUT norm. We also wanted to test whether there were speaker group related differences between the early peak (H!H*L) use and the use of all other categories combined. To test this, we conducted a Chi-Square test with *speaker group* and dependent variable *percentage of nuclear pitch accents* (H!H*L vs. non-H!H*L), which revealed significant results ($\chi^2[2] = 150.8, p < 0.001$). Post-hoc comparisons with Bonferroni correction showed that this difference is significant for SSBE vs. AUT and SSBE vs. BIL (each $p < 0.001$), but not for BIL vs. AUT ($\chi^2[1] = 0.72, n.s.$).

Finally, while it can be seen that the L*H is the only nuclear accent used by AUT speakers in sentence-internal continuations, SSBE speakers alternate between a L*H and a H*L nuclear accent (in 53.1% vs. 46.9% of cases, respectively). Once again, BIL speakers were found to show patterns of use in their L1 that are in-between the monolingual groups, with 87.5% usage of L*H (i.e., a much higher frequency of use than that of monolingual SSBE speakers) and 12.5% usage of H*L (much lower than that of the monolingual SSBE speakers). We tested this difference statistically by means of a Chi-Square test with *speaker group* and the dependent variable *nuclear pitch accents* (H*L vs. L*H) and found it to be statistically significant ($\chi^2[2] = 74.3, p < 0.001$). All three pairwise-comparisons (AUT vs. SSBE, SSBE vs. BIL, and AUT vs. BIL) were also highly significant ($p < 0.001$ each).

3.3. Realizational Dimension

We then analyzed two aspects of the realizational dimension of intonation, i.e., pitch range and prenuclear alignment. Based on the literature, we expected pitch range to be wider in SSBE than in AUT speakers, and the BIL speakers to be somewhere in between. The mean values for the 90% pitch range in semitones by the three groups shown in Figure 5 appear to confirm this. We ran an ANOVA with *90% pitch range* as the dependent variable and *speaker group* (as above) as independent variable. No significant effect of *speaker group* was found ($F[2,21] = 2.0, n.s.$). Subsequent post-hoc *t*-tests with Bonferroni correction also revealed no statistically significant pairwise difference between the group pairs, although there are tendencies for AUT vs. BIL and for AUT vs. SSBE.

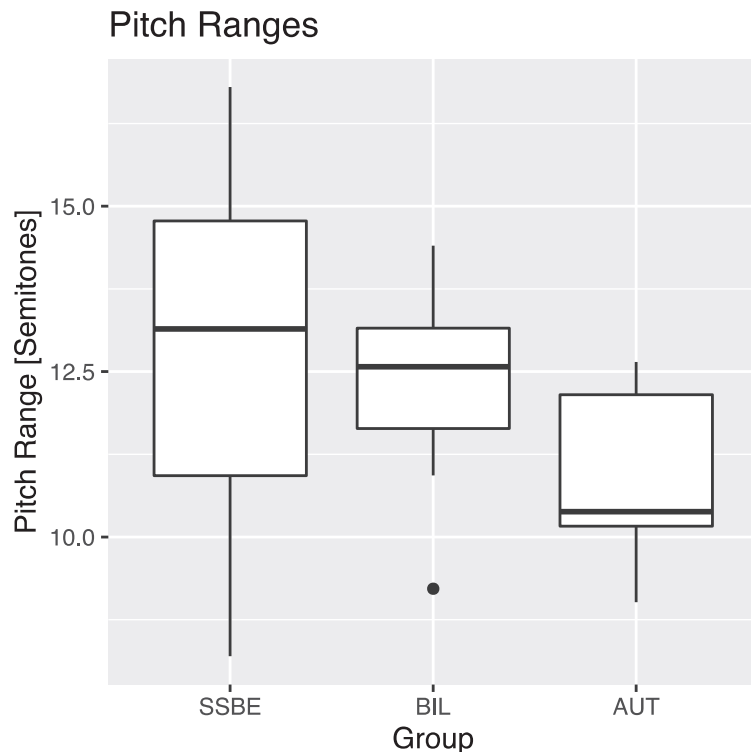


Figure 5. 90% pitch range in semitones of the three participant groups.

For alignment, two one-way ANOVAs with *speaker group* (SSBE, BIL, AUT) as independent variable were run, one with the start of the rise (measured as the temporal difference between the start of the rise and the initial consonant of the syllable bearing the prenuclear accent) as dependent variable, and the other with the end of the rise (measured as the temporal difference between the end of the rise and the beginning of the post-accentual vowel) as dependent variable. Results are plotted in Figure 6a,b, respectively. For alignment of the start of the rise, it was confirmed that the differences were highly significant ($F[2,21] = 64.4, p < 0.001$). Post-hoc *t*-tests with Bonferroni-correction showed significant differences between the monolingual SSBE and monolingual AUT groups ($p < 0.001$), between SSBE monolinguals and BIL ($p < 0.001$), and between the AUT and BIL ($p < 0.01$) groups. On average, SSBE controls start their prenuclear rises 9.8 ms after the consonant onset (C0) of the accented syllable, whereas AUT controls start the rises well into the accented vowel (137.3 ms after C0). The BIL speakers show intermediate values, with an alignment of 78.8 ms. For alignment of the end of the prenuclear rise (measured as the temporal difference between H and V1), we found no significant differences between the three speaker groups ($F[2,21] = 2.2, n.s.$).¹⁰

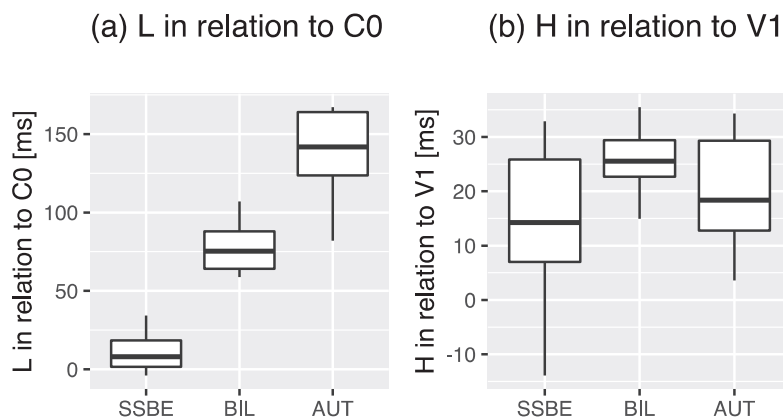


Figure 6. Alignment of (a) the start (L) of the rise in relation to the initial consonant of the syllable bearing the prenuclear accent (C0); and (b) alignment of the end (H) of the rise in relation to the start of the post-accentual vowel (V1) in ms for the three groups of speakers.

In addition to the planned analyses of pitch range and prenuclear alignment, we decided to also investigate a third aspect of phonetic realization, i.e., how the early peak (H!H*L) was realized. According to the LILt, building on assumptions from the SLM(-r), an L2-on-L1 effect is likely to be completely absent when a pitch accent is part of the inventory of the L2 but not the L1. Yet, our results showed that the BILs in our study used the H!H*L in their L1, despite the fact that this particular pitch accent was completely absent from the tonal inventory of the monolingual SSBE control group. While this suggests that the bilinguals may have fully transferred the L2 category into their L1, it is possible, as suggested by the SLM(-r) and LILt, that it may be realized differently from how it is realized by monolingual speakers because of a need to maintain contrast within a shared phonetic space. For this particular pitch accent, however, we think this scenario is unlikely given that a pitch accent where the pitch maximum is reached on a metrically weak syllable immediately *preceding* the accented syllable makes it sufficiently different from any other existing categories in the L1-L2 shared phonetic space, and therefore a need to ‘exaggerate’ its realization seems unnecessary. However, to make sure, we decided to nevertheless examine whether there were possible realizational differences between early peaks in the AUTs’ Austrian German productions and the early peaks in the BILs’ English productions. This was examined by calculating the f_0 differences (in semitones) between the metrically weak syllable bearing the early peak (measured as the mean f_0 value in this syllable) and the following accented syllable (again, measured as the mean f_0 in this syllable). Given the tendencies for differences in general pitch range between the BIL and the AUT groups reported above, we normalized the H!H*L ranges for their speaker-specific pitch range by expressing the f_0 differences (in semitones) as a proportion of the same speakers’ pitch range (also in semitones). Figure 7 shows a comparison of the normalized early peak ranges. A t -test revealed no significant differences between the groups in the normalized early peak ranges ($t[12.7] = 1.4$, n.s.). In other words, no differences between the BILs’ realization of H!H*L in their L1 and the realization by AUT speakers were found, suggesting that the BILs have fully transferred this L2 category into their L1.

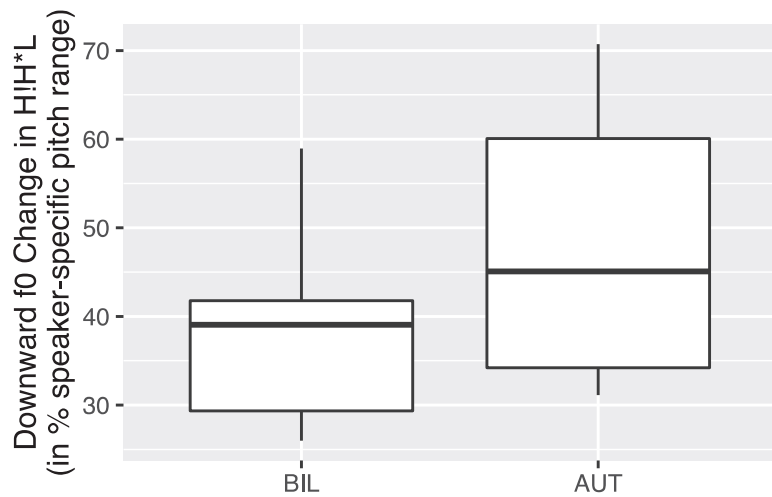


Figure 7. Downward f0 change in H!H*L pitch accents, normalized to speaker-specific pitch ranges.

3.4. Semantic Dimension

Based on the literature discussed earlier, we expected SSBE speakers to show a preference for a falling pitch accent to indicate sentence-internal continuation, AUT speakers to prefer a rising pitch accent, and the BIL speakers to be intermediate between the two monolingual groups. A Chi-Square test with *percentage of use of falls or rises* as dependent variable, and with the three *speaker groups* as independent variable showed significant cross-language differences ($\chi^2[1] = 74.3, p < 0.001$). These confirmed that whereas AUT speakers show a preference for the use of L*H (100%), the SSBE-controls used rises and falls in approximately equal measure (53.1% L*H vs. 46.9% H*L). The BILs were once again in-between both monolingual groups with 87.5% L*H and 12.5% H*L (cf. also Table 2 in Section 3.2). Post-hoc chi-squared tests with Bonferroni correction showed that all three pairwise comparisons were significant (SSBE vs. BIL and SSBE vs. AUT: $p < 0.001$, AUT vs. BIL: $p < 0.01$).

4. Discussion

This study aimed to gain a better understanding of the plasticity of native language intonation due to long-term immersion in an L2-speaking environment. To this end, we examined the four dimensions of intonation in the LILt proposed by Mennen (2015) in the read L1 speech of late English–Austrian sequential bilinguals who emigrated to Austria in adulthood. As such, this study is the most comprehensive investigation of L1 attrition of intonation, as no previous studies have considered whether and how L1 modifications manifest in all dimensions of intonation. The results revealed widespread L2-induced influences on the L1 intonation of the bilinguals in each dimension of intonation and in each sentence type, although the extent of the L2-on-L1 effect varied. The form the observed L1 modifications in intonation took was consistently one of assimilation, with intermediate values between the L1 and the L2, or complete assimilation to L2 properties. We will consider the implications of these findings below.

First, let us consider the pervasiveness of the L2-on-L1 effect and the fact that it was evidenced on each dimension of intonation. The LILt assumes that cross-language influences in intonation are—at least to some extent—influenced by the existence of cross-linguistic differences in the intonation systems of the L1 and L2, such that no L2-on-L1 effect would be expected where there are no cross-language differences. This is confirmed by the results, which suggest that the extent of the L2-on-L1 effect largely depends on the degree of cross-language differences. Without exception, wherever the L1 and L2 showed cross-

language differences, an L2-on-L1 effect was found in the bilinguals' first language, and this was apparent in both pitch accents and boundary tones. That is, the L2-on-L1 effect for the boundary tones was restricted to the frequency of use of high (%H) and low (%L) boundary tones at the start of their intonation phrases, reflecting the cross-language differences that only occurred in initial boundary tones. For pitch accents, the L2-on-L1 effect was restricted to just one pitch accent (the early peak) on the systemic dimension, whereas the effect was more pervasive on the frequency and realizational dimension and in some sentence types compared to others, reflecting the extent of cross-language differences. The findings of our study therefore confirm that LILt's method of classifying and characterizing cross-language intonation differences on four intonation dimensions is not only an effective method for predicting where cross-linguistic influences in the L2 are likely to occur (Albin 2015; Busà and Stella 2015; Graham and Post 2018; Pešková 2020; Sanchez 2020; Schauffler 2021) but can also predict where L2-induced influences in L1 intonation may be likely.

Let us now turn to the form of the observed L1 modifications in intonation. Based on assumptions from the LILt and the SLM(r), we predicted that the form L1 modifications would take would depend on whether the cross-language differences in intonation are categorical or gradient in nature, with gradient aspects being more vulnerable to cross-language influences than categorical ones (Graham and Post 2018; Jun and Oh 2000; Mennen et al. 2010; Sanchez 2020). We hypothesized that when cross-language differences are gradient, bilinguals are likely to identify them as variants or 'allotones' of the L1 tonal category. This would result in the L2-on-L1 effect to be one of assimilation. This is indeed what we found: the L2-on-L1 effect observed for the gradient cross-language differences in our study was consistently one of assimilation. While we have no data for how the bilinguals actually produced related tonal categories in their L2, the occurrence of assimilatory L2-on-L1 effects suggests that the bilinguals are likely to have merged the gradient cross-language differences between the L1 and L2 into a composite L1-L2 phonetic category, given that cross-language influences "provide a reflex that is diagnostic of L2 category formation or its absence" (Flege and Bohn 2021, p. 42). The values that we found for gradient aspects of intonation were sometimes half-way between the two monolingual control groups, other times closer to the L1 or approximating the L2 norms, reflecting different degrees of the L2-on-L1 effect. Interestingly, in a few cases we found that the bilingual's frequency of use fell within the Austrian German monolingual norms, suggesting that the bilinguals had fully merged the L1 and L2 properties. This was found for the frequency of use of nuclear rises (L*H) in statements and yes-no questions, and the use of early peaks (H!H*L) in declarative questions, where the bilinguals were found to use these pitch accents in English as often as the monolingual Austrian controls did in Austrian German. Full assimilation has also been reported for segments, although it is considered to be unusual. For instance, de Leeuw et al. (2018b) present a case of a German-English bilingual whose realization of the L1 rhotic had assimilated entirely into the monolingual norm of the English retroflex. While this was presented as a case of "extreme phonetic attrition" (de Leeuw et al. 2018b, p. 163) caused by prolonged reduced L1 use, our results suggest that it may not be as unusual as previously thought, at least not for intonation. Possible reasons for why full merging may have been observed in our data will be explored below.

For categorical cross-language differences, the LILt assumes that bilinguals are likely to establish a new L2 category as long as it is sufficiently different from any other category available in the L1, as would be the case when a category is part of the inventory of the L2 but not the L1. The only categorical cross-language difference between Austrian German and SSBE is on the systemic dimension and concerns the early peak (H!H*L), which is present in the inventory of monolingual Austrian German speakers but is not used by the monolingual SSBE speakers. The early peak is different from any other pitch accents in Austrian German and SSBE due to the association of a high tone with a metrically weak syllable immediately preceding the accented syllable. This particular language-specific tune-text association, where a high tone occurs on an a metrically weak syllable preceding

the syllable that is actually accented, is thought to be unusual in most Western European languages (Ladd 1996; Mennen 2015), although it also occurs in German German. Given this difference from any other pitch accents in the bilinguals' L1, we expected that the early peak would not be identified by bilinguals as an instance of one of their already existing L1 pitch accents, and we therefore assumed that there would be no L2-on-L1 effect. To our surprise, all bilinguals in our study transferred the early peak into their L1, although they used it significantly less often than the monolingual Austrian German controls. Moreover, a comparison of the phonetic realization of early peaks by the bilingual speakers in their L1 and the early peaks produced by the monolingual Austrian German speakers in our study showed no significant differences in phonetic realization. This suggests that the early peak has been transferred fully into the L1 of the bilingual speakers.

Even so, why did the bilingual speakers transfer the early peak into their L1 in the first place? Neither the LILt nor the SLM(-r) can explain this in the current version of these models. Perhaps Markedness Theory can provide a possible explanation. The Markedness Differential Hypothesis (Eckman 1977) proposes that aspects from the L2 that are different and marked, i.e., infrequent in the world's languages, will pose more difficulties for L2 learners than aspects that are different but less marked. Conversely, "those forms that are less marked in the L2 are more likely to replace more marked forms in the L1" (Gürel 2004, p. 54). As, according to Ladd (1996) and Mennen (2015), a tonal target on a metrically weak syllable such as in the early peak (H!H*L) is unusual in West European languages, it may therefore be more marked than other pitch accents. Yet, the marked early peak is transferred to the bilinguals' L1. We therefore conclude that typological markedness cannot explain our findings.

