THE PERCEPTION AND PRODUCTION OF PROMINENCE IN SPANISH
BY HERITAGE SPEAKERS AND L2 LEARNERS

BY

JI YOUNG KIM

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Spanish
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral committee:

Professor José Ignacio Hualde, Chair
Professor Silvina Montrul
Associate Professor Chilin Shih
Associate Professor Rebecca Foote

ABSTRACT

Heritage speakers (i.e., descendants of immigrants that speak an ethnic minority language in a society where a different language is spoken as the majority language) are linguistically a unique population, because, while their early and immersed exposure to the heritage language puts them in an advantageous position with regard to their linguistic knowledge of the language, compared to those who learned it as adults, lack of intergenerational transmission is evident in this population, as demonstrated in the shift to the majority language of the society. Recently, the attempt to examine heritage speakers' linguistic knowledge from an acquisitional point of view has increased significantly, which has led to the creation of "heritage language acquisition" as an independent field of study. However, compared to other linguistic subfields, such as syntax and morphology, phonology has been an understudied area in heritage language research. Although phonology is the field in which heritage speakers have the most noticeable advantage over second language (L2) learners, their speech is nonetheless often perceived as accented by monolingual native speakers, which makes their speech unique in its own right. Thus, research on heritage language phonology is needed to gain a comprehensive understanding of heritage speakers' linguistic system.

This study examines U.S. Spanish heritage speakers' perception and production of two types of prosodic prominence in Spanish: word-level prominence (i.e., lexical stress), marking paradigmatic contrast (e.g., *Canto*. 'I sing.' vs. *Cantó*. 'He/She/You (formal) sang.'), and sentence-level prominence (i.e., nuclear stress), marking information status and focus (e.g., *¿Qué hizo Mariana? - Mariana cantó.* 'What did Mariana do? - Mariana sang.' vs. *¿Quién cantó? - Cantó Mariana.* 'Who sang? - Mariana sang.'). Although lexical stress and nuclear stress exist in both Spanish and English, due to cross-linguistic differences between the two languages, English L2 learners are found to have great difficulties acquiring them in Spanish. The overarching goal of this study is to examine whether transfer from English (i.e., the majority language) is also observed in heritage speakers.

Spanish monolingual native speakers, Spanish heritage speakers, and English L2 learners of Spanish participated in four experimental studies: two forced-choice identification tasks for the perception of lexical stress and nuclear stress, a reading aloud task for the production of lexical stress, and a simulated interactive elicitation task for the production of nuclear stress. Results showed that, while the heritage speakers performed similarly to the monolingual speakers in the perception of lexical stress and nuclear stress, they showed a deviant pattern in their production, such as early alignment of f0 peak and elongation of unstressed final vowels in the production of paroxytones (lexical stress), and early alignment of f0 peak in the production of focused constituents (nuclear stress). Heritage speakers' discrepancy between perception and production is likely to be due to asymmetry in their use of Spanish. That is, heritage speakers speak Spanish much less frequently than they hear it. The L2 learners, on the other hand, showed divergent patterns from the monolingual speakers in both the perception and the perception. This suggests that, thanks to early exposure to the heritage language, heritage speakers have an advantage over L2 learners in their perception of heritage language speech sounds, but such an advantage is not guaranteed in their production. As nuclear stress is higher in the stress hierarchy and acquired later in life than lexical stress, the present study predicted that heritage speakers' use of nuclear stress would be affected to transfer from English to a larger degree than lexical stress. However, the nuclear stress results bore unexpected results that are not necessarily phonological in nature (e.g., bias toward focus on subject, use of cleft constructions to mark focus), making it difficult to make direct comparisons between the two linguistic features. Possible explanations to the unexpected results and suggestions for future research are presented.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank everyone who has given me immense support in completing this dissertation. I owe my deepest gratitude to my adviser Professor José Ignacio Hualde. Without his guidance, immediate feedback, continuous optimism (with a touch of linguistic humor), and his encouragement whenever I felt doubtful about my work, this dissertation would not have been possible. He is the best adviser a student could ask for and my true academic role model. I am also extremely grateful to my dissertation committee, Professor Silvina Montrul, Professor Chilin Shih, and Professor Rebecca Foote, who, not only gave me constructive criticisms for the improvement of my dissertation, but also made me feel truly supported, which helped me immensely in completing my dissertation. It was an honor to work with the greatest experts in heritage/second language acquisition and phonology. I would also like to thank Professor Anna María Escobar, who has supported me from the first day of graduate school and who is one of the biggest contributors that helped me get this far. I am also in debt to Professor Jaehak Lee (Korea University) for encouraging me to apply for graduate schools and further pursue my research interests in linguistics.

