

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

Prosodic effects of code-switching in Spanish-Basque bilinguals

**Permalink**

<https://escholarship.org/uc/item/1pz4v3bn>

**Author**

Aly, Ann Marie

**Publication Date**

2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Prosodic effects of code-switching in Spanish-Basque bilinguals

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of  
Philosophy in Linguistics

by

Ann Marie Aly

2017



## ABSTRACT OF THE DISSERTATION

Prosodic effects of code-switching in Spanish-Basque bilinguals

by

Ann Marie Aly

Doctor of Philosophy in Linguistics

University of California, Los Angeles 2017

Professor Sun-Ah Jun, Chair

This dissertation investigates the prosodic aspects of code-switching in Spanish-Basque bilinguals, which is not currently represented in the bilingualism literature. Bilinguals from the Basque Country in Spain (aged 21-31) from a Basque-dominant region (Lekeitio,  $n = 3$ ) and a Spanish-dominant region (Bilbao,  $n = 4$ ) participated in the present study and provided spoken data in unilingual Spanish, Spanish with Basque code-switches, and Basque with Spanish code-switches. The Matrix Language Framework (Myers-Scotton (1993)) predicts that in the presence of language-specific differences, those present in the dominant language of the interaction will override those of the less-dominant language. Spanish and Basque have several prosodic

differences, including latency of peak alignment in pitch accents. The current study investigates whether peak alignment differs between unilingual contexts (Spanish only), and code-switched contexts (Spanish code-switching and Basque code-switching). Additionally, previous work on code-switching in Spanish-English communities (e.g. Olson (2012, 2016b) and Fricke et al (2016)) provide evidence that code-switched words may be hyperarticulated or perceived as more prominent than non-switched words. In order to investigate this claim, the  $f_0$  maximum (a reliable acoustic correlate of prominence in Spanish and Basque) is measured in code-switched and non-switched words in the present study.

The results of the dissertation provide weak support for the Hyperarticulation account, as higher  $f_0$  during code-switching was only found for one group (Lekeitio) during phrase-final positions of one task (Discourse Completion Task). However, there was strong support for the Matrix Language Framework, as there was evidence that the peak alignment patterns of the dominant (or Matrix) language overruled the non-dominant language in code-switching contexts. However, the Matrix Language effects also interacted with speaker language dominance and language usage. Specifically, a speaker's dominant language was more sensitive to deviation from unilingual norms when embedded in code-switching contexts. Effects of language usage on were also seen, in which the use of Basque socially resulted in more conservative peak alignment production and lower  $f_0$  when Basque was the Matrix Language. These results suggest a complex interaction of language context, language dominance, and language switching costs which contribute to variations in prosodic realizations of code-switches.

The dissertation of Ann Marie Aly is approved.

Patricia Keating

Bruce Hayes

Gorka Elordieta

Sun-Ah Jun, Committee Chair

University of California, Los Angeles

2017

*for Itxaso and Unai*

## TABLE OF CONTENTS

|           |                                                                             |    |
|-----------|-----------------------------------------------------------------------------|----|
| <b>1.</b> | <b>INTRODUCTION</b>                                                         |    |
|           |                                                                             | 1  |
| 1.1       | Overview of the dissertation                                                | 1  |
| 1.2       | Introduction to code-switching                                              | 1  |
| 1.3       | Previous studies on code-switching                                          | 3  |
| 1.3.1     | Morphosyntax and the Matrix Language Framework                              | 3  |
| 1.3.2     | Code-switching and cognitive control: Insights from neuro/psycholinguistics | 5  |
| 1.3.3     | Code-switching and sound systems: Research on phonetics and phonology       | 7  |
| <b>2.</b> | <b>SPANISH AND BASQUE CONTEXTUALIZED IN THE BASQUE COUNTRY</b>              |    |
|           |                                                                             | 13 |
| 2.1       | Language status, use, and demographics                                      | 13 |
| 2.2       | Basque prosody and intonation                                               | 19 |
| 2.2.1     | Standard Basque                                                             | 19 |
| 2.3       | Spanish prosody and intonation                                              | 27 |
| 2.3.1     | Basque Country Spanish intonation                                           | 32 |
| 2.4       | The current study's hypotheses                                              | 33 |
| 2.4.1     | Hypothesis 1: Peak alignment                                                | 33 |
| 2.4.2     | Hypothesis 2: f0 maxima                                                     | 34 |
| 2.4.3     | Hypothesis 3: Language dominance                                            | 35 |
| <b>3.</b> | <b>METHODOLOGY</b>                                                          |    |
|           |                                                                             | 36 |
| 3.1       | Participant recruitment and demographics                                    | 36 |
| 3.2       | Data collection procedure                                                   | 43 |
| 3.3       | Analysis procedure                                                          | 48 |
| 3.3.1     | Code-switching definitions and criteria                                     | 48 |
| 3.3.2     | Native speaker judgments survey                                             | 55 |
| 3.3.3     | Phonetic measurements                                                       | 63 |
| 3.3.4     | Statistical analyses                                                        | 68 |
| <b>4.</b> | <b>RESULTS</b>                                                              |    |
|           |                                                                             | 75 |
| 4.1       | Distribution of code-switches                                               | 75 |

|       |                                                                                                               |     |
|-------|---------------------------------------------------------------------------------------------------------------|-----|
| 4.2   | Peak alignment .....                                                                                          | 79  |
| 4.2.1 | Unilingual Spanish baseline (Reading task) .....                                                              | 80  |
| 4.2.2 | Discourse Completion Task (Discourse Completion Task): ML Spanish, EL Basque .....                            | 84  |
| 4.2.3 | Interview Task: ML Basque, EL Spanish .....                                                                   | 89  |
| 4.2.4 | Summary of peak alignment results.....                                                                        | 95  |
| 4.3   | F0 max.....                                                                                                   | 97  |
| 4.3.1 | Monolingual Spanish baseline (Reading Task) .....                                                             | 97  |
| 4.3.2 | Discourse Completion Task (Discourse Completion Task): ML Spanish, EL Basque .....                            | 99  |
| 4.3.3 | Interview Task: ML Basque, EL Spanish .....                                                                   | 101 |
| 4.3.4 | Summary of f0 max results .....                                                                               | 105 |
| 5.    | ..... INTERPRETATION OF RESULTS AND CONCLUSION .....                                                          | 106 |
| 5.1   | Hypotheses addressed .....                                                                                    | 106 |
| 5.2   | Summary of hypotheses .....                                                                                   | 116 |
| 5.3   | Implications of the study's results.....                                                                      | 118 |
| 5.4   | Directions for future study and concluding remarks .....                                                      | 120 |
| 6.    | ..... APPENDICES .....                                                                                        | 123 |
| 6.1   | Appendix A: Participant Language Questionnaire (English and Spanish versions available for participants)..... | 123 |
| 6.2   | Appendix B: Reading task and English translation (Filler sentences in italics) .....                          | 126 |
| 6.3   | Appendix C: Discourse Completion Task (Basque trigger words in bold).....                                     | 128 |
| 6.4   | Appendix D: Native speaker judgment survey .....                                                              | 139 |
| 7.    | ..... REFERENCES .....                                                                                        | 142 |

## LIST OF FIGURES

|                                                                                                                                                                                                         |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1: Map of the Basque Country.....                                                                                                                                                                | 14 |
| Figure 2: Map of the regions of the Basque Country.....                                                                                                                                                 | 14 |
| Figure 3: 2011 Census statistics on native language in the Basque Country by region .....                                                                                                               | 16 |
| Figure 4: 2011 Census statistics on level of Basque fluency in the Basque Country .....                                                                                                                 | 17 |
| Figure 5: Fluent Basque speakers by age, 2011 Census .....                                                                                                                                              | 18 |
| Figure 6: Schematic representation of L+H* pitch accent.....                                                                                                                                            | 21 |
| Figure 7: Example of L+H* (early peak) pitch accent on the word Andaluciako (Andalucian)..                                                                                                              | 22 |
| Figure 8: Prosodic structure of Standard Basque according to Elordieta & Hualde (2014).....                                                                                                             | 23 |
| Figure 9: Northern Bizkaian Basque dialect region.....                                                                                                                                                  | 24 |
| Figure 10: Schematic representation of H*+L pitch accent.....                                                                                                                                           | 25 |
| Figure 11: Example of H*+L pitch accent on the name Nerea.....                                                                                                                                          | 26 |
| Figure 12: Northern Bizkaian Basque prosodic structure .....                                                                                                                                            | 27 |
| Figure 13: Prosodic structure of Spanish, according to Estebas-Vilaplana & Prieto (2008).....                                                                                                           | 29 |
| Figure 14: Schematic representations of H+L* (left) and L*+H (right).....                                                                                                                               | 30 |
| Figure 15: Schematic representations of L+H* (left) and L+<H* (right) .....                                                                                                                             | 31 |
| Figure 16: Pitch tracks containing L+<H* (delayed peak) pitch accent on the name Lorena,<br>L*+H on the Spanish word miraba ('s/he saw'), and L+H* on the Spanish word trilingue<br>(‘trilingual’)..... | 31 |
| Figure 17: Language dominance scores, Semantic Verbal Fluency task.....                                                                                                                                 | 43 |
| Figure 18: Survey interface for rating ambiguous code-switches for the word bueno, ‘well, good’<br>.....                                                                                                | 57 |
| Figure 19: Survey respondents' native language by region .....                                                                                                                                          | 59 |
| Figure 20: Survey respondents' educational system by region .....                                                                                                                                       | 60 |
| Figure 21: Survey respondents' level of education by region .....                                                                                                                                       | 61 |
| Figure 22: Peak alignment measurement example .....                                                                                                                                                     | 67 |
| Figure 23: Histograms of f0 max with and without outliers for Reading Task (left), Discourse<br>Completion Task (middle), and Interview (right).....                                                    | 70 |
| Figure 24: Histograms of peak alignment before and after outliers in the Reading Task (left),<br>Discourse Completion Task (middle), and Interview (right).....                                         | 70 |
| Figure 25: Code-switches by Task and Region .....                                                                                                                                                       | 76 |
| Figure 26: Code-switches by Task and Speaker (1-4 = Bilbao; 5-7 = Lekeitio) .....                                                                                                                       | 77 |
| Figure 27: Interview task code-switches by Lexical Category .....                                                                                                                                       | 78 |
| Figure 28: Code-switches by Prosodic Position .....                                                                                                                                                     | 79 |
| Figure 29: Peak alignment in Reading Task by Prosodic Position .....                                                                                                                                    | 81 |
| Figure 30: Peak alignment in the Reading Task (monolingual Spanish).....                                                                                                                                | 82 |
| Figure 31: Peak alignment in the Discourse Completion Task, Bilbao group.....                                                                                                                           | 87 |
| Figure 32: Peak alignment in the Discourse Completion Task, Lekeitio group .....                                                                                                                        | 88 |
| Figure 33: Peak alignment by Social Language in Interview Task .....                                                                                                                                    | 91 |

|                                                                                                                    |     |
|--------------------------------------------------------------------------------------------------------------------|-----|
| Figure 34:Peak alignment by Region and Context in the Interview Task .....                                         | 92  |
| Figure 35: Interaction between Context and Language Dominance (Semantic Verbal Fluency) in the Interview Task..... | 94  |
| Figure 36: f0 max in Reading Task.....                                                                             | 98  |
| Figure 37: f0 max for Lekeitio speakers by Position and Context in Discourse Completion Task .....                 | 100 |
| Figure 38: f0 by Language Dominance (Semantic Verbal Fluency) in the Interview Task .....                          | 103 |
| Figure 39: f0 max by Social Language in the Interview Task.....                                                    | 104 |

## LIST OF TABLES

|                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 1: Basque dialects in present-day Spain according to Trask (1997) and Zuazo (2013) .....                                | 19  |
| Table 2: Expected differences in peak alignment, according to Elordieta & Hualde (2014) and Elordieta & Irurtzun (2016) ..... | 34  |
| Table 3: Speaker demographics and self-reported language use .....                                                            | 38  |
| Table 4: Self-reported language competence and dominance; “S” = Spanish, “B” = Basque .....                                   | 39  |
| Table 5: Average responses per language in the Semantic Verbal Fluency task.....                                              | 42  |
| Table 6: Basque trigger words for code-switches in the Discourse Completion Task .....                                        | 47  |
| Table 7: Tasks administered to participants.....                                                                              | 48  |
| Table 8: Unambiguous Spanish code-switches in Interview task.....                                                             | 52  |
| Table 9: Ambiguous code-switches from Interview task .....                                                                    | 53  |
| Table 10: Hybrid code-switches from the Interview task.....                                                                   | 55  |
| Table 11: Unambiguous filler words for survey .....                                                                           | 57  |
| Table 12: Code-switch survey scores by dialect region.....                                                                    | 63  |
| Table 13: Differences in peak alignment for Spanish and Basque (Elordieta & Irurtzun (2016))                                  | 64  |
| Table 14: Break indices used in current study .....                                                                           | 65  |
| Table 15: Outlier check for outcome variables .....                                                                           | 69  |
| Table 16: Predictor variables used in regression diagnostics for multicollinearity .....                                      | 73  |
| Table 17: Variables used in regression models.....                                                                            | 73  |
| Table 18: Peak alignment means for Reading Task.....                                                                          | 82  |
| Table 19: Regression results for peak alignment in the Reading Task.....                                                      | 83  |
| Table 20: Hypothesized Spanish peak alignment in ML and EL contexts .....                                                     | 84  |
| Table 21: AIC and BIC of peak alignment models for the Discourse Completion Task .....                                        | 85  |
| Table 22: Regression results for peak alignment in the Discourse Completion Task .....                                        | 89  |
| Table 23: AIC and BIC of peak alignment models for the Interview Task.....                                                    | 89  |
| Table 24: Peak alignment means by Region and Context in the Interview Task .....                                              | 92  |
| Table 25: Regression results for peak alignment in the Interview Task.....                                                    | 95  |
| Table 26: Regression results for f0 max in Reading Task .....                                                                 | 98  |
| Table 27: AIC and BIC of f0 max models for the Discourse Completion Task.....                                                 | 99  |
| Table 28: Regression results for f0 max in Discourse Completion Task .....                                                    | 101 |
| Table 29: AIC and BIC comparing f0 max models for the Interview Task .....                                                    | 102 |
| Table 30: Regression results for f0 max in the Interview Task .....                                                           | 104 |
| Table 31: Summary of hypotheses and relevant findings .....                                                                   | 117 |

## ACKNOWLEDGMENTS

There are many people who have contributed to the inception, execution, and completion of this dissertation and I am thankful for each and every one of them. First and foremost, I am grateful for the help and guidance of my advisor, Sun-Ah Jun, who has been incredibly supportive of me during my time at UCLA. I have learned so much about intonation, research, and life from her that I will never forget. I am also very grateful for my committee members, Pat Keating, Bruce Hayes, and Gorka Elordieta, who took the time to advise my work and provide feedback even when on sabbatical and/or nine hours of time zones away. I have learned so much from each of you and truly thank you for your time and knowledge.

I am also tremendously grateful to Itxaso del Olmo and Unai Nafarrate Errasti, both of whom played a huge part in my dissertation. Itxaso's help in collecting my data and Basque-language consultations were absolutely invaluable to my work. More than that, Itxaso invited me to Lekeitio, where I got to meet her family, friends, and experience my favorite town in the Basque Country. Unai's hard work transcribing my Basque language data not only provided the basis of my data analysis, but also the start of a great friendship. My dissertation is dedicated to you both. Thank you for your help and for sharing Basque language and culture with me.

The other crucial ingredient that made my dissertation possible was the strong support system consisting of my family and friends. Y'all kept me sane through some rough and transformative times I experienced as a graduate student and I cannot adequately express my gratitude for the times you've listened, helped, fed me, brought me coffee, proofread things, and reminded me how fun life can be despite hard times. After this dissertation is done, those memories and events will remain to remind me of what is truly important in my life.

## VITA

|      |                                                             |
|------|-------------------------------------------------------------|
| 1987 | Born                                                        |
| 2009 | B.A., Music, Florida State University                       |
| 2012 | M.A., Spanish, Florida State University                     |
| 2014 | M.A., Linguistics,<br>University of California, Los Angeles |

## PUBLICATIONS

- Aly, Ann, Stephanie Jacobson, Nancy Ward, & Megha Sundara (in preparation).  
Acoustic correlates of stress in Spanish infant-directed speech.
- Aly, Ann (in preparation). Is less more? Predictors of academic performance in bilingual  
and monolingual US college students.
- Glewwe, Eleanor & Ann Aly. (in press). Palatalization in Llogoori. Proceedings of  
ACAL 46, 2014.
- Aly, Ann (in press). Code-switching in Miami Cuban Spanish: Segmental and  
Suprasegmental effects. In Cuza, A. (ed) *Contemporary Approaches to Cuban Spanish  
Dialectology*. Georgetown University Press.
- Aly, Ann. (2015). The multiple prosodic cues differentiating questions and statements in  
Miami Cuban Spanish. In *Proceedings of The 18<sup>th</sup> International Congress of Phonetic  
Sciences*. University of Glasgow Press: Glasgow.
- Aly Bailey, Ann. (2014). Cuban Spanish Intonation: A preliminary AM model. In  
Campbell, N., Gibbon, D., & D. Hirst. *In Proceedings of Speech Prosody 7* (pp. 809-  
813). Trinity College, Dublin, Ireland.
- Aly Bailey, Ann. (2013). Similar, yet different: Acquisition of Brazilian Portuguese nasal  
vowels by native speakers and Spanish-English bilinguals. In J. Aaron, J. Cabrelli  
Amaro, G. Lord, & A. de Prada Pérez. *Selected Proceedings of the 16th Hispanic  
Linguistics Symposium* (pp. 128-142) Somerville, MA: Cascadilla Proceedings Project.
- Aly Bailey, Ann & Anel Brandl. (2013). Incorporating pronunciation instruction in the

beginning Spanish language classroom: A perceptual study. In J. Levis & K. Levelle (ed.) *Proceedings of the 4<sup>th</sup> Pronunciation in Second Language Learning and Teaching Conference, 2012* (pp. 207-223). Ames, IA: Iowa State University.

# 1 INTRODUCTION

## 1.1 Overview of the dissertation

The current dissertation investigates how the differences between two languages are consolidated in contexts which involve switching between these languages (also known as *code-switching*). More specifically, the current study is concerned with two main questions related to code-switching: how are differences in prosody (i.e., sounds above the segmental level) realized during code-switching and how do competency-related factors such as language use and dominance affect the prosodic realizations of code-switches? These two questions will be examined using a Spanish-Basque speech corpus, which contains bi-directional code-switches (i.e., Spanish to Basque and Basque to Spanish). Spanish and Basque contain several key prosodic differences, which will be discussed in more detail in the following two chapters in order to contextualize the prosodic variables chosen for analysis. Additionally, the sociopolitical climate in northern Spain over the past century has resulted in asymmetrical levels of fluency in Spanish and Basque as well as semi-diglossic language environments. These latter issues will be addressed by operationalizing language fluency and domains of use for both Spanish and Basque to account for their effects on the prosodic variables chosen.

The following sections of this chapter will provide a brief introduction to code-switching as a discipline, including previous studies and relevant theoretical frameworks. In the next chapter, background information about the sociopolitical environments and prosodic systems of Spanish and Basque will be discussed.

## 1.2 Introduction to code-switching

*Code-switching* (also referred to as *code-mixing*) refers to the use of two or more languages, language styles, or dialects within a single discourse or interaction. Many interactions involving

code-switching contain a dominant language (or dialect), which contributes the majority of the linguistic material of the interaction, and one or more guest languages, which contribute less material (i.e. the code-switched material). Although less common, it is also possible for there to be equal contributions of linguistic material from all languages or varieties involved in the interaction, in which case there is no clear dominant language. Code-switching is common practice in bilingual and bidialectal communities and serves many linguistic, social, and pragmatic purposes, as will be discussed in detail in the present dissertation. Within bilingual communities, there are often colloquial names for the pairings of languages that are frequently code-switched, such as *Spanglish/Espanglish* (Spanish and English), *Taglish/Barok/Carabao* (Tagalog and English), *Portuñol/Portunhol* (Spanish and Portuguese), and *Manglish* (a mixture of Malay, English, Tamil, and/or Cantonese).

In many communities, code-switching is a highly stigmatized or politicized practice (e.g. Chana & Romain (1984) for Punjabi-English bilinguals, Bentahila (1983) for French-Arabic bilingual Moroccans, Lantto (2014) for Spanish-Basque bilinguals). Common beliefs about code-switching (attested in the aforementioned studies as well as anecdotal experience) include that speakers who code-switch lack competence in one or more languages they speak; that code-switching creates its own dialect or creole (e.g. *Spanglish*) that is viewed as distinct (or in some cases inferior) to the monolingual varieties of those language(s); and that code-switching is less prestigious than monolingual interactions.

Due to these commonly held beliefs, research on code-switching can be challenging to conduct. Furthermore, the variation in code-switching observed between language communities (and even within the same community) makes cross-linguistic generalizations on code-switching difficult, as will be discussed in the following section. Despite these difficulties, studies

beginning in the late 1970s have contributed numerous findings and insights to the linguistics literature, spanning disciplines such as formal linguistics (e.g. syntax, phonology), sociolinguistics, psychology/psycholinguistics, and linguistic anthropology. The following sections will discuss the previous research on code-switching, beginning with the early studies focusing on morphosyntactic properties of code-switching and continuing with more recent studies on the psycholinguistic and phonological properties of code-switching.

### 1.3 Previous studies on code-switching

#### 1.3.1 Morphosyntax and the Matrix Language Framework

Early work on code-switching focused primarily on describing the morphosyntactic patterns found in code-switch corpora and subsequently developing grammatical constraints based on the observed patterns. These studies were pivotal in establishing that code-switching contained systematic patterns (similar to monolingual varieties), which challenged stereotypes about code-switching as random, incompetent language mixing. Whereas the syntactic studies of the 1980s and 1990s (Poplack (1980), Belazi et al (1994), DiScuillo et al (1986), Myers-Scotton (1993), *inter alia*) proposed constraints specific to code-switching contexts (e.g., dictating which patterns of code-switching were (un)grammatical), later studies within the Minimalist framework (Chomsky (1995)) such as MacSwan (1999) and López (2017) propose that code-switching patterns should be accounted for within the existing syntactic constraints of Minimalism (e.g., Merge and Move). Although most of the morphosyntactic proposals are specific to that subfield, the Matrix Language Framework, proposed by Myers-Scotton (1993) has the potential to apply to other linguistic subfields, such as the phonological and phonetic questions raised by the current study.

The Matrix Language Framework's central construct is that there is a *matrix language* (ML) and an *embedded language* (EL) during code-switching whose roles differ in their capacities. The ML is the language that Myers-Scotton (1993) claims plays a more dominant role, which may be defined differently depending on the study and discipline to which it belongs. Most commonly, the ML is defined through a frequency-based criterion, in which the language which contributes the highest number of morphemes in the interaction is the ML. Less-frequent but attested criteria for defining the ML include identifying the language in which speakers are most fluent; L1 status (in which the L1 is the ML); and identifying which language is “socially unmarked” for a specific context and defining it as the ML. Following the most common and empirically-verifiable method, the present study will use the frequency-based criterion when discussing ML contexts; the ML is that which contributes more morphemes and the EL is that which contributes fewer.

The asymmetry between the ML and EL with respect to frequency has consequences for the constructs for the Matrix Language Framework. Myers-Scotton's (1993) *Matrix Language Hypothesis* states that the ML sets the morphosyntactic “frame” of the interaction, determining the morpheme order as well as providing all *system morphemes* (system morphemes are defined by Myers-Scotton as any morpheme which is [+Quantification] as opposed to *content morphemes*, which are [-Quantification]). Thus, in the Matrix Language Framework, the ML has a privileged role, setting the morphosyntactic patterns of the interaction, which are only violated in the case of *EL islands* (usually idioms or other “peripheral” EL material that does not obey the ML frame).

The present study will test the claims of the Matrix Language Framework on prosodic variables to see if the prosodic patterns of the Matrix Language override those of the Embedded

Language when cross-linguistic prosodic differences arise. If the privileged status of the Matrix Language seen in Myers-Scotton (1993) for morphosyntactic patterns also applies to phonological properties, it is possible that the Matrix Language provides a *linguistic* frame (as opposed to a *morphosyntactic* frame) for the languages involved. The present study hypothesizes that the Matrix Language of the interaction will provide the prosodic frame for the utterance, to which any Embedded Language material will assimilate. Chapter 2 will present more detail on the specific prosodic features of Spanish and Basque that will be tested with the Matrix Language Framework.

### 1.3.2 Code-switching and cognitive control: Insights from neuro/psycholinguistics

Studies investigating code-switching from a cognitive or neurological perspective typically involve controlled *language switching* tasks which require the participant to either read or name specified stimuli in the language(s) that they speak. Although these types of tasks are controlled simulations of code-switching, they can provide insights into cognitive processes involved in switching languages and tasks as measured by reaction times and neuro-imaging. The findings of studies examining reaction times during language switching among bilinguals will be discussed in this section.

One major finding from language and task switching studies provides evidence for different reaction times depending on the speaker's level of fluency. For example, highly-fluent bilinguals who completed word-naming tasks in Gullifer et al. (2013) showed no difference in reaction times between experimental blocks which alternated languages (simulating code-switching) and single language blocks (simulating monolingual contexts). In Meuter & Allport (1999), L1 (first language) dominant bilinguals had longer reaction times (i.e. higher processing costs) on numeral-naming trials which switched languages when compared to trials which did not switch.

Within the language-switched trials, those switching from L2 (second language) to L1 had longer reaction times than trials switching from L1 to L2. These results suggest not only that level of fluency in bilinguals may mediate the level of processing cost involved in language switching, but also that the direction of switching may have different processing costs in asymmetrical (i.e. those more fluent in one language) bilinguals.

Processing costs related to language dominance and fluency can also be seen at the phonetic level in Olson (2012), which employs a picture-naming task with Spanish-English bilinguals of different proficiency levels (L1-dominant and balanced bilinguals). Instead of measuring reaction times as the previously discussed studies did, the voice onset timing (VOT) of voiceless stops (/p, t, k/) were measured. Spanish and English differ in their realization of voiceless stops, with English making use of long-lag (also referred to as *aspirated*) VOT in stressed positions, whereas Spanish VOT is always short-lag (or *unaspirated*). The actual duration of aspirated and unaspirated stops occupies a range that can depend on various dialectal and individual factors, so participants in Olson (2013) were compared to themselves (i.e., only within-speaker and not across-speaker comparisons of Spanish and English VOT).

The results showed that in monolingual mode (95% in English or Spanish; 5% in the other language), both Spanish-dominant and English-dominant participants showed VOT costs in their L1. That is, Spanish-dominant participants had *longer* VOTs in Spanish during code-switched trials (compared with their own Spanish non-switched production) and English-dominant participants had *shorter* English VOTs in English code-switched trials. This effect was not seen in bilingual mode (50% English, 50% Spanish), leading Olson (2013) to propose that monolingual and bilingual modes require different levels of inhibitory control which may affect one language (here, the L1, similar to Meuter & Allport (1999)) more than the other. In a follow-

up study, Olson (2016a), employing a similar methodology as Olson (2013), confirmed that language switching costs were most severe in (mostly) monolingual mode (95%), with no evidence for a “compound” effect (i.e. extra processing cost) of bilingual mode combined with code-switching.

The studies reviewed in this section reveal several valuable insights about how bilingualism interacts with cognitive functions and phonological realization while code (or language) switching. In particular, studies such as Gullifer et al (2013) and Meuter & Allport (1999) suggest that highly-fluent bilinguals may have more of a cognitive advantage on task switching than L1-dominant bilinguals. Furthermore, level of fluency may also affect the direction of influence; the L1-dominant bilinguals in Olson (2013) showed compromise (i.e. values different from their monolingual baseline) VOT durations for their L1 on code-switching trials in (mostly) monolingual blocks, but not their L2, indicative of difficulty with inhibitory control in mostly monolingual contexts. The language dominance and fluency levels of the present study’s participants will be considered in order to test whether the phonetic effects seen in Olson (2013, 2016b) also apply to suprasegmental (i.e. prosodic) variables in naturalistic code-switching. More information on the present study’s operationalization of fluency and prosodic variables will be discussed in Chapters 2 and 3.

### 1.3.3 Code-switching and sound systems: Research on phonetics and phonology

Studies on the effects of code-switching on the phonetics and phonology of the languages involved have lagged behind those on morphosyntax and processing until recently. Due to the combined difficulty of eliciting code-switching in an environment which is adequate for acoustic analysis (i.e. minimal background noise and controlled segmental contexts), most studies analyzing segmental features of code-switching (such as Bullock et al (2006) and Olson (2012))

use scripted tasks with code-switches in them. With the more recent development of corpora that include naturalistic and spontaneous speech, researchers such as Fricke et al (2016) and Balukas & Koops (2015) have investigated phonetic and suprasegmental aspects of code-switching in more naturalistic data. Major findings seen in common across code-switching studies on phonetics and phonology include effects of *language context* (e.g., monolingual or bilingual modes), *dominant language* (L1 or L2), *language-specific effects*, and *anticipatory/spillover effects* (e.g., modifications before or after the code-switching site). This section will discuss each of these major findings with respect to the studies which contribute to each of them.

Olson (2012, 2016b), Piccinini & Garellek (2014), and Aly (in press) examine f0 in code-switched and monolingual modes in Spanish-English bilinguals, revealing differences between these language contexts. However, their results differ in their details. Piccinini & Garellek (2014) examine the f0 of one female bilingual in monolingual Spanish and English modes as well as code-switched modes (i.e. sentences with either embedded Spanish or English words) using pre-scripted sentences. Their results show that overall (global) f0 is highest in English, followed by code-switched modes, and lowest in monolingual Spanish modes. These results suggest different strategies for monolingual and code-switched contexts as well as potential inherent language differences in f0 realization between Spanish and English. Olson (2012, 2016b) also measures f0 in pre-scripted speech (pseudo-dialogues) in Spanish-dominant and English-dominant bilinguals. Olson (2012) revealed higher f0 and longer vowel duration on code-switched English words when compared to monolingual English words, proposing that code-switched words are inherently focused or hyperarticulated. Perceptual support for Olson's (2012) Hyperarticulation account is found in Fricke et al (2016), in which listeners perceived code-switches as more prominent than non-switches, particularly if they were NPs (noun phrases).

Later work by Olson (2016b), which employed the same task type and bilingual populations, found evidence of a gradient in f0 range and vowel duration in different language modes, similar to the spectrum found in Piccinini & Garellek (2014). That is, insertional code-switches have the greatest pitch range in a mostly monolingual mode (i.e. few embedded or insertional morphemes code-switches) compared to bilingual mode (nearly equal contributions from both languages) and fully monolingual mode. Additionally, this spectrum was mediated by language dominance, in which f0 and duration were significantly greater in speakers' L1 (regardless of language mode) but not their L2.

Interactions between language mode and language dominance are also seen in Aly (in press). This study investigates the pitch accent type and peak alignment (distance in ms from the f0 peak to the offset of the associated syllable) in a semi-spontaneous corpus of Miami-Born and Cuba-Born Spanish-English bilinguals. When monolingual Spanish and code-switched (embedded English words) contexts were compared, only the Cuba-Born group, who are Spanish-dominant, showed significant differences in peak alignment based on language context in which f0 peak alignment was earlier in monolingual Spanish than code-switched mode. This difference in peak alignment was not seen in the Miami-Born group, who were English-dominant bilinguals. However, when pitch accent type was investigated, the Miami-Born group preferred to realize English code-switches with pitch accents exclusive to American English (e.g., H\* and H+!H\*, which are not in the tonal inventory of Miami Cuban Spanish) whereas the preferred pitch accent of the Cuba-Born group exists in both inventories, (e.g., L+H\*). This asymmetry suggests that language dominance may determine how bilinguals manage switching language modes,

including strategies involving phonetic (e.g. peak alignment) or phonological (e.g. pitch accent type) distinctions<sup>1</sup>.

Apart from effects related to the speakers themselves (i.e. language mode or dominance), language-specific effects on code-switching are seen in several studies which examine VOT duration in Spanish-English code-switching contexts. As discussed in the previous section, Olson (2013) found evidence for L1 effects (i.e. shorter English VOT and longer Spanish VOT) during a language switching task. These results were also corroborated in a later code-switching study by Olson (2016c) with a slight difference: whereas the English-dominant group still realized English code-switches with shorter VOT than in monolingual mode, Spanish-dominant speakers realized English code-switches with shorter VOT and Spanish code-switches with longer VOT than their respective monolingual modes. Shorter English VOT during code-switches is also reported in Bullock et al (2006) for both Spanish and English-dominant speakers and in Balukas & Koops (2015) and Piccinini & Arvaniti (2015) for early Spanish-English bilinguals. Additionally, Piccinini & Arvaniti (2015) report *shorter* Spanish VOT during code-switching, which is different from Olson (2016c). However, this difference may be accounted for when task and participants are taken into consideration: Piccinini & Arvaniti (2015) analyzed semi-spontaneous speech of early bilinguals whereas Olson (2016c) analyzed scripted speech of late bilinguals.

The last major finding in phonetic code-switching studies that will be discussed is anticipatory /spill-over effects. The existence of anticipatory cues in the acoustic signal (that differ from monolingual mode) would indicate that listeners have phonetic cues leading up to

---

<sup>1</sup> Although peak alignment is a phonological distinction in some Spanish varieties, such as Peninsular Spanish (Estebas Vilaplana & Prieto (2008)), it is a phonetic distinction in Miami Cuban Spanish (Aly (2014)).

code-switches and that code-switching may involve long-distance speech planning strategies. A spill-over effect after a code-switch suggests repair or lag in switching back into the matrix or dominant language after a code-switch. With respect to anticipatory effects, Bullock et al (2006) reported shorter English VOT before a (Spanish) code-switch and Fricke et al (2015) reported slower speech rate before a code-switch. Fricke et al (2015) also tested the effects of slowed speech rate before code-switching in an eye-tracking study, using habitual code-switchers as participants. Their results revealed more robust word recognition of code-switched words when anticipatory speech rate cues were present as opposed to speech without anticipatory cues. This suggests that habitual code-switchers may be able to utilize phonetic cues to activate other languages for faster processing during an interaction. With respect to spill-over effects, similar results are seen for Spanish-English VOT in which post-switch English VOTs are shorter (Balukas & Koops (2015), Wilson et al (2017)). Wilson et al (2017) also provides evidence for *longer* Spanish VOT after a code-switch and Elias et al (2017) report more centralized Spanish vowels after an English code-switch, suggesting that English-like features (such as longer VOT and vowel reduction) may “leak” onto Spanish words post-switch.

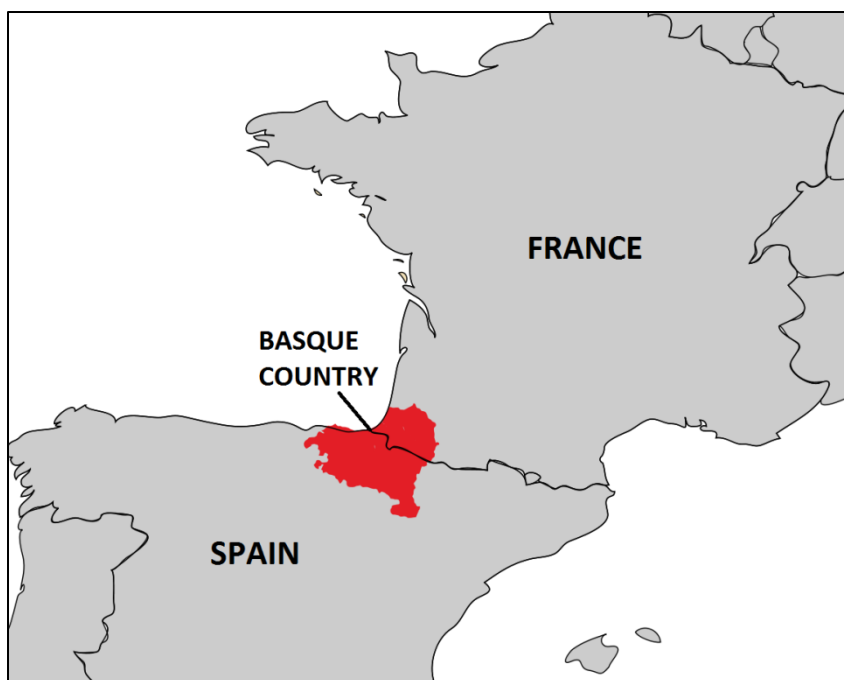
The studies in this section provide strong support for effects of code-switching on VOT in Spanish-English speech, although the direction of the effect may interact with language dominance, language mode, and task type. These studies also provide evidence that code-switching effects may be phonetic or phonological in nature and that phonetic cues that differ due to speech rate may be perceived by bilingual listeners as a means to facilitate word recognition. These valuable insights from the studies discussed are also met with several limitations: all the studies in this section examine Spanish-English bilinguals in the United States (with the exception of Wilson et al (2017), who examine Spanish-English bilinguals in Spain),

making generalizations to other bilingual communities difficult. The majority of the studies discussed analyze scripted speech (although more recent studies are now using corpora), which may not reflect naturalistic code-switching and the many socio-pragmatic functions that may constrain it. Lastly, the majority of studies discussed investigate phonetic properties, such as VOT or f0 maxima, which do not inform how speakers may make use of phonological or contrastive information which may differ between languages while code-switching. The following chapter will provide background information on Spanish and Basque in order to contextualize the variables measured in the present study, including the sociopolitical and prosodic information relevant to this study's analysis.