Perhaps we need to consider the possibility that L1 attrition of intonation is just different from what is typically observed in L1 attrition at the segmental level. While we occasionally may observe full merging at the segmental level when the L2 and L1 differ in gradient aspects of their pronunciation, full assimilation of categorical differences, i.e., where the L2 is sufficiently different from any already existing L1 categories, has to our knowledge never been reported for segments. The equivalent in segmental terms would be if, for instance, an L2 Welsh learner from England started to use a lateral fricative in their English. That is not something we consider likely to happen and it therefore suggests that the process of L1 attrition of intonation is different from that of segments. A possible explanation may be that intonation is more malleable—more so than segments—because of its weaker link with orthography as compared to segments. In fact, "intonation allows for a high degree of variation in the choice and distribution of tonal categories" or their phonetic realization, due to the fact that noticeable variations may not be perceived as foreign but only lead to "a slightly different interpretation" (Jilka 2000, p. 58). This would allow bilinguals more flexibility in the use of L2 pitch accents in their L1.

It is therefore not unreasonable to assume that the bilingual speakers in our study may have transferred the early peak to fulfil a semantic or pragmatic function that is expressed in the L2 but not the L1. We found that the monolingual Austrian German speakers used the early peak *only* in questions. In fact, when inspecting the use of early peaks in the different question types (see Table 2), we see that it is used increasingly more frequently when the number of syntactic and/or lexical markers of interrogativity in the question types decreases. In questions with a question word and inversion (WHQs), i.e., where there are two lexical/syntactical markers of interrogativity, the early peak is used least often (31.3%). In questions with inversion (YNQs), which is another marker of interrogativity, the early peak is used more often (62.5%). In declarative questions (DQs), where no lexical or syntactical markers of interrogativity are present, the early peaks are most frequently used (71.9). The same is true for the bilinguals, who also use early peaks least often in WHQs (12.5%), followed by YNQs (38.5%), and use it most often in DQs (78.1%). This suggests that, just like the monolingual Austrian German speakers, the bilingual speakers may be using the early peak as a marker of interrogativity, and that its use may be constrained by the number of other (i.e., syntactic or lexical) markers of interrogativity present in an

utterance (see Haan 2002, for a similar discussion on the trade-off between prosodic and syntactic and/or lexical markers of interrogativity). It is possible that bilingual speakers may have felt the need to mark these degrees of interrogativity in their L1 due to immersion in an L2 environment. This may also explain why the frequency of occurrence of early peaks in DQs by the bilinguals is on a par with that of the monolingual Austrian speakers. We are, however, unsure what could explain the equal frequency of use of rises (L*H) in the statements and YNQs of the bilinguals and monolingual Austrian speakers. Unlike the early peak, the L*H is not used exclusively in questions. Therefore, it is not—at least not on its own—a marker of interrogativity. While it is possible that there is a specific semantic or pragmatic meaning that is associated with the use of L*H in statements and YNQs that the bilinguals have attempted to transfer to their L1, further research is needed to explore what particular meaning this is and to what extent it differs from L*H in other sentence types.

Our study highlights a few areas that are in need of further research. We deliberately did not investigate the influence of predictor variables (such as AoA, LoR, L2 proficiency, amount of L1 use, and amount of L2 use, etc.) on L2-induced influences in L1 intonation, as a full exploration of their role would require larger participant numbers and datasets than were currently available. We know virtually nothing on how L2-induced changes to prosodic aspects of L1 pronunciation (or segmental features for that matter) are related to the production of similar features in the L2, how such influences progress over time (but see Kornder and Mennen 2021) and which factors may influence their occurrence. For instance, would assimilation be more apparent in bilinguals who have only recently moved to an L2-speaking country? Can we expect more frequent use of early peaks in bilinguals with high L2 proficiency? Such questions highlight the need for controlled studies into the effect of predictor variables on L2-induced influences on L1 intonation in its various dimensions. In addition, our study investigated intonation in read speech as this gave us control (in terms of, for instance, expected stress patterns, number of pitch accents, or phonetic content) over the utterances that we intended to compare within and between languages. Intonation in read speech is, however, different from intonation in spontaneous speech (e.g., Blaauw 1994; Howell and Kadi-Hanifi 1991; Laan 1997) and future studies are necessary to investigate to what extent L2-induced influences are also found in spontaneously produced L1 intonation. Another aspect that may need further investigation is the extent to which the typological relationship between the bilinguals' L1 and L2 may influence the observed L2-induced changes in L1 pronunciation. While we do not assume that languages from different language families necessarily differ more in their respective phonological features than closely related languages (after all, we also find considerable cross-language differences in the intonation of the two Germanic languages under investigation in our study), it would be important to also examine L2-induced influences in languages with more extensive cross-language differences in intonation. In particular, it would be interesting to investigate whether more extensive cross-language differences in the systemic dimension would exert the same effect on the L1 as that observed in our study, where the categorical cross-language differences on the systemic dimension were restricted to just a single pitch accent. We assume that when bilinguals use an L2 pitch accent which does not form part of the L1 inventory (such as the early peak), this will be perceptually salient to native listeners and may contribute to them being perceived as non-native. It is possible that when more categorical intonation differences are transferred into the L1, the impression of non-nativeness may increase. However, there are no studies that directly link listener judgements of non-nativeness to specific L2-induced deviances from the L1 norm and it remains an open question whether categorical changes contribute more to the impression of non-nativeness than gradient changes or whether this impression arises from an accumulation of the various changes that may be present in a bilingual's L1.

In closing, the present study demonstrates that an individual's native language intonation system is not protected against L2 influences. In fact, the permeability of L1 intonation is not restricted to its phonetic realization as might be suggested by previous studies, but is found to occur across the board, affecting every dimension of intonation. This highlights

the need for studies that go beyond investigations of just one or two aspects of L2-induced modifications in L1 speech. Instead, studies should compare a wider range of prosodic and segmental areas of pronunciation within the same group of individuals. Such studies will provide a more holistic view of the areas of pronunciation that are susceptible to L1 attrition and those that may be less permeable.

Author Contributions: Conceptualization, I.M., R.M. and U.R.; methodology, I.M., R.M., U.R. and K.E.; formal analysis, I.M., U.R. and K.E.; writing—original draft preparation, I.M. and U.R.; writing—review and editing, I.M., R.M., U.R. and K.E.; visualization, U.R. and I.M.; supervision, I.M.; project administration, I.M.; funding acquisition, I.M. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Austrian Science Fund (FWF), grant number P33007-G.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Ethics Committee of the University of Graz (protocol code GZ. 39/37/63 ex 2019/20; date of approval: 27 February 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are not publicly available due to ongoing data analyses and because the participants did not give permission for their data to be shared.

Acknowledgments: We would like to thank the Austrian Science Fund (FWF) for their financial support of this research. We would also like to thank Klaus Jansch for his support with WikiSpeech. Finally, we thank Rebecca Clift, Kathleen McCarthy, Paul Foulkes, Ghada Khattab, Jiaying Li, and Bettina Beinhoff for their help in accessing participants in the UK.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ Note that cross-language interaction in pronunciation is observed in all types of bilinguals (see for instance [Amengual 2019](#)), including simultaneous bilinguals (i.e., individuals growing up speaking both languages since birth), early sequential bilinguals (i.e., individuals brought up monolingually before attending school in the majority language when they become bilingual), and late sequential bilinguals (i.e., those individuals who become bilingual after the age of puberty). The focus of this paper is on cross-language influences in the pronunciation of late sequential bilinguals.
- ² In this study, listeners were asked to indicate on which aspects of pronunciation they based their judgement of non-nativeness. The comments were classified as referring to segmental (pertaining to 64.7% of comments) or prosodic aspects of pronunciation (pertaining to 35.3% of comments), and then further divided into particular segmental (e.g., specific vowels or consonants) or prosodic features (e.g., stress, speaking rate, intonation). For prosody, intonation was mentioned most (in 22.4% of the comments), whereas rhythm/stress and speaking rate were mentioned considerably less often (in 6.9% and 6%, respectively).
- ³ Note that the SLM originally referred to this as a common “phonological space”. However, in the SLM-r, [Flege and Bohn \(2021, p. 21\)](#) acknowledge “that use of this term was a misnomer” and now prefer to refer to it as “common phonetic space.” This is also the term we use in the present paper.
- ⁴ The original dimensions proposed by [Ladd \(1996, p. 119\)](#) were meant to describe four types of dialect differences: (1) the phonetic implementation (‘realizational differences’); (2) the inventory of boundary tones and pitch accents (‘systemic differences’); (3) the distribution of boundary tones and pitch accents (‘phonotactic differences’); and (4) functionality (‘semantic differences’). [Mennen \(2015\)](#) adapted them to describe types of cross-language differences in intonation. Therefore, the dimensions in [Mennen \(2015\)](#) differ from those proposed by Ladd.
- ⁵ Some researchers (e.g., [Féry 1993](#); [Peters 2018](#)) present this accent type not as a separate pitch accent but as a modification of the H* or H*L pitch accent which retracts the peak to a preceding syllable, and is used to express additional meaning. In their view, early peaks are not seen as part of the systemic dimension of intonation but rather belong to the realizational dimension of intonation. Other researchers (e.g., [Baumann and Grice 2006](#); [Grice and Baumann 2002](#); [Niebuhr 2007](#)), however, treat the early peak as a distinct pitch accent, and thus as belonging to the systemic dimension of intonation. In other words, there is disagreement as to whether the retraction of the peak results from the same phonological representation or from different phonological representations. Controversies of this kind are typically resolved by conducting perception experiments and relying on native speaker intuition in acceptability judgment tasks. [Niebuhr \(2007\)](#) showed that German listeners linked early peaks (H!H*L, which they represent as H+L* in their framework) to different contexts than medial peaks (H*L, which they represent as H*), showing that there is a clear meaning differentiation between these pitch accents, thus corroborating the existence of

separate categories in German intonation. We therefore adopt their interpretation of the early peak as a distinct pitch accent and treat it as belonging to the systemic dimension.

- 6 In languages such as English and German, nuclear prominence is typically rightmost, that is the last prosodic word in a phrase bears the nuclear accent and is therefore the most prominent within that phrase (e.g., Ladd 1996).
- 7 While this study was on cross-language differences in pitch range between speakers of English (SSBE) and German (Northern Standard German), it also took linguistic measures linked to pitch accents and boundary tones. Their results showed that the German speakers more often showed low pitch accents compared to the English speakers.
- 8 The only exception to this was for the statements where participants read out four different sentences once. This was done because the statements were also used to test the alignment of prenuclear rises, and we wanted our data to be as comparable as possible with previously reported data from studies on English and German (Atterer and Ladd 2004; de Leeuw et al. 2012; Ladd et al. 1999). The total number of sentences per sentence type, however, was the same given that there was one repetition of four sentences for the statements and two repetitions of two sentences for the other sentence types.
- 9 Before running ANOVAs or *t*-tests we always tested whether its requirements were met by testing the normal distribution of the data by means of a Shapiro-Wilk tests, and using Levene's Test for Homogeneity of Variance. We will not report this further, unless requirements were not met.
- 10 It is sometimes argued that alignment is best expressed as a proportional measure, for instance as a ratio of the accented syllable (e.g., Silverman and Pierrehumbert 1990). We therefore calculated the alignment of the start of the rise and that of the end of the rise as a proportion of the syllable duration. This did not change our results. An ANOVA with *proportional alignment of L* as dependent variable and speaker group as independent variable showed a significant effect of speaker group ($F[2,21] = 80.9$, $p < 0.001$) and post-hoc *t*-tests with Bonferroni correction showed significant differences between the monolingual SSBE and AUT ($p < 0.001$), between SSBE and BIL ($p < 0.001$), and between AUT and BIL ($p < 0.01$). For *proportional alignment of H*, no effect was found ($F[2,21] = 0.24$, n.s.).

References

- Albin, Aaron Lee. 2015. Typologizing Native Language Influence on Intonation in a Second Language: Three Transfer Phenomena in Japanese EFL Learners. Ph.D. dissertation, Indiana University, Bloomington, IN, USA; p. 3715823.
- Alharbi, Amirah, Anouschka Foltz, Lisa Kornder, and Ineke Mennen. Forthcoming. L2 acquisition and L1 attrition of VOTs of voiceless plosives in highly proficient late bilinguals. *Second Language Research*.
- Amengual, Mark. 2019. Type of early bilingualism and its effect on the acoustic realization of allophonic variants: Early sequential and simultaneous bilinguals. *International Journal of Bilingualism* 23: 954–70. [CrossRef]
- Atterer, Michaela, and D. Robert Ladd. 2004. On the phonetics and phonology of 'segmental anchoring' of F0: Evidence from German. *Journal of Phonetics* 32: 177–97. [CrossRef]
- Baumann, Stefan, and Martine Grice. 2006. The intonation of accessibility. *Journal of Pragmatics* 38: 1636–57. [CrossRef]
- Baumann, Stefan, Martine Grice, and Ralf Benz Müller. 2000. GToBI—A phonological system for the transcription of German intonation. In *Prosody 2000. Speech Recognition and Synthesis*. Edited by Stanislav Puppel and Grazyna Demenko. Poznan: Adam Mickiewicz University, pp. 21–28.
- Beckman, Mary E., and Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3: 15–70. [CrossRef]
- Beckman, Mary E., Julia B. Hirschberg, and Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In *Prosodic Typology. The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun. Oxford: Oxford University Press, pp. 9–54.
- Bergmann, Christopher, Amber Nota, Simone A. Sprenger, and Monika S. Schmid. 2016. L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics* 58: 71–86. [CrossRef]
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam: John Benjamins, pp. 13–34.
- Blaauw, Eleonora. 1994. The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication* 14: 359–75. [CrossRef]
- Boersma, Paul, and David Weenink. 2022. Praat: Doing Phonetics by Computer [Computer Program]. Version 6.1.36. Available online: <http://www.praat.org/> (accessed on 11 February 2022).
- Breen, Mara, Laura C. Dilley, John Kraemer, and Edward Gibson. 2012. Inter-Transcriber Reliability for Two Systems of Prosodic Annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory* 8: 277–312. [CrossRef]
- Busà, Maria Grazia, and Antonio Stella. 2015. The acquisition of English L2 prosody by Italian native speakers: Experimental data and pedagogical implications. In *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference*. Edited by John Levis, Rani Mohammed, Manman Qian and Ziwei Zhou. Santa Barbara: University of Santa Barbara, pp. 15–26.
- Chang, Charles B. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40: 249–68. [CrossRef]

- Chang, Charles B. 2013. A novelty effect in phonetic drift of the native language. *Journal of Phonetics* 41: 520–33. [CrossRef]
- Chang, Charles B. 2019. Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research. *Journal of Phonetics* 74: 96–113. [CrossRef]
- Chen, Aoju. 2007. Language-specificity in the perception of continuation intonation. In *Tones and Tunes II: Phonetic and Behavioural Studies in Word and Sentence Prosody*. Edited by Carlos Gussenhoven and Tomas Riad. Berlin: Mouton de Gruyter, pp. 107–42.
- Chen, Sally, and Janice Fon. 2008. The peak alignment of prenuclear and nuclear accents among advanced L2 English learners. In *Proceedings of the Speech Prosody 2008 Conference*. Edited by Plínio A. Barbosa, Sandra Madureira and Cesar Reis. Campinas: State University of Campinas, pp. 643–46.
- Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37–46. [CrossRef]
- Cruttenden, Alan. 1997. *Intonation*. Cambridge: Cambridge University Press.
- de Leeuw, Esther. 2019a. Phonetic Attrition. In *The Oxford Handbook of Language Attrition*. Edited by Monika S. Schmid and Barbara Köpke. Oxford: Oxford University Press, pp. 204–17.
- de Leeuw, Esther. 2019b. Native speech plasticity in the German-English late bilingual Stefanie Graf: A longitudinal study over four decades. *Journal of Phonetics* 73: 24–39. [CrossRef]
- de Leeuw, Esther, Aurela Tusha, and Monika S. Schmid. 2018a. Individual phonological attrition in Albanian- English late bilinguals. *Bilingualism: Language and Cognition* 21: 278–95. [CrossRef]
- de Leeuw, Esther, Ineke Mennen, and James M. Scobbie. 2012. Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism* 16: 101–16. [CrossRef]
- de Leeuw, Esther, Ineke Mennen, and James M. Scobbie. 2013. Dynamic systems, maturational constraints and L1 phonetic attrition. *International Journal of Bilingualism* 17: 683–700. [CrossRef]
- de Leeuw, Esther, Monika S. Schmid, and Ineke Mennen. 2010. The effects of contact on native language pronunciation in an L2 migrant context. *Bilingualism: Language and Cognition* 13: 33–40. [CrossRef]
- de Leeuw, Esther, Tusha Aurela, Zhao Hui, Helke Kyle, and Greenfield Alice. 2018b. A Case of Extreme Phonetic Attrition in the German Rhotic. In *Mind Matters in SLA*. Edited by Clare Wright, Thorsten Piske and Martha Young-Scholten. Bristol: Multilingual Matters, pp. 162–82.
- Dmitrieva, Olga, Allard Jongman, and Joan A. Sereno. 2020. The effect of instructed second language learning on the acoustic properties of first language speech. *Languages* 5: 44. [CrossRef]
- Draxler, Christoph, and Klaus Jänsch. 2008. WikiSpeech—A Content Management System for Speech Databases. Paper presented at Interspeech 2008, the Ninth Annual Conference of the International Speech Communication Association, Brisbane, Australia, September 22–26; pp. 1646–49.
- Eckman, Fred R. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–30. [CrossRef]
- Escudero, David, Lourdes Aguilar, Maria del Mar Vanrell, and Pilar Prieto. 2012. Analysis of inter-transcriber consistency in the Cat_ToBI prosodic labeling system. *Speech Communication* 54: 566–82. [CrossRef]
- Féry, Caroline. 1993. *German Intonational Patterns*. Tübingen: Niemeyer.
- Flege, James Emil. 1987. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65. [CrossRef]
- Flege, James Emil. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–72.
- Flege, James Emil. 2002. Interactions between the native and second-language phonetic systems. In *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Edited by Petra Burmeister, Thorsten Piske and Andre Rohde. Trier: Wissenschaftlicher Verlag, pp. 217–44.
- Flege, James Emil, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Rtree Wayland. Cambridge: Cambridge University Press, pp. 3–83.
- Gargiulo, Chiara, and Mechteld Tronnier. 2020. First language attrition on prosody in a foreign language environment: A speech production study on anaphora resolution. *Journal of Monolingual and Bilingual Speech* 2: 219–44. [CrossRef]
- Grabe, Esther. 1998. Comparative Intonational Phonology: English and German. Ph.D. dissertation, Ponsen et Looien, Wageningen, The Netherlands. MPI Series 7.
- Grabe, Esther. 2004. Intonational variation in urban dialects of English spoken in the British Isles. In *Regional Variation in Intonation*. Edited by Peter Gilles and Jörg Peters. Tuebingen: Niemeyer, pp. 9–31.
- Grabe, Esther, Brechtje Post, Francis Nolan, and Kimberley Farrar. 2000. Pitch accent realization in four varieties of British English. *Journal of Phonetics* 28: 161–85. [CrossRef]
- Graham, Calbert, and Brechtje Post. 2018. Second language acquisition of intonation: Peak alignment in American English. *Journal of Phonetics* 66: 1–14. [CrossRef]
- Green, David W. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1: 67–81. [CrossRef]
- Grice, Martine, and Stefan Baumann. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191: 267–98.
- Guion, Susan G. 2003. The vowel systems of Quichua-Spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica* 60: 98–128. [CrossRef]
- Gürel, Ayşe. 2004. Selectivity in L2-induced L1 attrition: A psycholinguistic account. *Journal of Neurolinguistics* 17: 53–78. [CrossRef]
- Gussenhoven, Carlos. 1983. *A Semantic Analysis of the Nuclear Tones of English*. Indiana: Indiana University Linguistics Club.

- Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Haan, Judith. 2002. *Speaking of Questions: An Exploration of Dutch Question Intonation*. Utrecht: LOT.
- Hopp, Holger, and Monika S. Schmid. 2013. Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics* 34: 361–94. [\[CrossRef\]](#)
- Howell, Peter, and Karima Kadi-Hanifi. 1991. Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication* 10: 163–69. [\[CrossRef\]](#)
- Hualde, José Ignacio. 2005. *The sounds of Spanish*. Cambridge: Cambridge University Press.
- Huang, Becky H., and Sun-Ah Jun. 2011. The effect of age on the acquisition of second language prosody. *Language and Speech* 54: 387–414. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jilka, Matthias. 2000. The Contribution of Intonation to the Perception of Foreign Accent. Ph.D. thesis, University of Stuttgart, Stuttgart, Germany.
- Jun, Sun-Ah, ed. 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- Jun, Sun-Ah, ed. 2014. *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
- Jun, Sun-Ah, and Mirah Oh. 2000. Acquisition of second language intonation. In *Proceedings of International Conference on Spoken Language Processing*. Edited by Baozong Yuan and Taiyi Huang. Beijing: University of Beijing, vol. 4, pp. 76–79.
- Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich H. Frauenfelder, and Narly Golestani. 2016. Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics* 57: 21–39. [\[CrossRef\]](#)
- Kartushina, Natalia, and Clara D. Martin. 2019. Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics* 77: 100920. [\[CrossRef\]](#)
- Kisler, Thomas, Uwe D. Reichel, and Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45: 326–47.
- Kornder, Lisa, and Ineke Mennen. 2021. Longitudinal developments in bilingual second language acquisition and first language attrition of speech: The case of Arnold Schwarzenegger. *Languages* 6: 61. [\[CrossRef\]](#)
- Laan, Gitta P. M. 1997. The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication* 22: 43–65. [\[CrossRef\]](#)
- Ladd, D. Robert. 1996. *Intonational Phonology*, 1st ed. Cambridge: Cambridge University Press.
- Ladd, Robert D. 2000. Bruce, Pierrehumbert, and the Elements of Intonational Phonology. In *Prosody: Theory and Experiment. Text, Speech and Language Technology*. Edited by Merle Horne. Dordrecht: Springer, vol. 14, pp. 37–50.
- Ladd, D. Robert, Dan Faulkner, Hanneke Faulkner, and Agnes Schepman. 1999. Constant “segmental anchoring” of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America* 106: 1543–54. [\[CrossRef\]](#)
- Lado, Robert. 1957. *Linguistics across Cultures. Applied Linguistics for Teachers*. Ann Arbor: University of Michigan Press.
- Landis, J. Richard, and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33: 159–74. [\[CrossRef\]](#)
- Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: John Wiley and Sons.
- Mackey, William F. 2000. The description of bilingualism. In *The Bilingualism Reader*. Edited by Li Wei. Oxford: Routledge, pp. 26–54.
- Major, Roy C. 1992. Losing English as a first language. *The Modern Language Journal* 76: 190–208. [\[CrossRef\]](#)
- Major, Roy C. 2010. First language attrition in foreign accent perception. *International Journal of Bilingualism* 14: 163–83. [\[CrossRef\]](#)
- Mayr, Robert, David Sánchez, and Ineke Mennen. 2020. Does teaching your native language abroad increase L1 attrition of speech? The case of Spaniards in the United Kingdom. *Languages* 5: 41. [\[CrossRef\]](#)
- Mayr, Robert, Sasha Price, and Ineke Mennen. 2012. First language attrition in the speech of Dutch-English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition* 15: 687–700. [\[CrossRef\]](#)
- Mennen, Ineke. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63. [\[CrossRef\]](#)
- Mennen, Ineke. 2007. Phonological and phonetic influences in non-native intonation. In *Non-Native Prosody: Phonetic Description and Teaching Practice*. Edited by Jürgen Trouvain and Ulrike Gut. Berlin: Mouton de Gruyter, pp. 53–76.
- Mennen, Ineke. 2015. Beyond segments: Towards a L2 intonation learning theory. In *Prosody and Languages in Contact*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin: Springer, pp. 171–88. [\[CrossRef\]](#)
- Mennen, Ineke, and Denise Chousi. 2018. Prosody in first-generation adult immigrants and second-generation heritage-language users: The timing of prenuclear rising accents. In *Proceedings of the 9th Speech Prosody Conference*. Edited by Katarzyna Klessa, Jolanta Bachan, Agnieszka Wagner, Maciej Karpinski and Daniel Śledziński. Poznan: University of Poznan, pp. 828–32.
- Mennen, Ineke, Aoju Chen, and Fredrik Karlsson. 2010. Characterising the internal structure of learner intonation and its development over time. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech (New Sounds)*. Edited by Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel and Małgorzata Kul. Poznan: University of Poznan, pp. 319–24.
- Mennen, Ineke, Felix Schaeffler, and Gerard Docherty. 2012. Crosslanguage difference in f0 range: A comparative study of English and German. *Journal of the Acoustical Society of America* 131: 2249–60. [\[CrossRef\]](#)
- Mennen, Ineke, Felix Schaeffler, and Catherine Dickie. 2014. Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition* 36: 303–29. [\[CrossRef\]](#)
- Moosmüller, Sylvia, Carolin Schmid, and Julia Brandstätter. 2015. Standard Austrian German. *Journal of the International Phonetic Association* 45: 339–48. [\[CrossRef\]](#)

- Niebuhr, Oliver. 2007. The signalling of German rising-falling intonation categories: The interplay of synchronization, shape, and height. *Phonetica* 64: 174–193. [CrossRef]
- Nolan, Francis. 2006. Intonation. In *Handbook of English Linguistics*. Edited by Bas Aarts and April McMahon. Oxford: Blackwell, pp. 433–57.
- Nooteboom, Sieb, and Jacques Terken. 1982. What makes speakers omit pitch accents? An experiment. *Phonetica* 39: 317–36. [CrossRef]
- Odlin, Terence. 1989. *Language Transfer: Cross-Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Odlin, Terence. 2006. Cross-Linguistic Influence. In *The Handbook of Second Language Acquisition*. Edited by Catherine J. Doughty and Michael H. Long. Malden: Blackwell, pp. 436–86.
- Osborne, Denise M., and Miquel Simonet. 2021. Foreign-language phonetic development leads to first-language phonetic drift: Plosive consonants in native Portuguese speakers learning English as a foreign language in Brazil. *Languages* 6: 112. [CrossRef]
- Penfield, Wilder, and Lamar Roberts. 1959. *Speech and Brain Mechanisms*. Princeton: Princeton University Press.
- Pešková, Andrea. 2020. L2 Spanish and Italian Intonation: Accounting for the Different Patterns Displayed by L1 Czech and German Learners. Habilitation thesis, Osnabrück University, Osnabrück, Germany.
- Peters, Jörg. 2018. Phonological and semantic aspects of German intonation. *Linguistik Online* 88: 85–107. [CrossRef]
- Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, Indiana University Linguistics Club, Bloomington, IN, USA.
- Reubold, Ulrich, Sanne Ditewig, Robert Mayr, and Ineke Mennen. 2021. The effect of dual language activation on L2-Induced changes in L1 speech within a code-switched paradigm. *Languages* 6: 114. [CrossRef]
- Sanchez, Covadonga. 2020. The Production and Perception of Subject Focus Prosody in L2 Spanish. Ph.D. dissertation, University of Massachusetts, Amherst, MA, USA. [CrossRef]
- Sancier, Michele L., and Carol A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25: 421–36. [CrossRef]
- Sanker, Chelsea, Sarah Babinski, Lyn Burns, Marisha Evans, Juhyae Kim, Slater Smith, Natalie Weber, and Claire Bower. 2021. (Don't) Try This at Home! The Effects of Recording Devices and Software on Phonetic Analysis. Lingbuzz: Lingbuzz/005748. Available online: <https://lingbuzz.net/lingbuzz/005748> (accessed on 7 February 2021).
- Schauffler, Nadja. 2021. Alternation Preferences and Focus Marking. Ph.D. dissertation, University of Stuttgart, Stuttgart, Germany.
- Schmid, Carolin, and Sylvia Moosmüller. 2013. Gender differences in the phonetic realization of semantic focus. In *Proceedings of the Prosody–Discourse Interface Conference 2013 (IDP 2013)*. Edited by Piet Mertens and Anne Catherine Simon. Leuven: University of Leuven, pp. 119–23.
- Sharwood Smith, Michael, and Eric Kellerman. 1986. *Crosslinguistic Influence in Second Language Acquisition*. New York: Pergamon Press.
- Siddins, Jessica, and Ineke Mennen. 2019. Pitch accent realisation in Austrian German. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Canberra: Australasian Speech Science and Technology Association Inc., pp. 2846–50.
- Silverman, Kim, and Janet Pierrehumbert. 1990. The timing of prenuclear high accents in English. In *Papers in Laboratory Phonology I*. Edited by John Kingston and Mary Beckman. Cambridge: Cambridge University Press, pp. 72–106.
- Stoehr, Antje, Titia Benders, Janet G. van Hell, and Paula Fikkert. 2017. Second language attainment and first language attrition: The case of VOT in immersed Dutch–German late bilinguals. *Second Language Research* 33: 483–518. [CrossRef] [PubMed]
- Tobin, Stephen J., Hosung Nam, and Carol A. Fowler. 2017. Phonetic Drift in Spanish-English Bilinguals: Experiment and a Self-Organizing Model. *Journal of Phonetics* 65: 45–59. [CrossRef] [PubMed]
- Trofimovich, Pavel, and Wendy Baker. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30. [CrossRef]
- Ueyama, Motoko. 1997. The phonology and phonetics of second language intonation: The case of “Japanese English”. In *Proceedings of the Fifth European Conference on Speech Communication and Technology 5*. Edited by George Kokkinakis, Nikos Fakotakis and Evangelos Dermatas. Rhodes: University of Patras, pp. 2411–14.
- Ulbrich, Christiane. 2005. *Phonetische Untersuchungen zur Prosodie der Standardvarietäten des Deutschen in der Bundesrepublik Deutschland, in der Schweiz und in Österreich*. Frankfurt am Main: Peter Lang.
- Ulbrich, Christiane, and Mikhail Ordin. 2014. Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics* 45: 26–42. [CrossRef]
- Van Hell, Janet G., and Ton Dijkstra. 2002. Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin and Review* 9: 780–89. [CrossRef]
- Willems, Nico. 1982. *English intonation from a Dutch point of view*. Dordrecht: Foris.
- Yoon, Tae-Jin, Sandra Chavarria, Jennifer Cole, and Mark Hasegawa-Johnson. 2004. Intertranscriber Reliability of Prosodic Labeling on Telephone Conversation Using ToBI. In *Proceedings of the International Conference on Spoken Language Processing (Interspeech 2004)*. Jeju Island: ISCA, pp. 2729–32.

Article

Intonation Patterns Used in Non-Neutral Statements by Czech Learners of Italian and Spanish: A Cross-Linguistic Comparison

Andrea Pešková

Department of Romance Studies, The Faculty of Arts, Osnabrück University, 49074 Osnabrück, Germany; apeskova@uos.de

Abstract: The objective of the study is to contribute to our understanding of the acquisition of second language intonation by comparing L2 Italian and L2 Spanish as produced by L1 Czech learners. Framed within the L2 Intonation Learning theory, the study sheds light on which tonal events tend to be successfully learnt and why. The study examines different types of non-neutral statements (narrow focus, statements of the obvious, *what*-exclamatives), obtained by means of a Discourse Completion Task. The findings show that the two groups diverge significantly in producing the nuclear pitch accents L+H* (L2 Spanish) and (L+)H*+L (L2 Italian), which is indicative of a target-like realization in each language. However, the learners struggle with the acquisition of the target boundary tones HL% and L!H% in L2 Spanish and prenuclear pitch accents in both Romance varieties. It is speculated that this is due not only to difficulties in acquiring semantic or systemic dimensions, but also to perceptual salience and frequency effects. In addition, the study explores individual differences and reveals no significant effects of the time spent in an L2-speaking country, the age of learning and the amount of active use of a foreign language on accuracy in L2 production.

Keywords: L2 intonation; L1-to-L2 transfer; L2 Intonation Learning Theory; AM model of intonational phonology; non-neutral statements

Citation: Pešková, Andrea. 2022. Intonation Patterns Used in Non-Neutral Statements by Czech Learners of Italian and Spanish: A Cross-Linguistic Comparison. *Languages* 7: 282. <https://doi.org/10.3390/languages7040282>

Academic Editors: Ineke Mennen and Laura Colantoni

Received: 27 March 2022

Accepted: 20 October 2022

Published: 2 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

There is a certain paradox in the acquisition of intonation. On the one hand, intonation is said to be very difficult if not impossible for L2 adult speakers to acquire (Chun 1998, p. 74), and on the other, anecdotal evidence suggests that intonation is a feature of language we pick up rapidly when we learn a new language or dialect. This apparent contradiction raises the question as to which features of L2 intonation are learnt and which are not, and why this is the case.

Many language contact studies claim that intonation is very sensitive to change or convergence (e.g., Matras 2009). Changes in intonation patterns are mostly understood as products of L2 pronunciation, “imperfect” acquisition and accommodation or imitation processes. A very often cited example here is the study by Colantoni and Gurlekian (2004) of contemporary *Porteño* Spanish, a Buenos Aires variety, which emerged from the contact between Spanish and various Italian dialects caused by mass immigration during the 19th and early 20th centuries. *Porteño* shares several features with Italian, such as earlier pitch alignment in prenuclear accents, prosodic focus marking and final “long falls” in declaratives (see also Gabriel et al. 2010, 2011; Gabriel and Kireva 2014). According to Colantoni and Gurlekian, all these features can be attributed to the socio-historical context, particularly the acquisition of Spanish as a second language by immigrants. Cross-linguistic influences are also reported in many other studies, which variously show that the contact-induced forms occur in the alignment of prenuclear pitch accents (e.g., O'Rourke 2004 on Peruvian Spanish influenced by Quechua), the realization of boundary tones of questions (e.g., Romera and Elordieta 2020 on Spanish in contact with Basque) or nuclear pitch accents in different types of sentences (e.g., Sichel-Bazin et al. 2015 for the Occitan dialect spoken in Cisalpine, which is in contact with Italian).