This dissertation would not have been possible without my participants, who not only willingly participated in my study, but also helped me recruit more participants and provided me with fresh new insights on bilingual speakers. I would like to express my deepest appreciation to Professor Eduardo Velázquez (Universidad Autónoma de Querétaro) for always receiving me with open doors whenever I visited Mexico to collect data and for permission to use his lab. I would also like to thank my research assistants, Amy Klobuchar and William Vélez, for helping me coding the data. Without their help, this dissertation would have been a much lengthier process. The travels to Mexico, data collection, and data analyses were made possible thanks to the financial support by the National Science Foundation Doctoral Dissertation Research Improvement Grant, the School of Literatures, Cultures, and Linguistics Dissertation Completion Fellowship, the Tinker Foundation Field Research Grant, and the Graduate College Dissertation Travel Grant.

Although completing this dissertation was challenging, I would not have done it any other way, because during the process I have met so many wonderful people whom I consider family now. I cannot express more how thankful I am to my Korean friends in Champaign, Sun Joo Chung, Hannah Kim, Eun Hee Kim, Sarah Choi, Joonhan Kim, Sylvia Choi, and Nayoung Kim, for making Champaign a second home. They are the people with whom I spent most of my time and shared so many unforgettable memories. They have helped me tremendously in feeling less homesick whenever I missed Korea. I also owe a great debt of gratitude to Ronald Choe who always makes me laugh with his sarcastic humor (which I love!) and inspires me in so many ways and to Itxaso Rodriguez for accompanying me during this entire process and keeping my sanity. I cannot imagine having completed this disssertation without her. Many thanks to Tuan Lam, Jeriel Melgares, Krista Evenson, Nicole Wong, Israel de la Fuente, Luján Stasevicius, Justin Davidson, and Daniel Scarpace for making my life as a graduate student so much more enjoyable. I also want to thank my Daewon friends, Eun Joo Park, Shin Young Park, Eun Hye Jeon, Seung-hye Hong, and Soyeon Ku for staying in touch with me even after being away for a long period of time and making me feel like I have never left home.

The biggest contributors to my dissertation, without any doubt, are my family, Edward Kim, Sungwon Ro, Sun Hee Kim, and David Kim, who have given me unconditional support and never stopped believing in me. This dissertation is dedicated to them.

TABLE OF CONTENTS

Chapter 1 Introduction

1.1 The situation of Spanish and other immigrant languages in the U.S.

Over the past few decades, the immigrant population in the U.S. has grown exponentially with the number of foreign-born population increasing from 9.7 million in 1960 (i.e., 5.4% of the total population) to 40 million in 2010 (i.e., 12.9 % of the total population), fueled primarily by Hispanic immigrants comprising more than half of all foreign-born population (Acosta & de la Cruz, 2011; Grieco et al., 2012). Now, Hispanics are the largest minority population in the U.S. (i.e., 17% of the total population) (Stepler & Brown, 2015). As a result, Spanish became by far the most spoken non-English language at home, outnumbering Chinese, the second most spoken non-English language, by more than 13 times more (Gonzalez-Barrera & Lopez, 2013; Lopez & Gonzalez-Barrera, 2013; Ryan, 2013). Despite recent decline of Hispanic immigrants since 2010, Hispanic population is expected to continue growing, as U.S.-born Hispanics became the new primary source of Hispanic population growth (Krogstad & Lopez, 2013; Taylor et al., 2012; Stepler & Brown, 2015). The number of U.S.-born Hispanics has increased from 21.1 million in 2000 to 34.9 million in 2013 (i.e., 65% increase), while the number of foreign-born Hispanics has increased from 14.1 million to 18.8 million over the same time period (i.e., 33% increase) (Stepler & Brown, 2015).

However, it is uncertain whether the vitality of Spanish will continue with the changing demographics of the Hispanic population, as first generation Hispanic immigrants are the ones who are serving as “a replenishing source of speakers lost through generations” (Carreira, 2013). Lack of intergenerational transmission is clearly demonstrated in the shift of language dominance from Spanish to English and more use of English in later generations (Brown & Patten, 2014; Hakimzadeh & Cohn, 2007; Hurtado & Vega, 2004; Rumbaut et al., 2006; Taylor et al., 2012). According to Taylor et al. (2012), although nearly all Hispanics acknowledge the importance of preserving the Spanish language for future generations, in reality, Spanish proficiency has greatly reduced across the generations. 91% of first generation Hispanics reported that they speak Spanish very well/pretty well, while 82% of second generation Hispanics and only 47% of third generation Hispanics responded that way. The gap in language dominance across the generations is highly related to asymmetrical use of the home language, which is frequently observed in immigrant homes. That is, foreign-born parents would speak to their U.S.-born children mostly in the home language, while the children would speak to their parents mostly in English (Hurtado & Vega, 2004; Kim, 2010). Such an asymmetrical pattern of language use becomes possible when "predominantly Spanish-speaking parents understand English, and when predominantly English-speaking children understand Spanish", leading to receptive bilingualism (Hurtado & Vega, 2004; Veltman, 1983). Indeed, large scale surveys on Hispanic immigrants report that for first generation Hispanic immigrants who were born outside of the U.S., Spanish is the pervasive language used at home, while their U.S.-born children are much more likely to speak English than Spanish, although they are bilingual speakers of the two languages (Brown & Patten, 2014; Hakimzadeh & Cohn, 2007; Taylor et al., 2012).