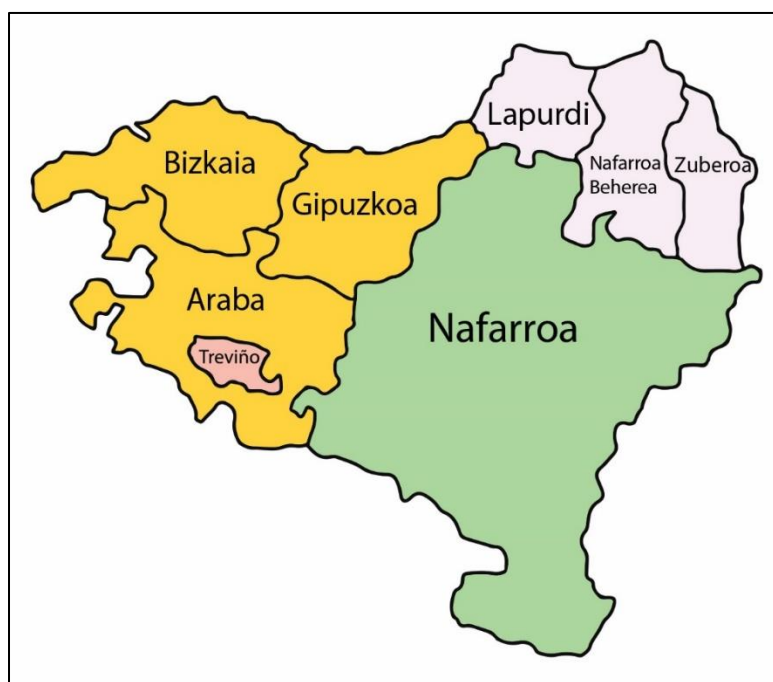
## 2 SPANISH AND BASQUE CONTEXTUALIZED IN THE BASQUE COUNTRY

### 2.1 Language status, use, and demographics

Basque, a language isolate, is spoken in current-day northern Spain and southern France. The region in the northern Spain where Basque is spoken is referred to as the Basque Autonomous Region or more generally known as the *Basque Country* (*País Vasco* in Spanish; *Euskal Herria* in Basque) and includes the regions of Araba/Álava (henceforth Araba), Gipuzkoa, and Bizkaia. The part of the Basque Country situated in present day Spain is also referred to as the Southern Basque Country. Basque is recognized as a co-official language (along with Spanish) in the Basque Autonomous Region as well as in Nafarroa, a separate autonomous region of Spain with historically Basque roots. In present-day France, the regions of Lapurdi, Nafarroa Beherea, and Zuberoa (also known as the Northern Basque Country) are historically Basque-speaking provinces where Basque does not have co-official status. The map in Figure 1 (Eddo (2010); text added by the current author) shows the Basque Country within the geographical context of present-day Spain and France; the map in Figure 2 (Willtron (2006); colors and labels edited by the current author) shows the Basque Country provinces in more detail. In Figure 2, the areas in yellow are Spain's Basque Autonomous Community; the green portion is the autonomous region of Nafarroa; and the lavender portions are the Basque Country regions of France.



*Figure 1: Map of the Basque Country*



*Figure 2: Map of the regions of the Basque Country*

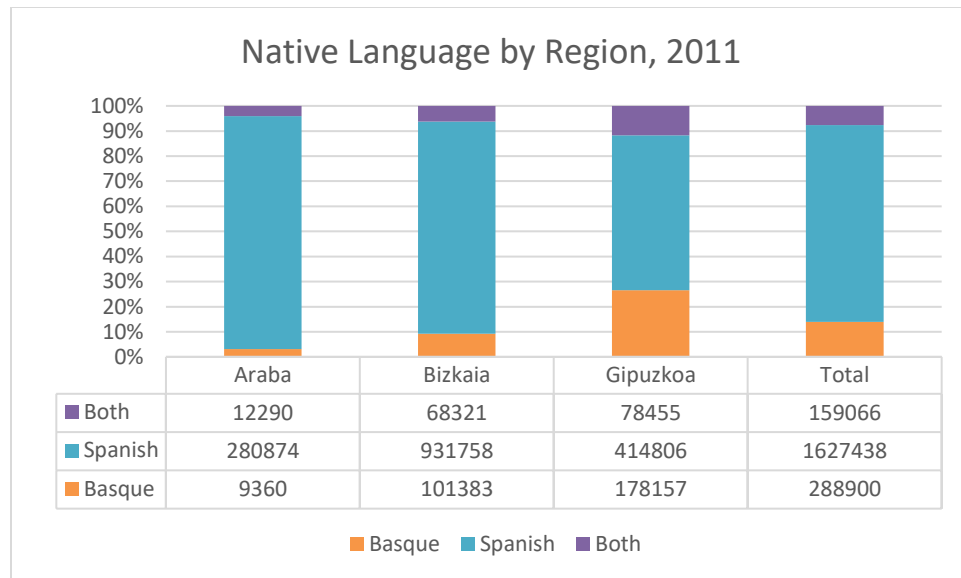
Under the dictatorship of Francisco Franco in Spain from 1936-1975, Spanish was the only official language permitted in Spain, with the use of any other language considered illegal. This suppressed much of Spain's linguistic diversity (such as Catalan, Galician, and Basque, among others), resulting in endangered statuses for many non-Spanish languages until after Franco's death in 1975. Since the end of the dictatorship, Spain recognizes several languages as co-official languages in bilingual regions, including Basque in the Basque Country.

Public education in Basque is now available in the Basque Country with multiple levels of immersion ranging from education primarily in Spanish with Basque as a second language (Model A), 50/50 Spanish-Basque bilingual education (Model B), Basque as the primary language with Spanish as a second language (Model D), an exclusively Spanish model for seasonal residents of Bizkaia (Model X) and a trilingual model (English-Spanish-Basque) which was piloted at 40 schools in Spain starting in 2010. In Nafarroa, the same Spanish-Basque bilingual models (A, B, and D) are also offered in addition to an exclusively Spanish language model (Model G) available to all residents<sup>2</sup>.

Despite language revitalization efforts, statistics from the Basque Country Population and Housing Census (Basque Government (2011)) reveal that the majority of Basque Country residents in Bizkaia, Araba, and Gipuzkoa are native Spanish speakers (acquired before the age of three). Bizkaia reported the highest number of native Basque speakers in 2011, although these speakers only represent 27% of those surveyed. Figure 3 shows the native language(s) reported by Basque Country residents in 2011 by region for comparison.

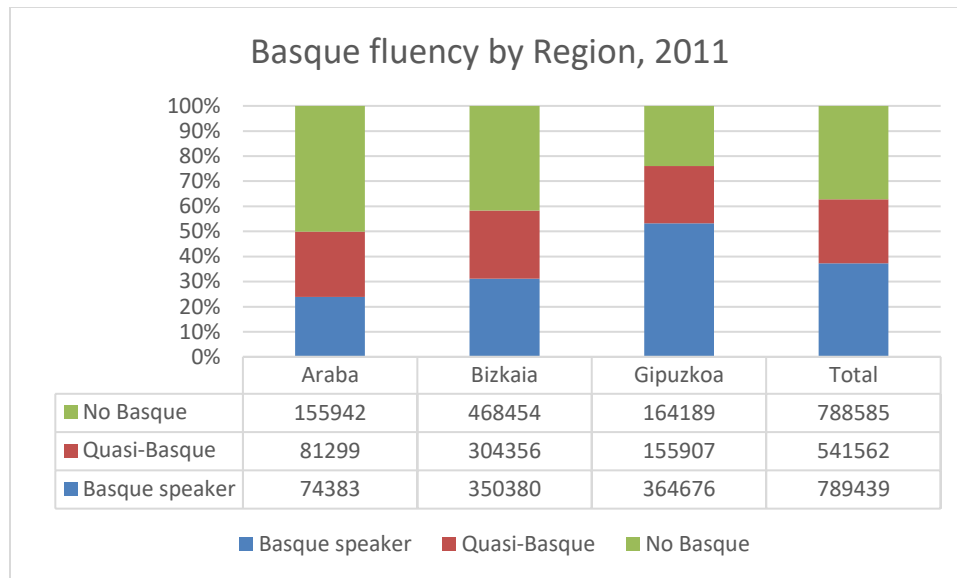
---

<sup>2</sup> For clarification purposes, there is no Model C in Spanish bilingual education.



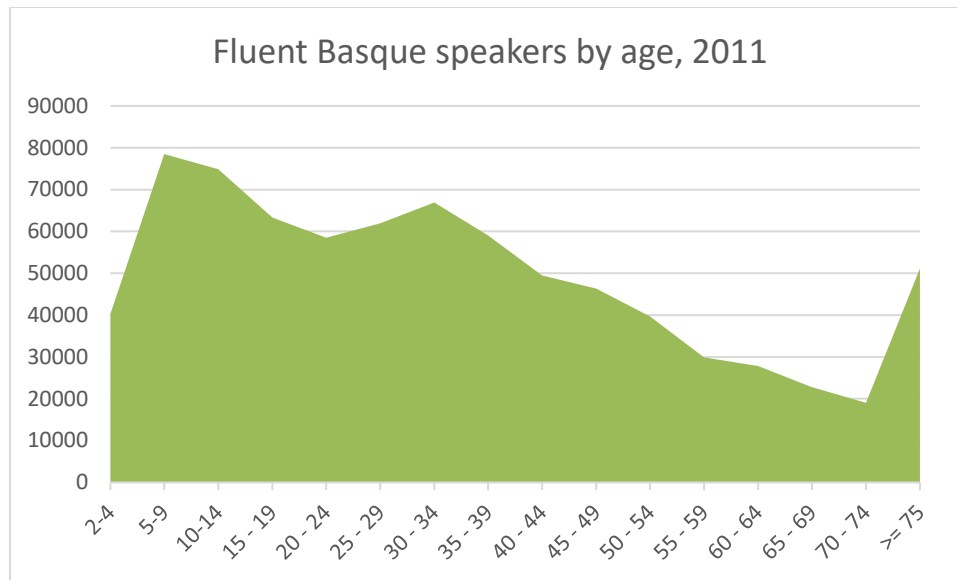
*Figure 3: 2011 Census statistics on native language in the Basque Country by region*

Although most Basque Country residents surveyed in the 2011 Census are native Spanish speakers, 63% of respondents reported having at least some knowledge of Basque. Gipuzkoa has the highest number of fluent Basque speakers, with 76% of respondents reporting that they understand, speak, read, and write “well” in Basque (followed by 58% in Bizkaia and 50% in Araba). These results imply that a number of respondents in the Basque country are L2 Basque speakers, having learned in a formal or educational setting (such as Model A or B). Figure 4 illustrates the levels of Basque fluency by region for comparison.



*Figure 4: 2011 Census statistics on level of Basque fluency in the Basque Country*

When Basque fluency and age are considered, the 2011 Census reveals that the highest concentration of fluent Basque speakers are those under the age of 34, as seen in Figure 5. This is expected, given the language revitalization initiatives (such as bilingual education) that began after the dictatorship and the recognition of Basque as a co-official language with Spanish in 1978. Another interesting observation from Figure 5 is the relatively high number of fluent Basque speakers over 75 years of age; it is possible that these respondents are speakers who acquired Basque as children (e.g. the youngest respondents in this category were born in 1936) and maintained it throughout the dictatorship.



*Figure 5: Fluent Basque speakers by age, 2011 Census*

These demographics reveal several important linguistic traits about the Basque Country. First, Spanish fluency is still widespread and is the most common native language. Second, the majority of Basque Country residents have at least some Basque fluency, indicating its status as a second language for most residents. Lastly, the highest concentration of fluent Basque speakers are 34 years old or younger, indicating at least some degree of effectiveness for Basque language revitalization. In order to sample from the demographic with the highest level of Basque fluency (and therefore more likely to code-switch), only speakers under the age of 35 (as of 2014) who are native Basque speakers or have attended bilingual schools are included in the present dissertation. More information on this dissertation's participants will be discussed in detail in Chapter 3. The next sections will provide information about the prosodic systems of Peninsular Spanish and Basque in order to contextualize the present study's hypotheses, which will be presented at the end of the chapter.

## 2.2 Basque prosody and intonation

### 2.2.1 Standard Basque

Standard Basque (known as *Euskara Batua* in Basque) is the written standard of Basque developed in the 1960s for use in formal contexts, with its origins in Gipuzkoan and Lapurdian literary registers (Zuazo (2013), Elordieta & Hualde (2014)). Standard Basque is also spoken by L2 Basque speakers or native Basque speakers who learned Basque from L2-speaking parents. Although based on Gipuzkoan (or Central, due to its location) Basque, it may be spoken in any region of the Basque country, often in addition to a local vernacular variety. The local varieties of Basque are categorized geographically; Vasconists (i.e. Basque linguists) such as Trask (1997) use a more historical approach to dialect classification, including obsolete and extinct dialects, whereas Zuazo (2013) employs a more modern approach in which only dialects currently in use are included. A comparison of the Trask (1997) and Zuazo (2013) dialect regions that pertain to present-day Spain is shown in Table 1. The local dialect that will be relevant to the present study is Bizkaian/Western Basque, which will be discussed along with Standard Basque in the following sections.

| <b>Trask (1997)</b> | <b>Relevant region(s)</b> | <b>Zuazo (2013)</b> | <b>Relevant region(s)</b>                 |
|---------------------|---------------------------|---------------------|-------------------------------------------|
| Bizkaian            | Bizkaia                   | Western             | Bizkaia, Northern Araba, Western Gipuzkoa |
| Gipuzkoan           | Gipuzkoa                  | Central             | Gipuzkoa, Western Nafarroa                |
| High Navarrese      | Nafarroa                  | Navarrese           | Nafarroa                                  |
| Aezkoan             | Nafarroa (Aezkoa Valley)  | Navarrese           | Nafarroa                                  |
| Salazarese          | Nafarroa (extinct)        | N/A                 | N/A                                       |

*Table 1: Basque dialects in present-day Spain according to Trask (1997) and Zuazo (2013)*

Elordieta & Hualde (2014) propose AM intonational models of Standard Basque and Lekeitio Basque, both of which will be discussed and referred to extensively in the subsequent sections. Standard Basque is a lexical stress variety, in which the position of stress is not

predictably by rules, although there are tendencies. For example, lexical stress in Standard Basque is typically found on the second syllable of a word, but can be on the first or third syllable in certain morphological environments (i.e., interactions with stems and affixes, such as imperfectives and plural nominals). Examples demonstrating each case are shown in (1), in which the stressed syllable is marked in bold font, with affixes separated from stems by a hyphen.

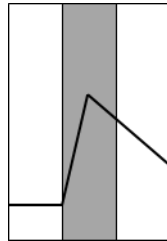
(1) Typical stress pattern: **gizon**-ak ‘man-plural’

b. Initial stress: **gizon**-a-i ‘man-plural definite-dative’ (*gizonai* also possible)

c. Stress on third syllable: gara**gar**do ‘beer’ (*garagardo* also possible)

Although Elordieta & Hualde (2014) clarify that the specific phonetic correlates of Basque stress have not been fully established, the stressed syllables of content words in Standard Basque bear rising pitch accents, notated as (L+H)\* with an f0 peak that occurs within the stressed syllable (*early peak*). (L+H)\* is the only pitch accent in the tonal inventory of Standard Basque. The pitch accent that precedes the verb (which also bears a pitch accent) in canonical Subject-Object-Verb order utterances is the *nuclear pitch accent*. The nuclear pitch accent is often perceived as the most prominent word to native speakers and carries the informational focus of the utterance. Elordieta & Hualde (2014:444) observe that (L+H)\* pitch accents in Standard Basque may also have f0 peaks realized on the post-tonic syllable (*delayed peaks*) when pre-nuclear (i.e., before the nuclear pitch accent), but that there is no evidence that peak alignment (i.e., delayed versus early f0 peaks) is contrastive in this dialect. Instead, they suggest that this alignment difference may surface in the presence of a following phrase-final pitch excursion, or boundary tone.

For ease of distinction and cross-linguistic comparison, pitch accents with early peaks will be referred to as L+H\* and pitch accents with delayed peaks will be referred to as L+<H\* in the present study, following the labeling conventions of the Peninsular Spanish AM model proposed by Estebas Vilaplana & Prieto (2008). However, phonologically, both early and delayed realizations are represented as (L+H)\* in the Standard Basque AM model proposed by Elordieta & Hualde (2014). L+H\* is shown schematically in Figure 6 with a black line representing f<sub>0</sub>. The gray portions of the schematic pitch accent represent the stressed syllable and the white portions surrounding it are (hypothetical) pre- and post-tonic syllables. Below the schematic pitch accent is a pitch track in Figure 7 showing the Basque word *Andaluciako* (‘Andalusian’) with a L+H\* pitch accent on the penultimate syllable (indicated with box). This word (*Andaluciako*) has penultimate stress as opposed to stress on the second syllable, as is most common in Standard Basque.



*Figure 6: Schematic representation of L+H\* pitch accent*

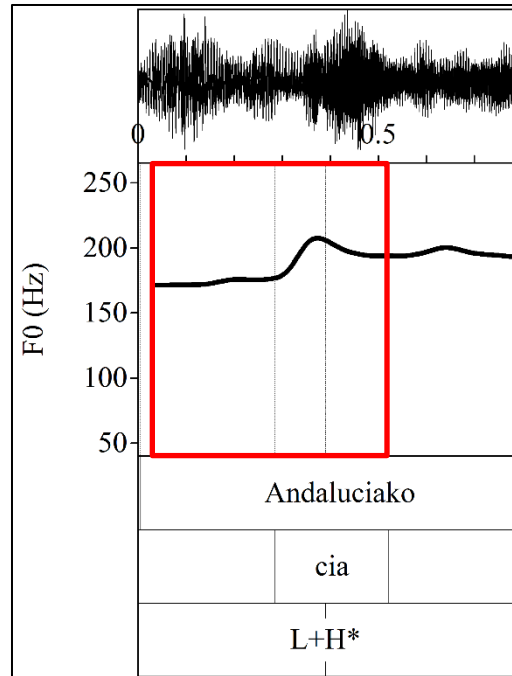
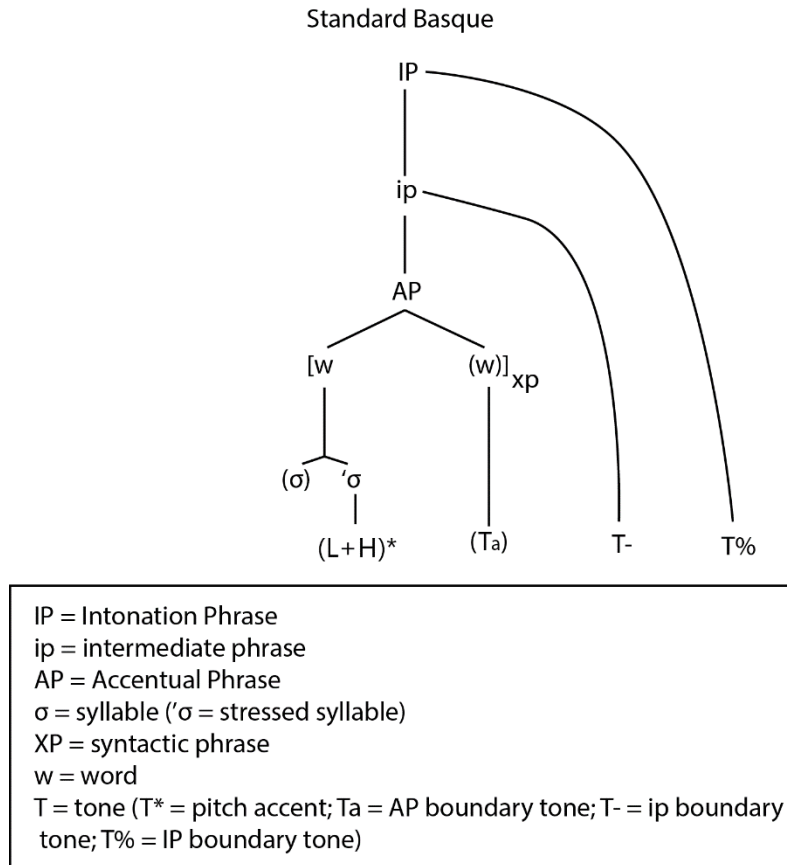


Figure 7: Example of L+H\* (early peak) pitch accent on the word Andaluciako (Andalucian)

Elordieta & Hualde (2014) propose three levels of prosodic phrasing in Standard Basque above the word: the AP (accentual phrase), ip (intermediate phrase) and IP (intonation phrase). The AP is typically marked by a rising boundary tone (indicated as H- or H<sub>a</sub>) at the right edge of a word or short syntactic phrase (usually a content word and an associated clitic, verb, or modifier), except when pre-verbal, in which cases a fall in f<sub>0</sub> is found. The AP is above the domain of the word, but below the domain of both the ip and IP. Each AP in Standard Basque must contain a minimum of one pitch accent. Intermediate phrases (ips) are marked by a boundary tone (H or L, similar to those found in Peninsular Spanish) at its right edge. These ips block (or decrease) downstep (the decline of f<sub>0</sub> across an AP) and must contain at least one AP. IPs are the highest prosodic level in Standard Basque (besides the utterance itself). IP breaks in Standard Basque also show syntactic dependencies, in which they can be bound by the right edges of adjunct clauses and at the left edges of parentheticals. Standard Basque IPs must contain

at least one ip and are marked by a boundary tone at its right edge, a pitch reset following the boundary tone, lengthening of the final syllable, and an optional pause. This hierarchy of prosodic levels in Standard Basque (word > AP > ip > IP) is shown schematically in Figure 8.



*Figure 8: Prosodic structure of Standard Basque according to Elordieta & Hualde (2014)*

#### 1.1.1. Northern Bizkaian Basque

The other variety of Basque that will be discussed is Lekeitio Basque, which is also called Northern Bizkaian Basque. As its name suggests, Northern Bizkaian Basque is spoken in Northern Bizkaia, in the coastal areas from Getxo to Ondarroa, including towns and cities 15-20 miles inland from the coast (Elordieta & Hualde (2015:408)). This dialectal area of Bizkaia is shown below on the map (Google Maps; altered by the author) in Figure 9, with the dialectal

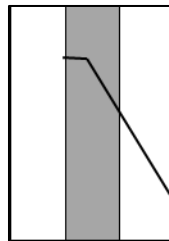
border towns of Getxo and Ondarroa indicated with red boxes. Amorebieta-Etxano (in the southern region of the map) approximates the inland boundaries of Northern Bizkaian Basque.



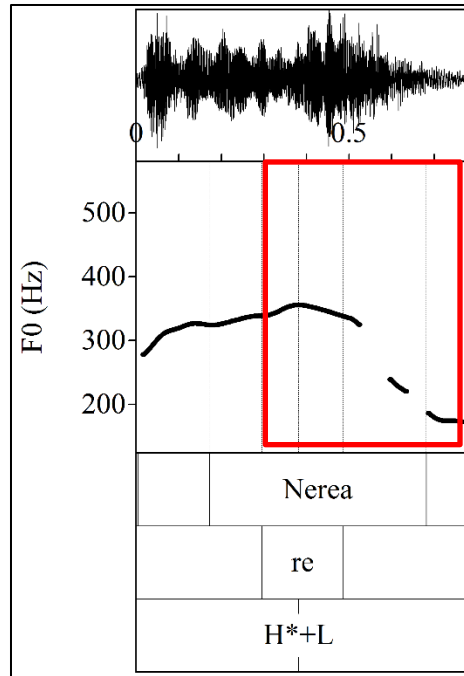
Figure 9: Northern Bizkaian Basque dialect region

Unlike Standard Basque, Northern Bizkaian Basque is a lexical pitch accent dialect, meaning that roots and affixes are lexically (i.e. underlyingly) specified for pitch accents. In Northern Bizkaian Basque, a word with at least one lexically specified accent will typically be realized with a falling pitch accent on the penultimate syllable. When roots and affixes not specified for lexical pitch accents are combined, the result is an unaccented word, but unaccented words receive an accent (called a *derived pitch accent* in Jun & Elordieta (1997)) only when they are produced in isolation or located in preverbal (object) position. Unlike a lexical accent which is realized on the penultimate syllable of a word, a derived accent is realized on the final syllable of a word.

Northern Bizkaian Basque has the same three levels of prosodic structure above the word as Standard Basque (AP, ip, and IP). Where these dialects differ (besides lexical stress vs. lexical pitch accent) is in the tonal inventory and structure within these levels of prosody. Lexical stress in Northern Bizkaian Basque are realized with a bitonal falling lexical pitch accent with the f0 peak in the penultimate syllable and low f0 on the final syllable (H\*+L), a pitch accent which is not attested in Spanish nor Standard Basque. The tonal patterns of the AP in Northern Bizkaian Basque is predictable, with typical APs containing the following structure: [%L H- H\*+L]. In cases in which lexical pitch accent is realized on the first or second syllable of a word (i.e., mono- or disyllabic word), the H\*+L pitch accent overrides the initial AP boundary tone (%L). APs in Northern Bizkaian Basque only contain one (lexically-specified or derived) accented word, which may contain unaccented words to its left. This differs from Standard Basque, which may have two words in an AP if they form a syntactic constituent. Figure 10 shows the schematic representation of the H\*+L pitch accent, followed by a pitch track of an H\*+L pitch accent on the Basque name *Nerea* in Figure 11.

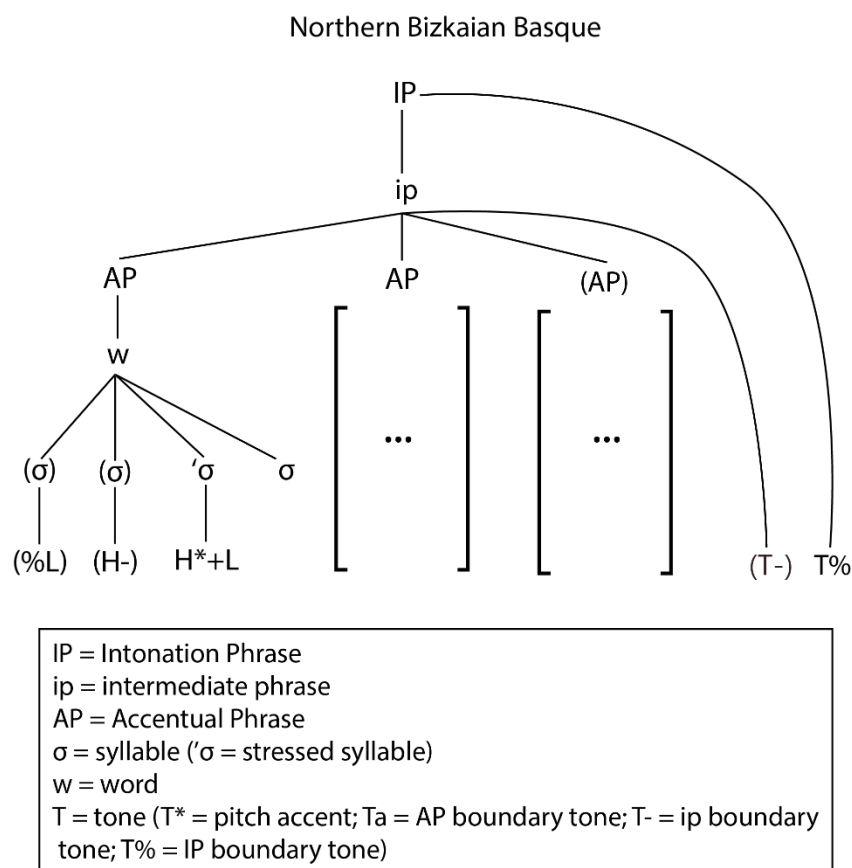


*Figure 10: Schematic representation of H\*+L pitch accent*



*Figure 11: Example of H\*+L pitch accent on the name Nerea*

Above the AP level, the ip in Northern Bizkaian Basque is not marked by a boundary tone, but only as the domain of downstep. Each ip in Northern Bizkaian Basque must contain a minimum of two APs, as opposed to the one-AP minimum in Standard Basque. The IP in Northern Bizkaian Basque, similar to that of Standard Basque and Spanish, has at least one ip, is marked with a tonal excursion at the right edge, may contain lengthening on its final syllable, and also blocks downstep. Additionally, the nuclear pitch accent in Northern Bizkaian Basque is located in a preverbal position in neutral (i.e., non-focused) utterances, as documented in Standard Basque. Figure 12 shows a schematic representation illustrating the prosodic hierarchy (word < AP < ip < IP) in Northern Bizkaian Basque as well the typical tonal contours present in neutral utterances. The second and third APs in the figure contain the same schematic representation as the first, but are marked with ellipses implying their identical shape.



*Figure 12: Northern Bizkaian Basque prosodic structure*

### 2.3 Spanish prosody and intonation

All varieties of Spanish have contrastive lexical stress, with stress most commonly occurring on the penultimate syllable of content words. Stress may also occur on the ultimate and antepenultimate syllables, although these cases are not as common as penultimate stress. Non-penultimate stress in Spanish is typically marked orthographically with a diacritic over the nucleus of the stressed syllable (such as the acute accent mark in *habló*, ‘he/she spoke’). Words with stress more than three syllables from the end of the word (i.e., preantepenultimate) are not

attested in Spanish (Hualde 2005:222)<sup>3</sup>. A minimal triplet which contrasts in stress is shown in (2), in which the stressed syllable of each word is bolded.

- (2) **tér**mino, ‘term’  
    term**ino**, ‘finish-pres-1sg’  
    termin**ó**, ‘finish-past-3sg’

Stressed syllables in Spanish attract pitch events (pitch accents, typically rising) in addition to being marked by other salient suprasegmental cues such as intensity and duration. However, segmental cues are not used to mark stressed or unstressed syllables, which differs from languages that employ segmental cues such as stop aspiration and vowel reduction (as in English) or vowel raising (as in Brazilian Portuguese). Due to the lack of segmental cues for stress in a syllable-timed language like Spanish, the difference between stressed and unstressed syllables is less prominent than in languages that undergo vowel reduction.

Within the Autosegmental-Metrical (AM) framework (Pierrehumbert (1980), Beckman & Pierrehumbert (1986), Ladd (1996)), two additional prosodic units have been proposed above the word level in Spanish. These levels are the *intermediate phrase* (ip) and the *intonation phrase* (IP), (Estebas-Vilaplana (2008) and Roseano & Prieto (2010)). The ip is typically marked with lengthening of the phrase-final syllable or a pitch reset, contains one or more pitch accents (i.e. pitch excursion on a stressed syllable), and has a boundary tone (pitch excursion) at its right edge after its final pitch accent (in Spanish, this is the *nuclear pitch accent*). The IP in Spanish is marked by lengthening of its final syllable, a boundary tone, and is followed by an optional pause. IPs typically coincide with a syntactic constituent as well, such as a clause. However, the status of the ip in Spanish is debated among Spanish intonationists (e.g. Sosa (1999), Nibert

---

<sup>3</sup> Gerunds with attached clitics in Spanish may have preantepenultimate stress, but as these cases are the result of morpheme combinations, they are not considered as single words.

(2000)). The prosodic structure of Spanish assuming the AM model is shown schematically in Figure 13, which demonstrates the hierarchy of the intonational constituents proposed for Spanish, including the controversial ip level (i.e., word < ip < IP). The current study assumes the intonational model of Spanish proposed in Estebas Vilaplana & Prieto (2008), which includes the ip level.

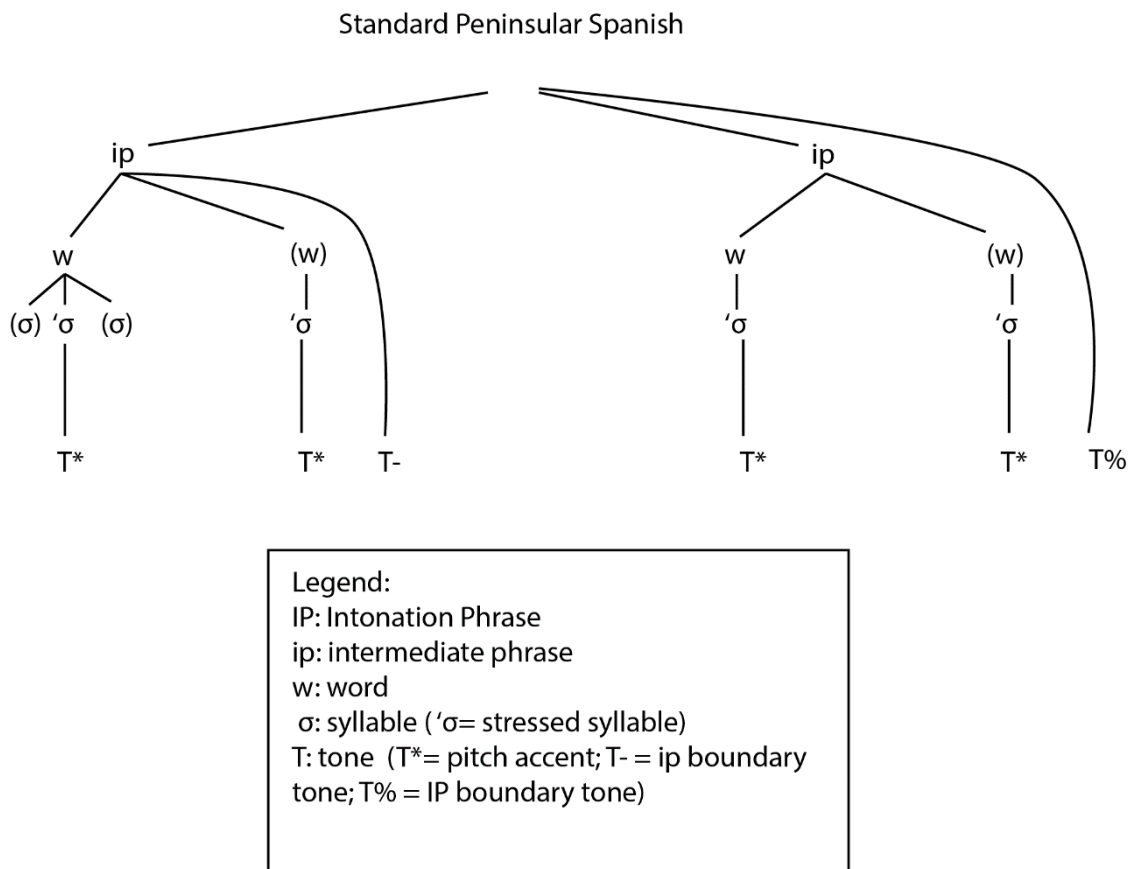


Figure 13: Prosodic structure of Spanish, according to Estebas-Vilaplana & Prieto (2008)

Estebas Vilaplana & Prieto (2008) propose a model of intonational phonology within the AM framework for Peninsular (or Castilian) Spanish, the standard dialect of Spain. Peninsular Spanish contains six pitch accents, whose distributions are determined by sentence type, information structure, and prosodic position. Two of the six pitch accents are monotonal (low

and high) and are characterized by a high (H\*) or low (L\*) tonal target f0 on the stressed syllable with a gradual f0 interpolation to and from the stressed syllable.

The remaining four pitch accents are bitonal rising and falling pitch accents, which differ in the alignment of the f0 peak. Unlike the monotonal pitch accents, the bitonal pitch accents have sharp transitions between f0 peaks and troughs leading into or out of the stressed syllable (called *leading tones* if pre-tonic or *trailing tones* if post-tonic), which differs from the more gradual interpolation before and after monotonal pitch accents. Starting with pitch accents that have a low f0 in the stressed syllable, L\*+H (or late peak) has an f0 trough in the stressed syllable, followed by a sharp rise, whereas H+L\* (or falling pitch accent) has an f0 trough or fall in the stressed syllable, preceded by a sharp fall from a high f0 target. This distinction in trailing tone and leading tone is shown schematically in Figure 14. A pitch track containing an example of a L\*+H pitch accent can be found in Figure 16<sup>4</sup>.

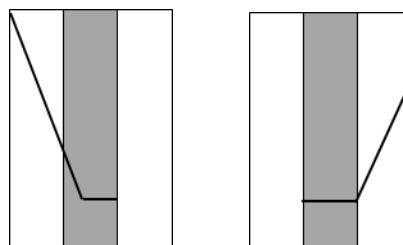


Figure 14: Schematic representations of H+L\* (left) and L\*+H (right)

The last two bitonal pitch accents contain f0 peaks in the stressed syllable but differ in alignment. L+H\* (or early peak) has an f0 peak in the stressed syllable preceded by an f0 valley and sharp rise. L+<H\* (or delayed peak) has an f0 rise in the stressed syllable, followed by an f0 peak on the post-tonic (or in some cases, the post-post tonic) syllable. The delayed peak pitch

---

<sup>4</sup> H+L\* pitch accents were not attested in the current data set. As a result, a pitch track containing this pitch accent will not be included.

accent ( $L+<H^*$ ) is the default (most common) pitch accent in prenuclear position in Peninsular Spanish, whereas the timed peak ( $L+H^*$ ) is most common in nuclear position. This difference in peak alignment is displayed schematically in Figure 15. Pitch tracks containing  $L+<H^*$ ,  $L^*+H$ , and  $L+H^*$  pitch accents are found in Figure 16.

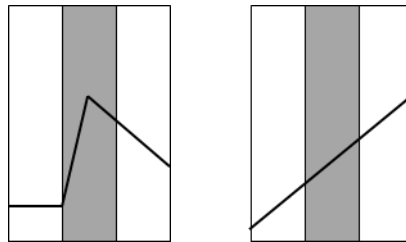


Figure 15: Schematic representations of  $L+H^*$  (left) and  $L+<H^*$  (right)

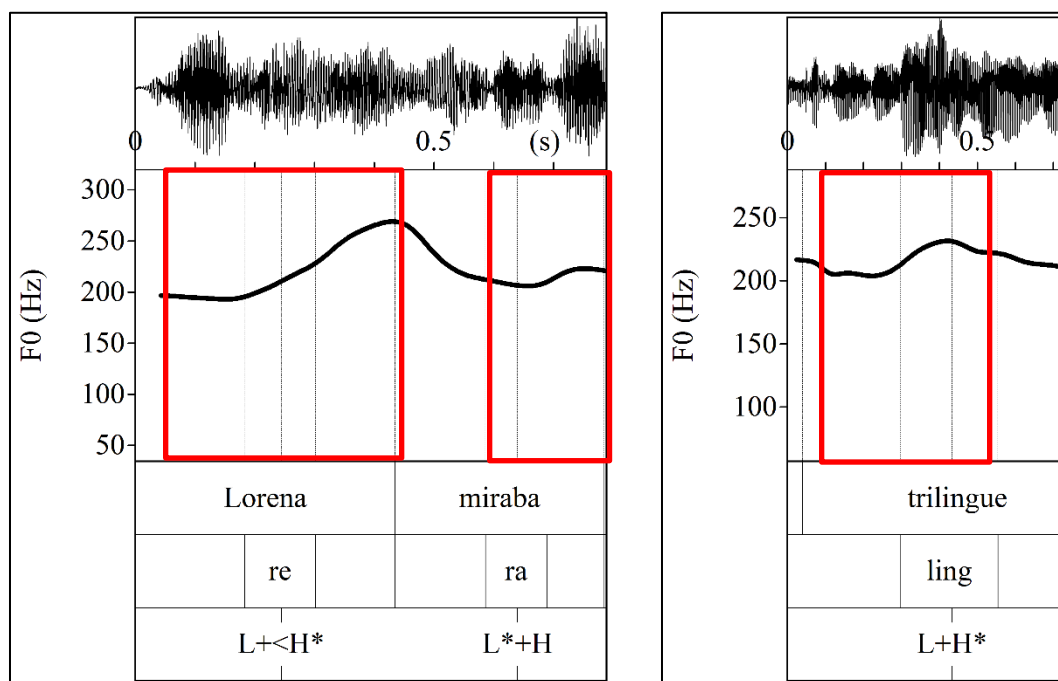


Figure 16: Pitch tracks containing  $L+<H^*$  (delayed peak) pitch accent on the name Lorena,  $L^*+H$  on the Spanish word *miraba* ('s/he saw'), and  $L+H^*$  on the Spanish word *trilingue* ('trilingual')

### 2.3.1 Basque Country Spanish intonation

Because the Basque Country is an area of varying degrees of language contact between Spanish and Basque, not only are contact phenomena like code-switching and borrowing expected, but phonological influence and transfer as well. Transfer of prosody and intonation in bilingual or diglossic language environments has been seen in Cuzco Spanish, which is in contact with Quechua (O'Rourke (2004)); Majorcan Catalan, which has contact with Peninsular Spanish (Simonet (2008), (2011); Romera & Elordieta (2013)); and Valencian Spanish, which is in contact with Valencian Catalan (Craft (2015)), amongst others. With respect to the Basque Country, Elordieta & Calleja (2005) and Elordieta & Irurtzun (2012, 2016) investigated the Spanish intonation of speakers in five different locations in the Basque country to determine whether there was evidence of influence from Basque. Three of the regions investigated (Vitoria-Gasteiz, Donostia-San Sebastián, and Bilbao) are Spanish-dominant communities, in which Basque is mostly spoken as a second language or for government purposes and the other two regions (Lekeitio and Goierri) are more Basque dominant environments, with majority native Basque speakers.