Similarly, previous research on L2 intonation reports that L2 production shows considerable transfer from L1, as well as mixed patterns, and that learners seem to experience particular difficulties in the semantic but also the phonetic dimension (e.g., Mennen 2015; Colantoni et al. 2016; Pešková 2019; see Section 3.1 for details). Many studies provide evidence that L2 intonation is characterized by L1 features (e.g., Ueyama 1997; Gabriel and Kireva 2014; Nicora et al. 2018; Méndez Seijas 2018; Pešková 2021) and that even advanced learners can still show influences of L1 prosody in their L2 (e.g., van Maastricht et al. 2016). Yet, it is important to point out that intonation is also—at least to a certain degree—learnable (e.g., Mennen 2004; Trofimovich and Baker 2006; De Leeuw et al. 2012; Mennen et al. 2014), with learners increasingly able over time to approximate the variety to which they are exposed. The current contribution adopts an innovative approach that attempts to uncover how L1-to-L2 intonational transfer works and what role prosodic similarities and dissimilarities between languages play. It does so by comparing the intonation of two different L2s, Italian and Spanish, as produced by learners who have the same L1, Czech. After obtaining spoken data by means of a Discourse Completion Task adapted for intonation research (Prieto 2001; Prieto and Roseano 2010; Frota and Prieto 2015), intonation contours are examined within the Autosegmental Metrical (AM) model of intonational phonology (Pierrehumbert 1980).

Non-neutral statements were selected for the present study for various reasons. According to the Interface Hypothesis (Sorace 2011), adult L2 learners tend to have particular difficulty acquiring phenomena located at the external interfaces, in this case the interface of syntax with pragmatics and prosody. As we will see, the two Romance languages under study and Czech differ in the intonational realization of non-neutral statements. In narrow focus statements, for instance, Spanish focus exhibits a rising pattern, whereas Italian focus exhibits a falling or rising–falling pattern of the nuclear pitch accents. In Czech, the phonetic realization of focus is more similar to Spanish. Hence, non-neutral statements provide an interesting source to verify the L1-to-L2 transfer hypothesis and to see to what extent the Czech learners of Italian would differ from the Czech learners of Spanish. The results, discussed within Mennen’s (2015) L2 Intonation Learning theory, help us understand the degree of success with which the learners are able to approximate the target patterns and whether they struggle with patterns that are either absent in their L1 or present but used to convey a different meaning. Section 2 provides details of this cross-linguistic comparison, which we use to make several predictions about the acquisition of L2 intonation.

The “learnability” of L2 sound patterns also depends on a range of internal and external factors. Language-dependent factors such as the position of tonal events within the utterance, the type of sentence and differences or similarities between the L1 and L2 can have either a positive or a negative impact on the production of native-like intonation. Among the language-independent factors which have been claimed to affect L2 speech are the age of acquisition, the quality and quantity of input, phonological awareness, length of residence in an L2-speaking country and a range of personal factors such as proficiency level, motivation or musicality (see, e.g., Colantoni et al. 2015; Derwing and Munro 2015 for a summary of such factors and further readings). The present study looks closer at individual differences among learners, focusing in particular on the age when L2 learning was initiated, time spent in an L2-speaking country and the amount of L2 exposure per week, factors that can have an impact on L2 speech production (see, e.g., Piske et al. 2001).

The paper is structured as follows. First, the intonation of Czech, Italian and Spanish non-neutral statements is compared in Section 2. Section 3 presents Mennen’s (2015) L2 Intonation Learning theory and formulates several research questions relevant for this study. Section 4 describes the experimental design and participants in the production experiment of the study and then explains the data analysis procedure. The results are offered in Section 5 and discussed in Section 6. The contribution ends with concluding remarks in Section 7.

2. Czech, Italian and Spanish Non-Neutral Statements in Contrast

This section introduces the main suprasegmental properties of the languages under study. All three languages are regarded as intonation languages and use pitch post-lexically for grammatical and pragmatic purposes. However, in contrast to Italian and Spanish, both lexical stress languages, Czech, as a fixed-stress language with initial prominence, has been proposed to be—in Féry’s (2010) terminology—a *phrase language* such as, for instance, French (Pešková 2017, forthcoming). According to Jun’s (2005, 2014) model of prosodic typology, Italian and Spanish are *head-prominence languages*, assigning phrase-level prominence by the phrase head, which is determined by a metrically strong syllable, whereas Czech is a *head/edge-prominence language*, since it marks phrase-level prominence by both the phrase head (stressed syllable, T*) and the edge at the phrase level (Ta), corresponding to Accentual Phrases (APs) (examples illustrating this feature are given in Section 2.1).

With regard to intonational properties, Italian and Spanish can be described as intonationally “richer” in comparison to Czech, in that they have a greater number of different pitch accents and boundary tones, which can be combined in different numbers of nuclear configurations (see Pešková 2017 for Czech; Gili Fivela et al. 2015 for Italian; Prieto and Roseano 2010 for Spanish). One nuclear configuration can convey one or more meanings in every language, but languages may differ substantially in this respect. For example, the L*+H pitch accent, which is phonetically realized as a F0 valley on the accented syllable with a subsequent rise on the postaccentual syllable, is a typical prenuclear pitch accent of information-seeking yes–no questions in (Peninsular) Spanish (Estebas-Vilaplana and Prieto 2010). The same tone represents a focus nuclear accent in Czech non-neutral statements or yes–no questions (Pešková 2017), whereas in various Italian regional varieties, it occurs in the nuclear position of different types of sentences, including yes–no questions, wh-questions and exclamatives (Gili Fivela et al. 2015).

It must be added that there is considerable variation in terms of intonation among the various regional varieties. It should therefore be clarified that the Spanish variety I refer to is that spoken on the Iberian Peninsula, known as Peninsular Spanish, because that is the variety to which most of the Spanish-learning participants in this study were exposed. As for Italian, most of the Italian-learning participants were exposed to various northern varieties of Italian (see Section 4.2 for details).

Since the comparison of intonation patterns serves as a point of departure for the examination of L2 pitch contours, the following subsections present the intonational contours of non-neutral statements in L1 Czech (Section 2.1), L1 Italian (Section 2.2) and L1 Spanish (Section 2.3).

2.1. Non-Neutral Statements in Czech

The realization of the nuclear configuration in non-neutral statements with contrastive focus typically consists of a low tone on the stressed syllable with a rise on the posttonic syllable (L*+H), followed by a low boundary tone (L%) (Figure 1). An alternative variant of this pattern is a rising tone on the accented syllable (labelled as L+H*), when the word is disyllabic (Pešková forthcoming). All items, contexts and examples are taken from the corpus of the present study, which is presented in Section 4.

Whereas we find the same nuclear configuration, L*+H L%, in statements of the obvious (Figure 2), the tonal structure of *what*-exclamative statements resembles that of wh-questions in Czech, in which the contour starts mostly with a high or rising tone at the very beginning of the utterance and ends with a fall (Figure 3).

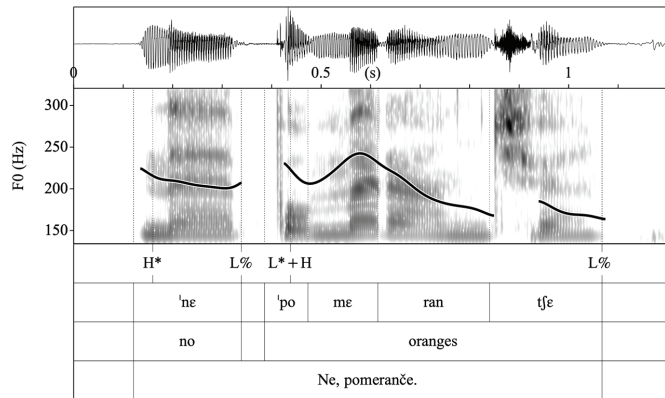


Figure 1. Waveform, spectrogram and F0 trace of the contrastive focus statement *No, oranges!* in L1 Czech produced with L*+HL%.

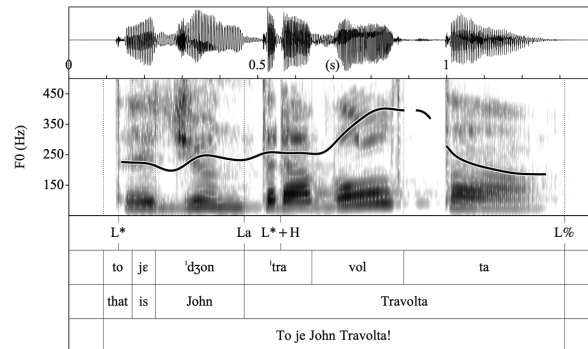


Figure 2. Waveform, spectrogram and F0 trace of the non-neutral statement of the obvious *It is John Travolta!* in L1 Czech produced with L*+HL%.

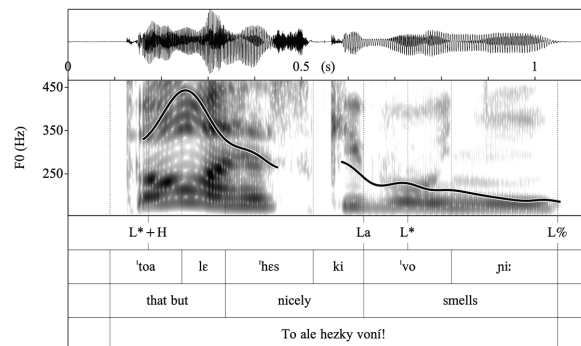


Figure 3. Waveform, spectrogram and F0 trace of the *what*-exclamative statement *What a lovely smell!* in L1 Czech produced with L* L%.

2.2. Non-Neutral Statements in Italian

In most Italian varieties, non-neutral statements of contrastive focus and statements of the obvious are characterized by a nuclear pitch accent that is phonetically realized as a rise and a fall located within the stressed syllable (Figure 4). I use a tritonal phonetics-based

label for this pattern, L+H*+L, in order to capture the exact movement of the pitch track within the stressed syllable. In Gili Fivela et al. (2015), different phonology-based labels, such as H*+L, L+H* or L*+>H, are proposed for this pattern, depending on the regional variety. In my analysis, the H*+L pitch accent is treated as another variant of L+H*+L, in which the high peak is aligned with the left edge of the stressed syllable (Figure 5). The phonetics-based labels applied in this study are presented in Section 4.

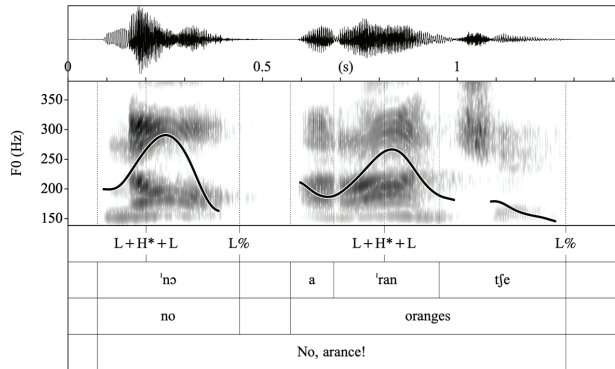


Figure 4. Waveform, spectrogram and F0 trace of the contrastive focus statement *No, oranges!* in L1 Italian produced with L+H*+L L%.

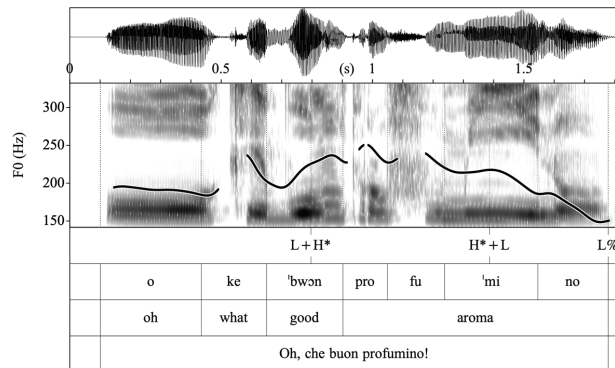


Figure 5. Waveform, spectrogram and F0 trace of the *what*-exclamative statement *What a lovely smell!* in L1 Italian produced with H*+L L%.

As for *what*-exclamative sentences, they show more variation in the nuclear configuration pattern; however, here again the most general pattern is what I label L+H*+L L% (for further details see Gili Fivela et al. 2015). The L*+>H pitch accent has been reported for various northern varieties in Gili Fivela et al. (2015) and was detected in the present L1 data too (Figure 6). This pitch accent is phonetically realized as a “F0 fall to the [tone bearing unit] followed by the rise to an early peak in the tonic syllable” (Gili Fivela et al. 2015, pp. 164–65).

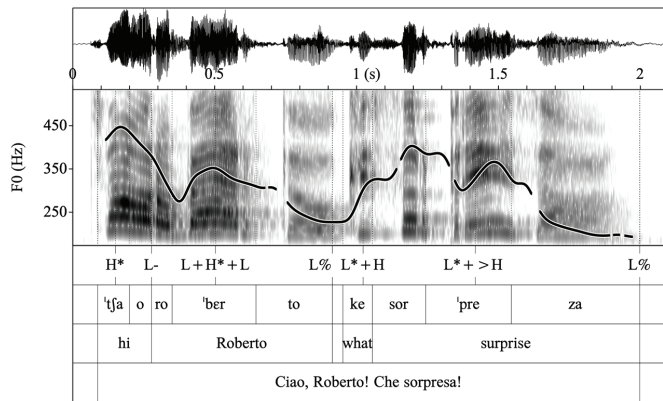


Figure 6. Waveform, spectrogram and F0 trace of the *what*-exclamative statement *Hi, Roberto! What a surprise!* in L1 Italian produced with L*+>H L%.

2.3. Non-Neutral Statements in Spanish

According to previous research (see Prieto and Roseano 2010), (Peninsular) Spanish non-neutral statements with a contrastive focus and exclamative statements are realized with the nuclear rising L+H* pitch accent and a low boundary tone L% (Figure 7). The L+H* focus accent has been observed in many Spanish varieties.

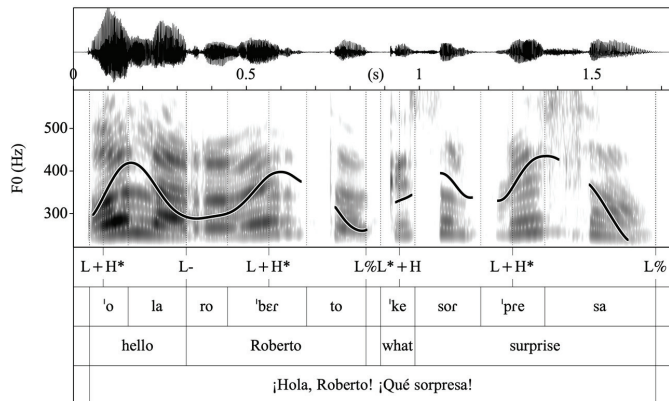


Figure 7. Waveform, spectrogram and F0 trace of the *what*-exclamative statement *Hi, Roberto! What a surprise!* in L1 Spanish produced with L+H* L%.

In contrast to Czech and Italian, statements of the obvious in Spanish show two different patterns: L+H* L!H% (Figure 8) and L+H* HL% (Figure 9). Though the latter contour is not attested in non-neutral statements in Prieto and Roseano (2010), it was produced systematically by the control participants in the present study in the context presented in Figure 9 (see Section 4.1 for details).

It should be noted that languages can be phonetically similar but phonologically different. An example is given with the trisyllabic name *Travolta*. In Czech, the accent is on the first syllable *tra* (Figure 2), but the rise coincides with the position of the stress in Romance languages, which is on the second syllable *vol* (see Spanish example in Figure 9). Furthermore, the languages differ in the realization of the prenuclear position, here attested in the *Travolta* sentence. Czech shows a tendency towards a pre-focal tonal compression (Figure 2) or it realizes an L*+H pitch accent in the very initial position, whereas Spanish

typically uses a prenuclear pitch accent realization, with a rising tone with the peak on the postaccentual syllable (L+<H*) of the name *John*. The Italian counterpart is very similar to Spanish, consisting of a rising–falling–rising–falling pitch track over the whole utterance. However, the tune–text association is different. In the prenuclear position, Italian uses an L+H* pitch accent and the nuclear configuration is L+H*+L L% (Figure 10). The copular verb is deaccented in both Italian and Spanish (marked with *).

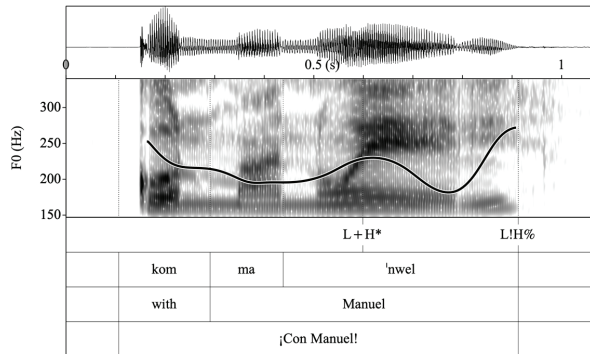


Figure 8. Waveform, spectrogram and F0 trace of the statement of the obvious *To Manuel!* in L1 Spanish produced with L+H* L!H%.

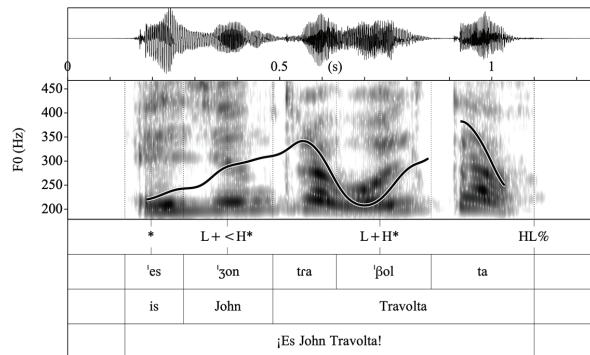


Figure 9. Waveform, spectrogram and F0 trace of the statement of the obvious *It is John Travolta!* in L1 Spanish produced with L+H* HL%.