Apart from the gap in language dominance and use across the generations, the loss of intergenerational transmission may also be due to the pressure that immigrants perceive to assimilate to the U.S. mainstream culture, which is English monolingualism (Fuller, 2012; Stevens, 1992). Unlike in Europe where bilingualism is seen favorably, in the U.S., speaking a language other than English as a home language is considered to be a disadvantage, because it marks a person as "outside of the mainstream and not a participant in the ideal trajectory of upward mobility in terms of income and social status" (Fuller, 2012). As bilingual children are constantly forced to negotiate between the majority culture and language, on the one hand, and the minority culture and language, on the other, during their identity formation, they end up conforming to the majority culture and language, as these are considered to be more prestigious and of higher status (Brown, 2003, 2009; Tse, 1998a).

Rumbaut et al. (2006) projected the linguistic life expectancy (measured by the ability to speak the ancestral language very well and the preference of using that language at the household) of Spanish and other minority languages spoken in the U.S. based on two large-scale surveys conducted in the 2000s on different generations of immigrant population residing in Southern California (i.e., the Immigration and Intergenerational Mobility in Metropolitan Los Angeles and the third wave of the Children of Immigrants Longitudinal study in San Diego). According to Rumbaut et al. (2006), despite greater retention of Spanish compared to other minority languages through the second generation, it is expected that the Spanish language will eventually "die out" by third generation of U.S. residence. The situation seems less fortunate for other minority languages. For instance, in the case of Asian languages (e.g., Chinese, Korean, Vietnamese), shift to English seems to occur more rapidly, at or near the second generation (Cho, 2000; Hing & Lee, 1996; Rumbaut et al., 2006). As Southern California is known as "the country's largest concentration of immigrants", the trend found in this study can be extended to other regions in the U.S. (Rumbaut et al., 2006), although the shift toward English may be less salient in rural areas, because the need for English is less pressing in these areas (Hurtado & Vega, 2004; Ramírez, 1988). Thus, with regard to the vitality of the minority languages in the U.S., Rumbaut et al. (2006) stated as follows:

"Historical and contemporary evidence indicates that English has never been seriously threatened as the dominant language of the United States and that – with well over 200 million monolingual English speakers – it is certainly not threatened today, not even in Southern California. What is endangered instead is the survival of the non-English languages that immigrants bring with them to the United States."

1.2 Learning/Re-learning the heritage language

In the U.S., the term "heritage language speaker" was introduced in the late 1990s when the National Standards in Foreign Language Education Project first released the "Standards for Foreign Language Learning" in 1996, which supports the study, maintenance, and revitalization of non-English languages in the U.S. (Pérez, 2010; Valdés, 2001). Simply speaking, heritage language speakers, or "heritage speakers" in short, are descendants of immigrants that speak an ethnic minority language in a society where a different language is spoken as the majority language. However, as it encompasses a broad spectrum of individuals of different background, such as nativity, generation, level of proficiency of the heritage language, age and context of acquisition of the heritage language, type of input received, schooling, etc. (Montrul, 2008; Valdés, 2001), the term has been used with varying definitions, depending on the perspective one adopts (e.g., second language acquisition, sociolinguistics, education policy) (Kondo-Brown, 2003; Polinsky & Kagan, 2007). The following is a list of definitions of heritage speakers that are commonly referenced in the literature:

Author(s)	Definition of heritage speakers
Valdés (2001)	"[An individual] who is raised in a home where a non-English language is spoken, who speaks or at least understands the language, and who is to some degree bilingual in that language and in English".
Polinsky and Kagan (2007)	"[P]eople raised in a home where one language is spoken who subsequently switch to another language".
Fishman (1991)	"[Individuals who speak] any ancestral language such as indigenous, colonial, and immigrant languages, […which] may or may not be a language regularly used in the home and the community".
Hornberger and Wang (2008)	"[I]ndividuals who have familial or ancestral ties to a particular language that is not English and who exert their agency in determining whether or not they are H[eritage] L[angauge]L[earner]s of that H[eritage]L[anguage] and H[eritage]C[ulture]".