In these four regions, peak alignment in stressed syllables of neutral declaratives was measured to see if the f0 peaks and valleys in the pitch accent were timed, or within the stressed syllable (more Basque-like) or delayed (more like Standard Peninsular Spanish). The results of Elordieta & Calleja (2005) revealed that Lekeitio Spanish has significantly earlier f0 peaks (within the stressed syllable, or L+H\*) than Vitoria Spanish, which showed the delayed peaks (L+<H\*) typical of Standard Peninsular Spanish. Later work by Elordieta & Irurtzun (2016) found that Spanish speakers from Bilbao and Donostia showed delayed peak alignment patterns

similar to Vitoria Spanish (and hence, to Peninsular Spanish), regardless of amount of regional language contact.

In order to test whether or not the results found in Lekeitio was due to a high number of native Basque speakers (as opposed to the native Spanish speakers from Vitoria, Bilbao, and Donostia), the Spanish of native (Standard) Basque speakers from Goierri was also measured for peak alignment, which showed delayed peaks similar to Peninsular Spanish. Therefore, the authors concluded that the potential influence of Basque on Spanish is only seen in Lekeitio Spanish, in which speakers showed early peaks timed within the stressed syllable.

## 2.4 The current study's hypotheses

### 2.4.1 Hypothesis 1: Peak alignment

In order to investigate how the prosodic systems of Spanish and Basque interact during code-switching, the peak alignment of pitch accents will be examined. As discussed in the previous sections, Peninsular Spanish differentiates between early peaks (i.e. f<sub>0</sub> peak within the stressed syllable) and delayed peaks (i.e. f<sub>0</sub> peak realized on the post-tonic syllable) based on prosodic position. That is, early peaks (in L+H\* pitch accents) occur in phrase-final positions (at ip and IP boundaries) and delayed peaks (in L+<H\* pitch accents) occur in phrase-medial positions. In Standard Basque, this positional difference between pitch accents is less strict: phrase-medial pitch accents may be delayed (when pre-nuclear) or early, whereas nuclear (i.e. pre-verbal) and phrase-final pitch accents contain early peak alignment. On the other hand, Northern Bizkaian Basque and Lekeitio Spanish do not distinguish between early and late peaks, containing pitch accents with early peak alignment only. These differences are summarized in Table 2.

| Language and Dialect     | Phrase-medial peak alignment        | Phrase-final peak alignment |
|--------------------------|-------------------------------------|-----------------------------|
| Spanish-Peninsular       | Delayed                             | Early                       |
| Spanish-Lekeitio         | Early                               | Early                       |
| Basque-Standard          | Delayed (when pre-nuclear) or Early | Early                       |
| Basque-Northern Bizkaian | Early                               | Early                       |

*Table 2: Expected differences in peak alignment, according to Elordieta & Hualde (2014) and Elordieta & Irurtzun (2016)*

The present study tests the claims of the Matrix Language Framework (Myers-Scotton 1993), which predicts that the patterns of the Matrix Language will override those of the Embedded Language in the presence of (morphosyntactic) differences. Applied to the prosodic differences in peak alignment, the current study hypothesizes that when Peninsular Spanish is the Matrix Language and Basque is the Embedded Language, all phrase-medial peaks will be *delayed*. Because Standard Basque can optionally have delayed pre-nuclear peak alignment, any phrase-medial delayed peaks in Standard Basque contexts will be examined for position (i.e. pre-nuclear or nuclear) to determine if it fits with expected Standard Basque values or Peninsular Spanish values. When Basque (Standard and Northern Bizkaian) is the Matrix Language and Peninsular Spanish is the Embedded Language, all phrase-medial peaks are predicted to be *early*, unless in pre-nuclear position in Standard Basque contexts (in which a delayed peak is possible). No difference in peak alignment is hypothesized for Northern Bizkaian Basque and Lekeitio Spanish pairings, as only early peak alignment is attested in these varieties.

#### 2.4.2 Hypothesis 2: f0 maxima

The findings of Olson (2012, 2016b) and Fricke (2016) discussed in Chapter 1, provide evidence that code-switches may be hyperarticulated (i.e. higher f0) or realized with narrow

focus in order to make a contrast between the two languages being used. Due to these findings, it is expected that code-switched words in the present study will have a *higher f0* than non-switched words. However, in the absence of an f0 difference between code-switched and non-switched words, as seen in Aly (in press), it is also expected that these modes will be contrasted using another phonetic or prosodic strategy.

#### 2.4.3 Hypothesis 3: Language dominance

The results of Olson (2015, 2016b) and Aly (in press) suggest that language dominance may interact with language mode (e.g. code-switched or not) to show asymmetrical phonetic or phonological effects on the languages involved during code-switching. It is hypothesized that the speaker's dominant language will be more susceptible to deviation from mono/unilingual norms when it is the embedded language, as the inhibitory control mechanisms required to switch back to the dominant language have been shown to be costlier than the reverse pattern, as seen in Reuter & Allport (1999) and Olson (2016b).

Now that the hypotheses for the present study have been discussed, the following chapter will provide details on the methodology employed by the current dissertation, including participant demographics, experimental tasks used, and analysis procedure.

### 3 METHODOLOGY

#### 3.1 Participant recruitment and demographics

This dissertation's corpus contains the speech of seven female Spanish-Basque bilinguals from the Basque Country. In order to appropriately communicate with participants in Spanish and Basque and to elicit code-switches in both of these languages, a female native Basque-Spanish bilingual (26 years old) from Lekeitio (henceforth the Interviewer) was hired to conduct all interviews. In addition to being a native Spanish-Basque bilingual, the Interviewer is also bi-dialectal in Northern Bizkaian Basque and Standard Basque. Participants were recruited via email through the University of Deusto in Bilbao ( $n = 4$ ) and through personal correspondence with the interviewer in Lekeitio ( $n = 3$ ). All participants, including the Interviewer, were asked to complete a questionnaire which asked for basic demographic information as well as questions about their language use and self-perception of language competence.

Participants were recruited in Bilbao and Lekeitio in order to have data from speakers in a historically Basque-dominant town (Lekeitio) and a historically Spanish-dominant city (Bilbao). Although located in northern Bizkaia, Bilbao is the capital of Bizkaia and has the largest population in the region. Consequently, Bilbao's demographics are less homogenous than smaller areas (such as Lekeitio) and the city has no true local or native variety. Rather, most Bilbao residents speak Standard Peninsular Spanish and are educated in Standard Basque. The local varieties of Spanish and Basque spoken in Bilbao are likely those of speakers who moved from different areas of the Basque country or who have a historical family tradition of speaking those varieties. Lekeitio residents are likewise educated in Standard Basque, but the local variety

of Basque in Lekeitio has a stronger historical and social tradition and is used in most informal contexts.

Dialectal membership of the present study's Lekeitio participants as Northern Bizkaian Basque speakers was determined by acoustic analysis (presence of H\*+L pitch accents in Basque language data) and by confirmation of the Interviewer, a native of Lekeitio. Dialect membership of the Bilbao participants as Standard Basque speakers was also determined by acoustic analysis (presence of L+H\* pitch accents in Basque language data) and by familial demographics (e.g. their parents were either native Spanish speakers, from Bilbao, or speak only Spanish at home).

The language questionnaires revealed that all participants and the Interviewer were born and raised in the Basque Country, ranged from 21-31 years old (average age of 24 years old), and each had either completed a Bachelor's degree or was currently an undergraduate university student. Younger participants were optimal for the present study, as younger speakers (i.e. those born after the Franco dictatorship) are more likely to be native or fluent Basque speakers, as seen with the Census statistics presented in Chapter 2. Five of the seven participants plus the interviewer identified both Spanish and Basque as their first languages, with only two participants (Speakers 3 and 4, both from Bilbao) reported Spanish as their first language and Basque as their second language (acquired at two and three years old, respectively).

With respect to language use, half of the speakers ( $n = 3$ , plus the Interviewer) reported using both languages at home and with friends; two speakers each reported using only Spanish (Speakers 3 and 4, both from Bilbao) or Basque (Speakers 5 and 6, both from Lekeitio) at home and with friends; and three speakers (Speakers 3, 6, and 7) reported using only Spanish at work/school whereas only one speaker (Speaker 5, a teacher at a Lekeitio Basque school)

reported using only Basque at work. Speaker 2 is the only speaker who reported living in another city; although born in Bilbao, she moved to Igorre (a town in Bizkaia, approximately 14 miles southeast of Bilbao) as a child and returned to Bilbao to attend university. A summary of the demographics and language use reported by all speakers is shown in Table 3.

| Speaker     | Region            | Age | L1(s)   | L2 Age | Language(s)<br>at home | Language(s)<br>with friends | Language(s) at<br>work or school |
|-------------|-------------------|-----|---------|--------|------------------------|-----------------------------|----------------------------------|
| 1           | Bilbao            | 23  | Both    | N/A    | Both                   | Both                        | Both                             |
| 2           | Bilbao/<br>Igorre | 23  | Both    | N/A    | Both                   | Both                        | Both                             |
| 3           | Bilbao            | 22  | Spanish | 3      | Spanish                | Spanish                     | Spanish                          |
| 4           | Bilbao            | 21  | Spanish | 2      | Spanish                | Spanish                     | Both                             |
| 5           | Lekeitio          | 31  | Basque  | 8      | Basque                 | Basque                      | Basque                           |
| 6           | Lekeitio          | 21  | Basque  | 5      | Basque                 | Basque                      | Spanish                          |
| 7           | Lekeitio          | 22  | Both    | N/A    | Both                   | Both                        | Spanish                          |
| Interviewer | Lekeitio          | 26  | Both    | N/A    | Both                   | Both                        | Both                             |

*Table 3: Speaker demographics and self-reported language use*

Measures of language dominance were assessed with a combination of self-reporting in the questionnaire and a Semantic Verbal Fluency task administered before the start of the interview. For the self-reported measures, speakers were asked to rate their speaking, listening, and writing skills in Spanish and Basque on a scale of 1-5 (5 as most fluent) and to also identify what they considered their dominant language. Table 4 gives a summary of the participants' answers to the self-reported language competence questions, including domain (i.e. reading, listening and writing) and mean scores for each language, indicated with an "S" for Spanish and "B" for Basque. Scores less than 5 are shaded.

| Speaker     | Region   | Reading |   | Writing |   | Speaking |   | Average |     | Dominant language |
|-------------|----------|---------|---|---------|---|----------|---|---------|-----|-------------------|
|             |          | S       | B | S       | B | S        | B | S       | B   |                   |
| 1           | Bilbao   | 5       | 5 | 5       | 5 | 5        | 5 | 5       | 5   | Both              |
| 2           | Bilbao   | 5       | 5 | 5       | 5 | 5        | 5 | 5       | 5   | Both              |
| 3           | Bilbao   | 5       | 5 | 5       | 4 | 5        | 4 | 5       | 4.3 | Basque            |
| 4           | Bilbao   | 5       | 5 | 5       | 5 | 5        | 5 | 5       | 5   | Both              |
| 5           | Lekeitio | 4       | 5 | 4       | 5 | 5        | 5 | 4.3     | 5   | Basque            |
| 6           | Lekeitio | 4       | 5 | 4       | 4 | 5        | 5 | 4.3     | 4.6 | Basque            |
| 7           | Lekeitio | 5       | 5 | 5       | 5 | 5        | 5 | 5       | 5   | Both              |
| Interviewer | Lekeitio | 5       | 5 | 5       | 5 | 5        | 5 | 5       | 5   | Spanish           |

Table 4: Self-reported language competence and dominance; “S” = Spanish, “B” = Basque. Scores less than 5 are shaded.

Table 4 reveals that the majority ( $n = 4$ ) of speakers (plus the Interviewer) ranked themselves with 5s in all modalities in both languages. These four speakers (1, 2, 4, 7) also indicated that they considered both languages their dominant language, whereas the Interviewer reported Spanish as her dominant language despite ratings of all 5s for both languages. Speakers 3, 5, and 6 were the only speakers who ranked themselves with scores of less than 5 in any category and also indicated Basque as their dominant language. Speaker 3 (Bilbao) ranked her writing and speaking abilities in Basque with 4s, resulting in the lowest average score for Basque (4.3) of all speakers. Speakers 5 and 6 (Lekeitio) ranked their reading and writing abilities in Spanish with 4s, resulting in the lowest average scores for Spanish (4.3) of all speakers. Speaker 6 was the only speaker to rank herself with 4s in Spanish (reading and writing) and Basque (writing), which may indicate *linguistic insecurity*. Linguistic insecurity, a term coined by Labov (1972) in reference to non-standard dialects of American English in New York City, may occur when a speaker is aware that their dialects or speech styles do not match the perceived prestige or

standard dialect(s) of their country or region. In the case of speakers from Lekeitio, the varieties of both Spanish and Basque spoken there are notably different from Standard Basque and Peninsular Spanish in orthography, lexical items, and their sound systems (cf. Elordieta & Hualde (2014) and Elordieta & Irurtzun (2016) for phonological differences; lexical differences in code-switching between regions will be discussed in Section 3.3.2 of this chapter).

An interesting note is that for Speakers 3, 6, and the Interviewer, self-rankings of 5s are not necessarily associated with their choice of dominant language. Speakers 3 and 6 indicated that Basque was their dominant language, despite rankings less than 5. In the case of Speaker 6, this may be due to usage, as she indicated speaking only Basque at home and with friends, whereas Speaker 3 indicated speaking Spanish at home, with friends, and at work/school. For Speaker 3 and the Interviewer, it is possible that their choice of dominant language is related to factors other than self-perceived competence or usage, such as ethnic or national identity. It is beyond the scope of this dissertation to interpret or assign intention to the definition that this data set's speakers may have used when choosing to report their dominant language, but the current study acknowledges that usage and identity may be two possible interpretations that speakers used when answering this question.

The Semantic Verbal Fluency task, commonly used in psychological research, asks participants to name as many items in a semantic category (such as foods or animals) in one or more languages during a specified amount of time (such as 30 seconds). In order to avoid domain or culturally-specific categories that may affect the number of responses given, eight culturally-neutral categories that are commonly used in psycholinguistic research were chosen for the Semantic Verbal Fluency task (fruits, vegetables, animals, musical instruments, clothing, furniture, body parts, and colors). Participants completed four categories in Spanish or Basque

(order of languages were counterbalanced with participants) before switching to the other language for the remaining four categories. Each category lasted 60 seconds (timed by the researcher<sup>5</sup>), even if participants stopped responding before then. Due to differences in the quantity of responses across speakers, each speaker's response rate in both languages will be evaluated individually. Repeated or invented responses (i.e. those not in a dictionary or verified by a native speaker) as well as responses in the wrong language (i.e. a Spanish word during a Basque category) were not counted. Table 5 shows each participant's average scores in each language for the Semantic Verbal Fluency task, which was calculated as a sum of each speaker's total responses per language divided by the number of categories answered (e.g. 40 total responses in Spanish divided by four categories would result in a score of 10).

The variability in response rate across speakers as shown in Table 5 made it difficult to compare without a method of standardization. In order to compare speakers' response rates, the number of Basque responses was divided by the number of Spanish responses to produce a score representing the proportion of their responses relative to Spanish. If the score is less than 1, this indicates more responses in Spanish, as seen Speaker 1 whose percentage is .52 or 52%, meaning her Basque responses contained just over half as many items as her Spanish responses. Conversely, a score larger than 1 indicates more responses in Basque, as seen with Speaker 6, whose score was 1.1. A smaller score will indicate that a speaker is more Spanish-dominant, whereas a larger score (i.e., over 1) will indicate that the speaker is less Spanish-dominant. This dissertation does not attempt to establish an arbitrary cut-off for language dominance based on these scores, nor compare them to the self-reported language dominance responses due to the potential discrepancies in measurement and definition. The Semantic Verbal Fluency score

---

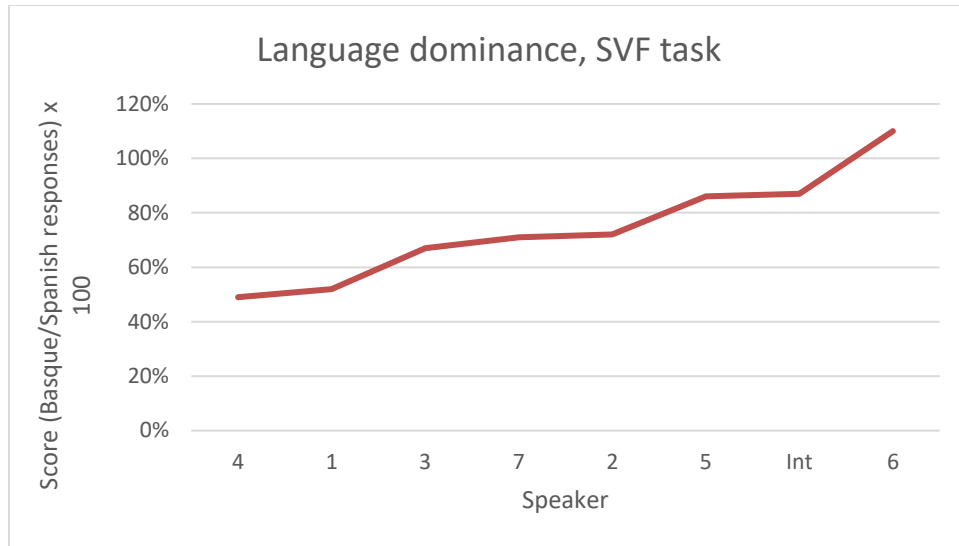
<sup>5</sup> The researcher mentioned throughout the methodology refers to the current study's author

(converted to a percent by multiplying by 100) will be used in conjunction with the self-reported language dominance variables in the statistical models that will be discussed in Section 3.3.4.

| Speaker     | Region   | Spanish average | Basque average | Basque/Spanish proportion |
|-------------|----------|-----------------|----------------|---------------------------|
| 1           | Bilbao   | 12.5            | 6.5            | .52                       |
| 2           | Bilbao   | 8.75            | 6.25           | .72                       |
| 3           | Bilbao   | 8.25            | 5.5            | .67                       |
| 4           | Bilbao   | 7.6             | 3.75           | .49                       |
| 5           | Lekeitio | 14              | 12             | .86                       |
| 6           | Lekeitio | 12.25           | 13.75          | 1.1                       |
| 7           | Lekeitio | 11.25           | 8              | .71                       |
| Interviewer | Lekeitio | 15.5            | 13.5           | .87                       |

*Table 5: Average responses per language in the Semantic Verbal Fluency task*

The graph in Figure 17 visually illustrates the continuum of language dominance within this group of speakers, with lower scores (to the left) representing more Spanish-dominant speakers and higher scores toward the right representing less Spanish-dominant speakers. It should be noted that the three leftmost (i.e. most Spanish-dominant) speakers (4, 1, 3) are from Bilbao and the three rightmost (i.e. least Spanish-dominant) speakers (5, Interviewer, 6) are from Lekeitio. The two speakers in the middle of the group (with scores in the 70% range, which represents the median of the group) are speakers 7 (Lekeitio) and 2 (Bilbao/Igorre). As mentioned previously, the self-reported responses may represent factors such as competence, usage, or identity, whereas the scores from the Semantic Verbal Fluency use real-time object naming as a proxy for competence.



*Figure 17: Language dominance scores, Semantic Verbal Fluency task*

The next section of this chapter will provide the reader with information on the data collection procedure of the linguistic portion (i.e. Spanish and Basque language data) of the corpus before continuing to the following sections on data analysis and coding.

### 3.2 Data collection procedure

All participants were recorded in Bilbao (speakers 1-4) or Lekeitio (speakers 5-7). All recordings were made with an Olympus LS-11 portable PCM recorder with a 44.1 kHz sampling rate and 16-bit rate. Participants from Bilbao were recorded in a small conference room at the University of Deusto with the Interviewer, participant and researcher present. Recordings in Lekeitio were made in the homes of participants or in study rooms at the public library in which the Interviewer, participant, and researcher were present. The entire procedure lasted about one hour per participant and participants were paid 10 euros (~\$12 USD) for their participation.

After the interviewer recruited a participant, the researcher confirmed that she spoke both Spanish and Basque before proceeding with an informed consent form and a demographic questionnaire. Four experimental tasks were administered after the consent and questionnaire forms: the Semantic Verbal Fluency task detailed in Section 3.1, a reading task completed in Spanish (henceforth Reading Task), a Discourse Completion Task edited from that of Roseano & Prieto (2010) for Basque Country Spanish (verified by the Interviewer), and a semi-spontaneous interview (henceforth Interview task) on the participant's daily life and opinions about current events. The order of the tasks was designed to progress from most formal (i.e., reading sentences) to least formal (i.e. unstructured interview) and to progress from primarily Spanish mode to mixed-mode (Spanish and Basque), and to primarily Basque mode. This progression allows for an analysis that considers which language is the dominant or Matrix Language of the task while also eliciting formal and informal styles of speech.

All materials presented to participants are available with an English translation in the Appendix. The consent process, demographic questionnaire (Appendix A), Reading Task (Appendix B), and Semantic Verbal Fluency tasks (Appendix C) were administered by the researcher in Spanish. The Interviewer administered the Discourse Completion Task in Spanish, but administered the Interview task in Basque. The participants were told that they could speak in either Spanish, Basque, or a mixture of both, for the Interview. Nonetheless, all participants completed the Interview task in Basque. During the Discourse Completion Task and the Interview task, the researcher sat quietly in the room but did not intervene in the tasks until they were completed. Upon completion of all tasks, the researcher thanked and paid each participant for her time.

The Reading Task was designed to elicit careful, formal speech from participants and consisted of Spanish sentences ( $n = 43$  target sentences) written on a piece of paper. For the purposes of generating as continuous a pitch track as possible to facilitate analysis, words containing sonorous segments (such as vowels, nasals, liquids, voiced approximants) were chosen. This task was divided into four parts containing diverse types of utterances: (1) declaratives with differing word order; (2) declaratives with increasing numbers of interstress syllables; (3) questions; and (4) utterances eliciting narrow focus by using a pseudo-dialogue. A pseudo-dialogue contains a short, written conversation between two speakers. One “part” is for the participant to read out-loud and the other “part” contains contextualizing information prompting her response (and is not read out loud). The Reading Task will serve as a baseline for the participants’ Spanish intonation patterns without code-switches into Basque. All sentences used in the Reading Task can be found in Appendix B.

The Discourse Completion Task provides prompts ( $n = 70$ ) to participants about different social and pragmatic contexts in Spanish (such as asking someone for the time or reacting to an unlikely event) and is designed to elicit targeted sentence types (such as declaratives, imperatives, or vocatives) and lexical items while allowing the participants to answer freely. Code-switching from Spanish to Basque is less common than switching from Basque to Spanish, as Spanish is typically the *pragmatically dominant* language (Lantto (2015)) in addition to being socio-politically dominant in Spain. This asymmetry of power between Basque and Spanish results in code-switching into Spanish being more common than code-switching into Basque. In order to deliberately elicit code-switching from Spanish to Basque (the less common direction of code-switching in the Basque Country) in the Discourse Completion Task, 15 Basque nouns (such as names, places, and traditional dishes) were included in 20 prompts to act as code-

switching triggers, as seen in studies such as de Bot et al (2009). All code-switching triggers for the Discourse Completion Task were verified by the Interviewer as culturally and linguistically felicitous before the start of the study. Some of the prompts attempted to solicit a recast (or repetition) of the trigger word in the answers, whereas others do not solicit a recast, but attempt to elicit novel code-switches from the participant. An example of a prompt eliciting an exclamative statement with a Basque trigger word that does not solicit a recast is shown in (3). Example (3) contains an example of a prompt that solicits a recast of the Basque trigger word in an interrogative utterance. Following both (2) and (3) is Table 6, which contains a list of all Basque trigger words from the Discourse Completion Task. All trigger words are indicated in Appendix C using bolded text.

- (3) Te invitan a un *marmitako* y es el mejor que has comido en tu vida. Estás encantada. ¿Qué dices?  
*They invite you to a marmitako and it's the best you've ever eaten in your life. You're in awe. What do you say?*

- (4) Estás hablando de *Begoña* con alguien y oyes que entra una persona. Pregúntale si es *Begoña* la persona que está entrando.  
*You are talking about Begoña with someone and you hear someone enter. Ask if it is Begoña that is entering.*

| Basque trigger word (orthography) | English gloss                                               | Prompt number | Type of prompt                      |
|-----------------------------------|-------------------------------------------------------------|---------------|-------------------------------------|
| Ane                               | Proper name (female)                                        | 2             | Declarative                         |
| Donostia                          | Place (city; also known as <i>San Sebastián</i> in Spanish) | 7             | Declarative                         |
| Amaia                             | Proper name (female)                                        | 8             | Declarative                         |
| Maialen                           | Proper name (female)                                        | 10, 69, 70    | Declarative (10), vocative (69, 70) |
| Getxo                             | Place (city)                                                | 13            | Declarative                         |
| Itziar                            | Proper name (female)                                        | 15            | Obviousness statement               |
| Andoni                            | Proper name (male)                                          | 15            | Obviousness statement               |
| Marmitako                         | Basque tuna stew                                            | 16            | Exclamative                         |

| Basque trigger word (orthography) | English gloss        | Prompt number | Type of prompt                                                |
|-----------------------------------|----------------------|---------------|---------------------------------------------------------------|
| Nerea                             | Proper name (female) | 21, 44        | Absolute interrogative (21), interrogative word question (44) |
| Begoña/Begoina                    | Proper name (female) | 27            | Absolute interrogative                                        |
| Jon                               | Proper name (male)   | 30            | Interrogative-word question <sup>6</sup>                      |
| Laga                              | Place (city)         | 43            | Interrogative-word question                                   |
| Iker                              | Proper name (male)   | 46            | Interrogative-word question                                   |
| Mikel                             | Proper name (male)   | 59            | Absolute interrogative with disbelief                         |
| Iratxe                            | Proper name          | 63, 64, 65    | Imperative                                                    |

*Table 6: Basque trigger words for code-switches in the Discourse Completion Task*

The final task, the Interview, was designed to elicit more natural speech from each of the participants in the language of their choice. Because code-switching from Basque into Spanish is more common among Spanish-Basque bilinguals (cf Lantto (2015)), specific prompts and trigger words to elicit code-switches (as seen in the Discourse Completion Task) were not necessary to elicit code-switching in this task. Some sample questions about daily life and current events were given to the Interviewer before the experiment, in which she was instructed to start with these general questions and follow up with additional questions if the participant wished to continue talking. The Interviewer told participants that during the Interview task, they could speak in either Spanish or Basque and could switch languages whenever they wished, but that she would be addressing them in Basque. As code-switching is a natural part of the Interviewer's speech, she was encouraged to speak naturally with participants in order to establish a context in which code-switching was welcome. Most of the responses in this task were longer declarative sentences or exclamatives. Due to different levels of familiarity and extroversion with each

---

<sup>6</sup> Also known as “wh-questions”

participant, the interviews ranged from ten to twenty minutes per participant and all participants chose to speak in Basque with Spanish code-switches. Data from the Interview task will be discussed in more detail in the analysis and results sections of this dissertation. Table 7 provides a summary of the tasks administered to participants and the type(s) of data they each elicited.

| <b>Task</b>                                           | <b>Type of data elicited</b>                                                     | <b>Language</b>                                                |
|-------------------------------------------------------|----------------------------------------------------------------------------------|----------------------------------------------------------------|
| Semantic verbal fluency (Semantic Verbal Fluency)     | Category naming task (4 categories per language) to determine language dominance | Basque and Spanish (order counterbalanced across participants) |
| Reading task                                          | Read speech of three word utterances ( $n = 43$ )                                | Spanish                                                        |
| Discourse Completion Task (Discourse Completion Task) | Semi-spontaneous speech ( $n = 70$ prompts with 20 Basque nouns)                 | ML Spanish, EL Basque (specifically elicited code-switches)    |
| Interview                                             | Semi-spontaneous, naturalistic speech (10-20 minutes per participant)            | ML Basque, EL Spanish (naturalistic code-switches)             |

*Table 7: Tasks administered to participants*

The following sections will discuss the analysis procedure for this dissertation's data set, including establishing criteria for code-switches, acoustic measurements, and concluding with statistical analyses.

### 3.3 Analysis procedure

#### 3.3.1 Code-switching definitions and criteria

For the purposes of this study, it is crucial to have a clear and consistent definition for what constitutes a code-switch. Due to the variety of languages that may be used in code-switching, including closely-related languages, languages in high contact, and creole languages, this definition may vary depending on the languages in question and contexts in which they exist. As discussed in Chapter 2, Spanish and Basque, although unrelated, have historically had a high level of contact resulting in bidirectional grammatical borrowings and influence (e.g. Spanish *izquierda*, 'left' from Basque *esker*; Basque *teklatu*, 'keyboard' from Spanish *teclado*).

Additionally, the existence of numerous dialects and registers of Spanish and Basque throughout the Basque country creates different lexical inventories, norms, and code-switching practices in each respective dialect group. As the current dissertation analyzes the speech of two diverse dialect groups (Northern Bizkaian Basque for Lekeitio participants and Standard Basque for Bilbao participants), these differences will be accounted for using both formal (dictionary) and informal (native speaker judgments) criteria in order to categorize potential code-switches.

In the current data set, the Basque portions of the recordings were transcribed orthographically by a native Spanish-Basque bilingual (henceforth the Transcriber) who was not otherwise involved with the study. The Transcriber is a 29-year-old male from Arrasate/Mondragón, Spain (located in Gipuzkoa) who speaks both the standard and local varieties of Spanish and Basque. In each transcription provided, he included a Basque to Spanish translation and notes on potential code-switches, colloquial terms, or other linguistic phenomenon (such as errors, dialectal features, phonetic features, etc.). The Spanish portions of the recordings were transcribed orthographically by three undergraduate research assistants who self-identified as fluent in Spanish and verified by the researcher, who is a near-native Spanish speaker. The tagged code-switches from the Transcriber plus any other potential Spanish code-switches in the Interview task noted by the researcher were then further investigated. The Basque code-switches from the Discourse Completion Task (see Table 6), as they were all recasts of the trigger words verified by the Interviewer, did not need to be investigated further for code-switch status.

As a first step to establishing code-switch status, potential code-switches were searched for in three different Basque language dictionaries: *Elhuyar Hiztegia* (Basque Government (2006)), a Standard Basque-Spanish online dictionary; *Morris Student Plus 2<sup>nd</sup> edition* (Morris Academy

(2005)), a Standard Basque-English online dictionary; and the Lekeitio Basque-Spanish-English dictionary found in Chapter 6 of Hualde et al (1994). Potential code-switches from Bilbao speakers were searched for in both *Elhuyar* and *Morris* dictionaries whereas potential code-switches from Lekeitio speakers were searched for in all three dictionaries. If a potential code-switch was found in both Standard Basque dictionaries for a Bilbao speaker or in the Lekeitio dictionary for Lekeitio speakers, it was eliminated as a code-switch and labeled as a *borrowing*. *Borrowings* are words or phrases that have been incorporated into the lexicon of a language ( $L_x$ ) via influence of another language ( $L_y$ ) and may in some cases include morphological or phonological adaptations into the  $L_y$  as discussed in Poplack et al (1988). Borrowings will not be considered for analysis in the present study, as they can exist in monolingual lexicons and/or interactions<sup>7</sup>.

Proper names and place names (e.g. cities, towns) are a controversial topic in code-switching literature with no common consensus on their status as code-switches or borrowings, as proper nouns often have no alternative form in mixed-language contexts. Whereas studies such as Lindsay (2006) exclude all proper nouns, studies such as Park (2006) include proper nouns, as they maintained the morphosyntactic properties of the Embedded Language. In the present study, proper nouns (names and places) will be included as code-switches if they do not contain morphosyntactic or phonological properties of the Matrix Language. For example, place names from the data set such as *Andaluciako* (from Andalucia), which has the Basque genitive suffix –*ko*, or *Holandara* (to Holland), which has the Basque allative suffix –*ra*, will not be considered code-switches to Spanish. However, a place name such as *Velilla de la Peña* is considered a code-switch to Spanish in the present study, as it does not contain any Basque morphosyntax or

---

<sup>7</sup> It should be clarified that although monolingual Basque speakers are rare in the Basque country today, many monolingual interactions or domains exist (i.e. Basque-only conversations or contexts)

phonological properties, but contain Spanish morphosyntax (e.g. preposition and feminine definite article *de la*) as well as Spanish phonology (use of Spanish-like [j] instead of Basque-like [ʎ]).

If a potential code-switch was not found in any of the relevant dictionaries, it was labeled as a code-switch ( $n = 49$  unique words or phrases). These code-switches are listed below in Table 8. The frequencies of code-switches by region, speaker, and category will be presented and discussed in Chapter 4.

| Word (orthography)        | Gloss                         | Lexical category           |
|---------------------------|-------------------------------|----------------------------|
| (costa) este              | East (coast)                  | Noun                       |
| (costa) oeste             | West (coast)                  | Noun                       |
| (Fiestas de la) virgen    | (Holidays of the) Virgin      | Proper noun (holiday name) |
| (Fiesta) de las angustias | (Holiday) of the Anguish      | Proper noun (holiday name) |
| a lo grande               | very large                    | Prepositional phrase       |
| ahí                       | over there                    | Preposition                |
| analizar                  | To analyze                    | Verb                       |
| aparte                    | Apart, separate               | Adverb                     |
| biomecanicamente          | Biomechanically               | Adverb                     |
| claro                     | Right, yes                    | Discourse marker           |
| cliente                   | Client, customer              | Noun                       |
| de España profunda        | from deep Spain               | Prepositional phrase       |
| de los críos              | of the children               | Prepositional phrase       |
| de paso                   | passing through               | Prepositional phrase       |
| ejemplo                   | Example                       | Example                    |
| en principio              | Beginning                     | Prepositional phrase       |
| entre que                 | between which                 | Preposition                |
| es que                    | I mean, it's that             | Discourse marker           |
| Fiestas de nuestra señora | Holidays of our lord          | Proper noun (holiday name) |
| historia                  | History                       | Noun                       |
| humanidades               | Humanities                    | Noun, plural               |
| igual                     | Like                          | Discourse marker           |
| jo(der)                   | Expletive (i.e. <i>fuck</i> ) | Discourse marker           |
| justo                     | Fair                          | Adjective                  |
| la calle esperanza        | Hope Street                   | Noun phrase                |
| la plaza nueva            | the new plaza                 | Noun phrase                |

| Word (orthography)    | Gloss                    | Lexical category    |
|-----------------------|--------------------------|---------------------|
| linguística           | Linguistics              | Noun                |
| listo                 | Ready                    | Adjective           |
| master                | Master's degree          | Noun                |
| menos                 | Least                    | Superlative         |
| Modelo A              | Model A (education)      | Noun                |
| no sé qué             | I don't know what        | Utterance           |
| ocho                  | Eight                    | Numeral             |
| pintxos               | typical Basque food      | Noun, plural        |
| porque                | Because                  | Conjunction         |
| Puente Arenas         | city in Spain            | Proper noun (place) |
| salida                | Exit                     | Noun                |
| segun                 | According to             | Preposition         |
| seguramente           | Surely, securely         | Adverb              |
| sobran                | (They) are in excess     | Verb                |
| supuestamente         | Supposedly               | Adverb              |
| tabernas <sup>8</sup> | Taverns, bars            | Noun, plural        |
| tumbarme              | To lay down              | Verb, reflexive     |
| vascos                | Basques (people)         | Noun, plural        |
| Velilla de la Peña    | Region in northern Spain | Proper noun (place) |
| ver                   | To see                   | Verb                |
| vidas                 | Lives                    | Noun, plural        |
| vocabulario           | vocabulary               | Noun                |
| yo que sé             | (that) I know            | Utterance           |

Table 8: Unambiguous Spanish code-switches in Interview task

Words that were only found in one dictionary (in the case of Bilbao speakers), were not in the Lekeitio dictionary (in the case of Lekeitio speakers), or were judged as ambiguous by a native speaker consultant<sup>9</sup> were labeled as *ambiguous code-switches* ( $n = 19$ , see Table 9). This process of determining code-switches and borrowings on a dialectal basis was made to distinguish between words that may be established borrowings in one dialect, but still considered code-switches in another.