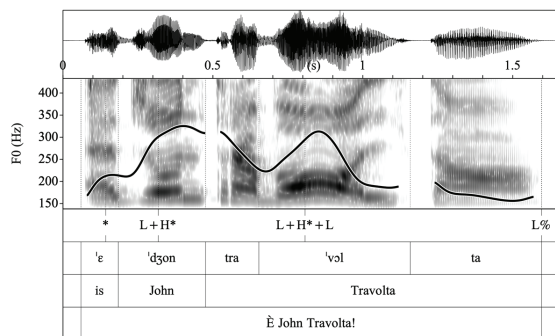


Figure 10. Waveform, spectrogram and F0 trace of the statement of the obvious *It is John Travolta!* in L1 Italian produced with L+H*+L L%.

2.4. Brief Summary

The most important characteristics of non-neutral statements in the three languages can be summarized as follows (see Table 1): Czech displays two main patterns, L*+H L% and L* L%, Italian varieties exhibit mostly a (L+)H*+L L% nuclear configuration and, finally, Spanish presents a nuclear pitch accent L+H* with three different boundary tones, L%, L!H% and HL%.

Table 1. Summary of the main tonal patterns of non-neutral statements in Czech, Italian and Spanish.

Tonal Patterns	L%	L!H%	HL%
L*+H	Czech		
L*	Czech		
(L+)H*+L	Italian		
L+H*	Spanish	Spanish	Spanish

Based on these tonal differences, Czech learners of Italian have to learn a new pattern, (L+)H*+L, which does not exist in Czech. Boundary tones of non-neutral statements in Italian should not present any difficulty for learners, since they end with a low pattern (L%) in both languages. In contrast, Czech learners of Spanish might have more difficulties with the acquisition of the boundary tones given that HL% does not exist in Czech and L!H% (or its variant LH%) is used in different types of yes–no questions. As for the Spanish nuclear pitch accent, it can be expected that Czech learners of Spanish are able to approximate the target pattern quite well, since the L+H* tone exists as a phonetic variant of L*+H in Czech.

Before presenting the research questions that arise from this cross-linguistic comparison in Section 3.2, the L2 Intonation Learning theory (Mennen 2015) is presented in Section 3.1.

3. Acquisition of L2 Intonation

3.1. L2 Intonation Learning Theory

This section provides an overview of the main ideas and postulations of Mennen’s (2015) L2 Intonation Learning theory (LILt), the theoretical framework for the present study. Cross-linguistic comparison is an essential point of departure for this model. This is because, in order to understand the acquisition of L2 intonation and predict potential difficulties, it is necessary to know how the L1 and L2 differ from each other. As already noted in Section 2, Czech has fewer patterns than Italian or Spanish, and Czech learners may thus be assumed to face difficulties in acquiring some of the intonation patterns of these two languages. Adapted from Ladd (1996, 2008), LILt assumes that languages differ across four dimensions: (1) systemic, (2) phonetic, (3) semantic and (4) frequency. These dimensions can help to understand where L2 deviations from the target L1 norm are likely to occur.

The *systemic* dimension refers to the inventory of categorical phonological elements. Here, the question is whether the L2 learners can produce those tonal events that do not form part of their L1 tonal inventory. For example, in Mennen et al. (2010), Italian and Punjabi learners of London English do not use the target complex pitch accents H*!LH and L*HL, which are present in the variety of English they were learning but not in their respective L1s.

The *phonetic* dimension is about how tonal tunes are implemented phonetically. Previous research has shown that learners have particular problems with the realization of target-like tonal alignment, tonal scale and tonal slope. For example, pitch accents in the initial position of yes–no questions in L1 (Peninsular) Spanish exhibit a wider pitch range when compared with L2 Spanish produced by Czech learners, reflecting transfer from their L1 (Pešková 2020). As another example, Mennen (2004) found that Dutch learners of Greek tended to align the prenuclear peaks in declarative sentences much earlier than L1

Greek speakers, in spite of very long exposure to the L2. Nevertheless, learners can also *overshoot* the L1 norms, as [Santiago and Delais-Roussarie \(2015\)](#) showed for L1 German and L1 Spanish learners of French, who tended to exaggerate the rises at the right edge of non-final clauses in French.

The *semantic* dimension refers to the functionality or distribution of the tones. The same tone can be used in different contexts and radically change the meaning. An example is the high-rising terminal in English statements (*uptalk*), which should be avoided in any L2 where statements require a falling pattern (see, e.g., [Méndez Seijas 2018](#) for L2 Spanish). There is also evidence for differences across regional varieties of a single language (of course, this holds for other dimensions too). For instance, in most Spanish varieties, yes–no questions are signaled by rising pitch, whereas Caribbean varieties use falling patterns ([Hualde and Prieto 2015](#)). Many studies report difficulties in this dimension and reveal patterns transferred from learners' L1s. For example, [Nicora et al. \(2018\)](#) identify the difficulties that Irish English-speaking learners of Italian have with yes–no questions, in that they tend to apply L1-based H+H* H% and H+H* L% patterns instead of the target contours, H*+L LH% and L*+H L%. The study attributes this to a low phonological/semantic awareness on the part of the learners; however, the same study demonstrates that explicit intonation training can help to improve L2 productions.

Finally, the *frequency* dimension concerns how frequently certain intonation patterns occur. Languages differ substantially in this regard. For instance, German learners of L2 Spanish have been found to use rising boundary tones more often in neutral wh-questions in L2 Spanish than do Czech learners of Spanish, who realize more falls ([Pešková 2021](#)). This result is clearly due to influences from the respective L1s. The falling pattern in Czech is also common in vocatives (initial calls), another feature that is easily transferred to L2 Italian and L2 Spanish ([Pešková 2019](#)).

Apart from a characterization of cross-language differences along four dimensions of intonation, LILt makes five theoretical generalizations and predictions that seek to explain why difficulties in learning foreign intonation acquisition arise and which principles “govern the acquisition process of intonation” ([Mennen 2015](#), p. 178). I summarize the five predictions very briefly below.

1. The first assumption is that many difficulties are connected with perceptual difficulties, much as [Flege's \(1995\)](#) Speech Learning Model and [Best and Tyler's \(2007\)](#) Perceptual Assimilation Model of L2 speech learning assume for L2 segments. In addition, Mennen offers further clarification for individual learner difficulties in L2 intonation acquisition, suggesting that they may be due to factors such as an inability to articulate certain aspects of the L2 and store them in acoustic memory (p. 179).
2. LILt postulates that the position and context in which intonation contrasts between L1 and L2 occur are important and need to be controlled for (p. 179). [Pešková \(2020\)](#) shows that L1 Czech learners of L2 Spanish or L2 Italian have more difficulty acquiring the initial and medial prenuclear positions when compared with nuclear positions in neutral statements. Possible explanations might be transfer from the L1, but also include the fact that the positions are less perceptually salient and that nuclear configurations are more important bearers of meaning.
3. The model predicts fewer intonation deviations when exposure to the L2 starts at an earlier age (pp. 180–81). The age of learning (AOL) has been shown to have an important impact on the degree of foreign accent and L2 production in general.
4. The fourth assumption is connected to developmental sequences and the prediction that, in the course of time, learners learn to approximate the targets more closely. For instance, [Mennen et al. \(2010\)](#), so far one of very few longitudinal studies on intonation development, observed an improvement after 30 months in the ability of Punjabi and Italian learners of English to approximate intonational targets.
5. The last assumption of LILt is that L1 and L2 categories exist in a common phonological space and that linguistic influence is bi-directional (p. 184). Interaction between the L1 and L2 in the form of assimilation or mixed patterns has been observed, especially

in individuals who have been living in an L2-speaking country for a long period (e.g., [Flege 1987](#) for segments; [De Leeuw et al. 2012](#) for intonation).

3.2. Research Questions

First of all, the present study aims to test in what way the two groups of learners differ from each other and how they are able to approximate the target patterns. As such, we seek to answer whether all tonal events are acquired equally or whether certain tonal events present specific difficulties for learners. The study tests the position (prenuclear, nuclear, boundary) in which intonation contrasts appear and examines whether the intonation dimensions constitute disparate degrees of difficulty, as LILt suggests ([Mennen 2015](#), p. 183). Taking into consideration the cross-language comparison presented in Section 2 and focusing on systemic and semantic dimensions, we can predict that Czech learners will have problems with (1) the acquisition of Spanish HL% and Italian (L+)H*+L, tones absent in their L1 inventory (systemic dimension), and (2) the appropriate use of L!H% in L2 Spanish non-neutral statements, since this tone exists in Czech yes–no questions (semantic dimension).

Secondly, given the large amount of variability in the data, we examine whether the age of learning (AOL), the length of residence (LOR) and the amount of active use of L2 (AUL) can explain L2 deviations across speakers. As mentioned above, LILt predicts a better performance when language exposure starts earlier. By the same token, having spent time in an L2-speaking country is reported in many previous studies to have enhancing effects on L2 speech (e.g., [Flege et al. 1997](#); [Henriksen 2012](#)), although there is no consensus on exactly how long the period should be. Some studies on L2 phonology even report a weak impact or no impact at all (e.g., [Oyama 1976](#); see [Piske et al. 2001](#) for an overview of research on this variable). The present study discusses whether the general assumption “the younger, the better” is justified and whether learners with a longer LOR perform more accurately in terms of intonation than those learners with less or no experience abroad. And, finally, it is tested whether the amount of input, here the active use of L2, correlates with intonational accuracy (meaning that the learner can appropriately produce a tonal pattern pertaining to the target language, regardless of whether it is present in their L1 or not).

Summarizing the above, the following questions are addressed:

- Q1: In what way do L2 Italian learners differ from L2 Spanish learners?
- Q2: In what way do the two L2 groups differ from L1 Spanish and Italian controls?
- Q3: Is it more difficult for L2 learners to acquire new patterns (systemic dimension) or put known patterns to new uses (semantic dimension)?
- Q4: Which individual factors (AOL, LOR, AUL) explain accuracy in L2 patterns?

4. The Production Experiment on the Intonation of L2 Italian and L2 Spanish

4.1. Experimental Design

The data were collected within the scope of a larger study on L2 Italian and L2 Spanish phonology, which comprised a combination of five different tasks (see [Pešková 2020](#), following [Pustka et al. 2018](#)). For the purposes of the present study, only data from the Discourse Completion Tasks (DCTs) were selected and analyzed. DCTs are a data-gathering method that was originally developed for the study of pragmatics ([Blum-Kulka et al. 1989](#)) and later became popular in L1 intonation research (see, e.g., [Frota and Prieto 2015](#) for Romance languages). In the task, participants are presented with a set of daily situations and asked to react accordingly (e.g., “Imagine that you see Natalia, a friend of yours, on the other side of the street. Call her”). The selection of items for the present study was based on the Spanish version of the DCT employed in [Prieto and Roseano \(2010\)](#), but it was adapted for L2 research and performed in two steps. In the first step, the participant reacted to the situations presented spontaneously (as expected in the original DCT procedure). In the second step, the participant was given a prepared answer embedded in the context and was asked to say it aloud as naturally as possible. Only the answers from the second step

were included in the subsequent analysis, since they showed more fluency and sounded more natural in comparison to the spontaneous reactions produced in the first step. The advantage of this process was also that the utterances were identical across all participants tested, allowing a greater degree of control over factors other than intonation, and thus ideal for comparison purposes.

The full set of DCTs used in Pešková's (2020) prompt scenarios included 25 situations (in Spanish or Italian) intended to elicit a variety of different sentence types, such as statements, vocatives, exclamatives, imperatives, yes–no questions and wh-questions, with either neutral or biased meanings (e.g., counter-expectational echo questions, confirmatory questions, statements of the obvious or command imperatives). The five following scenarios leading to the production of non-neutral statements represent the focus of the present study.¹

6. Contrastive focus on the object
7. Situation (S04): "You enter a store where the shop attendant is a little hard of hearing. You tell her that you would like a kilo of oranges, but she doesn't hear you well and asks you if you want lemons. Tell her that you want oranges."
8. Answer:
9. It. *No, arance.*
10. Sp. *No, naranjas.*
11. 'No, oranges.'
12. Statement of the obvious (surprise nuance)
13. Situation (S05): "You are with a friend and you tell him/her that Mary, a mutual friend of yours, is getting married. Your friend asks you who she is marrying. You're surprised that s/he doesn't know, because everyone knows that Mary is planning to marry her long-time boyfriend, Manuel. Tell him/her that she's getting married to Manuel."
14. Answer:
15. It. *Con Manuele!*
16. Sp. *¡Con Manuel!*
17. 'To Manuel!'
18. Statement of the obvious (exclamative nuance)
19. Situation (S15): "You show a picture of a very famous actor to your friend. S/he asks you who it is. This astonishes you, because everybody knows who the actor is. How do you react?"
20. Answer:
21. It. *(Ma come!) È John Travolta!*
22. Sp. *¡Es John Travolta!*
23. 'It's John Travolta!'
24. *What-exclamative sentence (positive evaluation)*
25. Situation (S16): "You are invited for dinner at your friend's place and when you arrive you smell a delicious aroma. What do you say to your friend?"
26. Answer:
27. It. *Che buon profumino!*
28. Sp. *¡Qué rico olor!*
29. 'What a lovely aroma!'
30. *What-exclamative sentence (positive surprise)*
31. Situation (S18): "Somebody knocks on the door. You open it and there is your friend Robert. You have not seen him for years. How do you react?"
32. Answer:
33. It. *Ciao, Roberto! Che sorpresa!*
34. Sp. *¡Hola, Roberto! ¡Qué sorpresa!*
35. 'Hi, Roberto! What a surprise!'

Before the experiment, all learners gave written consent to be recorded and filled out a questionnaire to provide information about their language background such as the age

they had started learning the L2 in question, how much of the L2 they used per week, how long they had lived in a country where that L2 was spoken and their knowledge of other foreign languages.

After they had carried out the DCT experiment in the L2, the learners were recorded performing the DCT again, this time in their L1, Czech.

4.2. Participants

The study included 52 participants: 20 Czech learners of Italian, 20 Czech learners of Spanish and 12 controls (six L1 Italian speakers, six L1 Spanish speakers). All learners had grown up with Czech as their only L1 and had started to learn Italian or Spanish in a formal setting—mostly in secondary school or university—in the Czech Republic. None of the learners had received any pronunciation training and were not aware of the aims of the production experiment in which they participated for this study.

The learners were selected according to their L2 proficiency levels, as indicated by the level of the courses they were attending at the time of the experiment; their proficiency ranged from B1 to C2 according to the *Common European Framework of Reference for Languages* (CEFR 2018). I use data from a previous study (Pešková 2020) that aimed to measure differences in intonation acquisition across different proficiency levels which might suggest improvement over time. Basic-level learners (A1–A2) were not included in the study because their spoken output is mostly limited to very simple structures and is more strongly characterized by transferred phenomena. It should be added that Pešková (2020) reports that proficiency level shows only a slight correlation with L2 intonation acquisition. The present study also reveals a non-significant effect between B and C levels ($\chi^2(13) = 13.57$, $p = 0.406$) and speculates that this may be due to the fact that the B2 and C1 levels of the participants were not in reality very far apart and the assignation of learners to a particular proficiency level is generally based upon grammatical and lexical skills, rather than phonological competencies. Another possible explanation is that intonation becomes fossilized at the B2 level or even earlier. For this reason, the students' purported language proficiency level is not taken into account in the present paper.

Let us now turn our attention to inter-participant variability. As we will see, L2 classes, especially in the learners' home country, are characterized by great irregularities (see Table A1 in Appendix A for details). Starting with the variety of Italian or Spanish to which learners had been exposed, they reported having non-native, as well as native, instructors. Spanish-speaking instructors mostly came from mainland Spain (mostly Madrid), though a few came from the Canary Islands or various Latin American countries (Mexico, Chile, Peru). As for Italian-speaking instructors, the majority came from the northern dialect areas of Italy (Turin, Verona, Milan). For this reason, the two groups of control speakers consisted of six L1 speakers of mainland Iberian Spanish (4F, 2M) (Madrid, Ciudad Real) and six L1 speakers from the north of Italy, mostly Turin (4F, 2M). It should be mentioned that it was very difficult to find learners who had been exposed to a single dialect. Although the majority of participants had had more contact with Peninsular Spanish, or northern Italian varieties, they also reported having native-speaking contacts (mostly friends) from other dialect areas. With regard to the learners' L1, the participants spoke Standard Czech and came from two main dialectal areas (Bohemian and Moravian). According to the available descriptions (see, e.g., Palková 1994) and our L1 control data of the present study, there are no substantial differences between the two varieties in terms of intonation that are relevant for the present study.