Table 1 Definitions of heritage speakers in the literature

Due to the heterogeneous nature of heritage speakers (Montrul, 2008; Valdés, 2001), this dissertation takes a rather narrow definition of heritage speakers and limits its scope to first generation U.S.-born heritage speakers with foreign-born parents (i.e., individuals of Depth 4, in Hakuta and D'Andrea's (1992) study). Several characteristics of these speakers are stated below:

Generally speaking, heritage speakers are early bilingual speakers of English (i.e., majority language) and a home language (i.e., minority language) who were born and raised in the U.S.. The parents of these speakers are native speakers of the heritage language, thus, heritage speakers acquire this language as the first language (L1) and, for many of them, systematic exposure to English, which is their second language (L2), does not occur until they enter institutional settings (e.g. daycare, kindergarten). Since English is the majority language of the society and the heritage language is a minority language, heritage speakers' use of English increases as they grow up, whereas their use of the heritage language becomes limited to familial settings. This subsequently results in a gradual shift of language dominance from the heritage language (L1) to English (L2). Therefore, while heritage speakers have very strong command of English, their command of the heritage language is usually short of the native speaker level of their parents or peers raised in their home countries (Montrul, 2008; Valdés, 2001).

Despite the language shift to English evidenced across the generations, recently there has been a growing interest among heritage speakers who aspire to maintain or improve their linguistic skills of the home language (Lynch, 2014; Oh & Nash, 2014). Stevens (1992) argued that it is low fluency in English and dependence on the minority language, rather than speaking a minority language *per se* that may result in educational and economic disadvantages. In fact, being fluent in a minority language along with English can be “a key to academic and economic success in the United States” (Stevens, 1992). The learning/re-learning of the heritage language is appreciated especially among adult heritage speakers, as there are countless academic and economic benefits of being fluent in both English and a minority language (e.g., more job opportunities, enhancement of professional success) (Grosjean, 2010; Stevens, 1992; Tadmor et al., 2012). Thus, the more socioeconomic value a minority language has in the global economy and workplace, the more the heritage speakers will have positive attitude toward that language and stronger motivation to learn/re-learn it (He, 2006; Heller, 2007; Lynch, 2014; Martínez & Schwartz, 2012; Stevens, 1992). For instance, with regard to Spanish and Chinese, the two most spoken non-English languages in the U.S. (Gonzalez-Barrera & Lopez, 2013; Lopez & Gonzalez-Barrera, 2013; Ryan, 2013), heritage language learners tend to show a high degree of motivation to study their home language as these languages have become very popular as L2s in the U.S. (Beaudrie & Ducar, 2005; He, 2006; Lynch, 2014). Stevens (1992) stated the critical factors that contribute to a continued use of the heritage language as follows:

“What language(s) people use in their daily lives is affected by the incentives and opportunities to use one or another language with other people. The relative frequency with which adult Americans with a non-English first language use English vis-à-vis their non-English language is the outcome of two sets of factors. The first set includes indicators of the resources and incentives encouraging or allowing them to use English. The second set includes the demographic context that underlies the opportunities for non-English Americans to participate in social situations in which their minority language is a possible means of communications. [… A] minority language survives in a setting dominated by a majority language only if the minority language is used often as the means of communication.”

Martínez and Schwartz's (2012) study provides a good example of Stevens' (1992) argument. Martínez and Schwartz (2012) conducted a survey on Hispanic medical students participating in the Medical Spanish for Heritage Learners (MSHL) program, which requires completion of an internship, as part of the curriculum, in a local community health center where the majority of the patients are undocumented Hispanics and have limited proficiency in English. They found that the MSHL students "gained a heightened awareness" that Spanish was the language prevalently used among the patients and that there is a huge demand for Spanish-speaking health care providers. Realizing the usefulness of the home language in professional settings is likely to lead to stronger commitment to maintaining that language (Cho, 2000; Martínez & Schwartz, 2012; Stevens, 1992).