<sup>8</sup> *Taberna* is also used in Basque, but due to the presence of the Spanish plural marker instead of the Basque plural marker (i.e. *tabernak*), this will be considered a Spanish code-switch in this context

<sup>9</sup> The relevant native speaker consultants here are the Interviewer, Transcriber, and Committee Member Elordieta

| Word (orthography) | Gloss                        | Lexical category              |
|--------------------|------------------------------|-------------------------------|
| a ver              | Well; let's see              | Discourse marker              |
| Asignatura         | Academic subject             | Noun                          |
| Bueno              | Well, good                   | Discourse marker, adjective   |
| Faltan             | Lack                         | Verb                          |
| Fisioterapeuta     | Physical therapist           | Noun                          |
| Hombre             | Exclamative (i.e. wow); well | Exclamative, discourse marker |
| Necesario          | Necessary                    | Adjective                     |
| o sea              | I mean; that is              | Discourse marker              |
| Super              | Super                        | Superlative                   |
| típico             | Typical                      | Adjective                     |
| tipo               | Person, type                 | Noun                          |
| trilingüe          | Trilingual                   | Noun                          |
| mochilero          | Backpacker                   | Noun                          |
| podología          | Podiatry                     | Noun                          |
| pues               | Well, so                     | Discourse marker              |
| japonesa           | Japanese, fem.               | Noun                          |
| maneja             | Drive                        | Verb                          |
| necesario          | Necessary                    | Adjective                     |
| ya                 | Already                      | Adverb                        |

Table 9: Ambiguous code-switches from Interview task

In addition to the ambiguous code-switches seen in Table 9, there were also hybrid code-switches, in which morphological features of Basque were combined with a Spanish root ( $n = 10$ ). Table 10 contains these ten hybrid words in their IPA form, as transcribed by the researcher (column 1); the equivalent Spanish root, IPA transcription, and gloss (column 2); and the Basque morphological suffixes (column 3) and equivalent Basque word (column 4). There were no examples of hybrid code-switches in which the root was a Basque word which had Spanish morphological markers attached. This pattern of combining Spanish roots with Basque morphology (but not Basque roots and Spanish morphology) is also observed in Lantto (2012), who reports forms such as *novioarekin* (Spanish root *novio*, ‘boyfriend’ and Basque definite singular comitative suffixes *-a-rekin*). Seven of the ten hybrid words were nouns in the

absolutive (definite) singular or plural form (e.g. [kont̪enið̞o-a]; see rows 1-7 in Table 10), comprising the most common type of hybrid word. The remaining three words consisted of two nouns (one inessive singular noun, [desaroʎuan] and one nominalized verb [faʎaʎen]) and one perfective verb ([desið̞ið̞u]).

A comparison between the Spanish root IPA (column 2 of Table 10) and the hybrid word IPA (column 1) reveals several phonetic characteristics that reflect a mix of both Spanish and Basque phonological processes. For example, none of the hybrid words contain the voiceless dental fricative (θ) that is present in the Spanish roots in rows (3, 5, 6, and 8), which is more indicative of a Basque-like pronunciation. The hybrid words in rows 2-8 (including the perfective suffix -tu in row 8) maintain the voiced approximant allophones [β, ð̞] of /b, d/, which reflects a Spanish phonological process. Row 1 contains the only hybrid word that does not maintain the voiced approximant allophones, as the voiced dental approximant [ð̞] is realized as [t], which is more analogous to Basque (cf Basque words such as *unibersitate*, [unibersiʔate], ‘university’ compared with Spanish *universidad*, [uniβersið̞að̞]). Word 3 (produced by a speaker of Northern Bizkaian Basque) contains the voiced post-alveolar fricative [ʒ], which is part of the phonological inventory of Northern Bizkaian Basque, but not Spanish. As the present study’s focus is with the prosodic elements of code-switching, the segmental aspects of these code-switches will not be discussed further, but future studies that examine contact phenomenon such as hybrid or portmanteau forms are much needed.

| Word (IPA) <sup>10</sup> | Spanish root, IPA transcription, and gloss           | Basque morphology and gloss | Equivalent Basque word (orthography) |
|--------------------------|------------------------------------------------------|-----------------------------|--------------------------------------|
| 1. [oportunit̪ate̪a]     | ‘Oportunidad’, [oportunið̪að̪]<br><i>Opportunity</i> | -a (ABS.SING)               | ‘aukera’                             |
| 2. [kont̪enið̪oa]        | ‘Contenido’, [kont̪enið̪o]<br><i>content</i>         | -a (ABS.SING)               | ‘eduki’                              |
| 3. [influent̪si̪za]      | ‘Influencia’, [influenθ̪ja]<br><i>influence</i>      | -a (ABS.SING)               | ‘eragin’                             |
| 4. [moβ̪iment̪ua]        | ‘Movimiento’, [moβ̪imjent̪o]<br><i>movement</i>      | -a (ABS.SING)               | ‘mugimendu’                          |
| 5. [fasilið̪að̪eak]      | ‘Facilidad’, [faθ̪ilið̪að̪]<br><i>ease</i>           | -ak (ABS.PL)                | ‘erraztasun’                         |
| 6. [duresak]             | ‘Dureza’, [dureθ̪a]<br><i>hardness</i>               | -ak (ABS.PL)                | ‘gogorgune’                          |
| 7. [eʃ̪trukt̪urja]       | ‘Estructura’, [eʃ̪trukt̪ura]<br><i>structure</i>     | -a (ABS.SING)               | ‘egitura’                            |
| 8. [deðið̪u]             | ‘Decidir’, [deθ̪ið̪iɾ]<br><i>To decide</i>           | -tu (PERF)                  | ‘erabaki’                            |
| 9. [faɬ̪at̪en]           | ‘Fallar’, [faɟ̪ar]<br><i>To fail</i>                 | -te-n (NOM.INE)             | ‘huts egin’                          |
| 10. [desaroɬ̪uan]        | ‘Desarrollo’, [desaroɟ̪o]<br><i>development</i>      | -an (IN.SING)               | ‘garapen’                            |

Table 10: Hybrid code-switches from the Interview task

The following section will discuss the procedure and results for the native speaker judgment surveys, which will determine the status of the ambiguous and hybrid words as either *code-switches* or *borrowings*.

### 3.3.2 Native speaker judgments survey

The ambiguous and hybrid code-switches were compiled into a list for use in a native speaker judgment survey that asks Spanish-Basque bilinguals about their intuitions on whether a word can be used in Spanish, Basque, or both languages. Due to the dialectal differences present in the Basque and Spanish dialects of this dissertation’s consultants discussed in Section 2.2, two versions of the survey were administered (one for Northern Bizkaian Basque speakers and one

<sup>10</sup> These transcriptions reflect how the current study’s participants realized these words, not the expected phonological or phonetic realizations of the languages/variety in question

for non-Northern Bizkaian Basque speakers<sup>11</sup>). Both surveys were identical except for orthography; the Northern Bizkaian Basque (Northern Bizkaian Basque) survey used local orthographical norms (verified by the Interviewer), whereas the Non-Bizkaian Basque (non-Northern Bizkaian Basque) survey used Standard Basque orthographical norms, (verified by a native speaker of Standard Basque from Bilbao, henceforth Bilbaina).

In the cases in which there was a more Spanish-like or Basque-like way to spell a word, the Basque-like spelling was chosen in order to minimize the Spanish-language bias present in the survey (e.g. the instructions, questions, and scale were all in Spanish). For example, the word *mochilero* (“backpacker”) is a more Spanish-like spelling with its use of ‘ch’ and it would be more Basque-like if it used a ‘tx’ instead (*motxilero*). For the hybrid code-switches, for which there is no local or standardized spelling, both the Interviewer and Bilbaina were consulted to help create a spelling that fit within the orthographic conventions of the relevant dialect groups. In order to avoid revealing the objective of the survey, 6 non-ambiguous Spanish and Basque words (verified by a native Basque-Spanish bilingual) were also included as filler words in the surveys. These are shown below in Table 11. The inclusion of the ambiguous, hybrid, and filler words resulted in a total of 35 lexical items in the survey. All orthographic forms presented in the surveys can be found in Appendix D.

| Word<br>(orthography) | Gloss    | Language |
|-----------------------|----------|----------|
| alaba                 | daughter | Basque   |
| arrain                | fish     | Basque   |
| ikaslea               | teacher  | Basque   |

<sup>11</sup> Although “non-Northern Bizkaian Basque” is a broad category, the distinction here is mainly between the coastal and rural pitch-accent varieties of Basque (i.e. Northern Bizkaian Basque) and the Standard, Western, and Central Basque varieties that are post-lexical pitch accent dialects

| Word<br>(orthography) | Gloss   | Language |
|-----------------------|---------|----------|
| perro                 | dog     | Spanish  |
| cabeza                | head    | Spanish  |
| pulpo                 | octopus | Spanish  |

Table 11: Unambiguous filler words for survey

Both surveys were made using Google Forms online software and included a brief section on the demographic information of the respondent (such as age, city, native language(s), etc. See Appendix D for complete list of questions) and a five point Likert scale for each word in which respondents rated whether they thought the word was more likely to be heard in a Spanish-only conversation (a score of 1 indicated a “totally Spanish” word, “*totalmente español*”), a Basque-only conversation (a score of 5 indicated “totally Basque”, “*totalmente euskara*”), or if a word could be used equally in both languages (score of 3). Responses were not mandatory, so respondents also had the option to not answer any of the survey questions (the Northern Bizkaian Basque survey had 48 missing answers and the non-Northern Bizkaian Basque survey had 28 missing answers). All survey questions (except Basque filler words and ambiguous lexical items) were in Spanish. Figure 18 shows an example of what this interface looked like to the respondent for rating words.

bueno

1

2

3

4

5

Totalmente español

☐
☐
☐
☐
☐

Totalmente euskara

Figure 18: Survey interface for rating ambiguous code-switches for the word bueno, ‘well, good’

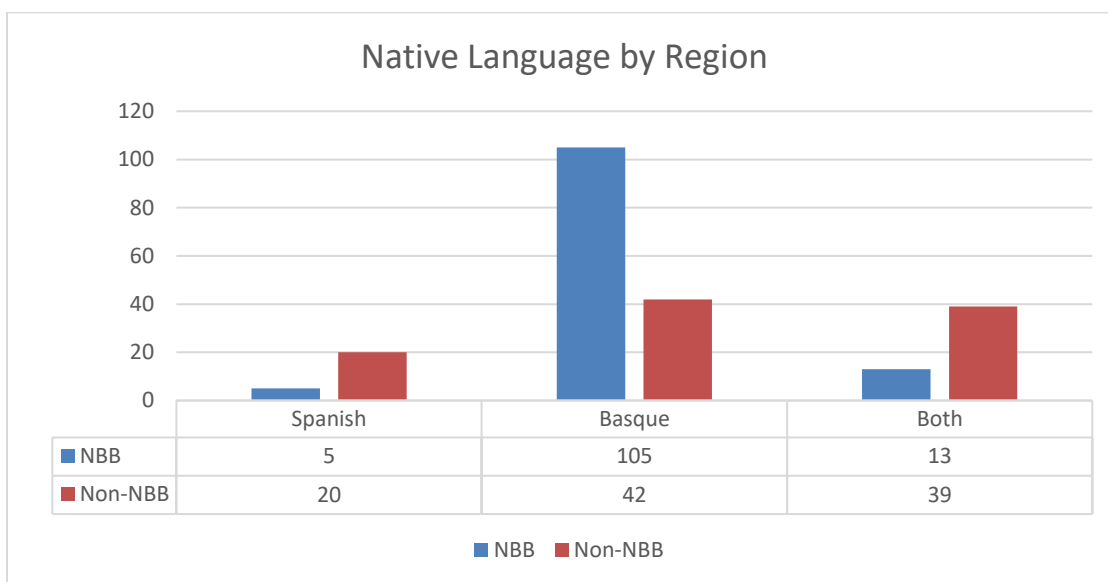
Respondents were recruited via social media and email by the researcher, Interviewer, Transcriber, and several professors at the University of the Basque Country and the University of Deusto. Respondents were able to click a link and anonymously participate in the study, which recorded their responses as well as time stamp without indicating any other personal or identifying information (such as name, email, location, etc.). The survey took approximately fifteen minutes to complete. Upon review of the responses, respondents were eliminated if they met any of the following criteria: (1) if they indicated that they did not speak either Spanish or Basque; (2) if they indicated that their native language was not Spanish and/or Basque; (3) if they indicated that they were under the age of 18<sup>12</sup>; (4) if they indicated that they did not live in the Basque Country. These criteria eliminated 27 respondents and yielded a total of 123 respondents from Northern Bizkaian Basque regions and 101 respondents from non-Northern Bizkaian Basque regions. Regional divisions between Northern Bizkaian Basque and non-Northern Bizkaian Basque regions were made using the geographical specifications established in Elordieta & Hualde (2014:408) in which Northern Bizkaian Basque regions range from Getxo to Ondarroa on the northern Bizkaian coast, including towns 15 miles inland. Locations outside the Northern Bizkaian Basque region (e.g. southern Bizkaian or Gipuzkoa) were categorized as non-Northern Bizkaian Basque.

Both surveys revealed several demographic and linguistic differences that varied by region. For example, the mean age of the Northern Bizkaian Basque respondents was 37.7 years whereas the mean age of Non-Northern Bizkaian Basque respondents was 25.6 years. The majority of respondents in both groups reported that they code-switched between Spanish and Basque, although this percentage was higher for Non-Northern Bizkaian Basque respondents (75%) than for Northern Bizkaian Basque respondents (64%). Similarly, although both groups reported

---

<sup>12</sup> The IRB approval for the current study only permits adult participants (18 years of age or older)

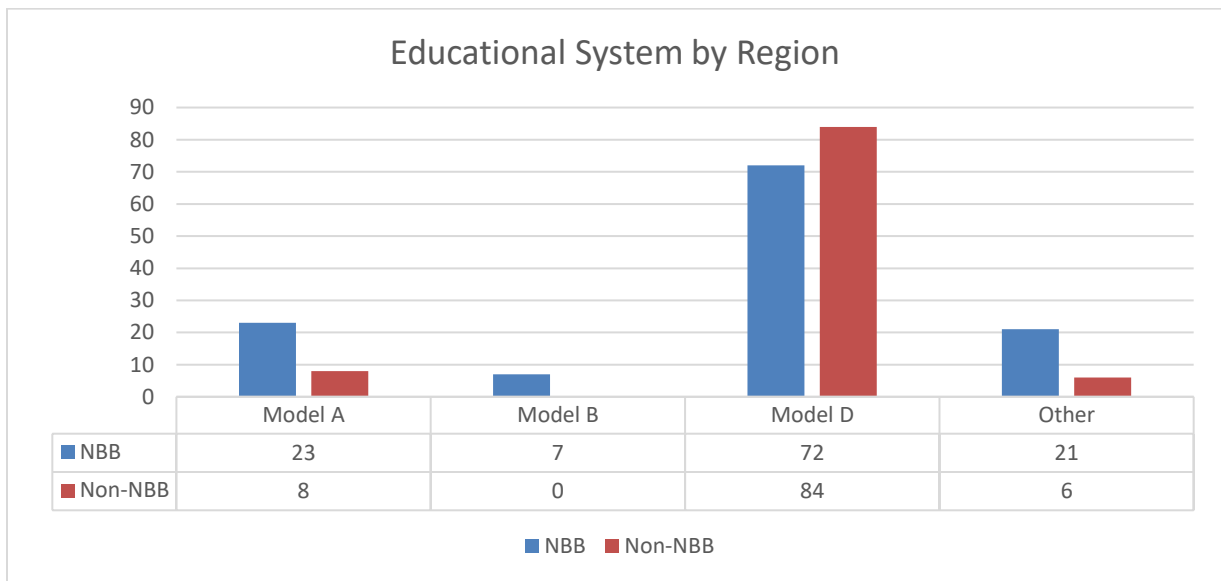
“Basque” as the most common native language, this percentage was much higher in the Northern Bizkaian Basque group (85%) than in the Non-Northern Bizkaian Basque group (42%). Figure 19 gives a more detailed visual representation of the respondents’ native language.



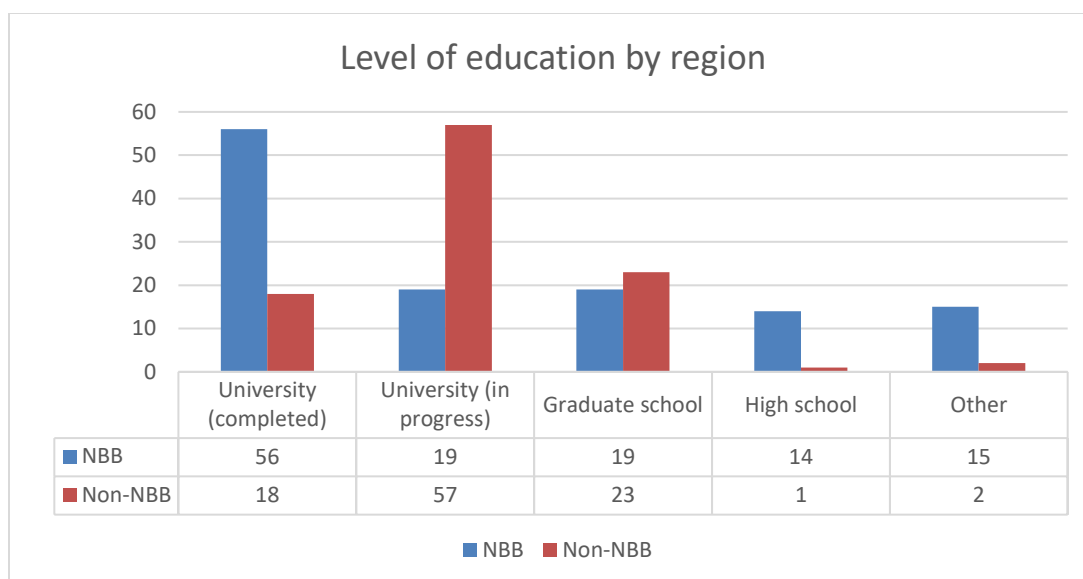
*Figure 19: Survey respondents' native language by region*

The majority of survey respondents from both groups (86% of Non-Northern Bizkaian Basque respondents and 59% of Northern Bizkaian Basque respondents) indicated that they were educated in the bilingual Model D method (see Section 2.1 for definitions), followed by the all-Spanish Model A method (8% of Non-Northern Bizkaian Basque respondents and 19% of Northern Bizkaian Basque respondents). Northern Bizkaian Basque respondents had a higher response rate of “other” for educational system when compared to Non-Northern Bizkaian Basque speakers (17% versus 6%, respectively), which may include private schools and Basque schools (*ikastolak*). These responses can be seen in Figure 20. With respect to highest level of education achieved, the majority of respondents in each group had at least some college education (76% of Northern Bizkaian Basque and 97% of non-Northern Bizkaian Basque),

although, due to the age ranges of the respondents, there were more college graduates in the Northern Bizkaian Basque group (61%) than in the non-Northern Bizkaian Basque group (41%). The Northern Bizkaian Basque group had the highest amount of “other” responses (12%) when compared to the Non-Northern Bizkaian Basque group (2%). These responses are displayed in Figure 21.



*Figure 20: Survey respondents' educational system by region*



*Figure 21: Survey respondents' level of education by region*

The next part of the survey (word judgments) was analyzed by dialect group and by word. Because the middle range of the survey (“3”) indicated a word that could be used equally in Spanish or Basque, a mean rating of 3 (scores between 2.5 and 2.9 were rounded to 3) will serve as the score for a borrowing<sup>13</sup> and also as the reference or center for the other categories (i.e. Spanish or Basque words). Words with a score below 2.5 were considered Spanish words (i.e. code-switches) and words with mean scores above 4 (scores between 3.5 and 3.9 were rounded to 4) were considered Basque words. In the context of the Interview task, which was in Basque and contained Spanish code-switches, only words with a score below 2.5 were labeled as code-switches and kept as target words for the analysis. Words from the survey with a score of 2.5 or higher were eliminated from the analysis ( $n = 13$  words, 17 data points). Words were eliminated on a dialect-specific basis; that is, if in the Northern Bizkaian Basque survey, a word had a mean score of 2.5 or higher (but not in the non-Northern Bizkaian Basque survey), the word in

<sup>13</sup> In this context, a word that could equally be used in both Spanish and Basque will be considered a borrowing, although no claims about direction of borrowing or the historical accuracy of the respondents' judgments is being made

question would be eliminated only from Lekeitio speakers in the analysis. Table 12 gives a summary of the mean score for each word by region and the resulting classification of each word.

| Word<br>(orthography <sup>14</sup> ) | Non-<br>Northern<br>Bizkaian<br>Basque | Category         | Northern<br>Bizkaian<br>Basque | Category         | Tokens<br>eliminated |
|--------------------------------------|----------------------------------------|------------------|--------------------------------|------------------|----------------------|
| jhombre! <sup>15</sup>               | 2.16                                   | Code-switch      | 1.84                           | Code-switch      |                      |
| a ver                                | 1.72                                   | Code-switch      | 1.44                           | Code-switch      |                      |
| asignatura                           | 2.19                                   | Code-switch      | 1.83                           | Code-switch      |                      |
| bueno                                | 2.39                                   | Code-switch      | 2.40                           | Code-switch      |                      |
| decididu                             | 1.97                                   | Code-switch      | 2.56                           | <b>Borrowing</b> | 1                    |
| desarroiuan                          | 2.32                                   | Code-switch      | 2.39                           | Code-switch      |                      |
| durezak                              | 1.94                                   | Code-switch      | 2.10                           | Code-switch      |                      |
| estrukturia                          | 2.73                                   | <b>Borrowing</b> | 2.91                           | <b>Borrowing</b> | 2                    |
| fallaten                             | 2.75                                   | <b>Borrowing</b> | 2.61                           | <b>Borrowing</b> | 1                    |
| faltan                               | 3.48                                   | <b>Borrowing</b> | 3.23                           | <b>Borrowing</b> | 1                    |
| fazilidadeak                         | 2.32                                   | Code-switch      | 2.36                           | Code-switch      |                      |
| fisioterapeuta                       | 2.87                                   | <b>Borrowing</b> | 2.52                           | <b>Borrowing</b> | 1                    |
| influentzia                          | 3.42                                   | <b>Borrowing</b> | 3.04                           | <b>Borrowing</b> | 1                    |
| japonesa                             | 2.96                                   | <b>Borrowing</b> | 3.25                           | <b>Borrowing</b> | 1                    |
| kontenidoa                           | 1.95                                   | Code-switch      | 2.39                           | Code-switch      |                      |
| manaja                               | 1.61                                   | Code-switch      | 2.35                           | Code-switch      |                      |
| mobimientua                          | 2.46                                   | Code-switch      | 2.82                           | <b>Borrowing</b> | 1                    |
| motxilero                            | 2.92                                   | <b>Borrowing</b> | 2.57                           | <b>Borrowing</b> | 3                    |
| necesario                            | 1.02                                   | Code-switch      | 1.03                           | Code-switch      |                      |
| o sea                                | 1.56                                   | Code-switch      | 1.39                           | Code-switch      |                      |
| oportunidadia                        | 2.42                                   | Code-switch      | 2.71                           | <b>Borrowing</b> | 1                    |
| podologia                            | 2.86                                   | <b>Borrowing</b> | 2.71                           | <b>Borrowing</b> | 1                    |
| pues                                 | 1.62                                   | Code-switch      | 1.34                           | Code-switch      |                      |
| super                                | 2.31                                   | Code-switch      | 2.19                           | Code-switch      |                      |
| tipiko                               | 2.64                                   | <b>Borrowing</b> | 2.41                           | Code-switch      | 2                    |
| tipo                                 | 2.54                                   | <b>Borrowing</b> | 2.37                           | Code-switch      | 1                    |
| trilingüe                            | 1.72                                   | Code-switch      | 1.66                           | Code-switch      |                      |
| ya                                   | 2.15                                   | Code-switch      | 1.84                           | Code-switch      |                      |

<sup>14</sup> The spellings in this table are those used in the non-Northern Bizkaian Basque survey; those used in the Northern Bizkaian Basque survey can be found in Appendix D.

<sup>15</sup> The exclamation marks were included for this word to indicate its colloquial meaning ('wow!') instead of the non-colloquial meaning ('man').

Table 12: Code-switch survey scores by dialect region

Overall, both dialect groups had similar scores on the words in the survey, showing similar scores (i.e., falling within the current study’s designated range of “code-switch” or “borrowing”) for 24 total words (15 code-switches and 9 borrowings). The dialect groups diverged on the scores of 5 words, in which the Northern Bizkaian Basque group rated *decidudu*, *mobimentua*, and *oportunidadia* with scores that designated borrowings (whereas the non-Northern Bizkaian Basque group rated these with scores that categorized them as code-switches) and the non-Northern Bizkaian group rated the words *tipiko* and *tipo* with scores that categorized them as borrowings (while the Northern Bizkaian Basque group rated these with scores that categorized them as code-switches). It is interesting to note that several of the hybrid forms with Basque morphology, such as *kontenidoa* and *fazilidadeak* were rated as code-switches by both dialect groups, indicating that some hybrid words are considered Spanish code-switches by bilinguals, despite the presence of Basque morphology. Now that the ambiguous and hybrid words from the data set have been explored and categorized, the following section will describe the phonetic measurements made in the current study, concluding this chapter with details about the statistical analyses used on this dissertation’s data set.

### 3.3.3 Phonetic measurements

Two main phonetic variables were measured in the present study: *peak alignment* and *f0 maximum* (henceforth *f0 max*). As discussed in Section 1.3.3, there is evidence that code-switched words are produced with a higher *f0* (Olson (2012), Fricke et al. (2016)) and that dialects of Spanish and Basque differ in where the *f0* peak is aligned. To summarize, Elordieta & Irurtzun (2016) established that peak alignment is within the stressed syllable (early) for Basque

and Lekeitio Spanish and after the stressed syllable (delayed) phrase-medially for Standard Peninsular Spanish. Table 13 summarized these anticipated dialectal differences below.

| Language-Dialect                              | Phrase-medial alignment           | Phrase-final alignment |
|-----------------------------------------------|-----------------------------------|------------------------|
| Spanish-Standard Peninsular                   | Delayed                           | Early                  |
| Spanish-Lekeitio                              | Early                             | Early                  |
| Basque-Standard                               | Delayed (if pre-nuclear) or early | Early                  |
| Basque-Northern Bizkaian<br>Basque (Lekeitio) | Early                             | Early                  |

*Table 13: Differences in peak alignment for Spanish and Basque (Elordieta & Irurtzun (2016))*

To measure peak alignment, the syllable boundary for the target word was manually placed on an interval tier in *Praat* and the f0 max was located using the “find f0 maximum” command. Words that did not contain pitch accents (i.e. function words or content words without evidence of an f0 excursion) were not included in the analysis. Each interval on the syllable tier was labeled with the syllable’s orthography and the word juncture break it has with respect to the following word. Word juncture breaks in the current study (see Table 14) are based on the definitions described for the English Tone and Break Indices (ToBI) transcription conventions described in Beckman et al (2005). The word juncture break indices in the current study are not meant to be language-specific, due to the presence of code-switching, but to be applicable to the types of word junctures possible in Spanish and Basque.

| Break Index | Corresponding boundary                                                                    | Use in current study                                             |
|-------------|-------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| 0           | very close phrase-medial word juncture, usually containing a clitic group or contractions | combined with 1                                                  |
| 1           | typical phrase-medial word juncture                                                       | phrase-medial word juncture                                      |
| 2           | mismatch between breaks 1 and 3                                                           | not used; only clear instances of 1 and 3 are used               |
| 3           | ip boundary                                                                               | ip or AP boundary; (considered <i>phrase-final</i> along with 4) |
| 4           | IP boundary                                                                               | IP boundary (considered <i>phrase-final</i> along with 3)        |

Table 14: Break indices used in current study

If the amount of juncture was perceptually and acoustically judged as a typical phrase-medial word boundary with no syllable lengthening, a “1” was used; a “3” was used at an intermediate phrase boundary (ip) or AP boundary; and a “4” boundary was used for intonation phrases (IP). For the purposes of this analysis, words with a “1” juncture will be considered *phrase-medial* and words “3” or “4” juncture will be considered *phrase-final*. This distinction in prosodic position will be used to account for pitch accent alignment differences in Standard Spanish (an ‘early’ peak in phrase-final position and a ‘delayed’ peak in phrase-medial position) Standard Basque (early or delayed peaks may occur phrase-medially in pre-nuclear position) as discussed in Chapter 2. Words that contained creaky voice, background noise, or other noise-related perturbations (such as laughter or coughing) were not analyzed, as the relevant acoustic information is obscured or absent in these contexts.

The range for the f0 max search consisted of the stressed syllable and post-tonic syllable in order to account for both early and delayed peaks. Contexts with tonal crowding (i.e. two adjacent stressed syllables) were not included in the analysis, as it would be difficult to discern a delayed peak from the following pitch accents in these cases. In the case of microprosodic

effects caused by plosives in onset position, the f0 maximum range excluded this segment to avoid f0 perturbations that could be erroneously classified as an f0 excursion.

Once the f0 maximum was located, a label was placed on a separate interval tier (from the syllable labels) which contained the f0 maximum as one boundary and the syllable offset (for early peaks) or syllable onset (for delayed peaks) as the other boundary. This distance (between f0 peak and syllable offset (early peaks) or syllable onset (delayed peaks)) will be referred to as *peak duration*. A different reference point was chosen for early and delayed peaks so that the normalized ratio (peak duration divided by syllable duration) of peak alignment between early peaks would be more comparable.

An example of the peak alignment measurement procedure is shown below in Figure 22. In this figure, an utterance by Speaker 8 (Lekeitio) from the Discourse Completion Task shows a Spanish sentence with a Basque proper name (*Hoy, he conocido a Maialen, la morena*, “Today, I met Maialen, the brunette”). This utterance contains two examples of early peaks (marked with “H” on the peaks tier) and one delayed peak (indicated with “<H” on the peaks tier). The stressed syllables of content words marked in the syllable tier show the word juncture label in the subscript (e.g. a “3” or ip boundary for *hoy*). The phrase-final word *morena* is not labeled in the syllable tier, as it has no perceptual or acoustic evidence of prominence in this utterance.

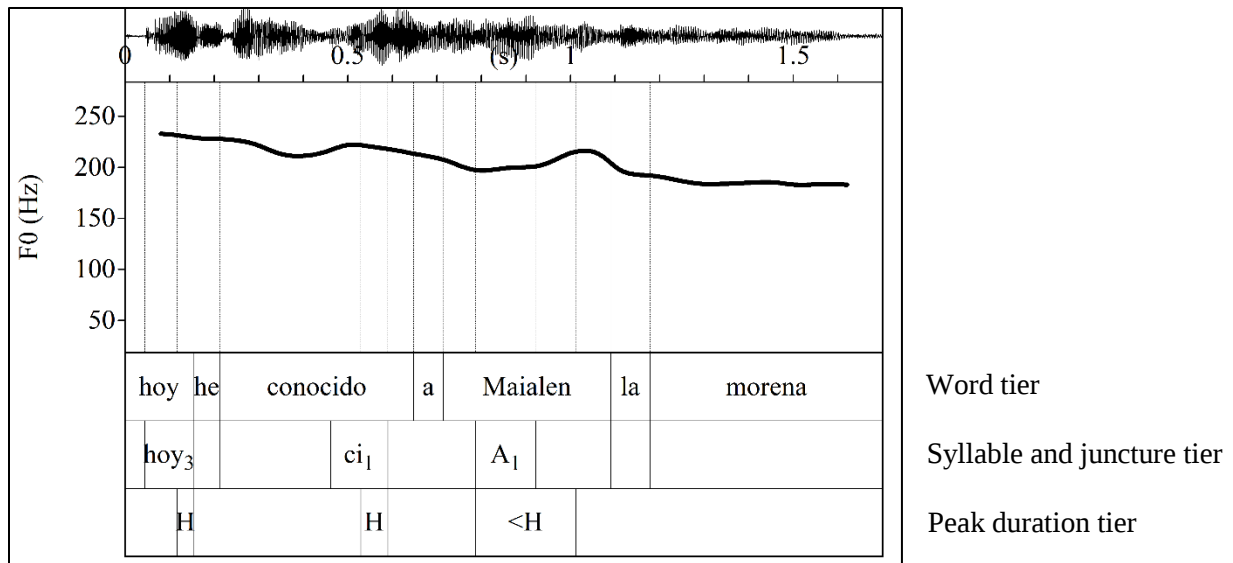


Figure 22: Peak alignment measurement example

After these tiers were labeled, a *Praat* script (Lennes (2004), edited by the current author) was used to extract duration in milliseconds (ms) from the syllables labeled in the second tier, the peaks labeled in the third tier, as well as the f0 max (in Hz) from the peaks labels in the third tier<sup>16</sup>. Once these values were calculated by the script, the peak alignment (early versus delayed) was determined by dividing the peak duration in the third tier by the syllable duration in the second tier (e.g. a peak duration of 50 ms was divided by a syllable duration of 100 ms, resulting in .50). This method allowed for a more standardized comparison of peak-to-syllable ratio of early and delayed peaks than values in raw milliseconds. Values under 1 indicate early peaks (e.g. a peak occurred at .50 or 50% of the total syllable duration) and values over 1 indicate delayed peaks (e.g. a peak occurred at 1.1, or 110% of the total syllable duration). The f0 max values from the peak labels (third tier) were kept in Hz as reported by the script. Because the

<sup>16</sup> The script recorded the f0 max in Hz in each interval, whereas the *Praat* “find f0 maximum” command was used to locate the f0 max in order to measure peak alignment, but not to record what the f0 value was at that point.

current study's participants were all adult females, no f0 normalization or transformation methods were used.

### 3.3.4 Statistical analyses

In order to investigate the predictors of peak alignment and f0 max in the current data set, multiple linear regressions were performed in R (R Core Team (2015)) for each outcome variable and task, resulting in six total models (peak alignment in the Reading Task, Discourse Completion Task, and the Interview Task; f0 max in the Reading Task, Discourse Completion Task, and the Interview Task). The choice to perform four models by sub-setting the data by task (Discourse Completion Task or Interview) instead of coding the tasks as an independent variable was made to facilitate interpretation of multi-way interactions. Before performing the analysis, data diagnostics were utilized to check for outliers and multi-collinearity between variables to prepare the data set for the final regression models.

The distribution of the data was checked using an R script (Dhana (2016)) that checks for outliers by using Tukey's method when an outcome variable is specified. Tukey's method assesses outliers using the Interquartile Range (IQR), which uses the median as the measure of central tendency and therefore makes it more resistant to extreme outliers than methods that use the mean or standard deviation. Since the current study's data set examines semi-spontaneous speech that likely has a high level of variability, Tukey's method for assessing outliers is the most appropriate, as it does not assume a normal distribution. The R outlier script was performed on both outcome variables (peak alignment and f0 max) for each task (Reading Task, Discourse Completion Task and Interview) and the outliers for each condition (e.g. f0 max in Discourse Completion Task or peak alignment in the Interview) were removed. Values were considered outliers if they were below or above 1.5 times (\*) the IQR (i.e. quartile 1 – 1.5\*IQR or quartile 3

+ 1.5\*IQR). Table 15 shows in more detail the amount and percentage of outliers for each condition as well as the adjusted means after removing outliers.

| Variable and Task                         | Number of outliers | Mean with outliers | Mean without outliers |
|-------------------------------------------|--------------------|--------------------|-----------------------|
| F0 max, Reading Task                      | 11 (3.5%)          | 237 Hz             | 237 Hz                |
| F0 max, Discourse Completion Task         | 17 (3.1%)          | 245 Hz             | 242 Hz                |
| F0 max, Interview                         | 24 (4.5%)          | 223 Hz             | 222 Hz                |
| Peak alignment, Reading Task              | 16 (5%)            | 1.5                | 1.2                   |
| Peak alignment, Discourse Completion Task | 5 (.9%)            | .81                | .78                   |
| Peak alignment, Interview                 | 29 (5.4%)          | .75                | .67                   |

*Table 15: Outlier check for outcome variables*

The following histograms in Figure 23 give a visual representation of the distribution of the data before and after removing outliers for each variable and task. In Figure 23, the histograms for f0 max are shown for the Reading Task (left), Discourse Completion Task (middle), and Interview (right). Below Figure 23 is Figure 24, which contains the histograms for peak alignment in the Reading Task (left), Discourse Completion Task (middle), and the Interview (right). When variables and tasks are compared, the histograms in Figure 23 show that the Interview task was affected most by removing outliers, as 5.4% of the data was excluded for peak alignment, followed by 4.5% of data for f0 maximum (Figure 24). Overall, the Discourse Completion Task was more normally distributed when compared to the Interview and Reading Tasks, with 4% of data being excluded for both f0 max and peak alignment combined.

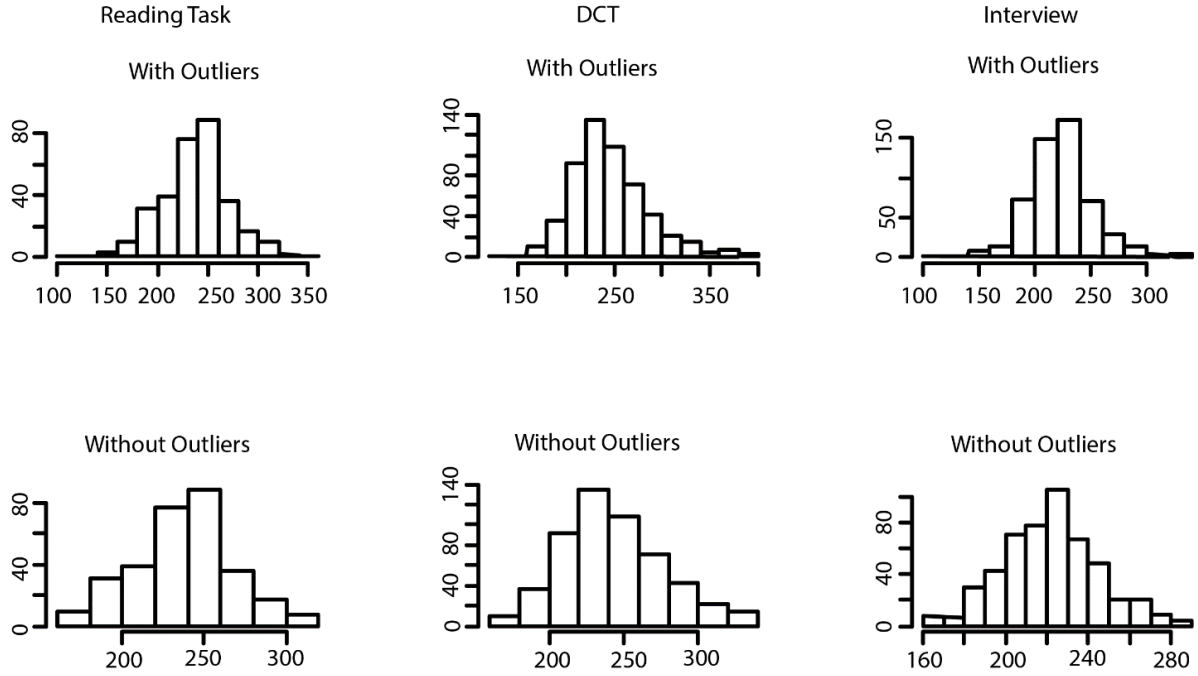


Figure 23: Histograms of  $f_0$  max with and without outliers for Reading Task (left), Discourse Completion Task (middle), and Interview (right). The x-axes are  $f_0$  in hertz (Hz).