Regarding the age of learning (AOL), the participants began to learn the L2 at or after puberty. The average AOL was 17 for L2 Spanish learners and 19 for L2 Italian learners, respectively. The lowest AOL was 10 (in the case of three learners) and the highest AOL was 33 and 35 (in the case of another three learners). Other factors that were difficult to control for were the sex of participants (female participants predominated), their knowledge of other foreign languages (mostly, but not exclusively, English)² and the degree of active exposure to or use of Italian/Spanish per week. About half of them used Italian/Spanish

less than three hours a week, whereas the other half used it more than ten hours weekly. The amount of time spent in an L2-speaking country also differed considerably from one participant to another. Some of them had spent only a short period of time abroad (e.g., holidays), whereas others had lived abroad for a full year or more. Few of the learners had no experience at all in an L2-speaking country.

All these factors can shape the features of an individual’s pronunciation (see Piske et al. 2001 for an overview and discussion). As already mentioned in Section 3.2, we next have a look at the role of individual differences and discuss three external factors that might have underlain L2 intonation deviations from target patterns.

4.3. Tonal Analysis

The data were transcribed first orthographically and then phonetically. The acoustic analysis was carried out with Praat software, version 6.1.48, (Boersma and Weenink 1992–2022) and the tonal annotation was done manually by the author, applying AM-based labels phonetically.³ This phonetic approach to labeling, which was oriented to the IPrA approach (Hualde and Prieto 2016), proved to be well suited for the analysis of L2 data. The broad tonal transcription (Elvira-García et al. 2016), in which the F0 course is described, was provided merely for practical purposes, that is, to help systematize and compare the patterns found in the L2 data. In the full corpus of recorded material, two monotonal pitch accents (H* and L*), five bitonal pitch accents (L*+H, L+H*, L+<H*, H+L*, H*+L) and two tritonal pitch accents (L+H*+L and H+L*+H)⁴ were identified (Figure 11). As for boundary tones, three monotonal boundary tones (H%, L%, !H%), three bitonal boundary tones (HL%, !LH%, LH%) and one tritonal boundary tone (LHL%) were identified in the data (Figure 12).

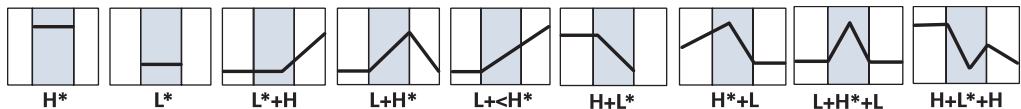


Figure 11. Schematic representation of pitch accents found in the data (Pešková 2020).

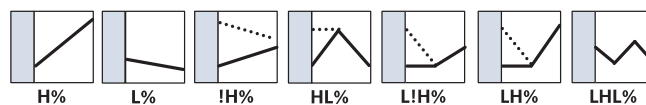


Figure 12. Schematic representation of boundary tones found in the data (the dotted line indicates alternative pitch tracks) (Pešková 2020).

The corpus for the present analysis consisted of 260 target non-neutral statements and 1040 tonal events, comprising prenuclear and nuclear pitch accents and boundary tones at the intonational phrase (IP) or intermediate phrase (ip) within the Prosodic Hierarchy (see, e.g., Selkirk 1984). The number also included the greeting (S18; ‘Hi, Roberto!’) and the negative particle *no* (S04; ‘No, oranges’), since they revealed some interesting tendencies in the two L2 varieties.

5. Results

5.1. Intonational Patterns

In the dataset, several substantial differences between the two L2 varieties were found. Given that we are interested in frequency distributions across groups and we are dealing with categorical data, differences are assessed using either chi-square or Fisher’s exact tests.

First of all, the frequency of use of particular pitch accents was calculated and showed different tendencies: while L+H* (a rising tone on the stressed syllable) predominated in both L1 and L2 Spanish (75% and 46%, respectively), more mixed patterns were produced in the two sets of Italian data. In L1 Italian data, 49% of all pitch accents were realized as either H*+L or its variant L+H*+L, followed by L+H* (24%). In the L2 Italian data, the

various patterns were more evenly distributed (Figure 13). It should be added that two cases of tritonal pitch accent H+L*+H were detected in the Italian data too, but they were clustered here into the L*+H group to keep the overall picture simple. The results revealed statistically significant differences between the two groups of learners ($\chi^2:(6) = 122,932$; $p < 0.001$), as well as between the L1 and L2 varieties (for Italian, $\chi^2:(5) = 16,287$; $p = 0.006$; for Spanish, $\chi^2:(6) = 14.87$, $p = 0.021$).

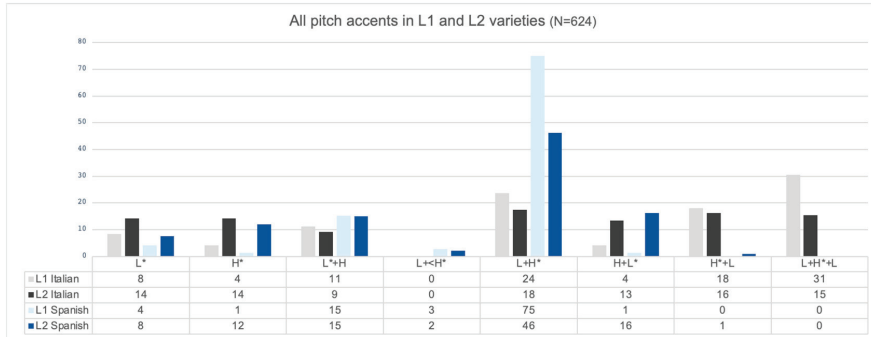


Figure 13. Tonal patterns of all pitch accents found in all L1 and L2 data (in %).

Now we take a closer look at the realization of pitch accents, grouping them according to whether they were located in prenuclear or nuclear position. Starting with the nuclear position, a crucial difference between Italian and Spanish varieties can be observed (Figure 14). Setting aside L*, H*, L*+H and L+<H*, which appeared with a very low frequency in the data or were completely absent, we can summarize the findings as follows: both Spanish groups showed a clear preference for a rising L+H* nuclear pitch accent (L1 Spanish: 100%; L2 Spanish: 62%). In the two Italian varieties, on the other hand, high–falling (H*+L) and rising–falling (L+H*+L) patterns predominated (L1 Italian: 73%; L2 Italian: 44%). This position also shows significant differences between L2 Spanish and L2 Italian ($\chi^2:(5) = 119,540$; $p < 0.001$), between L2 and L1 Italian ($\chi^2:(5) = 15,475$; $p = 0.009$) and between L1 and L2 Spanish (Fisher’s exact test; $p < 0.001$).⁵

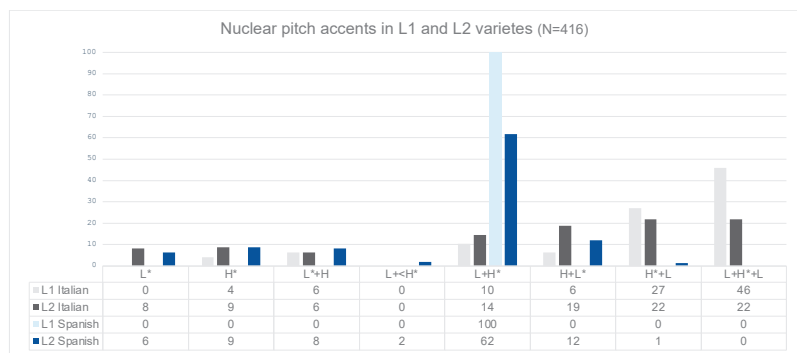


Figure 14. Tonal patterns of nuclear pitch accents found in all L1 and L2 data (in %).

There were also differences across groups in the production of prenuclear pitch accents, which displayed a great deal of variation (Figure 15). It should first be noted that the sample of prenuclear pitch accents in the data was very small, hence a certain degree of caution is required here. Prenuclear accents only occurred in the sentence *It is John Travolta* and in the very initial position of *what*-exclamative sentences. Nonetheless, some interesting

tendencies can be reported. First, L1 Italian controls produced at a high frequency two bitonal pitch accents, L+H* (50%) and L*+H (21%), and a monotonal L* pitch accent (25%). The latter was used in the sentence *È John Travolta* and indicates a deaccenting of the verb. L1 Spanish controls produced L*+H (46%) and L+H* (25%) with the highest frequency. In both control groups, L*+H was labeled at the beginning of the *what*-exclamative *What a surprise!* (S18). Considerable variation occurs in the L2 varieties too: L2 Spanish exhibited H+L* (25%), L*+H (23%) and L+H* (21%) with the highest frequency; L2 Italian learners produced L* (26%) with the highest frequency, closely followed in terms of frequency by the two monotonal patterns H* (25%) and L+H* (24%) ($\chi^2:(6) = 29,012; p < 0.001$; for all tonal realizations in the prenuclear position). With a few exceptions, the learners diverged from the target languages in this position; the difference was significant between the L1 and L2 Italian varieties (Fisher’s exact test; $p = 0.019$), as well as between the two Spanish groups (Fisher’s exact test; $p = 0.025$).⁶

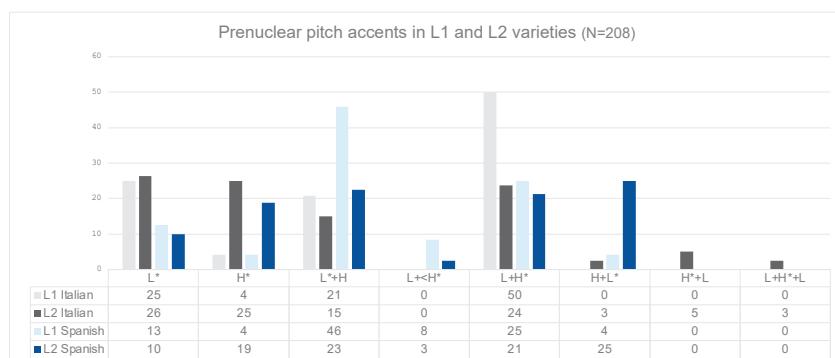


Figure 15. Tonal patterns of prenuclear pitch accents found in all L1 and L2 data (in %).

The results for boundary tones (IP) show much less variation. 97% of all non-neutral sentences were produced with an L% boundary tone in L1 Italian, followed by L2 Italian (94%) and L2 Spanish (86%). As expected, the non-neutral statements in L1 Spanish show a different pattern here (Section 2.3): 64% of the statements were produced with L% boundary tones, 19% with HL% and 14% with L!H% (Figure 16). No significant differences were obtained for L1 and L2 Italian varieties (Fisher’s exact test; $p = 0.064$) and between L2 Italian and L2 Spanish (Fisher’s exact test; $p = 0.884$). The only significant difference was found between L2 Spanish learners and L1 Spanish controls (Fisher’s exact test; $p < 0.001$).

With regard to the L1 Spanish data, it should be recalled that the bitonal L!H% pitch accent is a contour typical of statements of the obvious. The control participants also predominantly produced HL% in the sentences with nuances of surprise (*Es John Travolta* and *Qué sorpresa*). One statement, the greeting, was produced with LHL% at the end of the vocative (*Hola, Roberto*_{LHL%}) by one L1 Spanish participant. Interestingly, this tone is described as characteristically used for exhortative requests in the Spanish_ToBI (see Aguilar et al. 2009). L2 Spanish learners also produced slightly more different patterns than L2 Italian learners, but these patterns did not resemble those produced by the controls. For instance, the L!H% was produced once in the statement of contrastive focus (*No, naranjas*_{L!H%}) and only once in the expected statement of the obvious (*Con Manuel*_{L!H%}). The HL% boundary tone appeared seven times in the data with exclamative sentences or statements of the obvious. Recall that L!H% expresses different types of (mostly yes–no) questions in Czech, while HL% seems to be absent in Czech according to the L1 controls and previous empirical studies (see, e.g., Duběda 2014; Pešková forthcoming).

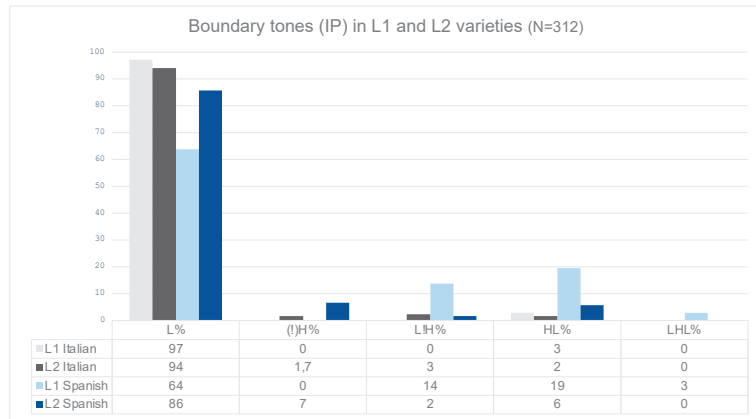


Figure 16. Tonal patterns of all boundary tones at the intonational phrase found in all L1 and L2 data (in %).

As for the boundary tones at the intermediate phrase, appearing only at the end of the greeting *ciao/hola* and after the negative particle *no* in our case, L1 Italian controls produced a low boundary tone (L-) in 100% of cases and L1 Spanish controls did so in 92% of cases. Both groups of Czech learners produced here an H- boundary tone in 20% and a L-boundary tone in 80% of cases, respectively. Interestingly, all boundary tones were combined with L+H* in L1 Spanish, whereas we find either L+H* or (L+)H*+L in L1 Italian. L2 Italian learners showed a mixed picture, using predominantly a falling pattern H+L* and, in three cases, a focus pattern H*+L after the particle *no* and the greeting *ciao*, whereas L2 Spanish learners produced a rising L+H* pattern in about half of cases and, in the other half, a falling H+L* pattern, aside from a few isolated cases of L* or H*.

I conclude this section by presenting some specific examples from the data (for further examples of narrow focus statements and statements of the obvious see Pešková 2020). The following four figures illustrate the main differences observed in the two L2 varieties. The first pair exemplifies non-neutral statements (exclamatives) in L2 Italian, in which the L+H*+L pitch accent was produced. Notice that we find the pattern in the prenuclear position of *buon* (Figure 17) and on the vocative *Roberto* too (Figure 18).

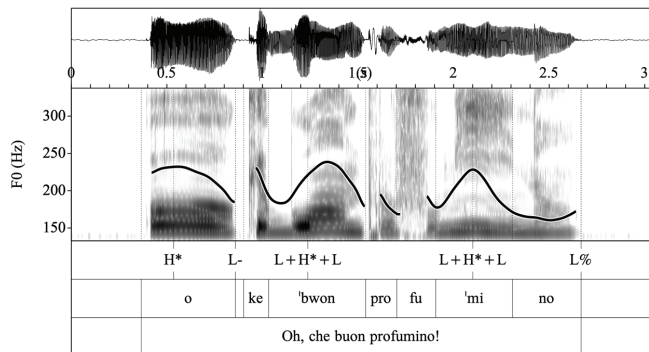


Figure 17. Waveform, spectrogram and F0 trace of the *what*-exclamative *Che buon profumino!* in L2 Italian (F32, level B1) produced with the target-like L+H*+L L% pattern.

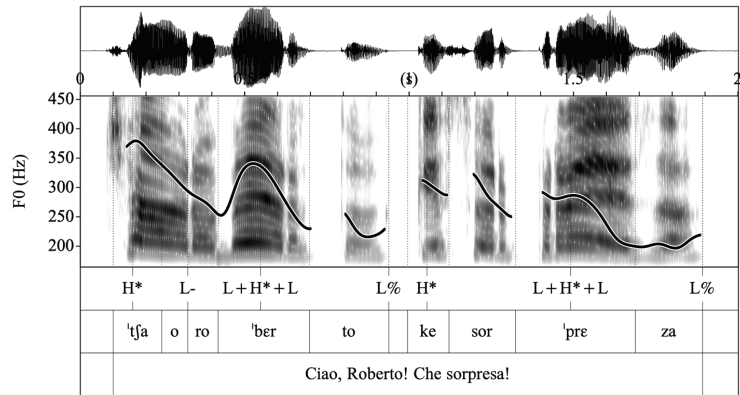


Figure 18. Waveform, spectrogram and F0 trace of the *what*-exclamative *Ciao, Roberto! Che sorpresa!* in L2 Italian (F34, level B2) produced with the target-like L+H*+L L% pattern.

The second pair offers the same sentences in L2 Spanish, realized with L+H* in the nuclear position in both cases; the latter one was produced with an upstep (j) (Figures 19 and 20).

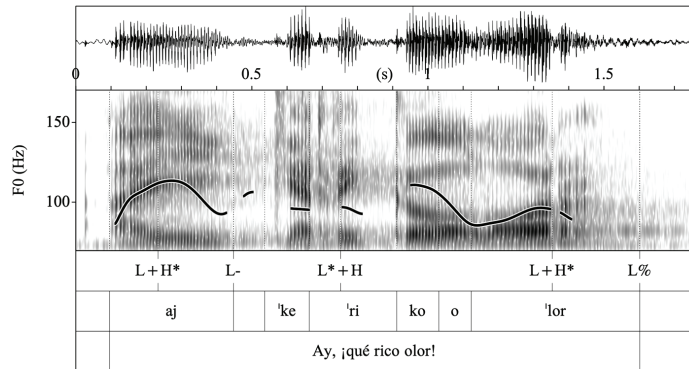


Figure 19. Waveform, spectrogram and F0 trace of the *what*-exclamative *¡Qué rico olor!* in L2 Spanish (M17, level C1) produced with the target-like L+H* L% pattern.