Apart from socioeconomic values of the heritage language, individual values, such as identity formation, also play an important role in heritage speakers' desire to learn/re-learn their home language (Campbell & Rosenthal, 2000; Cho, 2000; He, 2006; Lacorte & Canabal, 2003; Tallon, 2011; Tse, 1998b). As language takes a significant part in forming one's ethnic identity (Carreira, 2004; Cho, 2000; Guardado, 2014; Brown, 2009; Phinney et al., 2001; Tse, 1998b), heritage language learners decide to study their heritage language to stay connected to their ethnic group and culture, and have a better sense of who they are, through communicating and socializing with family members, with members of the heritage language community, and with speakers of the heritage language in the home country (Carreira, 2004; Cho, 2000; Lacorte & Canabal, 2003; Brown, 2009; Tse, 1998b). Indeed, studies have shown that heritage speakers who have a better command of their heritage language tend to have stronger ethnic identity than those who do not (Cho, 2000; Brown, 2009; Phinney et al., 2001). Therefore, it is unlikely that a complete shift to English monolingualism (i.e., death of the minority language) will occur among the U.S. immigrant population. Rather, there will be "different degrees of bilingualism, including receptive" (Hurtado & Vega, 2004).

Due to growing interest in learning/re-learning the home language, the demand for language classrooms designed for heritage language learners has increased greatly (Beaudrie & Ducar, 2005). As heritage speakers grow up in an immigrant household, in which a non-English language is spoken, they receive earlier and more immersed exposure to that language, which put them in a more advantageous position with regard to their linguistic knowledge of the heritage language, compared to those who started learning it as adults (Au et al., 2008; Au & Romo, 1997;

Polinsky & Kagan, 2007; Oh et al., 2003). However, it is yet questionable whether heritage language learners can be placed in the same classroom as L2 learners and, if not, what materials and teaching methods should be implemented that meet the linguistic needs of these learners (Peyton et al., 2001). In order to answer these questions, it is important to have a better understanding of the characteristics of heritage speakers' linguistic system and how they acquire the heritage language (Polinsky, 2008; Putnam & Sánchez, 2013). Recently, the attempt to examine heritage speakers' linguistic knowledge from an acquisitional point of view has increased significantly, which has led to the creation of "heritage language acquisition" as an independent field of study (Lynch, 2014; Polinsky, 2008). Heritage speakers are linguistically a unique population, thus, examining their linguistic system may provide an insight to finding "a crucial missing link between competent L1 learners, balanced bilinguals, and possibly L2 learners" (Polinsky, 2008).

1.3 General characteristics of the heritage speakers' linguistic knowledge

In the field of heritage language acquisition, heritage speakers are often compared with L2 learners (Lynch, 2014), with the idea that heritage speakers will perform better than L2 learners, yet worse than monolingual native speakers. However, it is important to not position heritage speakers simply somewhere in between the L2 learners and monolingual native speakers, because heritage speakers may differ from L2 learners not only quantitatively, but also qualitatively.

Despite being heterogeneous (Montrul, 2008; Valdés, 2001), there are several commonly observed characteristics of heritage speakers with the regard to their linguistic abilities of the heritage language that are quite different from L2 learners. The most noticeable feature in heritage speakers' linguistic abilities is that, among various aspects of grammar, heritage speakers generally have native- or near-native-like proficiency in phonology, while they have limited competence in morphosyntax, semantics, and pragmatics (Au et al., 2008; Campbell & Rosenthal, 2000; Ilieva, 2012; Martin et al., 2013; Montrul, 2012; 2013; Oh et al., 2003; Peyton et al., 2001; Polinsky & Kagan, 2007). Indeed, heritage speakers' native-like pronunciation is one of the main reasons that they are assigned to different tracks in the classroom from L2 learners who generally have a noticeable foreign accent (Campbell & Rosenthal, 2000; Peyton et al., 2001; Polinsky & Kagan, 2007). As foreign accents are detected as early as sometime between 4 and 6 years of age (Birdsong, 1999; Dupoux et al., 2010; Flege, 1991, 1995; Gildersleeve-Neumann et al., 2009; Long, 1990; Pallier et al., 1997; Strange, 1995), heritage speakers have an advantage over L2 learners, thanks to their earlier exposure to the heritage language.

It is also commonly noted that heritage speakers are better at oral skills than literacy skills (Campbell & Rosentha, 2000; Fishman, 2001; Klee, 1998; Montrul, 2011; Quintanar-Sarellana et al., 1993; Rumbaut et al., 2006; Tallon, 2011; Taylor et al., 2012; Valdés, 2001). In a survey on the language use among U.S. Hispanics, Taylor et al. (2012) reported that 82% of U.S.-born children of immigrants (i.e., heritage speakers) responded that they can “carry on a conversation in Spanish”, while 71% of them responded that “they can read a newspaper or book in Spanish”. Heritage speakers’ lack of literacy skills was also found in a linguistic study examining Spanish heritage speakers’ and L2 learners’ morphology in Spanish (Montrul, 2011). When testing three aspects of Spanish morphology (i.e., gender agreement, differential object marking, and tense-aspect-mood) in two domains (i.e., oral vs. written), Montrul (2011) found that Spanish heritage speakers performed significantly better in oral tasks than in written tasks, while L2 learners showed an opposite pattern. This is likely to be due to different language experience and practice that these two groups have, because heritage speakers usually use the heritage language orally, while L2 learners use it in classroom settings in which the emphasis is given on written skills (Campbell & Rosenthal, 2000; Montrul, 2011).