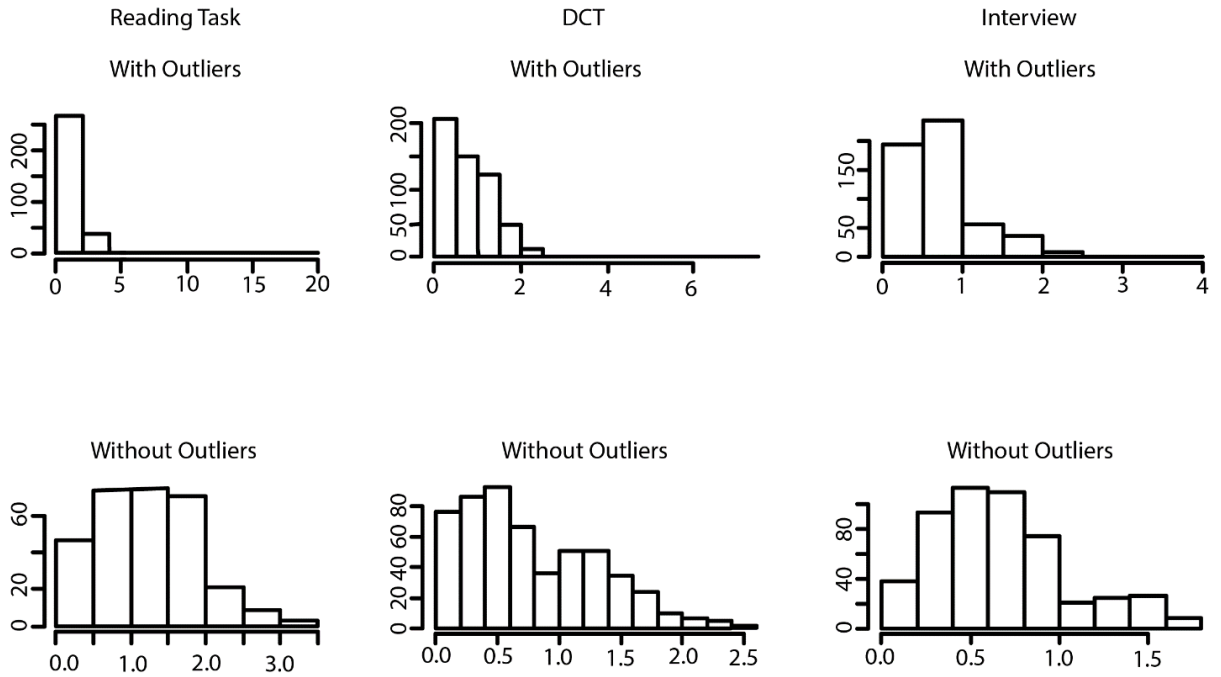


Figure 24: Histograms of peak alignment before and after outliers in the Reading Task (left), Discourse Completion Task (middle), and Interview (right). The x-axes are peak alignment ratios (peak to syllable edge/syllable duration)

After excluding the outliers from the analysis ( $n = 91$ ), the borrowings from the survey discussed in Section 3.3.4 ( $n = 7$ ; 9 tokens were already eliminated in the outlier analysis) were also excluded from the Interview, yielding a total of 529 analyzable tokens for the Interview task, 546 tokens for the Discourse Completion Task, and 305 for the Reading Task<sup>17</sup>. Before performing the final regression analyses, multi-collinearity between variables was examined, as variables that are highly correlated with one another may obscure the predictive power of individual variables in the model. This concern is relevant for the Discourse Completion Task and Interview Tasks, which both contain a large number of categorical variables with similar levels (i.e. variables whose levels are “Spanish”, “Basque”, or “Both”). Because the Reading Task only considers linguistic variables (Region, Context, and Prosodic Position), multicollinearity will not be tested for this dataset.

Due to the identical values reported by each speaker for *Language(s) spoken at home* and *Language(s) spoken with friends*, these variables will be combined into one variable called *Language (social)*. Additionally, because the only values reported in the *Spanish self-rating* variable were between 4 and 5 (and therefore, not a true ordinal variable), the *Spanish self-rating* variable was converted to a categorical variable with the levels “high” (average rating of 5), and “low” (average rating of 4.3). The *Basque self-rating* variable has three levels: “low” (rating of 4.3), “mid” (rating of 4.6), and “high” (rating of 5). For *Spanish self-rating*, there were two speakers (Speakers 5 and 6) in the “low” category and the remainder of the speakers (Speakers 1, 2, 3, 4 and 7) in the “high” category. For *Basque self-rating*, there was one speaker each in the “low” (Speaker 3) and “mid” (Speaker 6) with the remainder of speakers in “high” category

---

<sup>17</sup> This is the number of total tokens. This number will vary slightly by outcome variable, as the outliers excluded did not affect each outcome variable equally. Please refer to the regression tables in the results sections for the number of observations for each outcome variable and task.

(Speakers 1, 2, 4, 5, and 7). Similar, *Language(s) spoken at work/school* has a level (“Basque”) with only one speaker (Speaker 5). This imbalance between variable levels for these three self-rated fluency variables make comparisons between groups difficult and unreliable in a statistical analysis (particularly with *Basque self-rating*, which has two levels with one speaker only). For this reason, these three self-rated fluency variables (*Language(s) spoken at school/work*, *Spanish self-rating*, and *Basque self-rating*) will be eliminated from the statistical analyses.

In order to calculate the test for multicollinearity among the predictor variables in the Discourse Completion Task and Interview Task, a Variance Inflation Factor (VIF) function was performed (Beck (2013)) in R package fmsb (Nakazawa (2015)). This function performs a stepwise elimination procedure in which the variable with the highest VIF (a higher VIF is indicative of a predictor variable being highly correlated with another) is eliminated and then the VIF of the remaining variables is recalculated until all variables are below a specified threshold. Although VIF thresholds in the social sciences typically range between 5-10, the present study chooses the more conservative threshold of 5, as recommended in Pan & Jackson (2008). The following predictor variables in Table 16 were included in the multicollinearity diagnostic, which was performed for the Discourse Completion Task and Interview data sets.

| Predictor variable                                 | Type                                                                                                                   |
|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------|
| Language dominance (Semantic Verbal Fluency score) | Continuous (range = 44-110)                                                                                            |
| Region                                             | Categorical <ul style="list-style-type: none"> <li>• Lekeitio</li> <li>• Bilbao</li> </ul>                             |
| Prosodic Position                                  | Categorical <ul style="list-style-type: none"> <li>• Phrase-final</li> <li>• Phrase-medial</li> </ul>                  |
| Context                                            | Categorical <ul style="list-style-type: none"> <li>• Matrix language (ML)</li> <li>• Embedded language (EL)</li> </ul> |

| <b>Predictor variable</b>          | <b>Type</b>                                                                                                              |
|------------------------------------|--------------------------------------------------------------------------------------------------------------------------|
| Language dominance (self-reported) | Categorical <ul style="list-style-type: none"> <li>• Basque</li> <li>• (Spanish)<sup>18</sup></li> <li>• Both</li> </ul> |
| Language (social)                  | Categorical <ul style="list-style-type: none"> <li>• Basque</li> <li>• Spanish</li> <li>• Both</li> </ul>                |

*Table 16: Predictor variables used in regression diagnostics for multicollinearity*

The VIF diagnostic eliminated the variable *Language Dominance (self-reported)* (VIF of 5.36) from the Discourse Completion Task data set. This variable was eliminated from the analysis to allow for comparisons of the same variables across both data sets (Discourse Completion Task and Interview Task). Table 17 contains a revised list of the variables which will comprise the final regression models for the Discourse Completion Task and Interview tasks.

| <b>Regression models (Discourse Completion Task and Interview Tasks)</b> |
|--------------------------------------------------------------------------|
| <b>Outcome variables</b>                                                 |
| Peak alignment                                                           |
| F0 max                                                                   |
| <b>Predictor variables</b>                                               |
| Region                                                                   |
| Prosodic Position                                                        |
| Context                                                                  |
| Language dominance (Semantic Verbal Fluency)                             |
| Language (social)                                                        |

*Table 17: Variables used in regression models*

In each regression model, all predictor variables will enter the model simultaneously. The categorical variables will use the default reference group values specified by R (lowest coded

<sup>18</sup> No speakers reported “Spanish” as their dominant languages, so the only comparison in the statistical models will be between “Basque” and “Both”

group, either numerically or alphabetically) for consistent comparisons of variables across data sets. Reference groups for each variable will be specified for each regression model in the results section. Due to the expected differences in peak alignment and f0 by Region (Lekeitio, Bilbao), Prosodic Position (medial, final), and the hypothesized differences in Context (ML, EL)<sup>19</sup>, interaction terms for these variables will be specified for each model and assessed with model-fitting criteria. The specifics of interaction terms and model fitting will be discussed individually for each model in the results chapter.

The following chapter will present the results of the analysis, beginning with descriptive statistics about the code-switched portions of the Discourse Completion Task (Matrix Language Spanish) and Interview (Matrix Language Basque), before continuing to the results of the regression analyses for the peak alignment and f0 max for each task.

---

<sup>19</sup> For the purposes of the current study, the terms *Matrix Language* and *Embedded Language* will be used to refer to the language which contributes the largest number of morphemes to the utterance/context and does not assume any additional theoretical constructs apart from this descriptive purpose

## 4 RESULTS

This chapter will present the results of the dissertation. The descriptive statistics of the code-switched data set will be discussed first, followed by results for f0 max and concluding with the results for peak alignment.

### 4.1 Distribution of code-switches

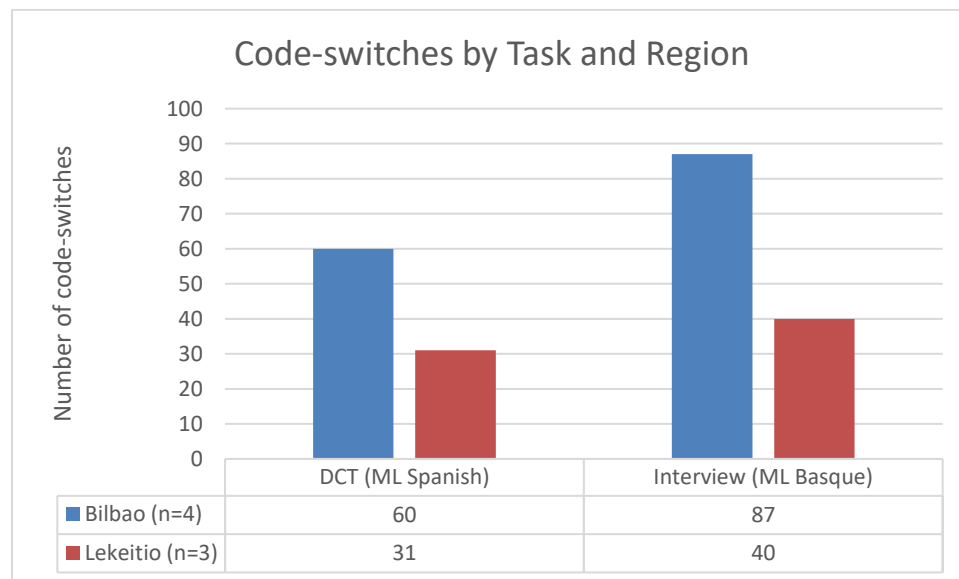
This section will report the distribution of code-switched (or embedded language) words by task, region, speaker, lexical category, and prosodic position for a thorough descriptive analysis of the corpus before expanding with the quantitative results in the next two sections. The corpus contained 218 analyzable code-switches (i.e. those that were submitted for acoustic analysis and not eliminated for being outliers, borrowings, or acoustically unfit tokens), 91 of which were contributed from the Discourse Completion Task and 127 from the Interview task.

When the code-switches are split by task (Discourse Completion Task and Interview) and region, Figure 25 shows that in both tasks, the Bilbao speakers contributed more than double the code-switches when compared to the Lekeitio speakers, with 66% of code-switches in the Discourse Completion Task and 69% of code-switches in the Interview task being realized by Bilbao speakers. This imbalance has two potential sources: one is the different number of speakers in each group (there are four Bilbao speakers and three Lekeitio speakers) and another is region itself. In the native speaker judgment survey discussed in Section 3.3.2, a higher percentage of non-Northern Bizkaian Basque speakers (75%) reported code-switching, compared to 65% of Northern Bizkaian Basque speakers<sup>20</sup>. As the Bilbao speakers in this study speak Standard or Central Basque (non-Northern Bizkaian Basque varieties), it is possible that the

---

<sup>20</sup> This survey question asked respondents if they code-switched when speaking Spanish and/or Basque, but did not ask about directionality (e.g. switching from Spanish to Basque or vice versa. All survey questions can be found in Appendix D).

tendency revealed in the survey also applies to the Bilbao (non-Northern Bizkaian Basque) and Lekeitio (Northern Bizkaian Basque) speakers in this study.



*Figure 25: Code-switches by Task and Region*

Further splitting the number of code-switches by speaker (see Figure 26) reveals that Speakers 1 and 2 (Bilbao) contributed the most code-switches for both tasks and Speaker 4 (Bilbao) contributed the fewest code-switches for both tasks. Speakers 3 (Bilbao), 5, and 6 (Lekeitio) produced similar numbers of code-switches in both tasks, with Speakers 2 (Bilbao) and 7 (Lekeitio) being the only speakers in the corpus to produce more than double the code-switches in the Interview task than the Discourse Completion Task. When the speakers are considered by their regional groupings, the Lekeitio group (speakers 5-7) are more evenly represented by each speaker's code-switching as opposed to the Bilbao group (Speakers 1-4), which is overrepresented with the data from Speakers 1 and 2. Although an ideal data set would contain (near) equivalent amounts of data from each participant or group, the semi-spontaneous nature of this study was not able to control for factors such as speaker extroversion, frequency of

code-switching, or discrepancies in code-switching between contexts (i.e. differences in frequency between tasks or language modes).

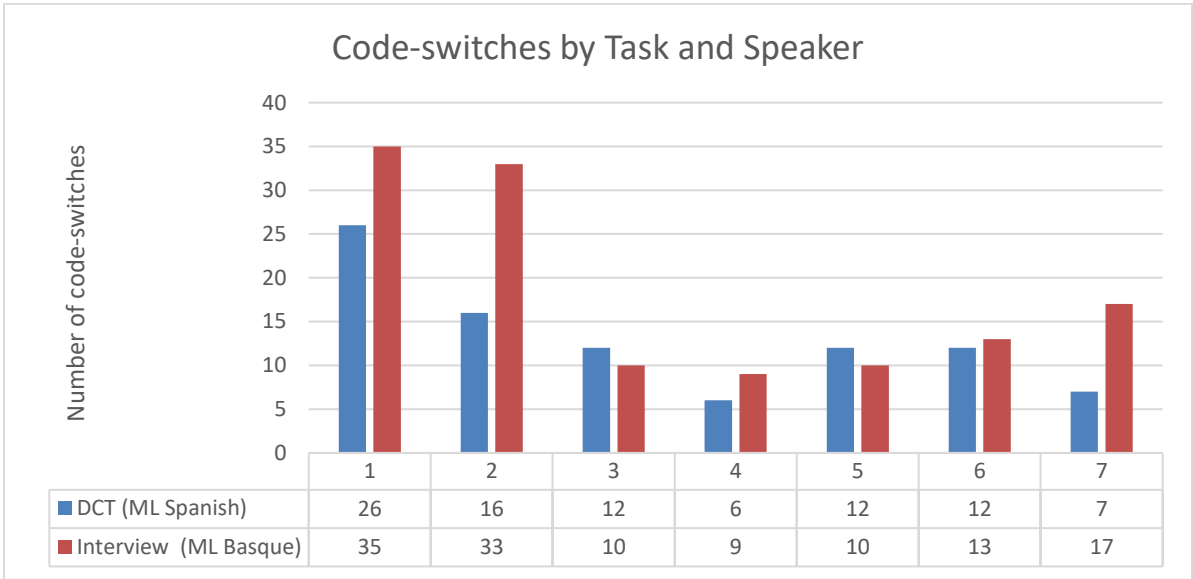
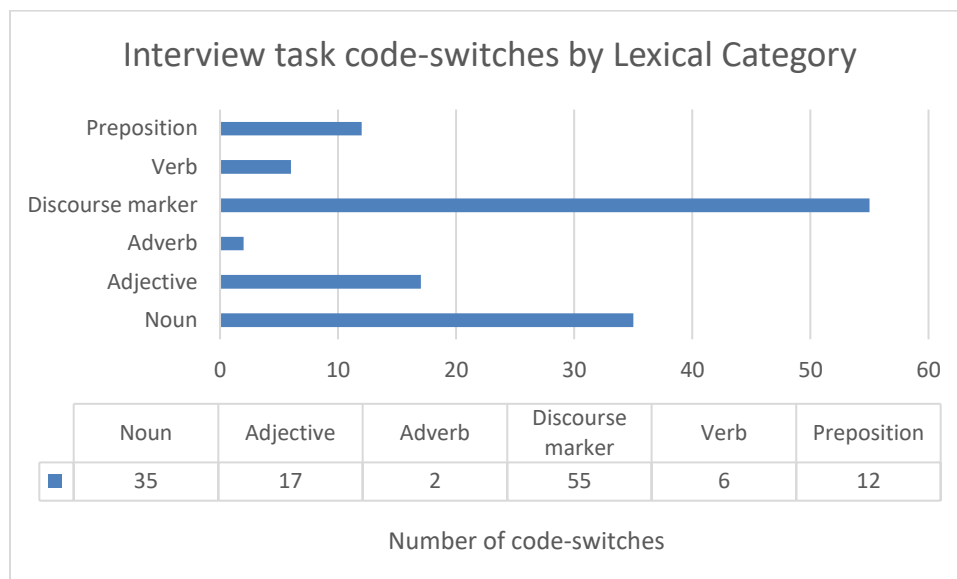


Figure 26: Code-switches by Task and Speaker (1-4 = Bilbao; 5-7 = Lekeitio)

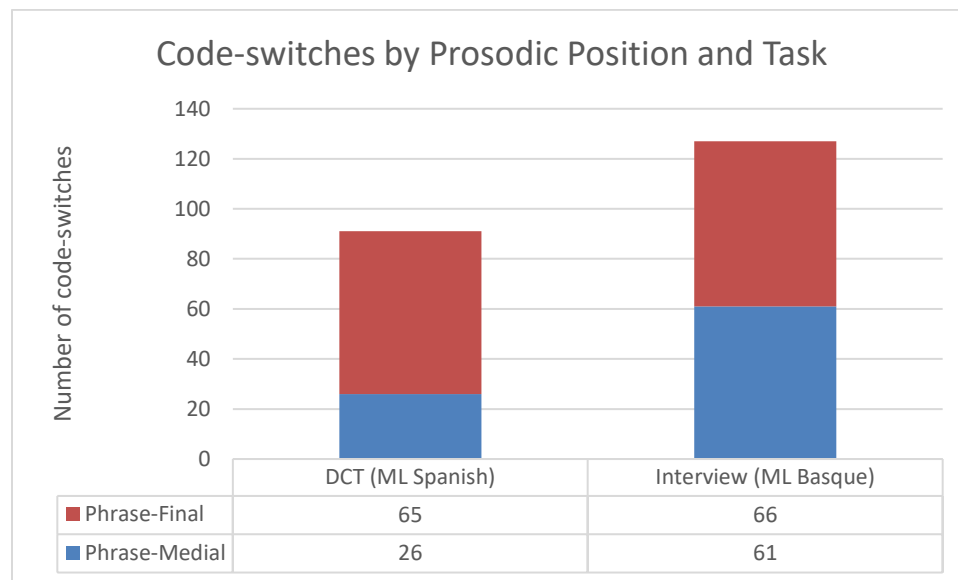
When the linguistic aspects of the corpus’ code-switches are considered, several trends emerge with respect to lexical category and prosodic position. As mentioned in Chapter 3, the code-switches in the Discourse Completion Task were all recasts of the Basque trigger words, providing no variation in lexical category for this task (all code-switches in the Discourse Completion Task were the nouns shown in Table 6). However, the Interview task contained a diversity of lexical categories, as shown in Figure 27. Nouns and discourse markers were the most frequently code-switched words of the lexical categories documented, accounting for 71% of all code-switches. These findings corroborate those of Poplack (1980, 1988), whose work provides evidence that nouns are among the most commonly code-switched words and also Lantto (2014, 2015), whose studies found that discourse markers and slang were the most frequently code-switched words by Spanish-Basque bilinguals.



*Figure 27: Interview task code-switches by Lexical Category*

Continuing the last descriptive category, prosodic position, Figure 28 reveals that the Discourse Completion Task and Interview had different proportions of code-switches in phrase-medial and final positions. More specifically, 71% of code-switches in the Discourse Completion Task occurred in phrase-final position, compared to only 52% in the Interview task. Previous work on prosody and code-switching by Fricke et al (2016) reported that code-switches typically contain a (phrase-final) boundary, which is only supported by the Discourse Completion Task data in the present study. This discrepancy between tasks has several possible sources. For one, the nature of the tasks may have been a contributing factor; the Discourse Completion Task contained situational prompts and lexical items that were recast by the speaker, which may have affected the prosodic position of the code-switches (as opposed to the Interview task, which had more open-ended questions and did not elicit recasts). Another reason could be the method of reporting prosodic boundaries in Fricke et al (2016): this study employed ratings by listeners for

determining prosodic boundaries, whereas the current study uses acoustic measurements (such as pause, lengthening, etc) to determine the presence of a prosodic boundary.



*Figure 28: Code-switches by Prosodic Position*

The following sections of this chapter will present the results of the quantitative analyses of the present study's code-switches, beginning with peak alignment and continuing to f0 maximum.

## 4.2 Peak alignment

This section will present the results of the peak alignment analyses by task, beginning with the reading task, continuing to the Discourse Completion Task, and concluding with the Interview task. In Chapter 2, the current study predicted that language context (i.e. ML or EL) would influence language-specific prosodic traits. More specifically, the prosodic patterns of the ML are hypothesized to override those of the EL. Under this hypothesis, Spanish peak alignment patterns (i.e. delayed for Standard Peninsular speakers) should surface on Basque code-switches in the Discourse Completion Task and Basque peak alignment patterns (early peaks) should

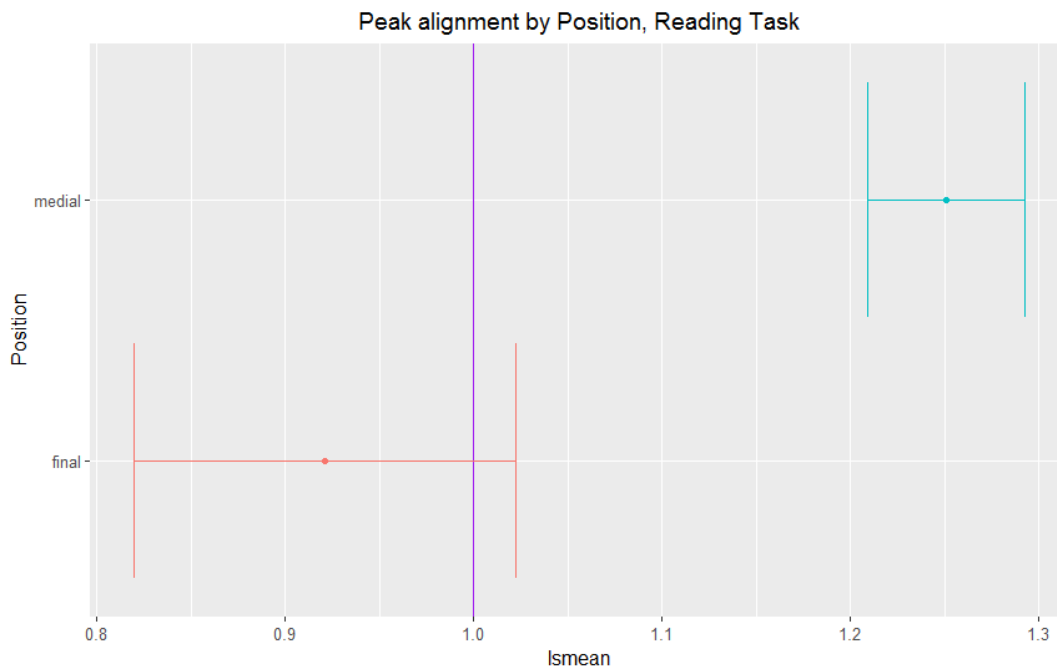
surface on Peninsular Spanish phrase-medial code-switches in the Interview task. Because Standard Basque can optionally have delayed peak alignment in phrase-medial pre-nuclear (i.e. not pre-verbal) position, delayed peaks on Basque words by Standard Basque speakers will be investigated for its position (i.e., a delayed peak in phrase-medial pre-nuclear position will not be considered a Matrix Language effect). Due to the early peak alignment expected in the Spanish and Basque varieties from Lekeitio, no difference in peak alignment is expected in this group for either task.

Additionally, dominant language is hypothesized to interact with peak alignment patterns in which a speaker's dominant language is expected to deviate from monolingual/unilingual norms when it is the EL compared to when it is the ML (e.g. a speaker whose dominant language is Spanish should show more influence from Basque when Spanish is the EL), as reported in Olson (2015, 2016b), Aly (in press), and Meuter & Allport (1999). Before presenting the results related to these hypotheses, the peak alignment results from the Reading task (Spanish language only) will be discussed in order to establish a unilingual baseline for comparison in code-switching tasks.

#### 4.2.1 Unilingual Spanish baseline (Reading task)

The purpose of the reading task was to establish a unilingual (i.e. not code-switched) Spanish baseline that would serve as a comparison to the code-switched tasks (Discourse Completion Task and Interview). The previous studies discussed in Section 2.3.1 (e.g. Elordieta & Calleja (2005), Elordieta & Irurtzun (2016)) reported that while most Basque Country Spanish dialects match Standard Peninsular Spanish norms with respect to peak alignment (delayed phrase-medial peaks and early phrase final peaks, as established by Estebas Vilaplana & Prieto (2008)), Lekeitio Spanish has early peaks in both prosodic positions (medial and final).

The multiple linear regression performed on the peak alignment data for the Reading Task was significant ( $F(3, 296) = 7.651, p < .001$ , adjusted  $R^2 = .063$ ) and revealed that Position (phrase-medial) was a positive predictor of peak alignment, signifying that phrase-medial peaks ( $M = 1.25, SE = .04$ ) were more delayed than phrase-final peaks ( $M = .92, SE = .10$ ). However, there was no significant effect of Region. This difference in peak alignment by prosodic position (for both regional groups) is shown in Figure 29, in which the x axis contains the peak alignment ratio (peak duration divided by total syllable duration) and the y axis contains prosodic position. Values above 1 on the x axis (indicated in all peak alignment plots with a purple vertical line for clear reference) indicate a delayed peak and values below 1 indicate an early peak.



*Figure 29: Peak alignment in Reading Task by Prosodic Position*

A further look by dialect group and prosodic position reveals that although not statistically significant, Lekeitio speakers have a smaller distinction in prosodic position (12%) when

compared to Bilbao speakers (54%). Furthermore, there is a trend in which Bilbao speakers have later phrase-medial peaks (1.4 or 140% of the total syllable duration) than Lekeitio speakers (1.1 or 110%). When peak alignment is considered with these interaction trends between dialect group and prosodic position in mind, the results of each group reveal the expected dialectal differences established in the previous studies. These differences can be seen in the plot in Figure 30. The means for each group and position is presented in Table 17.

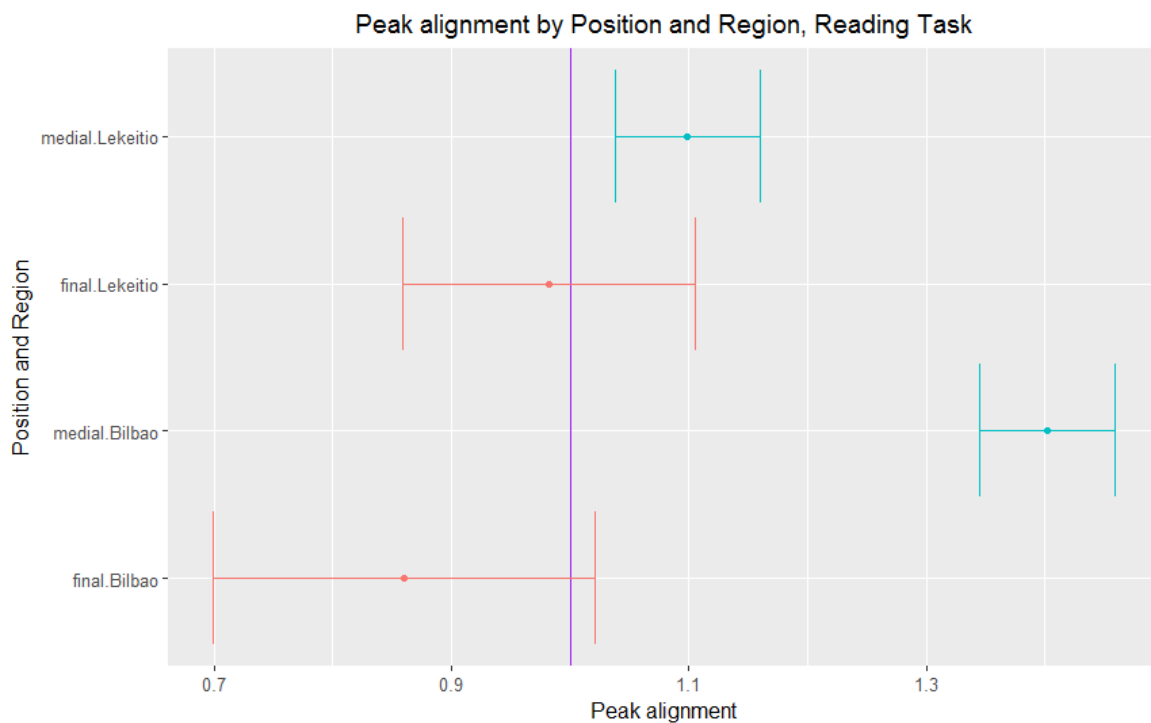


Figure 30: Peak alignment in the Reading Task (monolingual Spanish) by Position and Region

| Region   | Phrase-medial | SE  | Phrase-final peak | SE  |
|----------|---------------|-----|-------------------|-----|
| Bilbao   | 1.4 (delayed) | .06 | .86 (early)       | .16 |
| Lekeitio | 1.1 (delayed) | .06 | .98 (early)       | .12 |

Table 18: Peak alignment means for Reading Task

Table 19 contains the results of the regression model for peak alignment in the Reading Task, displaying the regression coefficients (unstandardized<sup>21</sup>), standard error, and t-values for all variables in the model.

| <i>Independent variables (predictors):</i>               |                 |                    |                 |
|----------------------------------------------------------|-----------------|--------------------|-----------------|
| <b>Simple effects</b>                                    | <b><i>B</i></b> | <b><i>SE B</i></b> | <b><i>T</i></b> |
| Region:Lekeitio (reference: Bilbao)                      | .122            | .202               | .605            |
| Position:Medial (reference: Final)                       | .542            | .170               | 3.184**         |
| <b>Interactions</b>                                      |                 |                    |                 |
| Region: Lekeitio x Position: Medial                      | -.426           | .219               | -1.946          |
| Constant                                                 | .860            | .161               | 5.357***        |
| Observations                                             | 300             |                    |                 |
| R <sup>2</sup>                                           | .0720           |                    |                 |
| Adjusted R <sup>2</sup>                                  | .0626           |                    |                 |
| Residual Std. Error                                      | .662            |                    |                 |
|                                                          | (df = 296)      |                    |                 |
| F Statistic                                              | 7.651***        |                    |                 |
|                                                          | Df (3, 296)     |                    |                 |
| <i>Note: * p &lt; .05 ** p &lt; .01 *** p &lt; 0.001</i> |                 |                    |                 |

*Table 19: Regression results for peak alignment in the Reading Task*

The results from the Reading Task revealed that in the current study's sample, speakers from Lekeitio and Bilbao produce delayed peaks in phrase-medial position and early peaks in phrase-final position, as predicted for Peninsular Spanish, but not for Lekeitio Spanish. Due to the presence of delayed peaks in the Lekeitio group, it is hypothesized that speakers from Lekeitio will be prone to the same ML effects of peak alignment as the speakers from Bilbao. This revised hypothesis on peak alignment is summarized below for clearer reference in Table 20. The hypothesis about speaker language dominance and peak alignment remains unaffected by the Reading Task results.

<sup>21</sup> Standardized coefficients are not reported for the regression analyses in this dissertation due to the presence of primarily categorical predictors in the models performed, which may obscure the interpretability of a standardized coefficient (compared to continuous predictors).

| Spanish Dialect     | Expected phrase-medial alignment | Actual phrase-medial alignment (from Reading Task) | Hypothesized phrase-medial alignment when ML | Hypothesized phrase-medial alignment when EL |
|---------------------|----------------------------------|----------------------------------------------------|----------------------------------------------|----------------------------------------------|
| Standard Peninsular | Delayed                          | Delayed                                            | Delayed                                      | Early                                        |
| Lekeitio            | Early                            | <b>Delayed</b>                                     | <b>Delayed</b>                               | <b>Early</b>                                 |

*Table 20: Hypothesized Spanish peak alignment in ML and EL contexts*

#### 4.2.2 Discourse Completion Task (Discourse Completion Task): ML Spanish, EL Basque

The regression models performed for the Discourse Completion Task contain Region, Context, Position, Language Dominance (Semantic Verbal Fluency), Language (social), and Spanish self-rating as predictor variables. Additionally, the models performed for the Discourse Completion Task and Interview Tasks consider the interactions between Region (Bilbao/Lekeitio) and (prosodic) Position (medial or final) as well as the interaction between Context (ML or EL) and all predictor variables entering the model (including with Region and Position, resulting in a three-way interaction term). In order to select the model which best fits the data, the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criteria) of the model with all interaction terms specified (i.e. full model) will be compared with a condensed model which eliminates non-significant interaction terms (simple effects will remain in the model). AIC and BIC both evaluate which model is more likely to be closer to the truth (i.e. which model is a better fit for the data). For both AIC and BIC, a smaller value is indicative of better fit, although the two criteria have slightly different assumptions (or priors). Most importantly for the current dataset, the BIC penalizes models with more parameters more severely than AIC. Hence, differences in model fit in AIC and BIC are most likely due to model complexity as opposed to likelihood. The differences in AIC and BIC in the present study will be interpreted with respect to the thresholds established by Raftery (1995) for BIC and Burnham

and Anderson (2002) for AIC in which differences  $< 6$  for BIC and  $< 4$  for AIC are considered strong evidence that the full model is not the best fit for the data. This procedure will be done for peak alignment and f0 max analyses in the Discourse Completion Task and Interview Task.

The model selection criteria indicated that the condensed model is a better fit for the Discourse Completion Task concerning peak alignment. Table 21 contains the degrees of freedom (df) and AIC and BIC for each model. Although the AIC did not reveal a difference between the models, the BIC (the more conservative of the two estimates with respect to larger models and free parameters) shows a difference of 18 between models, which is strong evidence that the condensed model is the best fit.

| Model           | df | AIC    | BIC    |
|-----------------|----|--------|--------|
| Full model      | 17 | 857.57 | 930.56 |
| Condensed model | 13 | 857.03 | 912.85 |

*Table 21: AIC and BIC of peak alignment models for the Discourse Completion Task*

The condensed regression model performed on peak alignment in the Discourse Completion Task reached significance ( $F(10, 530) = 3.75, p < .001$ , adjusted  $R^2 = .048$ ) and contained two significant interactions: a two-way interaction between Region and Position as well as a three-way interaction between Region, Position, and Context. There were no significant simple effects in the model apart from these two interactions, which have the largest effect sizes ( $B$ ) in the model ( $-.792$  for the three-way interaction and  $.664$  for the two-way interaction). This suggests that for peak alignment in the Discourse Completion Task, the predictors Region, Context, and Position only affect the outcome as cross-over interactions; that is, they depend on each other to be significant predictors of peak alignment but are not significant predictors independently.

The significant interactions were further explored by performing pairwise comparisons between the factors of the highest order (i.e. three-way) interaction (Region, Position, and Context) and lower-order interaction (Region and Position)<sup>22</sup> with a Tukey HSD adjustment for multiple comparisons. The pairwise comparisons of Region, Position, and Context revealed that the Bilbao group produced significantly later peaks in phrase-medial position ( $M = .95$ ,  $SD = .05$ ) than in phrase-final position for the ML (Spanish) condition ( $M = .67$ ,  $SD = .06$ ),  $p = .04$ . This difference is shown in Figure 31, in which the x-axis shows peak alignment ( $< 1$  is an early peak,  $> 1$  is a delayed peak) and the y-axis shows Context and Position for the Bilbao group. An additional observation of interest is that phrase-medial EL (Basque) contexts did not contain instances of delayed peak alignment, although this is possible for Standard Basque in phrase-medial pre-nuclear positions. The exclusive use of early peak alignment in Standard Basque in this task will be further discussed in the following chapter.