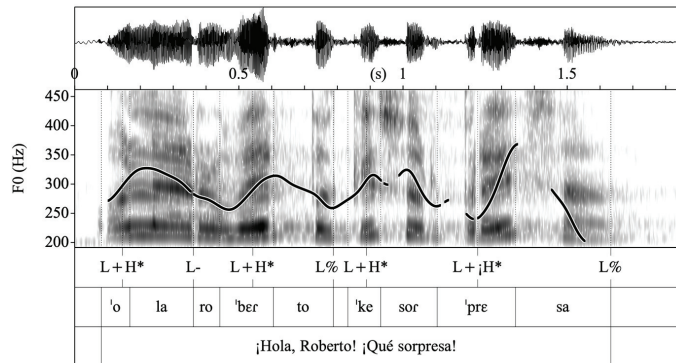


Figure 20. Waveform, spectrogram and F0 trace of the *what*-exclamative *¡Hola, Roberto! ¡Qué sorpresa!* in L2 Spanish (F16, level C2) produced with the target-like L+;H* L% pattern.

Not all sentences showed a target-like form. The L2 data yielded several patterns that were completely absent from the Italian or Spanish control data. The following two pairs of non-neutral statements represent what I would call a typical Czech intonation pattern. The exclamative statement in the first two examples begins with a high monotonal plateau on the wh-word *what* (It. *che* / Sp. *qué*), extended to the adjective, from which the pitch track simply falls (Figures 21 and 22).

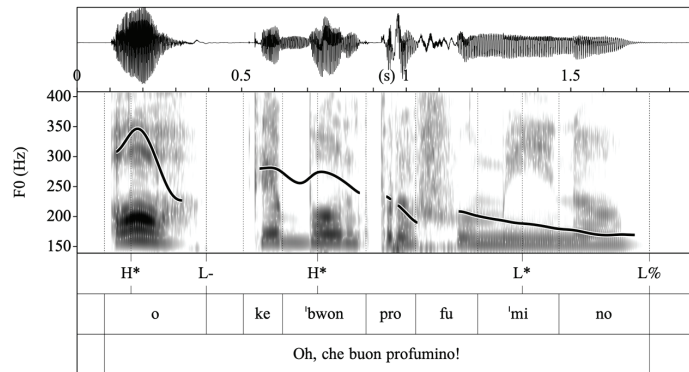


Figure 21. Waveform, spectrogram and F0 trace of the *what*-exclamative *Che buon profumino!* in L2 Italian (F38, level C2) produced with L* L%.

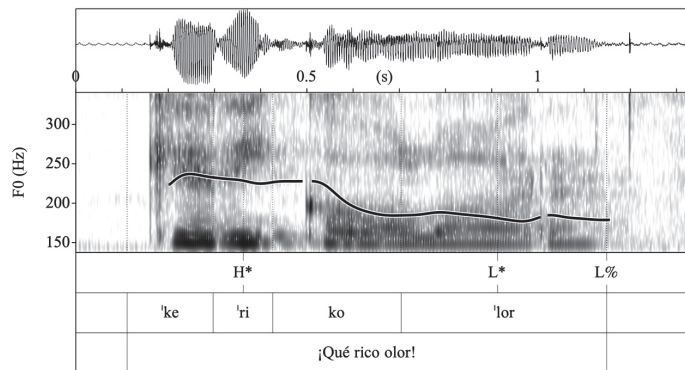


Figure 22. Waveform, spectrogram and F0 trace of the *what*-exclamative *¡Qué rico olor!* in L2 Spanish (F20, level B1) produced with L* L%.

We find a very similar intonation pattern (H* L* L%) in another L2 Italian example (Figure 23). Additionally, the vocative *Roberto* is realized here with a very low plateau that resembles the Czech pattern in such a position.

In the L2 Spanish counterpart (Figure 24), we find a contour that is almost identical to the greeting seen in the Italian example. As for the nuclear configuration, it was realized with a Czech focal pattern L*+H; as we can see, the speaker produced the word *sorpresa* ('surprise') with the accent placed on the first syllable instead of on the second syllable.

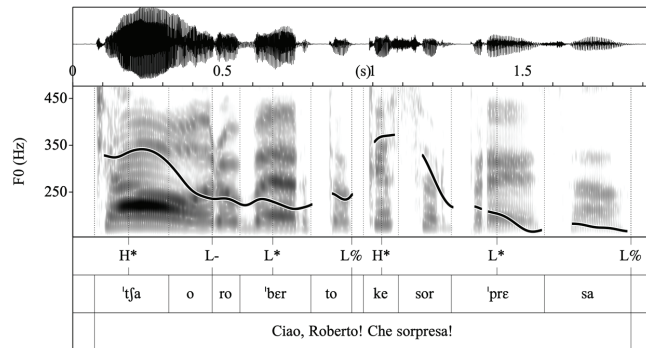


Figure 23. Waveform, spectrogram and F0 trace of the *what*-exclamative *Ciao, Roberto! Che sorpresa!* in L2 Italian (F35, level C1) produced with L* L%.

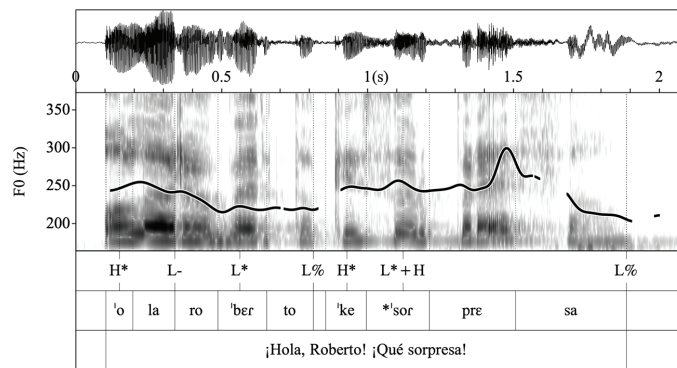


Figure 24. Waveform, spectrogram and F0 trace of the *what*-exclamative *¡Hola, Roberto! ¡Qué sorpresa!* in L2 Spanish (F4, level B2) produced with L*+H L%.

5.2. Individual Factors

In this subsection, I concentrate on the individual factors—the length of residence (LOR), the age of learning (AOL) and the amount of L2 use (AUL)—that might explain the variability observed in the L2 data. A binary logistic regression model in Generalized Linear Mixed Models (SPSS, IBM 2022) was performed to ascertain the effects of these factors on the likelihood that participants would produce accurate intonational patterns in their L2. In the model, Accuracy was the dependent variable, LOR, AOL and AUL were fixed effects and Speakers and Intercept were random effects. The model showed non-significant effects: AUL ($p = 0.954$), AOL ($p = 0.150$) and LOR ($p = 0.131$) (Table 2).

Table 2. Predictors of intonational accuracy for 20 L2 Italian and 20 L2 Spanish learners.

Variable	Learners (N = 40)			
	t	df	Sig.	Exp(B)
Intercept	0.565	3	0.572	2.841
LOR	−1.510	1	0.131	0.992
AOL	−1.442	1	0.150	0.976
AUL	0.058	1	0.954	1.001

All this indicates that there must be other factors that have a stronger effect on accuracy in L2 production, though caution is mandatory here, since the power of the present research design is relatively low (in the sense of Brysbaert 2021). Hence, it was tested whether Language (here L2 Italian or L2 Spanish) could have an effect on Accuracy. The analysis revealed a statistically non-significant result again, but, interestingly, the model was improved by the interaction of the variables Language ($p = 0.016$), LOR ($p = 0.055$) and AOL ($p = 0.038$), respectively.

Let us now turn to an illustrative qualitative analysis of accuracy involving four learners from each L2 group, two of the four with a B-level proficiency and two with C-level proficiency (Table 3). Though it covers only a small number of tonal events per speaker, it can be seen that the two learners with the C proficiency level (F38, F13), who had spent five years in Italy or Spain, respectively, made more errors than learners with less experience abroad. Nor does the amount of active use of an L2 seem to play a role either: For example, learner F37 (C2), with regular work trips to Italy and the highest weekly exposure to L2, performed less accurately than other learners. L2 deviations occurred in all positions and the learners also differed from each other in this respect. For example, learner F38 produced most errors in the nuclear position, omitting (L+)H*+L more frequently than learner F34. Interestingly, fewer errors were made by those learners who had started to learn Spanish at the age of 10 (M19), 13 (M17) and 14 (F34), that is, a bit earlier than the other learners: 15 (F08), 16 (F13), 18 (F38), 20 (F47) and 23 (F37). Although all the learners began to learn Italian/Spanish at or after puberty and not before, it seems that later AOL causes more L2 intonation deviations here. However, the factor AOL did not prove to be a significant predictor for the whole group of 40 learners, as already stated above. In the Discussion section we present some ideas of what may lie behind these results.

Table 3. Intonation errors of eight selected L2 Italian and L2 Spanish learners.

L2	Learner	Level	LOR ⁷	AOL	AUL (hours/week)	Errors	Prenuclear Accents (n = 4)	Nuclear Accents (n = 8)	Boundary Tones (n = 8)
Italian	F34	B2	Short visits to different locations	14	2	5	2	2	1
	F47	B1	5 months in Verona	20	0	7	1	4	2
	F37	C2	3 months in Perugia, regular work visits	23	40 (job)	8	2	4	2
	F38	C2	5 years in Tuscany and Rome	18	0	8	2	6	0
	M19	B2	1 month in the north of Spain	10	2	5	0	3	2
Spanish	F08	B2	No stay in an L2-speaking country	15	20	8	4	1	3
	M17	C1	3 months in Valencia	13	4	4	1	0	3
	F13	C2	5 years in different places (Spain)	16	3	13	3	4	6

6. Discussion

The first research question (Q1) was whether the two groups of L2 learners would differ from each other and to what extent. Since the results reveal several differences across the two L2 groups, we can confirm that intonation is learnable. To begin with, the learners diverged significantly in the realization of nuclear pitch accents. Whereas the L2 Spanish learners preferred an L+H* rising accent (62%), the L2 Italian group produced nuclear accents predominantly with (L+)H*+L realizations (44%). Interestingly, L2 Italian learners produced L+H* only in 10% of cases and L2 Spanish learners produced an Italian pattern in only two cases. Recall that H*+L and L+H*+L are typical realizations of Italian nuclear pitch accents, conveying emphasis or focus. These two patterns do not exist in Czech at all. In Spanish, it is L+H* that assumes this role. The Czech focal pattern is L*+H, but L+H* is present in Czech as a phonetic variant of L*+H.

The second and third questions (Q2, Q3) were related to the differences between learners and L1 controls and to the difficulties along the systemic and semantic dimensions.

Broadly speaking, the accuracy of reproducing the target tonal patterns in L2 Italian and L2 Spanish is relatively high in the nuclear position and the learners seem to have less difficulty with the systemic dimension in that they are able to learn tones absent in their L1 system such as (L+)H*+L. This finding is in line with what LILt would predict. It should be added that all L2 Italian learners with one exception produced the Italian target pattern (H*+L, L+H*+L), although at different frequencies. The only learner who completely omitted the Italian pattern was a proficient learner (level C1) with a relatively strong Czech accent and almost no experience in an L2-speaking area. Additionally, this learner seemed to be an introverted and quiet person, another aspect which might influence second language speech (see, e.g., Dewaele and Furnham 2000 on the role of personality in language learning). It is also interesting that some Czech learners of Italian used the nuclear patterns in the prenuclear position, and although the number of such instances was very low, this does point to *prosodic overgeneralization*, in other words, the inappropriate application of a tonal pattern or a pattern not seen in L1 speech. Overgeneralization belongs to the typical phenomena of interlanguage development (see, e.g., Ellis 1994; Gass and Selinker 2008) and shows that learners “identify that there is something to learn” (Gass 1988, p. 394).

As a next step, we predicted difficulties in the semantic dimension, such as for the acquisition of L!H% in Spanish statements of the obvious. Recall that this boundary tone exists in Czech in yes–no questions and was acquired, for instance, in L2 Italian yes–no questions (Pešková 2020). Nevertheless, the results of the present study showed only one production of L!H% in the L2 Spanish non-neutral statements of the obvious (*¡Con Manuel!*) (Figure 25). This means that only one learner was able to produce this boundary tone correctly.

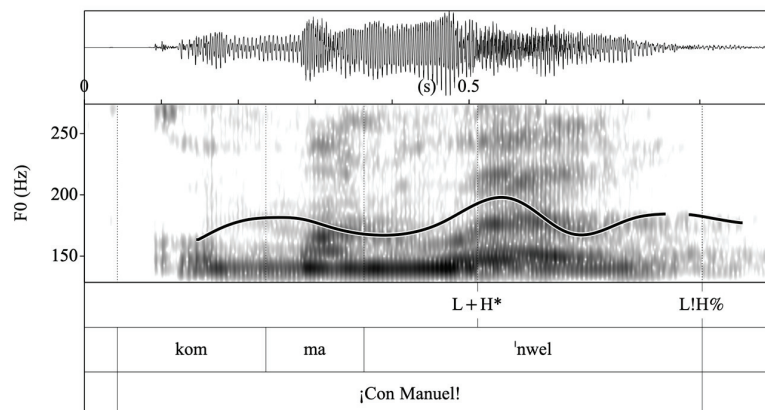


Figure 25. Waveform, spectrogram and F0 trace of the *what*-exclamative *¡Con Manuel!* in L2 Spanish (F15, level C1) produced with the target-like L+H* L!H% pattern.

At first sight, it seems easier to acquire a new tonal pattern ((L+)H*+L) (systemic dimension) than to re-apply a known pattern (L!H%) to a different context (semantic dimension). However, L2 learners also exhibited difficulty in L2 Spanish with HL%, which does not exist in L1 Czech either. This boundary tone was only produced by three learners of Spanish. Although I believe it is the semantic dimension that presents the greatest amount of difficulty for L2 learners, we should not rule out other explanations. For example, frequency can play an important role here too: the (L+)H*+L pitch accent is found in very different types of sentences in Italian and is more frequent than L!H% or HL% in Spanish. Moreover, as suggested in Pešková (2020), the Italian pattern is more prominent perceptually. Anecdotally, when people try to imitate Italian, they tend to use this pattern. This would also explain why L2 learners overuse the (L+)H*+L pattern in the prenuclear

position or in inappropriate contexts such as neutral vocatives, in which L1 speakers would produce L+H* (Pešková 2019).

Elsewhere the data show a range of variation and a sharper divergence from the L1 data in the prenuclear position. This suggests that prenuclear pitch accents tend to be much more susceptible to cross-linguistic influence because they have less semantic weight than the nuclear position, which conveys meaning, and because they also exhibit larger variation in L1 varieties, probably for the same reason.

The fourth and last question (Q4) was directed at the effect of individual factors (AOL, LOR, AUL) on the accuracy in L2 intonation. As we have noted, the statistical analysis revealed that LOR, AOL and AUL did not intervene very clearly. It could be hypothesized that the fact that learners were exposed mostly to one or more different L2 regional varieties would better explain the observed variation and the creation of mixed patterns. Additional factors potentially having an impact on L2 speech such as a general talent for pronunciation or individual aptitudes related to motor and music skills, mimicry or memory should not be ignored and deserve attention in future studies. And finally, follow-up research should also examine whether the relationship between production accuracy and foreign accent perceived by L1 listeners is symmetric or asymmetric.

7. Concluding Remarks

How much does transfer matter in the acquisition of L2 intonation? It is not very easy to quantify L1-to-L2 transfer in intonation, since L2 verbal output is not simply made up of target-like features or L1-transferred features, but also a range of mixed patterns, which are difficult to interpret. Despite such difficulties, both positive and negative influences can be detected across all learners and contexts, but to different degrees. The results of the present study, in which 20 Czech learners of Italian were compared with 20 Czech learners of Spanish, suggest that intonation is in fact “learnable”, albeit the two groups of learners diverged from L1 speakers in several ways. Whereas the learners were quite successful in learning new patterns in the nuclear position, they exhibited more difficulties with boundary tones, independently of whether the tone was present or absent in the learner’s L1. It may well be that factors such as perceptual prominence and frequency can explain this finding and predict the difficulties in L2 intonation learning. With regard to individual factors bearing on a learner’s ability to approximate L2 intonation targets, the findings exhibited no significant effects derived from the length of residence abroad (LOR), the age of learning (AOL) or the amount of L2 use (AUL).

These results constitute a step forward in our growing understanding of the study of the acquisition of L2 intonation. That said, considerable work deserves to be done in this area, particularly given its rich potential for practical application in the language learning classroom.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are not available to the public.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Overview of the learners who participated in the study (F = female, M = male).