Another important feature of heritage speakers’ linguistic abilities is that heritage speakers tend to comprehend the heritage language much better than speaking it (Au et al., 2008, Hulsen, 2000; Hurtado & Vega, 2004; Lynch, 2014; Montrul & Foote, 2014; Oh et al., 2013). In a study with Dutch immigrants in New Zealand, Hulsen (2000) compared Dutch heritage speakers’ performance in two tasks, a picture naming task, which tested participants’ productive language processes, and a picture-word matching task, which tested participants’ receptive language processes. The results showed that, although second generation Dutch immigrants (i.e., Dutch heritage speakers) performed worse than first generation immigrants in both tasks, the gap between the accuracy rate of the two groups was larger in the picture naming task (i.e., production) (first generation: 89%, second generation: 52%) than in the picture-word matching task (i.e., perception) (first generation: 98%, second generation: 90%), confirming that heritage speakers have better perception skills than production skills.

This asymmetry is likely to be due to the gap of heritage language use between the two domains. Despite the majority language of the society being English, at home heritage speakers are constantly exposed to the heritage language from early on in their lives, due to their parents and other household adults (e.g., grandparents, aunts, uncles) who talk to the heritage speakers or to each other in the heritage language (Beaudrie & Ducar, 2005; Brown & Patten, 2014; Hakimzadeh & Cohn, 2007; Hakuta & D'Andrea, 1992; Hurtado & Vega, 2004; Kim, 2010; Pearson, 2007; Potowski, 2004). Therefore, heritage speakers have "ample opportunities to overhear the [heritage] language" (Beaudrie & Ducar, 2005). However, compared to the amount of input they receive in the heritage language, their production of the language is less frequent. For instance, in a thorough survey of over 800 Hispanic high school and college students in Chicago, Potowski (2004) found that those who lived in the U.S. for 12 years or more, 82.9% of which were U.S.-born, spoke to their parents in Spanish 64.6% of the time, while their parents spoke to them in Spanish 76.9% of the time. Such a trend also found with other household adults (to uncle: 71.6%, from uncle: 76%; to aunt: 69.1%, from aunt: 76%; to grandmother: 88.4%, from grandmother: 89.9%; to grandfather: 76%, from grandfather: 78%). The use of Spanish is much less frequent with other family members of the same generation or outside of familial settings (siblings: 30%, cousins: 49.5%, friends: 29.2%) (Potowski, 2004). Au et al. (2008) argued that heritage speakers' production skills are highly related to whether they have regularly spoken the heritage language. In a study with Korean heritage speakers, Oh et al. (2013) divided the heritage speakers into two groups, childhood hearers who regularly heard Korean during childhood, but spoke the language minimally, and childhood speakers who not only regularly heard Korean during childhood, but also spoke it regularly, and compared their perception and production of denti-alveolar Korean stop consonants (i.e., /t/, /t^h/, and /t'/). Results showed that while childhood hearers only outperformed novice learners in the perception task, childhood speakers outperformed the childhood hearers and novice learners in both perception and production tasks. This finding supports the importance of regular use, not only in the perception, but also in the production, of the heritage language.

Apart from the three characteristics commonly observed in heritage speakers' linguistic abilities stated above (i.e., more native-like phonological grammar than other grammatical aspects, better oral skills than literary skills, and better perception skills than production skills), there are many other features that heritage speakers have that are different from L2 learners, such as the motivation of heritage language learning, knowledge of the heritage culture, and so on. Considering how different heritage speakers are from L2 learners, it is important to compare the two groups with caution and acknowledge them as separate groups, rather than simply positioning them along the same continuum. The asymmetries found in heritage speakers' linguistic abilities may be due to reduced input and use of the heritage language (Godson, 2004; Mikulski, 2010), leading to an "incomplete acquisition" of the heritage language (Montrul, 2012) or "fossilization" of incorrect language forms (Ilieva, 2012; Selinker, 1972). However, it is still unclear exactly how much language input and use are required to reach to a native-like proficiency (O'Grady et al., 2011) and whether they are the only main factors (Putnam & Sánchez, 2013).