---

<sup>22</sup> Lower-order interactions containing the same factors as higher-order interactions (for example, Region x Context versus Region x Context x Position) were only investigated when their results were not redundant with respect to the higher-order interaction

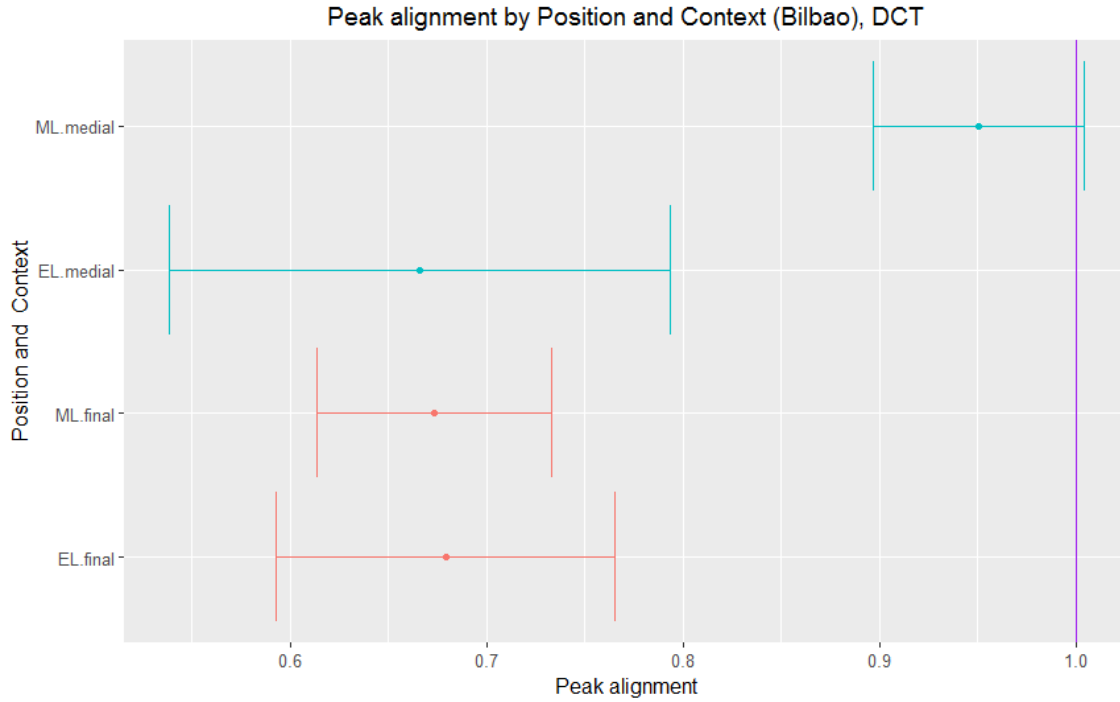


Figure 31: Peak alignment in the Discourse Completion Task, Bilbao group

The pairwise comparison of Region and Position<sup>23</sup> revealed that Lekeitio speakers produced significantly later peak alignment in phrase-medial position ( $M = 1.05$ ,  $SD = .11$ ) when compared to phrase-final position ( $M = .64$ ,  $SD = .08$ ),  $p = .005$ . Unlike the Bilbao group, this difference in position was not mediated by Context (ML or EL). Instead, Lekeitio speakers produced delayed peak alignment in phrase-medial positions in both Spanish (ML) and Basque (EL) contexts, whereas (unilingual) Northern Bizkaian Basque does not contain delayed peak alignment. The results seen in the Lekeitio group support the Matrix Language Hypothesis, which in the context of this study predicts that the peak alignment patterns of Matrix Language (in this case, the delayed peaks in Spanish) override that of the Embedded Language (early peaks in Basque). The difference in position in the Lekeitio group is shown in the plot in Figure 32.

<sup>23</sup> This two-way interaction was analyzed in addition to its involvement in higher-order interactions (i.e. Region, Context, and Position) due to the presence of non-redundant effects in the lower-order interaction that are not present in the higher-order interaction

However, this hypothesis is not supported in the Bilbao group, which showed no Matrix Language effects in the Embedded Language (Basque). This discrepancy between groups in Matrix Language effects could be due to language competency factors such as language dominance or usage, which will be revisited in more detail in the following chapter.

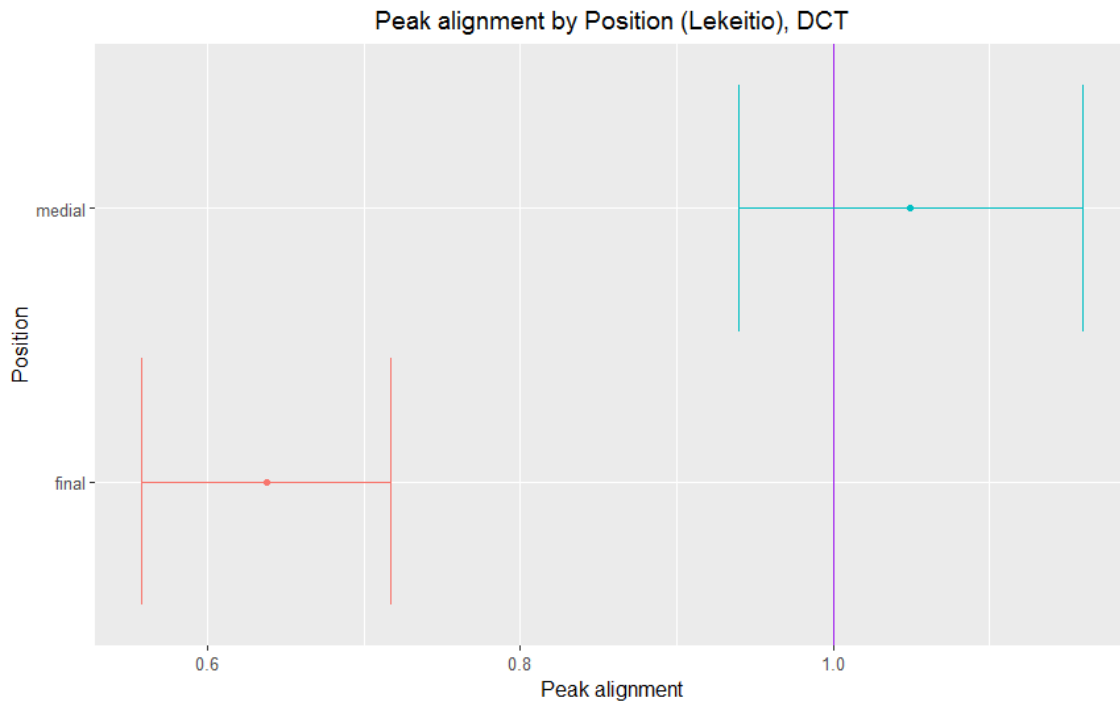


Figure 32: Peak alignment in the Discourse Completion Task, Lekeitio group

The regression coefficients, standard error, t-value, and significance levels can be found for each variable in Table 22.

| <i>Independent variables (predictors):</i>   |                 |                    |                 |
|----------------------------------------------|-----------------|--------------------|-----------------|
| <b>Simple effects</b>                        | <b><i>B</i></b> | <b><i>SE B</i></b> | <b><i>T</i></b> |
| Region:Lekeitio (reference: Bilbao)          | 0.002           | 0.158              | 0.018           |
| Context:ML (reference: EL)                   | -0.005          | 0.096              | -0.059          |
| Position:Medial (reference: Final)           | -0.013          | 0.148              | -0.088          |
| Language dominance (Semantic Verbal Fluency) | 0.002           | 0.002              | 1.254           |
| Language, social: Both (reference: Basque)   | 0.079           | 0.098              | 0.806           |
| Language, social: Spanish                    | -0.024          | 0.123              | -0.197          |

| <b>Interactions</b>                                   | <b>B</b>       | <b>SE B</b> | <b>T</b> |
|-------------------------------------------------------|----------------|-------------|----------|
| Region: Lekeitio x Context:ML                         | -0.083         | 0.169       | -0.489   |
| Region: Lekeitio x Position: Medial                   | 0.679          | 0.276       | 2.455*   |
| Context: ML x Position: Medial                        | 0.290          | 0.164       | 1.767    |
| Region: Lekeitio x Context:ML x Position: Medial      | -0.799         | 0.296       | -2.700** |
| Constant                                              | 0.449          | 0.233       | 1.928*   |
| Observations                                          | 541            |             |          |
| R <sup>2</sup>                                        | 0.066          |             |          |
| Adjusted R <sup>2</sup>                               | 0.048          |             |          |
| Residual Std. Error                                   | 0.528          |             |          |
|                                                       | (df = 530)     |             |          |
| F Statistic                                           | 3.746**        |             |          |
|                                                       | (df = 10; 530) |             |          |
| <i>Note: *p &lt; .05 **p &lt; .01 ***p &lt; 0.001</i> |                |             |          |

*Table 22: Regression results for peak alignment in the Discourse Completion Task*

#### 4.2.3 Interview Task: ML Basque, EL Spanish

The regression model performed on the Interview Task for outcome variable Peak Alignment contained predictor variables Region, Context, Position, Language Dominance (Semantic Verbal Fluency), Language Dominance (self-reported), and Language (social). The full model with all interaction terms specified was compared with a condensed model which only contained the significant interactions using AIC and BIC parameters. Both AIC and BIC revealed the condensed model to be a better fit with differences beyond the threshold of each criterion (< 4 for AIC, < 6 for BIC), which can be seen in Table 23.

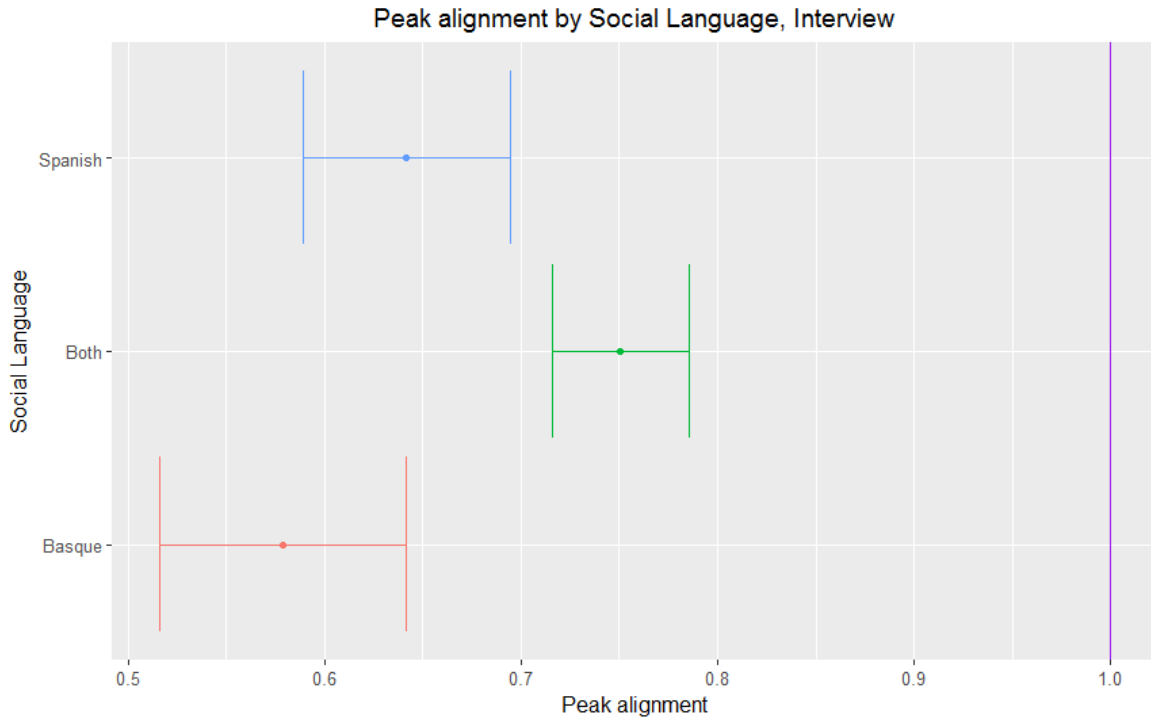
| Model           | df | AIC    | BIC    |
|-----------------|----|--------|--------|
| Full Model      | 17 | 444.03 | 515.72 |
| Condensed Model | 10 | 435.16 | 481.53 |

*Table 23: AIC and BIC of peak alignment models for the Interview Task*

The condensed model predicting peak alignment in the Interview Task reached significance, ( $F(8, 492) = 3.64, p < .001$ , adjusted  $R^2 = .04$ ) and revealed that Region, Context, Social

Language (Both), and Language Dominance (Semantic Verbal Fluency) were significant predictors of peak alignment. However, these all factors except Social Language were mediated by Context: the model contained a significant interaction between Region and Context as well as Language Dominance (Semantic Verbal Fluency) and Context. In the presence of these interactions, the simple effects will not be independently interpretable and will only be discussed with respect to their interactions.

Beginning with Social Language, participants who reported speaking Both Spanish and Basque socially ( $M = .75$ ,  $SE = .03$ ) produced significantly later peak alignment than participants who reported speaking Basque socially ( $M = .58$ ,  $SE = .05$ ), as shown in Figure 33. If speaking Basque socially (i.e. at home and with friends) is an indicator of language dominance (fluency or frequency-based), then the finding that those who speak Basque socially produce earlier peak alignment when Basque is the ML is expected. The finding that the dominant language is less susceptible to deviation from unilingual norms supports the present study's third hypothesis, as earlier peaks were maintained in the Interview task (ML Basque) by those who report speaking Basque in social domains. The general discussion of language dominance and its effects on  $f_0$  maximum and peak alignment will be discussed in more detail in the following chapter.



*Figure 33: Peak alignment by Social Language in Interview Task*

Continuing with the higher-order effects, the interaction between Region and Context was explored with a pairwise comparison between variables and a Tukey HSD adjustment. The pairwise comparisons revealed one between-group difference and two within-group differences. Beginning with the former (between-group), the Lekeitio group produced later peak alignment in ML (Basque) conditions when compared to the Bilbao group in ML conditions,  $p = .04$ . With respect to within-group differences, the Bilbao group produced significantly later peak alignment in EL (Spanish) conditions than ML (Basque) conditions,  $p = .04$  and the opposite effect occurred in the Lekeitio group, in which ML (Basque) conditions had later peak alignment than EL (Spanish) conditions,  $p = .01$ . Table 24 contains the means and standard deviations of peak alignment by Region and Context, followed by the plot in Figure 34, which provides a visual representation of these differences. Note that in the Interview Task, all peak alignment means are less than 1, indicating early peak alignment. Therefore, the differences in peak alignment by

Context and Dialect are not categorical (i.e. later peaks do not imply that they are *delayed* peaks).

| Region   | ML (Basque) | SE  | EL (Spanish) | SE  |
|----------|-------------|-----|--------------|-----|
| Lekeitio | .79         | .06 | .52          | .07 |
| Bilbao   | .57         | .04 | .75          | .05 |

Table 24: Peak alignment means by Region and Context in the Interview Task

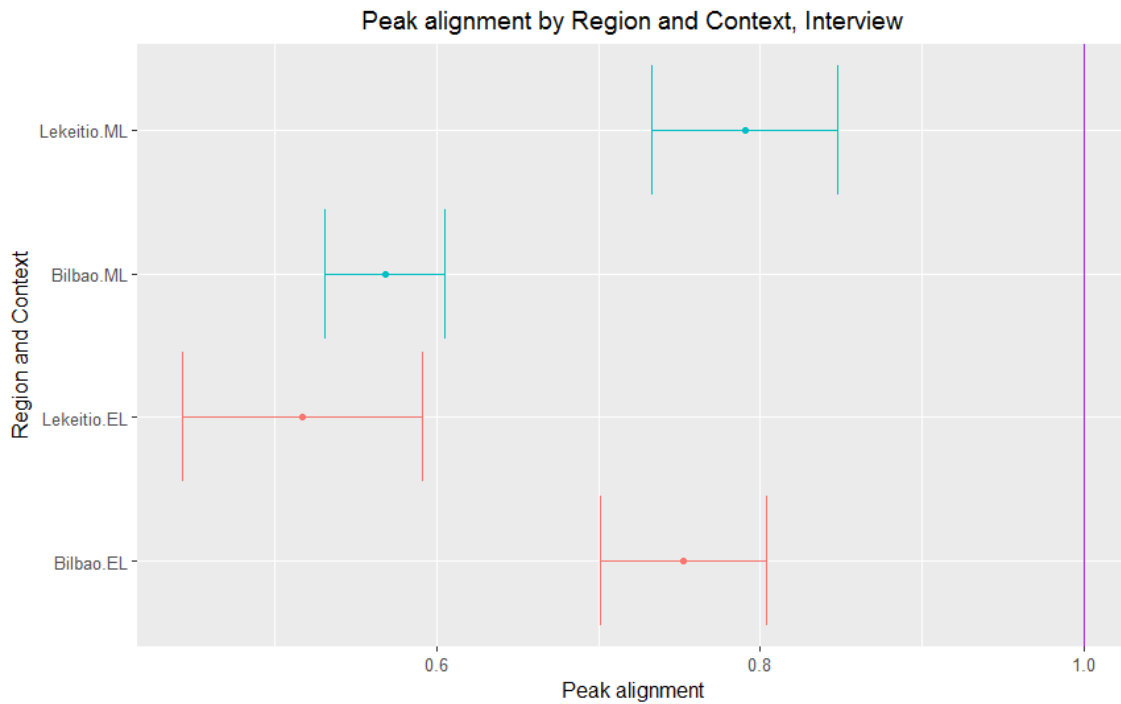
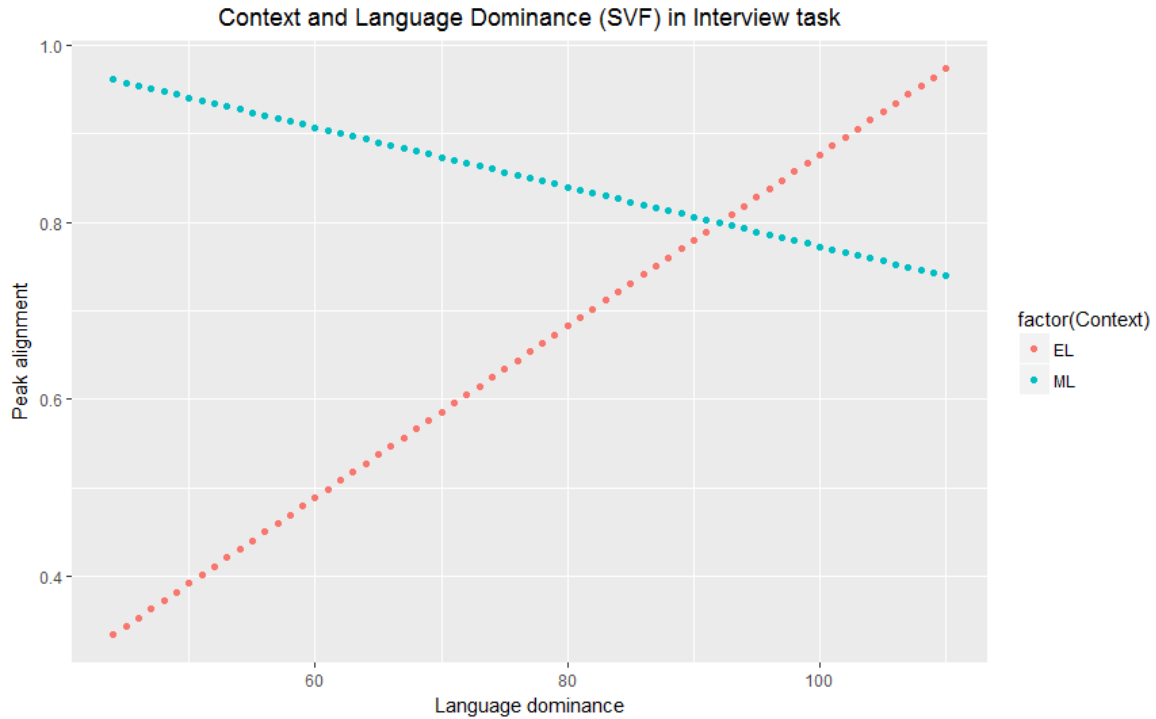


Figure 34: Peak alignment by Region and Context in the Interview Task

The presence of early peaks in both groups, contexts, and prosodic positions in the Interview Task provides support for the Matrix Language Hypothesis, as the peak alignment patterns of the Matrix Language (Basque) are surfacing in the Embedded Language (Spanish). However, the group-specific differences reveal two interesting findings. First, the Bilbao group still produced significantly later peak alignment in the Embedded Language (Spanish) than in the Matrix Language (Basque), which appears to be a truncated version of the expected unilingual peak alignment patterns (i.e. Spanish peaks later than Basque peaks, although not categorically

delayed). Second, the Lekeitio group showed the reverse pattern as Bilbao speakers, as they produced significantly later Matrix Language (Basque) peak alignment than Embedded Language (Spanish peak alignment). This later finding is the reverse of unilingual norms and the opposite of what is predicted by the Matrix Language hypothesis. Potential explanations for this finding in the Lekeitio group with respect to task type and sensitivity of the dominant language will be discussed in more detail in Chapter 5.

The interaction between Context and Language Dominance (Semantic Verbal Fluency) revealed that in EL (Spanish) contexts, Language Dominance (Semantic Verbal Fluency) was a positive predictor of peak alignment (i.e. predicting later peaks), with peak alignment increasing by 1 Hz for every unit increase in Language Dominance (Semantic Verbal Fluency). On the other hand, in ML (Basque) contexts, Language Dominance (Semantic Verbal Fluency) was a negative predictor of peak alignment, decreasing 1.3 Hz per unit increase in Language Dominance (Semantic Verbal Fluency). It is important to note that Language Dominance (Semantic Verbal Fluency) is a continuous predictor, with lower values indicating that a speaker is more Spanish-dominant and higher values indicating that a speaker is more Basque-dominant. This relationship between Context and Language Dominance (Semantic Verbal Fluency) is shown in Figure 35, which shows the predicted values of peak alignment based on the condensed regression model performed.



*Figure 35: Interaction between Context and Language Dominance (Semantic Verbal Fluency) in the Interview Task*

The interaction between Context and Language Dominance (Semantic Verbal Fluency) reveals that speakers who have lower Semantic Verbal Fluency (i.e. more Spanish-dominant speakers) produced earlier peak alignment in the Embedded Language (Spanish) than high Semantic Verbal Fluency (i.e. more Basque-dominant speakers). These results support both the third hypothesis posed by the current study, which predicted that a speaker's dominant language is more susceptible to deviation from unilingual norms when it is the Embedded Language. In this case, Spanish (the Embedded Language) deviated from unilingual norms for more Spanish-dominant speakers than for more Basque-dominant speakers, as peak alignment was significantly earlier for the Spanish-dominant speakers. Additionally, more Spanish-dominant speakers produced later Matrix Language (Basque) peak alignment than more Basque-dominant speakers,

which is not predicted by the hypotheses of this study. Issues related to language dominance and the Matrix Language Framework will be revisited in more depth in the next chapter.

The regression coefficients, standard error, t-value, and significance levels for this regression can be found for each variable in Table 25.

| <i>Independent variables (predictors):</i>                 |                           |                    |                 |
|------------------------------------------------------------|---------------------------|--------------------|-----------------|
| <b>Simple effects</b>                                      | <b><i>B</i></b>           | <b><i>SE B</i></b> | <b><i>T</i></b> |
| Region:Lekeitio (reference: Bilbao)                        | -0.234                    | 0.102              | -2.309*         |
| Context:ML (reference: EL)                                 | 0.741                     | 0.201              | 3.683***        |
| Position:Medial (reference: Final)                         | -0.046                    | 0.036              | -1.271          |
| Language dominance (Semantic Verbal Fluency)               | 0.010                     | 0.003              | 3.304**         |
| Language, social: Both (reference: Basque)                 | 0.172                     | 0.085              | 2.037*          |
| Language, social: Spanish (reference: Basque)              | 0.063                     | 0.099              | 0.639           |
| <b>Interactions</b>                                        |                           |                    |                 |
| Region: Lekeitio x Context:ML                              | 0.457                     | 0.124              | 3.692***        |
| Context: ML x Language dominance (Semantic Verbal Fluency) | -0.013                    | 0.003              | -4.061***       |
| Constant                                                   | 0.0007                    | 0.233              | 0.003           |
| Observations                                               | 501                       |                    |                 |
| R <sup>2</sup>                                             | 0.056                     |                    |                 |
| Adjusted R <sup>2</sup>                                    | 0.041                     |                    |                 |
| Residual Std. Error                                        | 0.559<br>(df = 492)       |                    |                 |
| F Statistic                                                | 3.642***<br>(df = 8; 492) |                    |                 |
| <i>Note: * p &lt; .05 ** p &lt; .01 *** p&lt;0.001</i>     |                           |                    |                 |

*Table 25: Regression results for peak alignment in the Interview Task*

#### 4.2.4 Summary of peak alignment results

In the Reading Task, which contained only Spanish language data, the statistical analyses revealed significantly later peak alignment in phrase-medial position when compared to phrase-final position for both Bilbao and Lekeitio speakers. Although anticipated for Bilbao speakers, who speak Standard Peninsular Spanish, this finding was unexpected for Lekeitio speakers, as

previous account for Lekeitio Spanish have reported early peak alignment only (Elordieta & Irurtzun (2016)). Additionally, a marginal interaction between Region and Position was found which showed that Bilbao speakers produced a greater difference in peak alignment between phrase-medial and final positions when compared to Lekeitio speakers. Due to the presence of delayed peak alignment in Lekeitio speakers, the hypothesis concerning peak alignment patterns (see Table 20) was revised to predict that Lekeitio speakers (as well as Bilbao speakers) would show evidence of Matrix Language effects if distinctions in positional (medial versus final) peak alignment were not maintained when speaking Spanish or were extended to Basque.

In the Discourse Completion Task (ML Spanish, EL Basque), there was a significant interaction between Region, Context, and (prosodic) position in which Bilbao speakers had significantly later peak alignment in phrase-medial position in ML (Spanish) contexts when compared to phrase-final ML (Spanish) contexts. There was also an interaction between Region and Position in which Lekeitio speakers showed significantly later peak alignment phrase-medially (compare to phrase-finally), but this difference did not interact with Context, as seen in the Bilbao group.

Finally, the Interview Task revealed that Social Language was a significant predictor of peak alignment, with those who reported speaking Basque socially producing significantly earlier peak alignment than those who reported speaking both Spanish and Basque socially. Additionally, there were significant interactions between Region and Context as well as Language Dominance (Semantic Verbal Fluency) and Context. Beginning with Region and Context, Lekeitio speakers produced significantly later peak alignment in ML (Basque) contexts when compared to their own EL (Spanish) peak alignment and the Bilbao group's ML peak

alignment. The Bilbao speakers showed the reverse within-group effect; they produced significantly later peak alignment in EL (Spanish) conditions when compared to their own ML (Basque) peak alignment. The interaction between Language Dominance (Semantic Verbal Fluency) and Context revealed that higher Semantic Verbal Fluency scores (i.e. more Basque-dominant) predicted later peaks in EL (Spanish) conditions and earlier peaks in ML (Basque) conditions. Further interpretation and the implications of these findings will be discussed in the following chapter, after the results for the f0 max regression models are presented in Section 4.3.

### 4.3 F0 max

#### 4.3.1 Monolingual Spanish baseline (Reading Task)

The regression model performed on f0 max in the Reading Task reached significance ( $F(3, 301) = 25.822, p < .001$ , adjusted  $R^2 = .20$ ) and contained one positive predictor of f0 max (Position:Medial). Phrase-medial f0 was significantly higher f0 (Hz) values ( $M = 243, SE = 1.69$ ) than the reference group (final;  $M = 203, SE = 4.14$ ). This result is expected, as gradual declination of f0 is typical in utterances (Cohen et al (1982), Ladd (1984)). Neither Region nor an interaction between Region, Context, and Position were significant predictors of f0 max in the Reading Task.

Figure 36 shows a plot of f0 max in the Reading Task by Position in which the x axis represents Position (medial or final) and the y-axis is f0 (Hz), visually demonstrating the significant difference in f0 between phrase-medial and final positions.

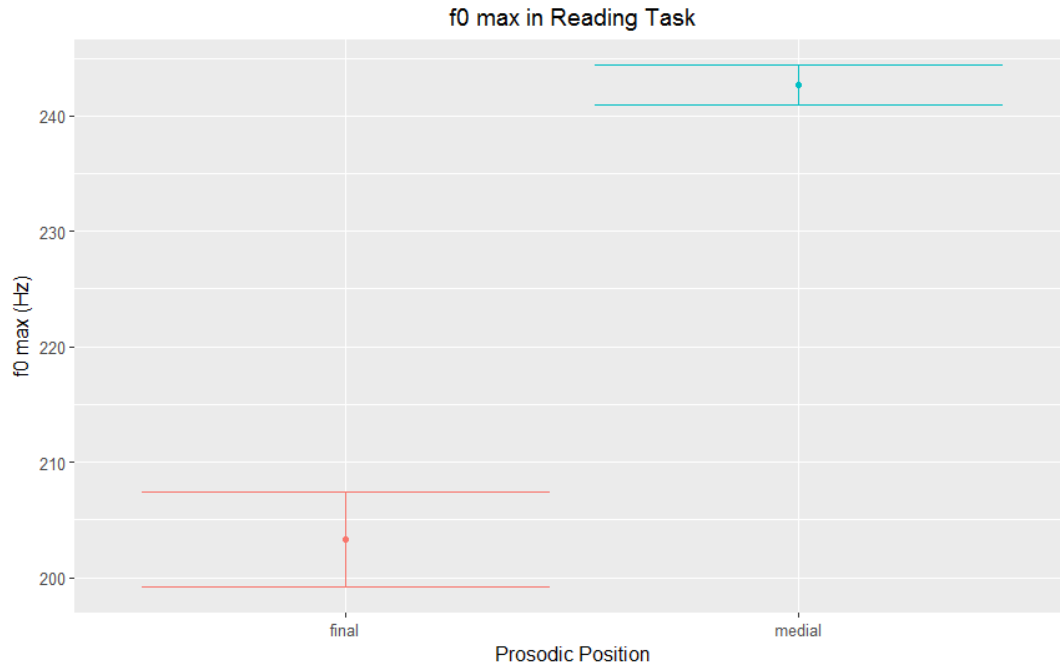


Figure 36: *f0 max in Reading Task*

Table 26 contains the regression coefficients, standard error, t-value, and significance levels for the model investigating f0 max in the Reading Task.

| <i>Independent variables (predictors):</i>             |                            |                    |                 |
|--------------------------------------------------------|----------------------------|--------------------|-----------------|
| <b>Simple effects</b>                                  | <b><i>B</i></b>            | <b><i>SE B</i></b> | <b><i>T</i></b> |
| Region:Lekeitio (reference: Bilbao)                    | 16.179                     | 8.287              | 1.952           |
| Position:Medial (reference: Final)                     | 47.135                     | 6.982              | 6.751**         |
| <b>Interactions</b>                                    |                            |                    |                 |
| Region: Lekeitio x Position: Medial                    | -15.535                    | 8.948              | -1.736          |
| Constant                                               | 195.211                    | 6.580              | 29.668**        |
| Observations                                           | 305                        |                    |                 |
| R <sup>2</sup>                                         | 0.205                      |                    |                 |
| Adjusted R <sup>2</sup>                                | 0.197                      |                    |                 |
| Residual Std. Error                                    | 27.129 (df = 301)          |                    |                 |
| F Statistic                                            | 25.822***<br>(df = 3; 301) |                    |                 |
| <i>Note: * p &lt; .05 ** p &lt; .01 *** p&lt;0.001</i> |                            |                    |                 |

Table 26: *Regression results for f0 max in Reading Task*

#### 4.3.2 Discourse Completion Task (Discourse Completion Task): ML Spanish, EL Basque

The regression models used to analyze f0 max in the Discourse Completion Task contained the same predictor variables and interaction terms as the peak alignment models for the Discourse Completion Task (Region, Position, Context, Language Dominance (Semantic Verbal Fluency), Language (social); interactions with Context and Context, Region, and Position). Model comparison between the full and condensed models (following the procedure outlined in the sections on peak alignment in the Discourse Completion Task and Interview Task) revealed that the condensed model was a better fit for the data set, with strong evidence for the condensed model from both the AIC and the BIC differences. AIC and BIC for the Discourse Completion Task models predicting f0 max can be found in Table 27.

| Model           | df | AIC     | BIC     |
|-----------------|----|---------|---------|
| Full Model      | 17 | 5189.22 | 5261.83 |
| Condensed Model | 13 | 5184.17 | 5239.70 |

*Table 27: AIC and BIC of f0 max models for the Discourse Completion Task*

The condensed regression model reached significance ( $F(10, 518) = 3.75, p < .001$ , adjusted  $R^2 = .049$ ) and contained one significant predictor of f0 max (Region) and two significant interactions (Region and Context; Region, Context, and Position). Due to their involvements in higher-order interactions, Region, Context, and Position will only be interpreted with respect to their 3-way interaction.

The three-way interaction between Region, Context, and Position was investigated using pairwise comparisons and a Tukey HSD adjustment. The pairwise comparisons revealed that Lekeitio speakers produced significantly higher f0 max in phrase-final EL (Basque) contexts ( $M = 262, SE = 7.98$ ) than in phrase-final ML (Spanish) contexts ( $M = 235, SE = 4.96$ ),  $p = .04$ , as

shown in Figure 37. This finding supports the Hyperarticulation hypothesis, which states that code-switched (EL) contexts are produced with higher  $f_0$  than non-switched (ML) contexts. However, this effect was not seen in phrase-medial prosodic positions for Lekeitio speakers nor in the Bilbao group for any context or position. This partial support for Hyperarticulation, similar to the peak alignment results seen, may be mediated by factors such as language dominance, which will be discussed in more detail in the following chapter.

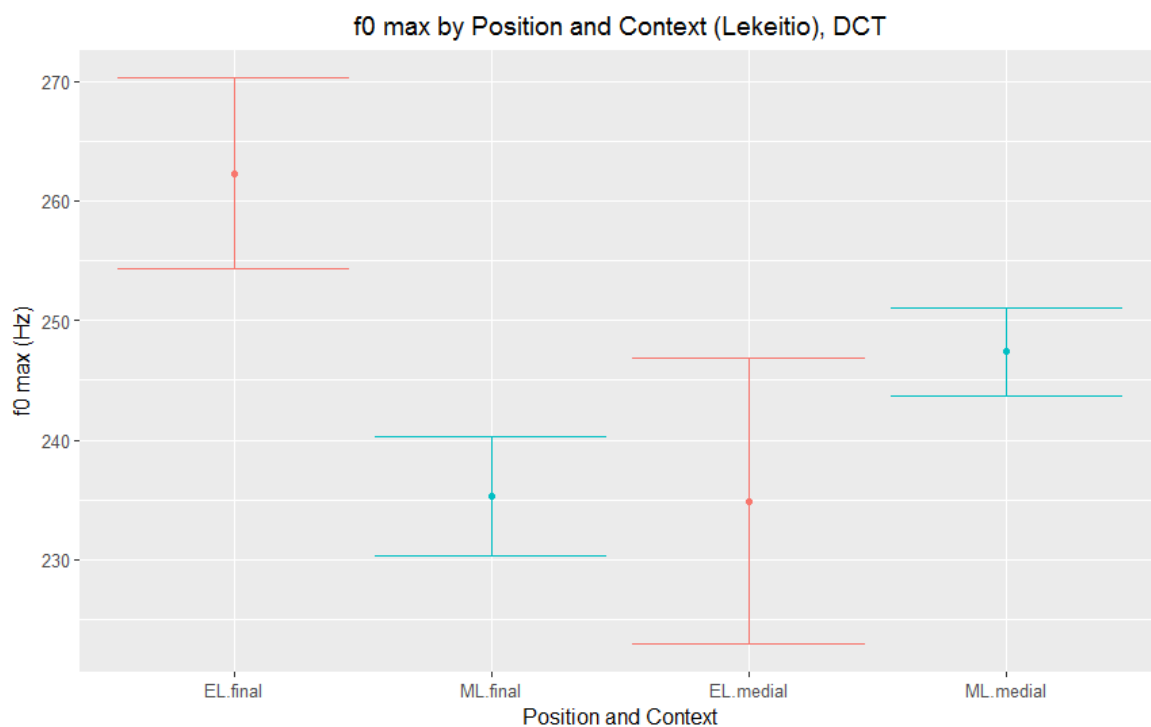


Figure 37:  $f_0$  max for Lekeitio speakers by Position and Context in Discourse Completion Task

Table 28 contains the regression coefficients, standard error, and t-values for all variables in the model predicting  $f_0$  max in the Discourse Completion Task.

| <i>Independent variables (predictors):</i>             |                            |                    |                 |
|--------------------------------------------------------|----------------------------|--------------------|-----------------|
| <b>Simple effects</b>                                  | <b><i>B</i></b>            | <b><i>SE B</i></b> | <b><i>T</i></b> |
| Region:Lekeitio (reference: Bilbao)                    | 24.260                     | 10.027             | 2.419*          |
| Context:ML (reference: EL)                             | 1.974                      | 6.112              | 0.323           |
| Position:Medial (reference: Final)                     | -6.053                     | 9.313              | -0.650          |
| Language dominance (Semantic Verbal Fluency)           | -0.611                     | 0.147              | 0.751           |
| Language, social: Both (reference: Basque)             | 9.556                      | 6.238              | 1.532           |
| Language, social: Spanish (reference: Basque)          | -2.934                     | 7.759              | -.0378          |
| <b>Interactions</b>                                    |                            |                    |                 |
| Region: Lekeitio x Context:ML                          | -28.906                    | 10.641             | -2.716**        |
| Region: Lekeitio x Position: Medial                    | -21.338                    | 16.715             | -1.277          |
| Context: ML x Position: Medial                         | 2.931                      | 10.290             | 0.285           |
| Region: Lekeitio x Context:ML x Position: Medial       | 36.524                     | 18.017             | 2.027*          |
| Constant                                               | 227.797                    | 14.789             | 15.403**        |
| Observations                                           | 529                        |                    |                 |
| R <sup>2</sup>                                         | 0.067                      |                    |                 |
| Adjusted R <sup>2</sup>                                | 0.049                      |                    |                 |
| Residual Std. Error                                    | 32.93 (df = 518)           |                    |                 |
| F Statistic                                            | 3.747***<br>(df = 10; 518) |                    |                 |
| <i>Note: * p &lt; .05 ** p &lt; .01 *** p&lt;0.001</i> |                            |                    |                 |

*Table 28: Regression results for f0 max in Discourse Completion Task*

#### 4.3.3 Interview Task: ML Basque, EL Spanish

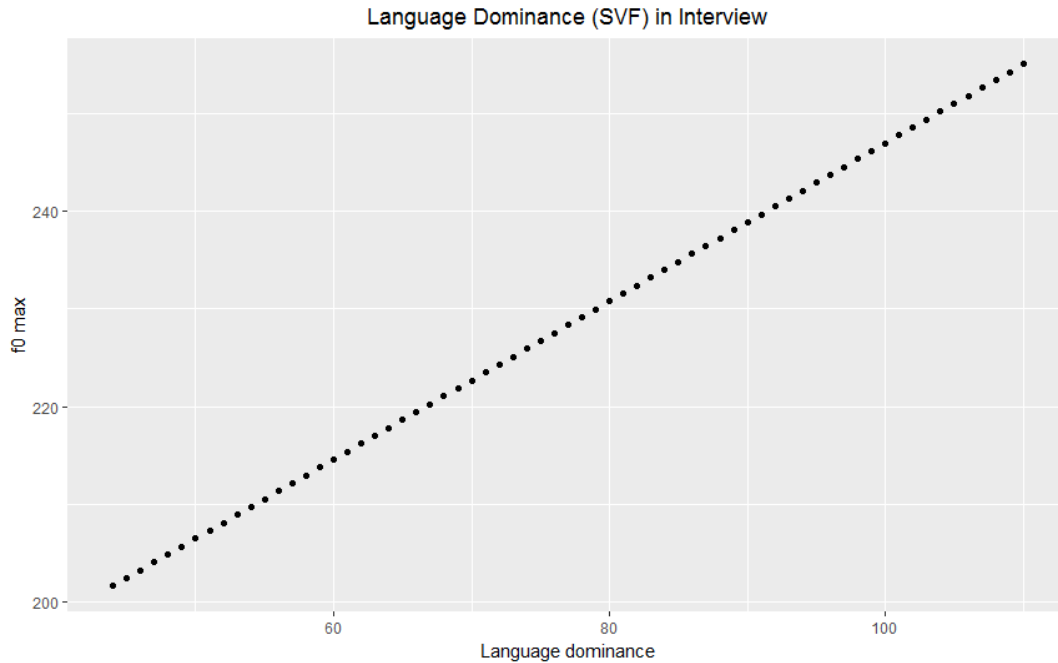
The full regression model investigating predictors of f0 in the Interview Task contained the same predictor variables as the model for peak alignment in the Interview (Region, Context, Position, Language Dominance (self), Language Dominance (Semantic Verbal Fluency), and Language (social)). This full model, which specified interaction terms for all predictors and Context as well as Region, Context, and Position was compared with a condensed model which eliminated all non-significant interactions. The AIC and BIC comparing these models both provided strong evidence that the condensed model was a better fit for the data, as seen in Table 29.

| Model           | df | AIC     | BIC     |
|-----------------|----|---------|---------|
| Full Model      | 17 | 4537.25 | 4609.10 |
| Condensed Model | 9  | 4533.12 | 4571.16 |

*Table 29: AIC and BIC comparing f0 max models for the Interview Task*

The condensed model predicting f0 max in the Interview Task reached significance ( $F(6, 499) = 14.11, p < .001$ , adjusted  $R^2 = .13$ ) and contained three significant predictors of higher f0 in the Interview: Language Dominance (Semantic Verbal Fluency), Social Language (Spanish), and Social Language (Both).