L2	Learner	Age	Level	LOR	AOL	AUL (hours/week)
Italian	F31	24	C1	1 month in the Dolomites, 5 months in Siena	18	2
	F32	38	B1	Short visits to different locations in Italy	20	1
	F33	26	C1	5 months in Padua	19	4
	F34	23	B2	Short visits to different locations in Italy	14	2
	F35	23	C1	16 months (Roma, Ravenna, Cervia)	18	28
	F36	29	B2	3 months in Parma, several weeks in Turin and Milan	20	0
	F37	30	C2	3 months in Perugia, regular work trips to different locations	23	40
	F38	37	C2	5 years in Tuscany and Rome	18	0
	F39	36	C2	6 years in San Benedetto del Tronto	19	2
	F40	23	B2	Different short visits to Italy	18	4
	F41	26	C1	6 months in Rome; many short visits to Ravenna	19	5
	F42	22	B1	2 months in Genoa	19	10
	F43	24	B2	Short visits to different locations in Italy	18	6
	M44	43	B1	2 weeks in Calabria	33	2
	F45	22	C1	4 months in Perugia	20	4
	M46	28	C1	Short visits to different locations in Italy	20	5–6
	F47	24	B1	5 months in Verona	20	0
	F48	26	B2	14 months in Florence	11	2
	F49	27	C1	Short visits to different locations in Italy	12	30
	F50	31	B1	3 months in Calabria	25	1
Spanish	M01	30	C2	1 year in different locations in Spain, 1 year in different Latin American countries	22	5
	M02	40	B1	Short visits to different locations in Spain	35	1
	F04	29	B2	Short visits to different locations in Spain	20	2
	F05	29	B1	Short visit in Catalonia	14	2
	F06	23	C1	Short visits to different locations in Spain, Mexico, Guatemala	17	20
	F07	21	B2	No stay in an L2-speaking country	22	1–2
	F08	20	B2	No stay in an L2-speaking country	15	20
	F09	24	B2	1 year in different locations in Spain	21	10
	F10	27	B2	Short visits to different locations in Spain, Mexico	13	10
	F11	19	C1	1 month in the south of Spain	17	8
	M12	33	B2	Short visits to different locations in Spain and Central America	24	2
	F13	31	C2	5 years in Spain	16	3

Table A1. Cont.

L2	Learner	Age	Level	LOR	AOL	AUL (hours/week)
	M14	22	C1	Short visits to different locations in Spain	17	2–3
	F15	24	C1	6 months in Barcelona and short visits to different locations in Spain	10	3
	F16	20	C2	Short visits to different locations in Spain	10	40
	M17	20	C1	3 months in Valencia	13	4
	F18	23	C1	10 months in Valencia; short visits to Cuba	17	8
	M19	34	B2	1 month in the north of Spain	10	2
	F20	45	B1	Short visits to different locations in Spain	19	1
	M21	30	C1	2 years in Barcelona	14	1

Notes

- Some of these non-neutral sentences (contexts 1–3) are also presented in Pešková (2020); however, the present paper includes additional contexts, fully discusses the results within the LILt and tests individual factors in order to explain the accuracy in L2 intonational patterns.
- The potential influence of other previously acquired L2s is not covered in the present study.
- The reliability of this manual labelling is discussed in Pešková (2020), in which the same annotation was carried out manually and then compared against automatic transcription of intonation using the Eti_ToBI tool (Elvira-García et al. 2016). Overall agreement was relatively high at 79% for all tonal events.
- The H+L*+H of the present study corresponds to the pitch accent L*+>H in Gili Fivela et al. (2015).
- To make a statistical analysis possible here, the low values for L+<H*, H*+L and L+H*+L were excluded and Fisher’s exact test was used.
- To make a statistical analysis possible, the low values for L+<H*, H*+L and L+H*+L were not included for the Spanish groups. As for the two Italian varieties, the analysis included L*, H*, L*+H, L+H* and (L+)H*+L.
- I use the term “residence” here to refer in general to the time spent in an L2-speaking country, independently of whether the learners had just made short visits or had spent a longer period of time there.

References

- Aguilar, Lourdes, Carme De-La-Mota, and Pilar Prieto. 2009. Sp_ToBI Training Materials. Available online: http://prosodia.upf.edu/sp_tobi/ (accessed on 22 June 2021).
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and Second-Language Speech Perception: Commonalities and Complementarities. In *Language Experience in Second Language Speech Learning*. Edited by Ocke-Schwen Bohn and Murray J. Munro. Amsterdam and Philadelphia: John Benjamins, pp. 13–34. [CrossRef]
- Blum-Kulka, Shoshana, Juliane House, and Gabriele Kasper. 1989. Investigating Cross-Cultural Pragmatics: An Introductory Overview. In *Cross-Cultural Pragmatics: Requests and Apologies*. Edited by Shoshana Blum-Kulka, Juliane House and Gabriele Kasper. Norwood: Ablex, pp. 1–34.
- Boersma, Paul, and David Weenink. 1992–2022. Praat: Doing Phonetics by Computer. Available online: <http://www.praat.org> (accessed on 13 September 2020).
- Brysaert, Marc. 2021. Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition* 24: 813–18. [CrossRef]
- CEFR. 2018. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. In *Companion Volume with New Descriptors. Language Policy Program*. Strasbourg: Council of Europe. Available online: <http://www.coe.int/lang-cefr> (accessed on 25 December 2018).
- Chun, Dorothy. 1998. Signal Analysis Software for Teaching Discourse Intonation. *Language Learning & Technology* 2: 74–93.
- Colantoni, Laura, and Jorge Gurlekian. 2004. Convergence and Intonation: Historical Evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition* 7: 107–19. [CrossRef]
- Colantoni, Laura, Gabrielle Klassen, Matthew Patience, Malina Radu, and Olga Tararova. 2016. Task-Effects in the L2 Perception and Production of English Sentence Types by L1 Spanish Speakers. Paper presented at 8th International Conference on Speech Prosody 2016, Boston, MA, USA, June 3; pp. 11–15.
- Colantoni, Laura, Jeffrey Steele, and Paola Escudero. 2015. *Second Language Speech. Theory and Practice*. Cambridge: Cambridge University Press. [CrossRef]

- De Leeuw, Esther, Ineke Mennen, and James M. Scobbie. 2012. Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism* 16: 101–16. [\[CrossRef\]](#)
- Derwing, Tracey M., and Murray J. Munro. 2015. *Pronunciation Fundamentals. Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdam and Philadelphia: John Benjamins. [\[CrossRef\]](#)
- Dewaele, Jean-Marc, and Adrian Furnham. 2000. Personality and speech production: A pilot study of second language learners. *Personality and Individual Differences* 28: 355–65. [\[CrossRef\]](#)
- Duběda, Tomáš. 2014. Czech Intonation: A Tonal Approach. *Slovo a Slovesnost* 75: 83–98.
- Ellis, Rod. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press. [\[CrossRef\]](#)
- Elvira-García, Wendy, Paolo Roseano, Fernández Planas, Ana Maria, and Eugenio Martínez Celdrán. 2016. A Tool for Automatic Transcription of Intonation: Eti-ToBI a ToBI Transcriber for Spanish and Catalan. *Language Resources and Evaluation* 50: 767–92. [\[CrossRef\]](#)
- Estebas-Vilaplana, Eva, and Pilar Prieto. 2010. Castilian Spanish Intonation. In *Transcription of Intonation of the Spanish Language*. Edited by Pilar Prieto and Paolo Roseano. Munich: Lincom Europa, pp. 17–48.
- Féry, Caroline. 2010. Indian Languages as Intonational 'Phrase Languages'. In *Problematising Language Studies. Cultural, Theoretical and Applied Perspectives. Essays in Honor of Rama Kant Agnihotri*. Edited by S. Imtiaz Hasnain and Shreesh Chaudhary. New Delhi: Aakar Books, pp. 288–312. [\[CrossRef\]](#)
- Flège, James E. 1987. The Production of "New" and "Similar" Phones in a Foreign Language: Evidence for the Effect of Equivalence Classification. *Journal of Phonetics* 15: 47–65. [\[CrossRef\]](#)
- Flège, James E. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Research*. Edited by W. Strange. Timonium: York Press, pp. 233–77.
- Flège, James E., Ocke-Schwen Bohn, and Sunyoung Jang. 1997. Effects of Experience on Non-Native Speakers' Production and Perception of English Vowels. *Journal of Phonetics* 25: 437–70. [\[CrossRef\]](#)
- Frota, Sónia, and Pilar Prieto, eds. 2015. *Intonation in Romance*. Oxford: Oxford University Press. [\[CrossRef\]](#)
- Gabriel, Christoph, and Elena Kireva. 2014. Prosodic Transfer in Learner and Contact Varieties: Speech Rhythm and Intonation of Buenos Aires Spanish and L2 Castilian Spanish Produced by Italian Native Speakers. *Studies in Second Language Acquisition* 36: 257–81. [\[CrossRef\]](#)
- Gabriel, Christoph, Ingo Feldhausen, and Andrea Pešková. 2011. Prosodic Phrasing in Portefño Spanish. In *Intonational Phrasing in Romance and Germanic: Cross-Linguistic and Bilingual Studies*. Edited by Christoph Gabriel and Conxita Lleó. Amsterdam and Philadelphia: John Benjamins, pp. 153–82. [\[CrossRef\]](#)
- Gabriel, Christoph, Ingo Feldhausen, Andrea Pešková, Laura Colantoni, Su-Ar Lee, Valeria Arana, and Leopoldo Labastia. 2010. Argentinian Spanish Intonation. In *Transcription of Intonation of the Spanish Language*. Edited by Pilar Prieto and Paolo Roseano. Munich: Lincom, pp. 285–317.
- Gass, Susan M. 1988. Second Language Acquisition and Linguistic Theory: The Role of Language Transfer. In *Linguistic Theory in Second Language Acquisition*. (= *Studies in Theoretical Psycholinguistics* 8). Edited by Suzanne Flynn and Wayne O'Neil. Dordrecht: Springer, pp. 384–403. [\[CrossRef\]](#)
- Gass, Susan M., and Larry Selinker. 2008. *Second Language Acquisition: An Introductory Course*, 3rd ed. New York: Routledge. [\[CrossRef\]](#)
- Gili Fivela, Barbara, Cinzia Avesani, Marco Barone, Giuliano Bocci, Claudia Crocco, Mariapaola D'Imperio, Rosa Giordano, Giovanna Marotta, Michelina Savino, and Patrizia Soriano. 2015. Intonational Phonology of the Regional Varieties of Italian. In *Intonation in Romance*. Edited by Sónia Frota and Pilar Prieto. Oxford: Oxford University Press, pp. 140–97. [\[CrossRef\]](#)
- Henriksen, Nicholas C. 2012. The Intonation and Signaling of Declarative Questions in Manchego Peninsular Spanish. *Language and Speech* 55: 543–76. [\[CrossRef\]](#)
- Hualde, José I., and Pilar Prieto. 2015. Intonational Variation in Spanish: European and American Varieties. In *Intonation in Romance*. Edited by Sónia Frota and Pilar Prieto. Oxford: Oxford University Press, pp. 350–91. [\[CrossRef\]](#)
- Hualde, José I., and Pilar Prieto. 2016. Towards an International Prosodic Alphabet (IPRA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7: 5. [\[CrossRef\]](#)
- IBM. 2022. *SPSS Statistics for macOS. (Version 28.0.)*. Armonk: IBM.
- Jun, Sun-Ah, ed. 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. [\[CrossRef\]](#)
- Jun, Sun-Ah, ed. 2014. *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. [\[CrossRef\]](#)
- Ladd, Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press. [\[CrossRef\]](#)
- Ladd, Robert. 2008. *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press. [\[CrossRef\]](#)
- Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press. [\[CrossRef\]](#)
- Méndez Seijas, Jorge. 2018. L2 Spanish Intonation in a Short-Term Study Abroad Program. In *The Routledge Handbook of Study Abroad Research and Practice*. Edited by Cristina Sanz and Alfonso Morales-Front. New York and London: Routledge, pp. 89–104. [\[CrossRef\]](#)
- Mennen, Ineke, Aoju Chen, and Fredrik Karlsson. 2010. Characterising the internal structure of learner intonation and its development over time. Paper presented at New Sounds 2010 (6th International Symposium on the Acquisition of Second Language Speech), Barcelona, Spain, April 20–22; Poznan: Adam Mickiewicz University, pp. 319–24.
- Mennen, Ineke, Felix Schaeffler, and Catherine Dickie. 2014. Second language acquisition of pitch in German learners of English. *Studies in Second Language Acquisition* 36: 303–29. [\[CrossRef\]](#)

- Mennen, Ineke. 2004. Bi-Directional Interference in the Intonation of Dutch Speakers of Greek. *Journal of Phonetics* 32: 543–63. [CrossRef]
- Mennen, Ineke. 2015. Beyond Segments: Towards a L2 Intonation Learning Theory. In *Prosody and Language in Contact: L2 Acquisition, Attribution and Languages in Multilingual Situations*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin and Heidelberg: Springer, pp. 171–88. [CrossRef]
- Nicora, Francesca, Laura Incalcaterra McLoughlin, and Barbara Gili Fivela. 2018. Impact of Prosodic Training on Italian as L2 by Hiberno-English Speakers: The Case of Polar Questions. Paper presented at 9th International Conference on Speech Prosody 2018, Poznań, Poland, June 13–16; Poznań: Adam Mickiewicz University, pp. 970–74.
- O'Rourke, Erin. 2004. Peak Placement in Two Regional Varieties of Peruvian Spanish Intonation. In *Contemporary Approaches to Romance Linguistics*. Edited by Julie Auger, J. Clancy Clements and Barbara Vance. Amsterdam and Philadelphia: John Benjamins, pp. 321–41. [CrossRef]
- Oyama, Susan. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5: 261–83. [CrossRef]
- Palková, Zdena. 1994. *Fonetika a fonologie češtiny. S obecným úvodem do problematiky oboru*. Prague: Karolinum.
- Pešková, Andrea. 2017. Czech ToBI. Paper presented at Phonetics and Phonology in Europe 2017, Cologne, Germany, June 12–14.
- Pešková, Andrea. 2019. L2 Italian and L2 Spanish Vocatives Produced by L1 Czech Learners: Transfer and Prosodic Overgeneralization. Paper presented at 19th International Congress of Phonetic Sciences 2019, Melbourne, Australia, August 5–9; pp. 1932–36.
- Pešková, Andrea. 2020. L2 Spanish and Italian Intonation: Accounting for the Different Patterns Displayed by L1 Czech and German Learners. Postdoctoral thesis, Osnabrück University, Osnabrück, Germany. Under review in *Open Romance Linguistics*, Language Science Press.
- Pešková, Andrea. 2021. Gibt es melodische Konstruktionen? Erwerb von W-Fragen im L2-Italienischen und L2-Spanischen. In *Konstruktionsgrammatische Zugänge zu Romanischen Sprachen*. Edited by Hans-Jörg Döhla and Anja Hennemann. Berlin: Frank & Timme (Strukturelle Dynamik und Sprachkontakt in der Romania), pp. 241–71.
- Pešková, Andrea. forthcoming. The Intonational Phonology of Czech. In *Prosodic Typology III: The Phonology of Intonation and Phrasing*. Edited by Sun-Ah Jun and Sameer ud Dowla Khan. Oxford: Oxford Scholarship.
- Pierrehumbert, Janet B. 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Piske, Thorsten, Ian R. A. MacKay, and James E. Flege. 2001. Factors Affecting Degree of Foreign Accent in an L2: A Review. *Journal of Phonetics* 29: 191–215. [CrossRef]
- Prieto, Pilar, and Paolo Roseano, eds. 2010. *Transcription of Intonation of the Spanish Language*. Munich: Lincom Europa.
- Prieto, Pilar. 2001. L'entonació dialectal del català: El cas de les frases interrogatives absolutes. In *Actes del Novè Col·loqui d'Estudis Catalans a Nord-Amèrica*. Edited by August Bover, Maria-Rosa Lloret and Mercè Vidal-Tibbits. Barcelona: Publicacions de l'Abadia de Montserrat, pp. 347–77.
- Pustka, Elissa, Christoph Gabriel, Trudel Meisenburg, Monja Burkard, and Kristina Dziallas. 2018. (Inter-) Fonología del Español Contemporáneo (I)FEC: Metodología de un programa de investigación para la fonología de corpus. *Loquens* 5: 1–16. [CrossRef]
- Romera, Magdalena, and Gorka Elordieta. 2020. Information-Seeking Question Intonation in Basque Spanish and its Correlation with Degree of Contact and Language Attitudes. *Languages* 5: 1–21. [CrossRef]
- Santiago, Fabian, and Elisabeth Delais-Roussarie. 2015. The Acquisition of Question Intonation by Mexican Spanish Learners of French. In *Prosody and Language in Contact: L2 Acquisition, Attribution and Language in Multilingual Situations*. Edited by Elisabeth Delais-Roussarie, Mathieu Avanzi and Sophie Herment. Berlin/Heidelberg: Springer, pp. 243–70.
- Selkirk, Elisabeth O. 1984. *Phonology and Syntax. The Relation between Sound and Structure*. Cambridge: MIT Press.
- Sichel-Bazin, Rafèu, Trudel Meisenburg, and Pilar Prieto. 2015. Intonational Phonology of Occitan: Towards a Prosodic Transcription System. In *Intonation in Romance*. Edited by Sónia Frota and Pilar Prieto. Oxford: Oxford University Press, pp. 198–234.
- Sorace, Antonella. 2011. Pinning down the concept of “interface” in bilinguals. *Linguistic Approaches to Bilingualism* 1: 1–33. [CrossRef]
- Trofimovich, Pavel, and Wendy Baker. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30. [CrossRef]
- Ueyama, Motoko. 1997. The Phonology and Phonetics of Second Language Intonation: The Case of “Japanese English”. Paper presented at Fifth European Conference on Speech Communication and Technology, Rhodes, Greece, September 22–25; pp. 2411–14.
- van Maastricht, Lieke, Emiel Kraemer, and Marc Swerts. 2016. Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. *Language Learning* 66: 124–58. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Languages Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34

www.mdpi.com



ISBN 978-3-0365-7455-4