1.4 Goals and outline of the dissertation

Heritage language acquisition is still an emerging field and not much research has been done in this area to meet the growing interest. Especially, there is scarce research on heritage speakers' phonological system, as studies in heritage language acquisition are mostly focused on syntax and morphology (Benmamoun et al., 2010, 2013; Polinsky & Kagan, 2007; Kupisch et al., 2014; Montrul, 2010). This is likely to be due to the fact that heritage speakers sound native-like, leading one to think that their phonological system would be the same as monolingual native speakers (Polinsky & Kagan, 2007). However, although it is true that heritage speakers have a more native-like pronunciation compared to L2 learners, this does not necessarily mean that the pronunciation of these two groups are indistinguishable, as anecdotal evidence has shown that heritage speakers' speech is perceived by monolingual native speakers as "'funny', 'off', and not like 'real' speakers of the language" (Polinsky & Kagan, 2007). Indeed, several foreign accent rating studies have shown that heritage speakers were perceived to have a stronger foreign accent than monolingual native speakers (Au et al., 2002; Au et al., 2008; Knightly et al., 2003; Kupisch et al., 2014). Thus, Polinsky and Kagan (2007) emphasized that "instrumental studies targeting the phonetics of heritage speech are badly needed, [since] virtually nothing is known about the nature of phonological representations in heritage speakers". It is important to take into account both segmental and suprasegmental features to understand the accent that is unique to heritage speakers. However, despite their immense value in the paucity of research on heritage language phonology, most studies have only focused on heritage speakers' segmental features (Amengual, 2012; Au et al., 2008; Boomershine et al., 2013; Chang et al., 2011; Godson, 2004; Knightly et al., 2003; Oh et al., 2003; Rao, 2015; Ronquest, 2013; Willis, 2005; Yang et al., 2015), and few studies have investigated heritage speakers' speech at the suprasegmental level (Gries & Miglio, 2014; Harris et al., 2015; Robles-Puente, 2014).

This dissertation examines heritage speakers' phonological system at the suprasegmental level, focusing on Spanish heritage speakers' perception and production of two types of prominence in Spanish: word-level prominence (i.e., lexical stress), marking paradigmatic contrast (e.g., *canto* 'I sing' vs. *cantó* 'he/she/you (formal) sang'), and sentence-level prominence (i.e., nuclear stress), marking information status and focus (e.g., *¿Qué hizo Mariana? - Mariana cantó.* 'What did Mariana do? - Mariana sang.' vs. *¿Quién cantó? - Cantó Mariana.* 'Who sang? - Mariana sang.'). The overarching goal of this study is to examine whether Spanish heritage speakers have their Spanish phonology affected by the shift of language dominance from Spanish to English. Spanish heritage speakers' performance is compared to those of Spanish monolingual speakers and English L2 learners of Spanish to see whether they show similar/different patterns from these two groups. It is of interest to examine how heritage speakers perform compared to monolingual native speakers, with whom they share the same L1 (i.e., Spanish), but differ in the dominant language (heritage speakers: English; monolingual speakers: Spanish), and English L2 learners of Spanish, with whom they share the same dominant language (i.e., English), but differ in the L1 (heritage speakers: Spanish; L2 learners: English). Moreover, as nuclear stress is higher in the stress hierarchy and acquired later in life than lexical stress (Vogel & Raimy, 2002; Wieman, 1976), this dissertation investigates whether the heritage speakers' use of nuclear stress is more strongly affected due to the language shift to English than lexical stress. Lastly, this study compares Spanish heritage speakers' perception and production with the intention to see whether discrepancy is found between the two domains, given that Spanish heritage speakers hear Spanish more frequently than they speak it.

This dissertation is divided into five chapters. Chapter 1, the present chapter, presents an overview of the situation of the minority languages in the U.S., with special focus on Spanish, and heritage language acquisition as an emerging field of study. Chapter 2 reviews previous studies on heritage language phonology and the theories and findings related to the acquisition of phonology. The following two chapters describe the research design and results of Spanish heritage speakers' perception and production of lexical stress (Chapter 3) and nuclear stress (Chapter 4). Finally, Chapter 5 summarizes the findings of Chapter 3 and 4, and provides a general discussion and future direction.

Chapter 2 Literature review

2.1 The “heritage” accent

Compared to other linguistic subfields, such as syntax and morphology, phonology has been an understudied area in heritage language research (Benmamoun et al., 2010, 2013; Kupisch et al., 2014; Montrul, 2010; Polinsky & Kagan, 2007). Although phonology is the field in which heritage speakers have the most noticeable advantage over L2 learners, heritage speakers’ speech is nonetheless perceived as accented by monolingual native speakers, which makes their speech unique in its own right (Benmamoun et al., 2010, Godson, 2004; Polinsky & Kagan, 2007).