Starting with Language Dominance (Semantic Verbal Fluency), for every unit of increase in Semantic Verbal Fluency score, f0 max increases by 1 Hz. Therefore, a higher Semantic Verbal Fluency (i.e. more Basque-dominant) predicts higher f0 whereas a lower Semantic Verbal Fluency (i.e. more Spanish-dominant) predicts lower f0. This effect is demonstrated visually in Figure 38, where predicted f0 max (from the condensed regression model) is shown on the y axis and Semantic Verbal Fluency score (range of 44-110) is shown on the x axis.



*Figure 38:  $f_0$  by Language Dominance (Semantic Verbal Fluency) in the Interview Task*

The last significant predictor of  $f_0$  in the Interview Task is Language (social). Speakers who reported speaking Spanish socially (at home or with friends) produced significantly higher  $f_0$  max ( $M = 237$ ,  $SE = 3.02$ ) when compared to those who reported speaking Basque socially ( $M = 203$ ,  $SE = 3.54$ ). Additionally, those who reported speaking both Spanish and Basque socially had significantly higher  $f_0$  max ( $M = 224$ ,  $SE = 1.98$ ) than those who reported speaking Basque socially. There were no significant differences between those who speak Spanish or Both (languages) socially. These differences are plotted in Figure 39, with social language on the x axis and  $f_0$  max on the y axis.

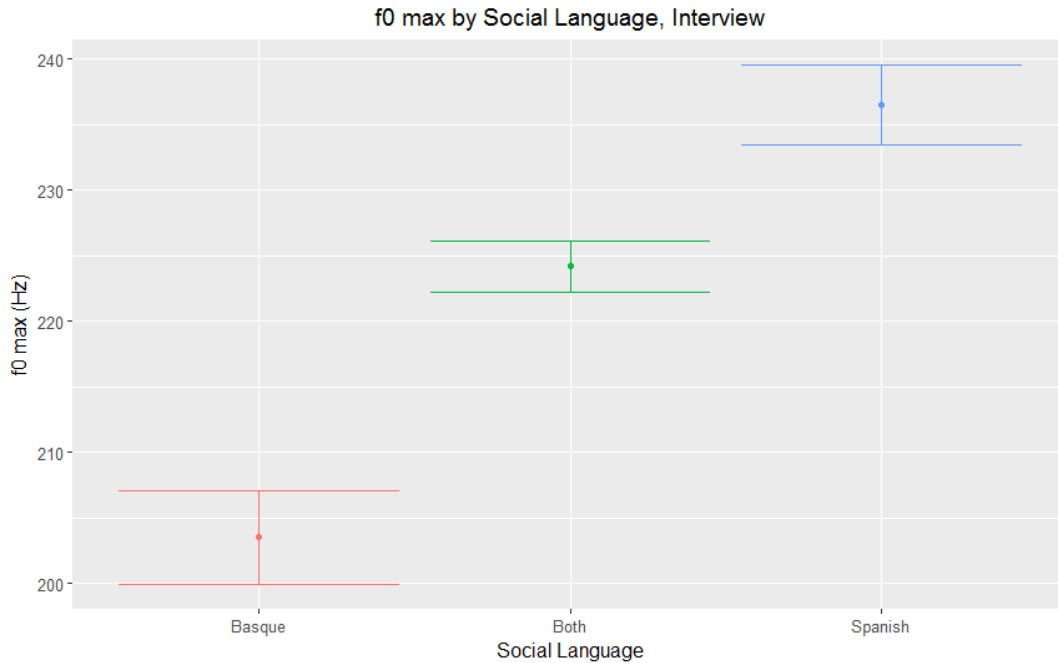


Figure 39: *f0 max by Social Language in the Interview Task*

Table 30 contains regression coefficients, standard error, and t-values for all variables in the model predicting f0 in the Interview Task.

| <i>Independent variables (predictors):</i>               |                        |                    |                 |
|----------------------------------------------------------|------------------------|--------------------|-----------------|
| <b>Simple effects</b>                                    | <b><i>B</i></b>        | <b><i>SE B</i></b> | <b><i>T</i></b> |
| Region:Lekeitio (reference: Bilbao)                      | -2.17                  | 3.878              | -.558           |
| Context:ML (reference: EL)                               | -0.765                 | 2.340              | -0.327          |
| Position:Medial (reference: Final)                       | 1.235                  | 2.096              | 0.589           |
| Language dominance (Semantic Verbal Fluency)             | .810                   | 0.096              | 8.408***        |
| Language, social: Both (reference: Basque)               | 20.679                 | 4.756              | 4.348***        |
| Language, social: Spanish (reference: Basque)            | 33.006                 | 5.590              | 5.905***        |
| Constant                                                 | 146.918                | 9.856              | 14.907***       |
| Observations                                             | 506                    |                    |                 |
| R <sup>2</sup>                                           | 0.145                  |                    |                 |
| Adjusted R <sup>2</sup>                                  | 0.135                  |                    |                 |
| Residual Std. Error                                      | 21.61 (df = 499)       |                    |                 |
| F Statistic                                              | 14.11*** (df = 6; 499) |                    |                 |
| <i>Note: * p &lt; .05 ** p &lt; .01 *** p &lt; 0.001</i> |                        |                    |                 |

Table 30: *Regression results for f0 max in the Interview Task*

#### 4.3.4 Summary of f0 max results

In the Reading Task (monolingual Spanish), phrase-medial positions contained significantly higher f0 max than phrase-final positions. Continuing to the Discourse Completion Task, there was a significant three-way interaction between Region, Context, and Position. Pairwise comparisons revealed that Lekeitio speakers produced significantly higher f0 in phrase-final EL (Basque) conditions than in phrase-final ML (Spanish) context. There were no significant f0 max differences for Bilbao speakers in the Discourse Completion Task.

In the Interview Task, both Language Dominance (Semantic Verbal Fluency), and Social Language were significant predictors of f0 maximum. Starting with Language Dominance (Semantic Verbal Fluency), a higher Semantic Verbal Fluency score (i.e. more Basque-dominant) predicted higher f0 max. Finally, those who reported speaking Spanish or both Spanish and Basque socially produced significantly higher f0 than those who reported speaking Basque socially.

The next chapter will interpret these results in more depth with respect to previous studies and linguistic theory, including implications for bilingualism and code-switching.

## 5 INTERPRETATION OF RESULTS AND CONCLUSION

### 5.1 Hypotheses addressed

In this section, the hypotheses posed in Chapter 2 will be re-addressed with respect to the results reported in Chapter 4. Following this discussion, the implications of the results found in this dissertation as well as ideas for future research will be considered before concluding the study.

The first hypothesis of this dissertation extended the Matrix Language Hypothesis (Myers-Scotton 1993) from morphosyntax to prosody. The current study hypothesized that in the presence of a language-specific prosodic difference, the pattern of the Matrix Language (ML) would override that of the Embedded Language (EL). In order to test the claims of the Matrix Language Framework on prosodic differences, differences in f0 peak alignment in stressed syllables between Spanish and Basque were considered. To reiterate, Spanish and Basque differ in the alignment of the f0 peak with respect to the stressed syllable. Early peaks (f0 peak aligned within the stressed syllable) are attested in Standard and Northern Bizkaian Basque, Lekeitio Spanish, and phrase-finally in Peninsular Spanish, whereas delayed peaks (f0 peak aligned after the stressed syllable) are attested in phrase-medially in Peninsular Spanish and as a less-common pre-nuclear (i.e. not pre-verbal) variant in Standard Basque (Estebas Vilaplana & Prieto (2008), Elordieta & Hualde (2015), Elordieta & Irurtzun (2016)). Therefore, peak timing distinctions between phrase-medial and phrase-final positions are of particular interest to this hypothesis.

However, the previous findings on peak alignment in Peninsular and Lekeitio Spanish were not corroborated in the present study. The results of the Reading Task (unilingual Spanish) revealed that members of both the Peninsular (Bilbao) and Lekeitio Spanish dialect groups distinguished between phrase-medial and phrase-final positions, with significantly later (delayed)

peaks in phrase-medial position (see Figure 29). This finding was unexpected for Lekeitio speakers, who were anticipated to have early peaks in phrase-medial and final positions, per the results of Elordieta & Irurtzun (2016). As Elordieta & Irurtzun (2016) also used read speech in their study, it is unlikely that the results found in the present study are due to task type. However, it may be possible that interactions between participants and the current author, who is not a speaker of Lekeitio Spanish, may have resulted in dialectal leveling during this portion of the study.

After establishing that both groups produce delayed phrase-medial peaks when speaking in unilingual Spanish contexts, the present study hypothesized that when Spanish is the ML, all phrase-medial EL (Basque) peaks will be delayed (even in pre-verbal positions), whereas when Basque is the ML, all EL (Spanish) peaks will be early. The hypothesis that the peak alignment patterns of the ML would override those of the EL was supported in both tasks, but not unanimously. In the Discourse Completion Task (ML Spanish, EL Basque), the Lekeitio group produced significantly later (delayed) peaks in phrase-medial position than in phrase-final position, regardless of context (ML or EL; see Figure 32), supporting the Matrix Hypothesis.

If the Matrix Language Framework hypothesis were not supported in this context, we would expect no difference in peak alignment by prosodic position in the EL contexts, as Northern Bizkaian Basque has early peaks in both prosodic positions. The Bilbao group had significantly later phrase-medial peaks *only* in ML (Spanish) contexts (Figure 32), suggesting that Bilbao speakers did not show ML (Spanish) effects on the EL (Basque) in the Discourse Completion Task. In the EL (Basque) contexts, Bilbao speakers produced early peaks in both prosodic positions. This asymmetry seen between groups with respect to peak alignment and language context (ML or EL) may be a result of language dominance: if the Lekeitio speakers are Basque-

dominant, switching *into* the dominant language from a non-dominant language may have higher “costs” (as seen with reaction times in Meuter & Allport (1999) and for VOT duration in Olson (2016c)) which may have made the EL more vulnerable to deviation from unilingual norms (i.e. early peaks).

When the three Lekeitio speakers (Speakers, 6, 7, 8) are considered with respect to language dominance, all three speakers had mid to high Semantic Verbal Fluency scores, indicating that they are less Spanish-dominant (than those who had lower Semantic Verbal Fluency scores). Additionally, Lekeitio is a Basque-dominant town in which Basque is spoken in informal contexts (such as shops, on the street, with strangers, with friends) more often than Spanish. The peak alignment results from the Discourse Completion Task suggest that ML effects may also be sensitive to language dominance, which will be discussed in more detail when the third research question is addressed.

In the Interview task (ML Basque, EL Spanish), both groups produced early peak alignment, regardless of context (ML or EL) or prosodic position (Figure 34). No instances of delayed peak alignment surfaced in this task, which is anticipated in unilingual conditions for Standard Basque in pre-nuclear positions (as an optional variant along with early peak alignment) and for Peninsular Spanish in phrase-medial positions. Unlike the Discourse Completion Task, both groups (Lekeitio and Bilbao) showed evidence of ML effects. However, there were still statistically significant phonetic peak alignment differences present in the Interview, despite the absence of categorically delayed peaks (i.e. means for delayed peaks) in either group. Beginning with the Bilbao group, EL (Spanish) peaks were significantly later (i.e. not categorically delayed, but with f0 peak realized later in the stressed syllable) than ML (Basque) peaks. This difference may be indicative of a truncated distinction between Spanish and Basque peak alignment, similar

to “compromised” or gradient values seen in code-switching contexts for VOT in studies such as Olson (2013) and for f0 in studies such as Piccinini & Garellek (2014) and Olson (2016b). An additional example of phonetic compromise during code-switching is the absence of delayed peak alignment seen in Standard Basque (i.e., Bilbao speakers) in the present study. In both the Discourse Completion Task and the Interview Task, only early peak alignment values were attested for Standard Basque speakers, which may indicate leveling of a non-phonological distinction as a strategy to maximally distinguish Spanish and Basque during code-switching contexts.

In the Discourse Completion Task, Bilbao speakers did not show ML effects and produced peak alignment consistent with unilingual norms for both Peninsular Spanish and Standard Basque, but this distinction was not maintained in the Interview. This asymmetry between tasks may be mediated by language dominance, as hypothesized for Lekeitio speakers in the Discourse Completion Task. For example, three of the four Bilbao speakers (Speakers 1, 3, 4) have the lowest Semantic Verbal Fluency scores, indicating that they are more Spanish-dominant. As previously mentioned, Bilbao is a Spanish-dominant city, with Basque typically spoken only in formal contexts or within native Basque-speaking families. Continuing with the hypothesis that switching into the dominant language has a higher cost than the reverse context, this may have affected the Bilbao group’s ability to maintain the peak alignment distinction between Spanish and Basque when Spanish was the EL.

There were also several significant differences in peak alignment in the Interview that did not support the Matrix Language Framework hypothesis. Namely, in the Interview, the Lekeitio group had significantly (phonetically) later peak alignment in ML (Basque) contexts than in EL (Spanish) contexts. Based on their peak alignment patterns in the Discourse Completion Task,

which showed evidence of ML effects, it is unexpected that the Interview would show evidence of EL effects (i.e. later peaks in the ML instead of the EL). These results are not consistent with unilingual norms for either Basque or Spanish (i.e. their Spanish baseline from the Reading Task) nor are they consistent with the language dominance hypothesis, as switching into the non-dominant language should be less costly than the reverse context. Lekeitio speakers also had significantly later ML (Basque) peaks than Bilbao speakers. The results from the Discourse Completion Task and Interview reveal that the Lekeitio group is consistently more sensitive in the code-switching contexts (i.e. unable to maintain or approximate unilingual peak alignment norms) created in the present study than are the Bilbao speakers.

One possible explanation for the Lekeitio group's sensitivity to these code-switching contexts is the order in which the tasks were presented. The study began with the current author conversing with the participants in Spanish (the current author has limited Basque proficiency) to administer the informed consent process and demographic questionnaire. The Discourse Completion Task (ML Spanish) and Interview (EL Basque) were administered by the Interviewer, who only spoke Basque to participants during the Interview task. This progression of tasks resulted in an overall Spanish-dominant interaction between participant, author, and Interviewer which may have been inherently costlier for Basque-dominant participants from Lekeitio. For the Spanish-dominant speakers from Bilbao, their dominant language was the embedded language only in the last task whereas Basque-dominant speakers were required to inhibit their dominant language until the final (Interview) task. If there are higher cognitive costs (and therefore, language-specific production costs) associated with switching into the dominant language, the progression of tasks in the current study would be more taxing on Basque-

dominant than Spanish-dominant individuals and may be a source of the sensitivity seen in Lekeitio speakers.

The current study's second hypothesis predicted that code-switched (i.e., Embedded Language contexts) would be produced with higher  $f_0$  than non-switched (i.e. Matrix Language contexts), based on studies that provide evidence that code-switched words may be realized with narrow focus (Olson (2012, 2016b) or perceived as more prominent than non-switched words (Fricke et al (2016)). This hypothesis was tested on the Discourse Completion Task (Matrix Language (ML) Spanish, Embedded Language (EL) Basque) and the Interview task (ML Basque, EL Spanish) and was only partially supported by the results. The only context in which code-switched words had higher  $f_0$  maxima than non-switched words was found in the Discourse Completion Task. In the Discourse Completion Task, Lekeitio speakers produced phrase-final code-switches (EL Basque) with significantly higher  $f_0$  maxima than phrase-final ML (i.e. non-switched) words (Figure 37). However, there were no differences in  $f_0$  maxima between the Lekeitio group's realizations of code-switched (EL) and non-switched (ML) realizations in phrase-medial position. Stronger support for the Hyperarticulation hypothesis would have included broader domains of  $f_0$  differences between code-switched and non-switched contexts, including phrase-medial positions and the presence of  $f_0$  differences between ML and EL contexts in the Bilbao group.

There are several possible reasons why the present study's results did not strongly corroborate those of Olson (2012, 2016b) and Fricke et al (2016), which showed evidence of hyperarticulated or focused code-switches. The first of these reasons is the type of data elicited: in Olson's studies, laboratory speech (contextualized reading passages containing the code-switched words) are used, which differs greatly from the semi-spontaneous nature of the

Discourse Completion Task. Olson's paradigm provides more controlled, consistent data, which the Discourse Completion Task in the present study lacks, but also does not reflect the nature of naturalistic code-switching. Fricke et al (2016) used a semi-spontaneous corpus which contains un-elicited code-switches. However, their measures of prosodic prominence were perceptual. It is possible that the perception of prominence on code-switches and the acoustic cues that typically indicate prominence (such as higher f0 range and longer duration) do not correspond in their data. The results in Frick et al (2016) were also mediated by syntactic category, in which code-switched nouns and noun phrases were perceived as more prominent than other categories, suggesting that there may be more cues to prominence in code-switching other than the prosodic cues (such as f0) considered in the present study. Given the differences in results, task type, and methods of analysis used in the present study, Olson (2012, 2016b) and Fricke et al (2016), future work that measures both production and perception of code-switches as well as non-acoustic cues to prominence within the same corpus could help consolidate these gaps.

The third and final hypothesis anticipated that a speaker's dominant language would be likely to deviate from unilingual norms when it is the Embedded Language, based on previous studies (such as Olson (2016a, 2016b), Aly (in press)). These studies also found asymmetrical results between groups differing in language dominance or proficiency. The effects of language dominance and usage variables on the realization of code-switched contexts will be discussed with respect to f0 maximum first and followed by peak alignment. Language dominance (Semantic Verbal Fluency score) and Social Language were significant predictors of higher f0 in the Interview task, although there was no interaction with Context (ML or EL). This refers to higher f0 across ML and EL contexts for the results discussed. In the Discourse Completion Task

(ML Spanish, EL Basque), there were no language dominance variables which reached statistical significance for either outcome variable.

In the Interview task (ML Basque, EL Spanish), a higher Semantic Verbal Fluency score (i.e. less Spanish-dominant) significantly predicts higher  $f_0$  (Figure 38). This is additional evidence against the Olson's Hyperarticulation account, which hypothesizes that a speaker's dominant language is realized with higher  $f_0$  when it is the EL. Support for the Hyperarticulation account would have been the opposite result, in which a *lower* SVF was a significant predictor of  $f_0$  max. The presence of higher global  $f_0$  in a speaker's dominant language is reported for L1-dominant Russian-English bilinguals in Altenberg & Ferrand (2006) and for Mandarin-English and Cantonese-English bilinguals in Lee & Sittis (2017). Additionally, results of Lee & Sittis (2017) were mediated by task type for each group (reading or monologue-style task), suggesting that language dominance and speech style are both factors in  $f_0$  range. In the present study, language dominance effects are only seen in the Interview task (informal and minimally structured) and not in the Discourse Completion Task (informal but highly structured), providing support for the interaction of task-type and language dominance on the production of  $f_0$ .

The other language dominance effect predicting  $f_0$  max in the Interview Task was Social Language. In the Interview task (ML Basque, EL Spanish), speakers who reported speaking Basque socially had lower  $f_0$  max than those who reported speaking Spanish or both Spanish and Basque socially (Figure 39). It is unanticipated that those who are more Basque-dominant (i.e. higher Semantic Verbal Fluency score) produce higher  $f_0$  while those who speak Basque socially produce lower  $f_0$  in ML Basque contexts. However, language dominance and language usage (i.e. language spoken at home and with friends) may result in different prosodic realizations as well as index different parts of a speaker's identity (i.e. language competence versus language

most frequently used in social contexts). The finding that those who speak Basque socially produce significantly lower f0 in ML Basque contexts than those who speak Spanish or both Spanish and Basque socially may be an example of prosodic style shift resulting in hypoarticulation. Lindblom's (1990) H&H (hypo- and hyper-articulation model), which also provided the basis for Olson's (2012) Hyperarticulation account, predicts that speakers vary their level of articulation depending on how much acoustic information is needed by their audience in a given context. In the case of Olson (2012), the author predicts that code-switching contexts require *more* articulation (in the form of f0 range and duration) compared to non-switched modes.

In the current study, there is evidence for the opposite (i.e. hypoarticulation) in ML Basque, EL Spanish contexts with respect to Social Language. As discussed in Chapters 2 and 3, Spanish has sociopragmatic dominance in the Basque Country, resulting in frequent code-switching *into* Spanish (as opposed to *into* Basque) among Spanish-Basque bilinguals. Hence, switching from Basque to Spanish is more frequent and pragmatically expected within this community with several lexical categories acting as code-switching defaults (such as profanity, discourse markers, and slang, cf Lantto (2015)). In addition to these sociopragmatic factors, the Interviewer and both speakers who reported speaking Basque socially (speakers 5 and 6) are from Lekeitio. In this speech contexts, where dialectal and community norms are established (i.e. code-switching into Spanish is expected), hypoarticulated speech (i.e. lower f0) may emerge as a result of fewer acoustic requirements by the listener. When dialectal membership or community speech practices differ (for example, a conversation between a speaker from Bilbao and a speaker from Lekeitio), more articulatory effort may be required, as seen with in ML Basque, EL

Spanish contexts in which those who reported speaking Spanish socially (speakers 3 and 4, both from Bilbao) had the highest f0 of all Social Language groups.

Language dominance and Social Language were also significant predictors of peak alignment in the Interview task (ML Basque, EL Spanish). Beginning with Language Dominance (Semantic Verbal Fluency score), the current study found an interaction between Language Dominance and Context in which more Basque-dominant speakers (i.e. higher Semantic Verbal Fluency score) produced (phonetically) earlier peak alignment in ML (Basque) conditions and later peak alignment in EL (Spanish) conditions (Figure 35). Similar to the results found for the Bilbao group in the Interview task, the presence of “compromised” peak alignment values (i.e. phonetically later peak alignment for Spanish that is not categorically delayed) may be an intermediate Matrix Language effect. That is, the distinction in peak alignment is still observed, but the Matrix Language (Basque) eliminates the categorical distinction between early and delayed peaks. These findings support the Language Dominance hypothesis of the current study, which predicts that deviation from unilingual norms will be greater when a speaker’s dominant language is the EL. In this case, those with *lower* Semantic Verbal Fluency scores (i.e. more Spanish-dominant) produced later peak alignment in the ML (Basque) than in the EL (Spanish). As opposed to the compromised values seen in the Basque-dominant speakers, who maintained a truncated distinction between Spanish and Basque unilingual norms, the presence of earlier Spanish peak alignment (when compared to Basque) is evidence for (further) deviation from unilingual norms when the dominant language (Spanish) is the EL.

The last language dominance variable that will be discussed is Social Language. In the Interview task, those who reported speaking Basque socially produced significantly earlier peak alignment than those who reported speaking both Spanish and Basque socially (see Figure 33).

To clarify, there were no categorical peak alignment distinctions in the interview task, but the presence of significant phonetic differences in peak alignment between groups may indicate a spectrum of peak alignment values that depend on factors such as language usage or dominance. Similar to the results seen for the Basque-dominant speakers' peak alignment, speaking Basque socially may facilitate more conservative productions of unilingual norms when compared to those who speak other languages socially. The link between social factors and prosodic realizations has also been documented in Elordieta & Romera (accepted), in which Spanish-Basque bilinguals who had more positive attitudes toward the Basque language and community also showed evidence of more Basque-like prosodic features in their production of Spanish. These results of these studies suggest that social factors such as domain of language usage and ethnolinguistic attitudes may affect the language in question (i.e. usage of Basque affecting Basque production) or transfer onto another language (i.e. transfer of Basque-like prosody into Spanish).

## 5.2 Summary of hypotheses

To summarize, the results of the current dissertation showed at least partial support for each of the hypotheses. The first hypothesis predicted that in the presence of language-specific (peak alignment) differences, the pattern of the ML would override that of the EL. The results supported this, though not unanimously: Lekeitio speakers were more sensitive to ML effects than the Bilbao group, suggesting that variables such as language dominance, switching costs, and order of tasks may also interact with vulnerability to ML effects during code-switching. The second hypothesis, which predicted higher  $f_0$  or Hyperarticulation in code-switches, had the least amount of support, with only one group and context producing significantly higher  $f_0$  during code-switches (i.e., Lekeitio speakers in the Discourse Completion Task produced higher phrase-

final f0 for code-switches in EL contexts). The final hypothesis, which predicted that factors such as language dominance and usage would interact with prosodic realizations of code-switches, had support in the Interview task only. More specifically, there was evidence in the Interview task of (truncated) maintenance of unilingual peak alignment norms in more Basque-dominant speakers and those who speak Basque socially, whereas peak alignment values deviated further from unilingual norms in Spanish-dominant speakers and those who speak both Spanish and Basque socially. Additionally, the finding that those who speak Basque socially produced lower f0 than those who speak Spanish or both languages socially provide evidence for *hypoarticulation* in congruent speech community conditions (i.e. similar dialect and language usage). The hypotheses and relevant results concerning them are summarized in Table 31. Several implications of these results will be briefly discussed in the next section.

| Hypothesis                                                                                                         | Results                                                                                                                                                                                           | Supported or Rejected                                            |
|--------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| 1. The peak alignment patterns of the Matrix Language will override the Embedded Language (Matrix Language Effect) | For Lekeitio speakers, Matrix Language Effects were seen in ML Spanish and ML Basque contexts. For Bilbao speakers, Matrix Language effect seen only in ML Basque context                         | Supported by Lekeitio group; Partially supported by Bilbao group |
| 2. Code-switched words will have higher f0 than non-switched words (Hyperarticulation)                             | Code-switched words had higher f0 in phrase-final positions for Lekeitio speakers in EL Basque; no difference elsewhere                                                                           | Weak support                                                     |
| 3. Language dominance and usage will affect prosodic realization of code-switching                                 | Spanish-dominant speakers' peak alignment strayed further from unilingual norms than Basque-dominant/Both-dominant in ML Basque; Social Basque speakers produce lower f0 than others in ML Basque | Supported in Interview Task (ML Basque)                          |

*Table 31: Summary of hypotheses and relevant findings*

### 5.3 Implications of the study's results

The first important implication of the present study's results is that there are numerous variables to consider when investigating code-switching in semi-spontaneous speech. The present study revealed interactions between variables such as Region, Context, Language dominance, and prosodic position, indicating the interaction of linguistic and proficiency-related variables that contribute to the prosodic realizations of code-switching. These results suggest that when studying bilinguals and code-switching, it is imperative to take into account variables that carefully assess fluency from several angles (for example, self-reported measures, Semantic Verbal Fluency score, proficiency exam scores) as well as language usage (in which domains languages are spoken, how often a language is spoken) and sociopragmatic variables (such as the social domains of code-switching, which language is usually the Matrix Language, community views on the languages in question). In order to further the study of naturalistic code-switching, studies that consider a more complete profile of the bilinguals in question (e.g. linguistic, social, pragmatic factors) as well as larger numbers of participants in each group may be able to address issues of language dominance and sociopragmatics more thoroughly.

The second important implication of the present study is that the Matrix Language Framework can be extended to phonological phenomena. The results of the Discourse Completion Task and Interview tasks showed evidence that the peak alignment patterns of the Matrix Language overrode those of the embedded language, particularly in the Lekeitio group. This asymmetry between groups with respect to Matrix Language effects also suggests that factors such as language dominance may mediate sensitivity to Matrix Language effects, which is not something previously considered in studies working within the Matrix Language Framework.

If the Matrix Language Framework can be applied to morphosyntax (as seen in Myers-Scotton (1993)) and phonology (the current study), this is evidence that code-switching effects permeate multiple subsystems of language planning and production. Recent work on lexical priming by Fricke & Kootstra (2016) also provides support for the Matrix Language Framework, which was used to account for lexical priming of code-switching based on the grammatical frame of the utterance (i.e. the Matrix Language). The peak alignment results in the present study also provide evidence for a continuum of peak alignment values that include unilingual norm and truncated or compromise values during code-switching, which have also been attested in code-switching research on VOT and f0 (Olson (2013, 2016c), Piccinini & Garellek (2014), Aly (in press)).

The last implication of the current findings is the asymmetry in code-switching that may be due to language inhibition or switching costs. The results of the current study revealed that Lekeitio speakers were more sensitive overall to code-switching contexts than the Bilbao group. As discussed in the previous section, the order of the tasks administered as well as the Matrix Language in each of them may have contributed to this sensitivity. If there are greater switching costs involved in switching into the dominant language from the non-dominant language (as discussed in studies such as Meuter & Allport (1999) and Olson (2016c)) as well as greater difficulty in inhibiting the dominant language during these tasks, then the current study's structure may have been biased (i.e. lower switching and inhibition costs) toward Spanish-dominant speakers. When this is combined with the sociopragmatic patterns of code-switching (for example, that code-switching from Basque to Spanish is more common in the Basque country), it is possible that established code-switching patterns in a community may have greater "costs" for speakers whose dominant language is typically the Embedded Language (EL) of

interactions. More research on code-switching preferences in bilinguals with differing dominant languages is needed to determine whether there are different code-switching strategies that are dependent on a speaker's dominant language (e.g. preferring a certain Matrix Language, truncation or compromised realizations during code-switching). As the current study's corpus only contains seven speakers, it is not conclusive whether the results found with respect to dominant languages and task types would generalize to a larger sample or other bilingual communities.

The following section will conclude the dissertation by acknowledging additional directions for future study and summarizing the contributions and limitations of the current dissertation.

#### 5.4 Directions for future study and concluding remarks

There are multiple avenues for future research on the topic of code-switching and prosody, several of which arise out of the limitations of the current study. Firstly, future work on code-switching in the Spanish-Basque community would benefit from a larger sample size, as the current data set contains data from only seven educated female speakers in their 20s and 30s. A corpus which contains not only more speakers, but a diversity of ages, education levels, proficiency levels, and dialects would be an ideal start to investigating the interactions of these variables in more quantitative detail. As discussed in the previous sections, the order of the tasks administered in the present study was rather Spanish-biased, which may have contributed to some of the asymmetrical effects seen in the Lekeitio and Bilbao groups. A more symmetrical study design which included more Basque-dominant tasks and counterbalanced the order in which Spanish-dominant and Basque-dominant tasks were presented to participants may provide results which can further address issues of language dominance and switching costs discussed in the previous sections.

Additionally, as code-switches from Spanish into Basque are not as common, the present study had fewer code-switches in the Discourse Completion Task than in the Interview and was not able to have a semi-spontaneous Matrix Language Spanish task that mirrored the Matrix Language Basque interview. This resulted in recasts of proper nouns and place names serving the only type of code-switch in the Discourse Completion Task, which is not optimal from a methodological standpoint as no non-noun code-switches were elicited and proper nouns have a controversial status in the code-switching literature. Future research may investigate the contexts and triggers that do facilitate a wider variety of code-switches from Spanish into Basque in order to understand the patterns present in this less common code-switching context.

Future research on this topic may also consider investigating pitch accent types used during code-switching in the Lekeitio group. Because Lekeitio Basque has a lexical pitch accent system with a different pitch accent type from those in Spanish, research that investigates how these phonological differences are resolved would provide an additional layer of insight into potential Matrix Language effects. However, more sonorous and controlled (i.e. less spontaneous) data may be needed for this task to avoid issues of micro-prosody, stress lapse, and stress clash, all of which affect the realization of pitch accents. Lastly, anticipatory or spill-over effects of prosodic variables such as peak alignment,  $f_0$ , or duration have yet to be investigated. The presence of anticipatory effects in phonetic variables such as VOT, vowel formants, and overall speech rate indicate that there are multiple acoustic cues to code-switching that may help facilitate processing in listeners (see Sections 1.3.2 and 1.3.3), yet there is currently no empirical data addressing if these cues extend to suprasegmental aspects of speech as well.

The current dissertation provided a preliminary prosodic analysis of Spanish-Basque code-switching by analyzing the peak alignment and  $f_0$  maxima in the semi-spontaneous speech of

seven female Spanish-Basque bilinguals. The results of the study provide limited support for the Matrix Language Framework on phonological variables (i.e. peak alignment in this study) as well as evidence for effects of language dominance (self-reported and quantitatively measured), language context, and regional (dialectal) differences. Although preliminary in nature, the current dissertation's findings provide a starting point for future research on semi-spontaneous code-switching, the prosody of code-switching, and research on code-switching on typologically unrelated language such as Basque and Spanish.

## 6 APPENDICES

### 6.1 Appendix A: Participant Language Questionnaire (English and Spanish versions available for participants)

1. Gender *Género*

- ☐ Male *Hombre*
- ☐ Female *Mujer*
- ☐ Other: \_\_\_\_\_ *Otro: \_\_\_\_\_*

2. Age: \_\_\_\_ years; Year of birth: \_\_\_\_\_ *Edad: \_\_\_\_ años; Año de nacimiento: \_\_\_\_\_*

3. Native country: *País nativo:*

4. Native city: *Ciudad nativa:*

5. Current city: *Ciudad actual:*

6. In which city have you lived the longest? \_\_\_\_\_ How many years? \_\_\_\_\_

*En cual ciudad ha vivido más:*

*Cuántos años:*

7. What is your native language? (choose all that apply)

*Cuál es su lengua nativa (indique todas las opciones relevantes)*

- ☐ English *Inglés*
- ☐ Spanish *Español*
- ☐ Basque *Euskara*
- ☐ Other: \_\_\_\_\_ *Otro: \_\_\_\_\_*

8. Do you speak other languages (apart from those indicated in 7)?

*Habla otras lenguas aparte que las que indicó en 7*

- ☐ Yes: *Sí*

i. Age you started speaking this/these language(s):

*Edad que empezó a hablar esa(s) lengua(s):*

☐ No No

9. In which language(s) do you feel most fluent?

*En cuál lengua se siente más cómodo/a?*

10. Occupation: *Trabajo:*

11. Level of education and place: *Nivel de educación y lugar*

12. Place of birth and native language(s) of your father *Lugar de origen y lengua nativa de su padre*

a. Place *Lugar*

b. Native language(s) *Lengua(s) nativa(s)*

13. Place of birth and native language(s) of your mother *Lugar de origen y lengua nativa de su madre*

14.

a. Place *Lugar*

b. Native language(s) *Lengua(s) nativa(s)*

15. Which language do you speak at home? (choose all that apply)

*Cuál lengua habla en casa? (Indique todas las opciones relevantes)*

a. English *Inglés*

b. Spanish *Español*

c. Basque *Euskara*

d. Other: \_\_\_\_\_ Otro: \_\_\_\_\_

16. Which language do you speak more at work? (choose all that apply)

*Cuál lengua habla más en el trabajo (indique todas las opciones relevantes)*

a. English *Inglés*

b. Spanish *Español*

c. Basque *Euskara*

d. Other: \_\_\_\_\_ Otro: \_\_\_\_\_

17. Which language do you speak with Friends? (choose all that apply)

*Cuál lengua habls con amigos (indique todas las opciones relevantes)*

a. English *Inglés*

b. Spanish *Español*

c. Basque *Euskara*

d. Other: \_\_\_\_\_ Otro: \_\_\_\_\_

18. Please indicate your level of fluency in the following languages (1= worst; 5=best)

*Indica su nivel de fluencia en las siguientes lenguas (1=peor; 5=mejor)*

a. Spanish: *Español*

i. Reading: *Leer*            1        2        3        4        5

ii. Speaking: *Hablar*       1        2        3        4        5

iii. Understanding (listening):    1        2        3        4        5  
*Entener (escuchar)*

b. Basque *Euskara*

i. Reading: *Leer*            1        2        3        4        5

ii. Speaking: *Hablar*       1        2        3        4        5

iii. Understanding (listening):    1        2        3        4        5  
*Entener (escuchar)*

c. English (if you speak English) *Inglés (si habla inglés)*

i. Reading: *Leer*            1        2        3        4        5

ii. Speaking: *Hablar*       1        2        3        4        5

iii. Understanding (listening):    1        2        3        4        5  
*Entener (escuchar)*

## 6.2 Appendix B: Reading task and English translation (Filler sentences in italics)

Lea las frases en voz alta como si estuviera hablando con un(a) amig(a).  
'Read the phrases out loud, as if you were talking with a friend'.

### Parte 1 'Part 1': (Contrasts in word order)

*Juanita tocaba perritos. 'Juanita petted puppies.'*  
Médicos veían ovejas. 'Doctors saw sheep.'  
*Científicos examinan girafas. 'Scientists examine giraffes.'*  
Lorena miraba iguanas. 'Lorena watched iguanas.'  
Bernabé veía limones. 'Bernabé saw lemons.'

Veían médicos ovejas. 'Doctors saw sheep.'  
*Están examinando girafas. 'They are examining giraffes.'*  
Miraba Bernabé iguanas. 'Bernabé watched iguanas.'  
*Compraban ellos pescado. 'They bought fish.'*  
Veía Marina limones. 'Marina saw lemons.'  
*Pagaba Juanita abogados. 'Juanita paid lawyers.'*

Veían iguanas médicos. 'Doctors saw iguanas.' \ 'Iguanas saw doctors.'  
*Cocinaba pescados Juanita. 'Juanita cooked fish.'*  
Miraba ovejas Lorena. 'Lorena watched sheep.'  
*Tocaban llamas científicos. 'Scientists petted llamas.'*  
Veía limones Bernabé. 'Bernabé saw lemons.'  
*Navegan barcos abogados. 'Lawyers sailed (on) ships.'*

### Parte 2 'Part 2': (Stress lapse and stress clash contexts)

Toma limonada horrible. 'He/she drinks horrible lemonade.'  
*Están tomando también. 'They are drinking too.'*  
Tomaba limonada horrible. 'She/He drank horrible lemonade.'  
*Estás tomando también. 'You are drinking too.'*  
Tomaría limonada horrible. 'She/He would drink horrible lemonade.'  
*Toman algo también. 'They drink something too.'*

Tomaríamos limonada horrible. ‘We would drink horrible lemonade.’  
 Ayudó el águila bonita. ‘She/He helped the beautiful eagle.’  
 Ayuda el águila bonita. ‘She/he helps the beautiful eagle.’  
 Mirábamos el águila bonita. ‘We looked at the beautiful eagle.’  
*Mirábamos arañas enormes. ‘We looked at enormous spiders.’*  
 Mirábamos marabús bonitos. ‘We looked at beautiful marabouts.’  
*Veían peces numerosos. ‘We saw numerous fish.’*  
*Capturaban peces numerosos. ‘They captured numerous fish.’*

### Parte 3 ‘Part 3’: (Questions and imperatives)

Marina bebió limonada. ‘Marina drank lemonade.’  
 ¿Quién bebió limonada? ‘Who drank lemonade?’  
 ¿Marina bebió limonada? ‘Did Marina drink lemonade?’  
 ¡Bebe limonada! ‘Drink lemonade!’  
 ¿Me das una limonada? ‘Could you give me lemonade?’  
 ¿Me darías una limonada? ‘Would you give me lemonade?’  
 Marina bebe limonada, ¿no? ‘Marina drinks lemonade, right?’  
 Marina no bebe limonada, ¿correcto? ‘Marina doesn’t drink lemonade, correct?’

Lorena mide limones. ‘Lorena measures lemonade.’  
 ¿Quién mide limones? ‘Who measures lemons?’  
 ¿Lorena mide limones? ‘Lorena measures lemonade?’  
 ¡Mide limones! ‘Measure lemons!’  
 ¿Me das un limón? ‘Give me a lemon?’  
 ¿Me darías un limón? ‘Would you give me a lemon?’  
 Lorena mide limones, ¿no? ‘Lorena measures lemons, right?’

### Parte 4 ‘Part 4’: Pseudo-dialogue for narrow and contrastive focus. Participants did not read portions in parentheses out loud.

Marina bebió limonada. ‘Marina drank lemonade.’  
 (¿Quién bebió limonada?) MARINA bebió limonada.  
 ‘(Who drank lemonade?) Marina drank lemonade.’  
 (¿Qué bebió Marina?) Marina bebió LIMONADA.  
 ‘(What did Marina drink?) Marina drank lemonade.’  
 (¿Qué hace Marina con la limonada?) Marina BEBIÓ limonada.  
 ‘(What does Marina do with the lemonade?) Marina drank lemonade.’  
 (¿Qué hace Marina?) Marina BEBIÓ LIMONADA.  
 ‘(What does Marina do?) Marina drank lemonade.’

¿Juan bebe limonada? (No,) MARINA bebe limonada.  
 ‘Juan drinks lemonade? (No,) Marina drinks lemonade.’  
 ¿Marina bebió café? (No,) Marina bebe LIMONADA.  
 ‘Marina drank coffee? (No,) Marina drinks lemonade.’  
 ¿Marina prepara limonada? (No,) Marina BEBE limonada.  
 ‘Marina prepares lemonade? (No,) Marina drinks lemonade.’  
 ¿Marina prepara café? (No,) Marina BEBE LIMONADA.  
 ‘Marina prepares coffee? (No,) Marina drinks lemonade.’

### 6.3 Appendix C: Discourse Completion Task (Basque trigger words in bold)

Adapted from: Prieto, P.; Roseano, P. (eds.) (2010). *Transcription of Intonation of the Spanish Language*, pp. 123-153. Lincom Europa: München.

#### 1. DECLARATIVAS (*Declaratives*)

##### 1.1. Neutra (*neutral*)

##### **Oraciones de una unidad tonal (*Sentences with one tonal unit*)**



1. Mira el dibujo y di lo que hace la chica.

*‘Look at the drawing and say what the woman is doing.’*

2. **Ane** te está contando que ayer se bebió una limonada. En este momento llega otro amigo y te pregunta qué dice **Ane**.

*‘Ane is telling you that yesterday she drank lemonade. At this moment, another friend arrives and asks you to tell him what Ane said.’*

##### **Oraciones de más de una unidad tonal (*Sentences with more than one tonal unit*)**

3. Mira el dibujo y di lo que ves. *‘Look at the drawing and say what you see.’*



4. Mira el dibujo y di lo que ves. *'Look at the drawing and say what you see.'*



### Enumeraciones (*Enumerations*)

5. Di los días de la semana. *'Say the days of the week.'*

6. Di qué almorzaste. *'Say what you had for lunch.'*

### Elementos periféricos (dislocaciones, vocativos, elementos parentéticos, aposiciones)

#### (*Peripheral elements-dislocations, vocatives, parthenteticals, apositions*)

7. Imagínate que acabas de conocer a alguien de **Donostia** y resulta que tú has vivido allí muchos años. ¿Cómo se lo dirías?

*'Imagine that you have just met someone from Donostia and, as it turns out, that you lived there for many years. How would you tell this person?'*

8. Estás en casa con tu hija, **Amaia**, que está mirando la tele. Dile que vas a salir un momento de compras.

*'You are at home with your daughter, Amaia, who is watching television. Tell her that you are going out for a while to shop.'*

9. Estás enfermo y esta mañana has tenido que ir al médico. Di que has ido a pesar de la lluvia.

*'You are sick and this morning you had to go to the doctor. Say that you have done this, despite the rain.'*

10. Conoces a dos chicas que se llaman **Maialen**, una rubia y otra morena. Di que hoy has visto a la morena.

*'You know two young women named Maialen, a blonde and a brunette. Say that you saw the brunette today.'*

## **1.2. No neutral (non-neutral)**

### **Focalización contrastive (contrastive focus)**

11. Entrás en una frutería y la mujer que trabaja ahí es un poco sorda. (diálogo)

*‘You enter a shopping center and the woman that works there is a bit deaf.’*

A. Quiero un kilo de limones *‘I want a kilo of lemons.’*

B. ¿De manzanas? *‘Of apples?’*

C: (respuesta) (answer)

### **Énfasis (emphasis)**

12. Entrás a una panadería y notas un buen olor a pan. Díselo a la chica que atiende la panadería.

*‘You enter a bakery and you note the rich scent of bread in the air. Comment on this to the woman who is working in the bakery.’*

### **Declarativa categórica (categorical declarative)**

13. Tú y una amiga estáis hablando de unos amigos que están buscando apartamento, y vosotros/as no estáis seguros/as de dónde van a vivir. Estás segura/o que se van a vivir a Bilbao pero tu amiga piensa – también bastante segura- que se van a mudar a **Getxo**. Dile, seguro, que no, que se mudan a Bilbao.

*‘You and a friend are talking about some individuals who are looking for an apartment and neither of you are sure about where they are going to live. You are sure that they are going to live in Bilbao, but your friend, who is also very sure, says that they are moving to Getxo. Tell her, with confidence, that they are not, that they are moving to Bilbao.’*

### **Declarativa dubitativa (Doubtful declarative)**

14. Te han encargado comprar un regalo para alguien que no conoces mucho y te preocupa no hacer una buena compra. Dile a la persona que te hizo el encargo que quizás no le guste el regalo que has comprado.

*‘They have left you in charge of buying a gift for someone you don’t know very well and you are worried about that you might not make a good buy. Tell the person who put you in charge that maybe they will not like the present that they bought.’*

### **Declarativa de obviedad (Declarative with obviousness)**

15. Estás con una amiga y le cuentas que **Itziar**, una amiga en común, está embarazada. Ella te pregunta de quién está embarazada y te extrañas mucho de que no lo sepa porque todo el mundo sabe que es de **Andoni**, su novio de toda la vida. ¿Qué le dices?

*'You are with a friend and you tell her that Itziar, a mutual friend, is pregnant. She asks you whose child she is having, and it is strange to you that she does not know because everyone knows it is Andoni, her lifelong boyfriend, with whom she is having the baby. What do you tell her?'*

### **Declarativa exclamativa (Exclamative statement)**

16. Te invitan a un **marmitako** y es el mejor que has comido en tu vida. Estás encantado. ¿Qué dices?

*'You are invited to eat stew (and it is the best you have had in your life). You are enchanted. What do you say?'*

17. Es la vez que más frío has tenido en tu vida. ¿Qué dices?

*'It is the coldest moment in your life. What do you say?'*

## **2. INTERROGATIVAS ABSOLUTAS (Absolute interrogatives)**

### **2.1. Neutra (neutral)**

#### **Oraciones de una unidad (Sentences with one (tonal) unit)**

18. Entrás a un supermercado y le preguntas al empleado si tiene galletas de chocolate .

*'You enter a market and ask the employee if they have chocolate chip cookies.'*

19. Estás en la calle y quieres preguntar la hora.

*'You are on the street and you want to ask for the time.'*

20. Pides permiso para pasar a la consulta donde te espera el médico.

*'You ask for permission to go to the office where the doctor is waiting for you.'*

#### **Oraciones de más de una unidad (Sentences with more than one (tonal) unit)**

21. Llamas por teléfono a casa de una amiga que se llama **Nerea** pero no está. Más tarde llamas de nuevo, pero ella no contesta al teléfono. ¿Cómo preguntas si ya ha llegado?

*'You call your friend Nerea's house number but she is not home. Later you call again but she does not answer the phone. How do you ask if she has arrived?'*

22. Has organizado una comida y decides cambiar la fecha para que todos los invitados puedan ir. Pregúntales si van a poder venir si la comida es el primer domingo de mayo.

*'You organized a dinner and you decide to change the date so that all the guests can come. Ask them all if they are going to be able to come if the dinner is the first Sunday of May.'*

### La disyunción (*Disjunction*)

23. Para tu cumpleaños has preparado pastel y helado. Pregúntales a los invitados si quieren pastel o helado.

*'For your birthday, you prepared cake and ice cream. Ask the guests if they want cake or ice cream'*

24. Tu primo quiere visitar a su tío, y lo quieres acompañar. Pregúntale si va a ir hoy o mañana.

*'Your son wants to visit his uncle and you want to go with him. Ask him if he is going today to tomorrow'*

### Enumeraciones (*Enumerations*)

25. Sabes que una prima lejana tuya tiene muchos hijos, pero no estás segura si son tres, cuatro, cinco o seis. Pregúntale a tu madre.

*'You know that your distant cousin has a lot of kids, but you aren't sure if she has three, four, five or six. Ask your mom'*

### Elementos periféricos (*Peripheral elements*)

26. Estás buscando a Marina pero no la puedes encontrar. Te cruzas con alguien que la conoce y después de hablar un poco sobre ella le preguntas si la ha visto.

*'You're looking for Marina but you can't find her. You run into someone else that knows her and after talking a bit about Maria you ask the person if she/he has seen her'*

## 2.2 No neutra (*non-neutral*)

### Focalización y énfasis (*Focus and emphasis*)

27. Estás hablando de **Begoña** con alguien y oyes que entra una persona. Pregúntale si es **Begoña** la persona que está entrando.

*'You are talking about Begoña about someone and you hear someone enter. Ask if it was Begoña who entered'*

### Preguntas exclamativas (*Exclamative questions*)

28. El electricista tenía que venir a las 10. Has tenido que salir a comprar y tu hija se ha quedado esperándole. Llegas de la compra y el electricista aún no ha llegado. ¿Cómo reaccionas?

*'The electrician was supposed to come at 10. You had to leave to go shopping and your daughter stayed at home to wait. You arrive back home from shopping and the electrician still hasn't come'*

29. Estás cenando en un restaurante. Hace mucho calor y el ambiente está cargado. Al lado tuyo está tu hermano temblando de frío. Sorprendida, le preguntas si tiene frío.

*'You're eating in a restaurant. It's really hot. At your side, your son is shivering. Surprised, you ask if he is cold.'*

### **Preguntas confirmatorias (Confirmation questions)**

30. **Jon** dijo que iba a venir a comer pero quieres confirmarlo. ¿Qué le dices a **Jon**?

*'Jon said he was going to come eat, but you want to confirm. What do you say to him?'*

31. Antes de ir a trabajar tu hermano dijo que no se sentía muy bien. Al volver, lo encuentras en la cama temblando de frío. Ves que no se siente bien, pero se lo preguntas, sabiendo cuál va a ser la respuesta.

*'Before going to work your brother said he didn't feel well. Upon returning, you find him in bed shivering. You see that he doesn't feel well, but you ask him, knowing what the answer will be.'*

32. Tienes muchas ganas de que alguien venga a una cena que has organizado. Se lo pides de modo que no pueda decir que no.

*'You really want someone to come to a dinner you organized. Ask him/her in a way that they can't say no.'*

33. Sabes que afuera está haciendo mucho frío. Entra alguien bien abrigado y le preguntas si tiene frío.

*'You know that it's really cold outside. Someone walks in all bundled up and you ask him/her if they are cold.'*

### **Preguntas imperativas (Imperative questions)**

34. Tus nietos arman mucho ruido y no te dejan oír la televisión. Les pides que se callen. (entre ruego y orden)

*'Your grandkids are making a lot of noise and you can't hear the TV. Ask them to be quiet.'*

35. No te hacen caso y esta vez lo pides más enfadada. (orden)

*'They (the grandkids) don't listen to you and this time you ask them, angrier.'*

36. Le preguntas a un amigo si se quiere venir a tomar un trago contigo. (invitación)

*'You ask a friend if they want to come to drink a beer with you.'*

37. Les preguntas a tus sobrinos si quieren caramelos. (invitación)

*'You ask your nephews if they want candy.'*

38. Organizas una fiesta en tu casa y tienes muchas ganas de que un compañero tuyo vaya. Pídele si quiere venir. (intención exhortativa: "me gustaría mucho que vinieras...")

*'You're organizing a party at your house and you really want a friend to go. Ask him/her to come'*

39. Necesitas subir tres pisos porque te dejaste el bolso. Vas con un niño pequeño y, para ganar tiempo, lo dejas abajo. Dile que no se mueva. (ruego-orden)

*'You need to go upstairs because you left your wallet there. You're with a small child and to save time, you leave him/her downstairs. Tell him/her to not move'*

40. Necesitas tranquilidad pero estás en medio de un gran ruido. Pregunta si alguna vez habrá tranquilidad en esta casa.

*'You need it to be quiet but you're in the middle of a lot of noise. Ask if there will ever be peace and quiet in your house'*

### **3. INTERROGATIVAS PARCIALES (*Biased questions*)**

#### **3.1. Neutra (*Neutral*)**

|                                                                         |
|-------------------------------------------------------------------------|
| <b>Oraciones de una unidad (<i>Sentences with one (tonal) unit</i>)</b> |
|-------------------------------------------------------------------------|

41. Pregunta qué hora es.

*'Ask what time it is.'*

42. Pregúntale la hora a una persona mayor.

*'Ask an elderly person for the time.'*

43. Te fuiste a la fiesta de **Laga**. Al llegar a la playa te encuentras con un amigo y le preguntas por dónde llegó.

*'You went to a party on Laga. When you get there you meet up with a guy and you ask him how he got there.'*

### Oraciones de más de una unidad (*Sentences with more than one (tonal) unit*)

44. Ves que **Nerea** se está yendo. Pregúntale dónde y cuándo va a volver. (Coordinación)

*'You see that Nerea is leaving. Ask her where and when she's going to return.'*

45. La vecina te cuenta que vino un señor a revisar la instalación del gas y que no le dejó entrar porque no tenía suficiente dinero en casa para pagarle. Él le dijo que volvería mañana. Pregúntale qué le va a decir si vuelve. (subordinación)

*'Your neighbor tells you that a utilities employee came but she didn't let him in because she didn't have enough money to pay him. He told her he was going to return tomorrow. Ask he what she's going to say to him if he returns.'*

### Elementos periféricos (*Peripheral elements*)

46. Encuentras un paquete en tu casa y le preguntas a tu hijo, **Iker**, quién ha traído esto. (posición final, vocativo)

*'You find a package in your house and you ask your son, Iker, who brought it.'*

### 3.2. No neutra (*non-neutral*)

#### Focalización contrastiva (*Contrastive focus*)

#### Énfasis y preguntas exclamativas (*Emphasis and exclamative questions*)

47. Tu primo te cuenta que el autobús que venía de Barcelona llegó con cuatro horas de retraso. Pregúntale, sorprendido/a, a qué hora terminó llegando.

*'Your cousin tells you that the bus that comes from Barcelona arrived four hours late. Ask him, surprised, at what time it finally came.'*

#### Preguntas dubitativas (*Doubtful questions*)

48. A las dos de la mañana llaman a la puerta. Estás dormida/o y te despiertan. Pregúntate quién será a estas horas.

*'At two in the morning, someone knocks at the door. You're asleep and it wakes you up. Ask yourself who it could be at this time.'*

#### Preguntas imperativas (*Imperative questions*)

49. Le pides a tu hijo/amigo que te haga unos arreglos en la casa y no estás seguro de que lo vaya a hacer, ya que no es la primera vez que se lo pides. Pregúntale, medio enfadado/a, cuándo lo va a hacer. (orden)

*'You ask your son to do some chores for you and you're not sure if he is going to do it, as it's not the first time you've asked. Ask him, half upset, when he's going to do it.'*

50. Tienes ganas de que unos amigos vengan a comer a tu casa. Medio suplicando (porque ya te han dicho que no pueden venir) les preguntas por qué no vienen. (Invitación, ruego)

*'You really want some friends to come eat at your house. Half begging (because they already told you that they couldn't come), ask them why they can't come.'*

51. Alguien te tira de la camisa un par de veces, pero cuando te das la vuelta no ves a nadie. Finalmente, a la tercera vez, ves que es un conocido tuyo muy pesado y charlatán que siempre que te ve no te deja ir. Dile qué quiere. (choque acentual, queja leve o protesta)

*'Someone taps you on the shoulder a couple of times and when you turn around you don't see anyone. Finally, the third time, you see that it's an annoying, chatty acquaintance that is hard to get away from. Ask him what he wants.'*

#### **Preguntas retóricas (Rhetorical questions)**

52. Le habías dicho a la gente que trabaja contigo que hicieran algo, pero cuando llegas descubres que no lo hicieron porque te estaban esperando. Pregúntales qué harían sin ti.

*'You told the people you work with to do something, but when you arrive at work you see that they didn't do it because they were waiting for you (to do it). Ask them what they would do without you'*

#### **4. INTERROGATIVAS REITERATIVAS (Reiterative/echo questions)**

##### **4.1. Neutra (neutral)**

|                                                                   |
|-------------------------------------------------------------------|
| <b>Preguntas reiterativas absolutas (Absolute echo questions)</b> |
|-------------------------------------------------------------------|

53. Has invitado un amigo al cine y te ha dicho que no puede venir. Te parece que no le has entendido bien. Le preguntas para aclararlo.

*'You invited a friend to the movies and he told you he can't come. It seems you didn't understand him. Ask him again to clarify'*

54. Te dicen la hora, pero no has oído bien. Piensas que te han dicho que son las nueve. Vuelves a preguntar.

*'They tell you the time, but you didn't hear it. You think they said 9. Ask them again.'*

|                                                                 |
|-----------------------------------------------------------------|
| <b>Preguntas reiterativas parciales (Biased echo questions)</b> |
|-----------------------------------------------------------------|

55. Te han preguntado dónde vas, pero no estás seguro si has entendido bien la pregunta. Averigua si es eso lo que habían preguntado.

*'They asked you where you're going, but you're not sure if you hear the question. See if that's what they asked you.'*

#### **Oraciones de más de una unidad tonal (Sentences with more than one (tonal) unit)**

56. Te han preguntado dónde vas y cuándo vas a volver. Pero no sabes si lo has entendido bien. Pregunta si es eso lo que te dijeron.

*'They asked you where you're going and when you're going to return. But you don't know if you heard correctly. Ask if that's that they asked you.'*

#### **La disyunción (Disjunction)**

57. Te han preguntado por dónde has llegado, pero no estás segura si te han pedido eso o si te han preguntado por dónde has entrado. Averigua si te han preguntado una cosa o la otra.

*'They asked you how you arrived, but you're not sure if they asked that or if they asked from where you entered. Find out if they asked one thing or the other.'*

#### **Elementos periféricos (Peripheral elements)**

58. Te comentan que una compañera tuya, Marina, quiere ir a bailar, y sabes que no le gusta salir de fiesta. No te lo crees y preguntas si es Marina la que quiere ir.

*'They tell you that your friend Marina want to go dance, but you don't that she doesn't like to go out. You don't believe it and you asked if it was Marina who wanted to go.'*

### **4.2 No neutra (non-neutral)**

#### **Focalización y énfasis (Focus and emphasis)**

59. Te dicen que un compañero tuyo, **Mikel**, se presenta para alcalde. No te lo crees y lo vuelves a preguntar.

*'They tell you that your friend Mikel is running for mayor. You don't believe it and you ask again.'*

#### **Preguntas reiterativas exclamativas (Exclamative echo questions)**

60. Tu vecina te cuenta que fue a un restaurante a comer y pidió pimientos con el pollo. Ella dice que le dieron patatas en lugar de pimientos. No lo puedes creer. Pregúntale qué le dieron muy sorprendida/o.

*'Your neighbor tells you that she went to a restaurant and ordered pimientos with chicken. She says that they gave her patatas instead of pimientos. You can't believe it. Ask her what they gave her, surprised.'*

## **5. ORACIONES IMPERATIVAS (Imperatives)**

### **5.1 Órdenes (Orders)**

61. Trabajas en la recepción de un hotel y entra una pareja que quiere una habitación. Diles que llenen un formulario.

*'You work in the lobby of a hotel and a couple enters who wants a room. Tell them to fill out a form.'*

62. Ves que están un poco distraídos y no lo completan. Díselo otra vez (con más insistencia).

*'You see that they are distracted and don't fill it out. Ask them again (with more insistence).'*

63. Estás en el parque con tu nieta/sobrina, **Iratxe**, y se te escapa. Dile que venga, que no se aleje tanto de ti.

*'You are in the park with your granddaughter/neice, Iratxe, and she gets away from you. Tell her to come back and not to go so far from you'*

64. Sales del parque y se te vuelve a escapar. Dile que venga (con más insistencia).

*'You leave the park and she gets away again. Tell her to come back (with more insistence).'*

65. Ahora estáis en la calle, por donde pasan coches, y se te vuelve a escapar. Estás muy nerviosa/o y le dices, enfadada/o , que no se separe de ti (con mucha más insistencia).

*'Now you guys are on the street where there's a lot of traffic and she gets away from you again. You're really nervous and you tell her, angry, not to separate from you (with way more insistence).'*

66. Estás paseando el perro, Bobi, y se te escapa. Llámalo.

*'You're walking the dog, Bobi, and he gets away from you. Call him.'*

### **5.2 Ruegos (Pleading)**

67. Quieres ir al cine con un amigo. Te dice que tiene trabajo, pero sabes que el trabajo lo puede hacer después. ¿Qué le dirías para convencerlo?

*'You want to go to the movies with a friend. He tells you he has work, but you know it can wait till later. What would you say to convince him?'*

68. Parece que quiere ir pero te dice que no. Insiste a ver si lo puedes convencer.

*It seems like he wants to go, but he tells you no. Insist on seeing if you can convince him.'*

## **6. VOCATIVOS (Vocatives)**

69. Entrás en la casa de una amiga tuya, **Maialen**, pero al entrar no la ves. Llámala.

*'You enter your friend Maialen house, but you don't see her. Call for her.'*

70. Pasan diez segundos y no sale nadie. Vuelve a llamarla.

*'Ten seconds go by and no one appears. Call for her again.'*

## **6.4 Appendix D: Native speaker judgment survey**

### Encuesta euskalduna/española Basque/Spanish Survey

¡Hola! En esta encuesta, vas a indicar tu opinión sobre palabras o frases en español y euskara. No hay respuestas correctas, solo tu opinión e intuición sobre las lenguas.

*Hi! In this survey, you will give your opinion over words and phrases in Spanish or Basque. There are no correct answers, only your opinion and intuition about the languages.*

#### Part 1: Información sobre ti

#### *Part 1: Information about you*

1. ¿Hablas español? *Do you speak Spanish?*
  - a. Sí *Yes*
  - b. No *No*
2. ¿Hablas euskara/vasca? *Do you speak Basque?*
  - a. Sí *Yes*
  - b. No *No*
3. ¿Cuál es tu lengua nativa? *What is your native language?*
  - a. Español *Spanish*
  - b. Euskara *Basque*
  - c. Las dos (español y euskara) *Both (Spanish and Basque)*
  - d. Otra: *Other:*
4. Cuando hablas con amigos o familia, ¿mezclas las lenguas (español y euskara) en la misma conversación o frase? *When you speak with friends or family, do you mix languages (Spanish and Basque) in the same conversation or sentence?*
  - a. Sí *Yes*
  - b. No *No*

5. ¿Dónde vives? \_\_\_\_\_ *Where do you live?*  
 6. Tu edad: \_\_\_\_\_ *Your age*

Parte 2: Indica si la palabra o frase es del **español, euskara, o las dos**. Piensa si sueles oír estas palabras en conversaciones en español, euskara, o una mezcla de las dos lenguas.

*Indicate if the word or phrase is in **Spanish, Basque** or **both languages**. Think about if you tend to hear these words in conversations in Spanish, Basque, or a mix of the two languages.*

(Words in this section, as indicated in Figure 19, were presented with a 1-5 scale)

|                                                             |
|-------------------------------------------------------------|
| <u>Survey items (orthography)</u>                           |
| Northern Bizkaian Basque/Non-Northern Bizkaian Basque forms |
| Filler words in bold                                        |
| failaten/ fallaten                                          |
| ¡hombre!                                                    |
| a ver                                                       |
| <b>alabia</b>                                               |
| <b>arrain</b>                                               |
| asignatura                                                  |
| bueno                                                       |
| <b>cabeza</b>                                               |
| desarroiluan                                                |
| desididu                                                    |
| durezak                                                     |
| estrukturia                                                 |
| faltan                                                      |
| fazilidadiak                                                |
| fisioterapeuta                                              |
| <b>Ikaslia/ikaslea</b>                                      |
| Influentzixa/ influentzia                                   |
| japonesa                                                    |
| kontenidua/kontenidoa                                       |
| <b>lorak</b>                                                |
| maneja                                                      |
| <b>manzanas</b>                                             |
| mobimientua                                                 |

|               |
|---------------|
| motxilero     |
| necesario     |
| o sea         |
| oportunidadia |
| <b>perro</b>  |
| podologia     |
| pues          |
| <b>pulpo</b>  |
| super         |
| tipiko        |
| tipo          |
| trilingüe     |
| ya            |

## 7 REFERENCES

- Altenberg, E. P. & Ferrand, C. T. (2006). Fundamental Frequency in Monolingual English, Bilingual English/Russian, and Bilingual English/Cantonese Young Adult Women. *Journal of Voice*, 20(1), 89–96.
- Aly, A. (in press). Code-switching in Miami Cuban Spanish: Segmental and Suprasegmental effects. In A. Cuza (Ed.), *Contemporary Approaches to Cuban Spanish Dialectology*. Georgetown University Press.
- Anderson, D., & Burnham, K. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management*, 66(3), 912–918.
- Balukas, C., & Koops, C. (2015). Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*, 19(4), 423–443.
- Basque Government. (2006). Elhuyar hiztegia (Standard Basque-Spanish online dictionary). Retrieved March 8, 2017, from [http://www.euskara.euskadi.net/r59-15172x/eu/hizt\\_el/index.asp](http://www.euskara.euskadi.net/r59-15172x/eu/hizt_el/index.asp)
- Basque Government. (2011). Statistics of the Basque Country. Retrieved March 8, 2017, from <http://en.eustat.eus/>
- Beck, M. W. (2013). Stepwise VIF selection (R script). Retrieved from <https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/><https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/>
- Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology*, 3(1), 255–309.
- Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The Original ToBI System and the Evolution of the ToBi Framework. In S. Jun (Ed.), *Prosodic Typology* (pp. 9–54). Oxford University Press.
- Belazi, H. M., Rubin, E. J. & Toribio, A. J. (1994). Code Switching and X-Bar Theory: The Functional Head Constraint. *Linguistic Inquiry*, 25(2), 221–237.
- Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer. Version 6.0.26. Retrieved from <http://www.fon.hum.uva.nl/praat/><http://www.fon.hum.uva.nl/praat/>
- Bullock, B., Toribio, A., Gonzalez, V., & Dalola, A. (2006). Language dominance and performance outcomes in bilingual pronunciation. *Proceedings of Generative Approaches to Second Language Acquisition (GASLA)*, 9–16.
- Chana, U., & Romaine, S. (1984). Evaluative reactions to Panjabi/English code-switching. *Journal of Multilingual and Multicultural Development*, 5(6), 447–473.

- Chomsky, N. (1995). *The minimalist program*. The MIT Press.
- Craft, J. (2015). Acquisition of Intonation by L2 Spanish Speakers While on a Six Week Study Abroad Program in Valencia, Spain (Master's thesis). Retrieved from DigiNole (FSU\_migr\_etd-9579).
- de Bot, K., Broersma, M., & Isurin, L. (2009). Sources of triggering in code-switching. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary Approaches to Code Switching* (pp. 103–128). Amsterdam: John Benjamins.
- Di Sciullo, A.-M., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, 22(1), 1-24.
- Dhana, K. (2016). Removing outliers (R script). Retrieved August 3, 2017, from <https://www.r-bloggers.com/identify-describe-plot-and-remove-the-outliers-from-the-dataset/https://www.r-bloggers.com/identify-describe-plot-and-remove-the-outliers-from-the-dataset/>
- Eddo. (2010). Basque Country location map. Retrieved from [https://upload.wikimedia.org/wikipedia/commons/a/a2/Basque\\_Country\\_Location\\_Map.svghttps://upload.wikimedia.org/wikipedia/commons/a/a2/Basque\\_Country\\_Location\\_Map.svg](https://upload.wikimedia.org/wikipedia/commons/a/a2/Basque_Country_Location_Map.svghttps://upload.wikimedia.org/wikipedia/commons/a/a2/Basque_Country_Location_Map.svg)
- Elias, V., McKinnon, S., & Milla-Munoz, A. (2017). The effects of lexical stress and distance from the codeswitching site on Spanish heritage speaker vowel production. Paper presented at *Bilingualism in the Hispanic and Lusophone World*, Tallahassee, Florida.
- Elordieta, G., & Calleja, N. (2005). Microvariation in accentual alignment in Basque Spanish. *Language and Speech*, 48(4), 397–439.
- Elordieta, G., & Hualde, J. I. (2014). Intonation in Basque. In S. J. Jun (Ed.), *Prosodic Typology II: The phonology of intonation and phrasing* (pp. 405–463). Oxford University Press.
- Elordieta, G., & Irurtzun, A. (2012). Estudio comparativo de alineamiento tonal en el castellano de Bilbao. In B. Camus-Bergareche & S. Gómez-Seibane (Eds.), *El Castellano del País Vasco*, 119-137). Bilbao: University of the Basque Country.
- Elordieta, G. and Irurtzun, A. (2016). Pitch accent tonal alignment in declarative sentences in the Spanish of the Basque Country: A study of language contact. In M. Armstrong, N. Henriksen & M. del M. Vanrell (eds.), *Intonational grammar in Ibero-Romance: Approaches across linguistic subfields*, 25-44. Amsterdam: John Benjamins.
- Elordieta, G. & Romera, M. (accepted). The influence of social factors on the prosody of Spanish in contact with Basque. In M. Bouzouita, R. Enghels and C. Vanderschueren (eds.), *Different perspectives on convergence and divergence in Ibero-Romance: language contact and contrasting national varieties*. Amsterdam: John Benjamins.

- Estebas-Vilaplana, E., & Prieto, P. (2008). Castilian Spanish Intonation. *Estudios de Fonética Experimental*, 17, 265–283.
- Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, 91, 181–201.
- Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of Memory and Language*, 89, 110–137.
- Fricke, M., Nadeu, M., & Maslowski, M. (2016). Prosodic organization of spontaneous Spanish-English bilingual speech. Poster presented at *Laboratory Phonology 15*. Ithaca, NY.
- Google. (n.d.). [Google maps view of northern Bizkaia].
- Gullifer, J. W., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology*, 4, 1–13.
- Hualde, J. I., Elordieta, G., & Elordieta, A. (1994). *The Basque dialect of Lekeitio*. Bilbao and Donostia-San Sebastián: Editorial Service of the University of the Basque Country and Diputación de Gipuzkoa.
- Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge University Press.
- Jun, S. J., & Elordieta, G. (1997). Intonational structure of Lekeitio Basque. In A. Botinis, G. Kouroupetroglou, & G. Carayannis (Eds.), *Intonation: Theory, models, and applications* (pp. 193–196). Proceedings of ESCA ETRW: Athens.
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Ladd, R. (1996). *Intonational phonology*. Cambridge University Press.
- Lantto, H. (2012). Grammatical code-switching patterns of early and late Basque-Spanish bilinguals. *Sociolinguistic Studies*, 6, 21–44.
- Lantto, H. (2014). Code-switching, swearing and slang: The colloquial register of Basque in Greater Bilbao. *International Journal of Bilingualism*, 18(6), 633–648.
- Lantto, H. (2015). Conventionalized code-switching: Entrenched semantic-pragmatic patterns of a bilingual Basque-Spanish speech style. *International Journal of Bilingualism*, 19(6), 1–16.
- Lee, B., & Sidtis, D. V. L. (2017). The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. *Speech, Language and Hearing*, 1–12.

- Lennes, M. (2004). Script for f0 measurement. Retrieved from <http://www.helsinki.fi/~lennes/praat-scripts/>
- Lindsey, B. (2006). Language loss and recovery in heritage language learner Spanish: A study on code-switching. *Divergencias: Revista de Estudios Linguisticos y Literarios*, 4(1), 3–22.
- López, L. (2017). Code-Switching and linguistic theory. Paper presented at *Bilingualism in the Hispanic and Lusophone World*, Tallahassee, Florida.
- MacSwan, J. (1999). A Minimalist approach to intrasentential code-switching (Unpublished doctoral dissertation). University of California, Los Angeles.
- Meuter, R. F. I., & Allport, A. (1999). Bilingual Language Switching in Naming : Asymmetrical Costs of Language Selection. *Journal of Memory and Language*, 40, 25–40.
- Morris Academy. (2005). Morris Student Plus (Standard Basque-English dictionary). Retrieved March 8, 2017, from <http://www1.euskadi.net/morris/>
- Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Clarendon Press.
- Nakazawa, M. (2015). CRAN - Package fmsb. Retrieved March 8, 2017, from <https://cran.r-project.org/web/packages/fmsb/index.html>
- Nibert, H. J. (2000). Phonetic and phonological evidence for intermediate phrasing in Spanish intonation (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.
- Olson, D. J. (2012). The phonetics of insertional code-switching: Suprasegmental analysis and a case for hyper-articulation. *Linguistic Approaches to Bilingualism*, 2(4), 439–457.
- Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41(6), 407–420.
- Olson, D. J. (2016a). The role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association*, 46(3), 1–23.
- Olson, D. J. (2016b). The impact of code-switching, language context, and language dominance on suprasegmental phonetics: Evidence for the role of predictability. *International Journal of Bilingualism*, 20(4), 453–472.
- Olson, D. J. (2016c). The gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics*, 37(3), 725–756.

- O'Rourke, E. (2004). Peak Placement in Two Regional Varieties of Peruvian Spanish Intonation. In Julie, J., Clancy Clements, & B. Vance (Eds.), *Selected papers from the 33<sup>rd</sup> Linguistic Symposium on Romance Languages* (pp. 321–341). Amsterdam: John Benjamins.
- Pan, Y., & Jackson, R. T. (2008). Ethnic difference in the relationship between acute inflammation and serum ferritin in US adult males. *Epidemiology and Infection*, 136(3), 421–31.
- Park, H.-S. (2006). Structural characteristics of proper nouns in Korean-Swedish discourse. *International Journal of Bilingualism*, 10(1), 17–36.
- Piccinini, P. E., & Garellek, M. (2014). Prosodic cues to monolingual versus code-switching sentences in English and Spanish. *Proceedings of the 7th Speech Prosody Conference*, 885–889. Trinity College, Dublin.
- Piccinini, P., & Arvaniti, A. (2015). Voice onset time in Spanish-English spontaneous code-switching. *Journal of Phonetics*, 52, 121–137.
- Pierrehumbert, J. (1980). The phonology and phonetics of English intonation (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en Espanol. *Linguistics*, 18(7/8), 581–618.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26(1), 47–104.
- Prieto, P., & Roseano, P. (Eds.). (2010). *Transcription of intonation of the Spanish language*. Lincom Europa.
- R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from [www.R-project.org/](http://www.R-project.org/)
- Raftery, A. (1995). Bayesian model selection in social research. *Sociological Methodology*, (25), 111–163.
- Romera, M & Elordieta, G. (2013). Prosodic accommodation in language contact: Spanish intonation in Majorca. *International Journal of the Sociology of Language*, 221, 127-151.
- Simonet, M. (2008). Language contact in Majorca: An experimental sociophonetic approach (Unpublished doctoral dissertation). University of Illinois, Urbana-Champaign.
- Simonet, M. (2011). Intonational convergence in language contact: Utterance-final F0 contours in Catalan–Spanish early bilinguals. *Journal of the International Phonetic Association*, 41(2), 157–184.

Sosa, J. M. (1999). *La entonación del español: Su estructura fónica, variabilidad y dialectología*. Cátedra: Madrid.

Trask, R. L. (1997). *The history of Basque*. Routledge: London and New York.

Willtron. (2006). Euskal Herria mapa mudo. Retrieved from [https://commons.wikimedia.org/wiki/File:Euskal\\_Herria\\_mapa\\_mudo.svg](https://commons.wikimedia.org/wiki/File:Euskal_Herria_mapa_mudo.svg)  
[https://commons.wikimedia.org/wiki/File:Euskal\\_Herria\\_mapa\\_mudo.svg](https://commons.wikimedia.org/wiki/File:Euskal_Herria_mapa_mudo.svg)

Wilson, D. (2017). How Spanish-English bilinguals produce code-switched sentences: an acoustic analysis. Poster presented at *Bilingualism in the Hispanic and Lusophone World*, Tallahassee, Florida.

Zuazo, K., & Branton, A. (2013). *The dialects of Basque*. Center for Basque Studies: Reno.