Au and colleagues (Au et al., 2002; Au et al., 2008; Knightly et al., 2003; Oh et al., 2003) investigated the degree of foreign accent in the speech of heritage speakers of Spanish and Korean in the U.S., and compared it with that of native speakers and L2 learners. In the case of Spanish heritage speakers (Au et al., 2002; Au et al., 2008; Knightly et al., 2003), Au and colleagues focused on heritage speakers’ production of Spanish stop consonants which are differentiated from the ones in English with regard to voice onset time (VOT) (i.e., Spanish /p, t, k/ with short lag-VOT vs. English /p, t, k/ with long-lag VOT; Spanish /b, d, g/ with voice-lead VOT vs. English /b, d, g/ with short-lag or voice-lead VOT) and the realization of intervocalic voiced stop consonants (i.e., lenition in Spanish vs. no lenition in English). Their results showed that while heritage speakers received significantly higher ratings than L2 learners of Spanish, their ratings were significantly lower than those of the native Spanish speakers. Similar patterns were also found in the global accent rating of heritage speakers’ semi-naturalistic productions extracted from the narration in a picture description task (Au et al., 2008; Knightly et al., 2003). With regard to Korean heritage speakers, Oh et al. (2003) focused on heritage speakers’ production of Korean denti-alevolar stop consonants (i.e., /t/, /t^h/, and /t’/). Apart from the number of phonemes (i.e., three-way contrast in Korean vs. two-way contrast in English), Korean denti-alveolar stop consonants are differentiated from English alveolar stop consonants (i.e., /d/ and /t/) mainly through VOT, as well as the pitch (f0) of the following vowel. Different patterns were found among Korean heritage speakers based on their use of Korean during their early childhood. Oh et al. (2003) distinguished childhood speakers (i.e., those who regularly spoke the heritage language during childhood) from childhood overhearers (i.e. those who regularly overheard the heritage language during early childhood, but rarely spoke it), and found that, while both groups received significantly lower ratings than native Korean speakers, only childhood speakers were perceived to have a better accent than L2 learners of Korean. Childhood overhearers, on the other hand, were not distinguishable from the L2 learners. These findings are contrary to what has been found in Au et al. (2008)'s study, in which childhood speakers and childhood overhearers of Spanish did not differ from each other in their global accent in Spanish. In their study, both childhood speakers and overhearers received significantly lower ratings than native Spanish speakers and significantly higher ratings than L2 learners of Spanish. Au et al. (2008) suspect that the difference found in their study and in Oh et al. (2003)'s study may be due to childhood overhearers' amount of relearning of the heritage language beyond early childhood. While the childhood overhearers of Spanish in Au et al. (2008)'s study had taken three to five years of Spanish classes prior to the study, the childhood overhearers of Korean in Oh et al. (2003)'s study had only started taking Korean classes at the time the study took place. Therefore, it is very likely that continued use of the heritage language beyond early childhood, apart from being exposed to it during this period, is an important factor in having benefits in the productive phonology of the heritage language. The importance of continued use of the heritage language will be discussed in more detail in Section 2.3.4.

Heritage accent has also been observed among heritage speakers in several countries in Europe in which the majority language is not English. Kupisch and colleages (Kupisch et al. 2014a, 2014b; Stangen et al., 2015) investigated the degree of foreign accent in both heritage speakers' majority and minority (i.e., heritage) languages, based on the global accent of speech samples extracted from naturalistic interviews. Kupisch et al. (2014b) compared the French pronunciation of German-French simultaneous bilinguals from bi-national families (2L1s) who predominantly spent their childhood and adolescence years in either Germany or France. That is, for those who lived in Germany, French was the minority language, while this was the majority language for the ones who lived in France. Results showed that those who lived most of their childhood and adolescence years in Germany (i.e., heritage speakers of French) were judged to have a foreign accent when they speak French (i.e., 51% of the time) more often than those who lived in France (i.e., 10 % of the time). Although to a lesser degree than what was found in Kupisch et al (2014b)'s study, Stangen et al. (2015) also found that foreign accent was observed among heritage speakers of Turkish who were either born in Germany or immigrated to Germany at a young age (i.e., around 50% of the time)[1]. Kupisch et al. (2014a) conducted an extensive study, by including a variety of heritage speakers, and provided a more comprehensive analysis of the comparison between heritage speakers' speech and that of monolingual speakers and L2 learners. Kupisch et al. (2014a) examined German-French and German-Italian 2L1s who grew up in Germany, France, or Italy during their childhood, in which one of their native languages was the majority language, and the other was a minority language (i.e., heritage language). They found that regardless of the language background, the bilingual speakers were perceived as native speakers less often in their heritage language than monolingual speakers and most of them were rated within the range of L2 learners. Kupisch et al. (2014a) also found that raters made their decisions with less certainty and revised their judgments more often when listening to bilinguals' speech files of the heritage language than those of monolingual speakers and L2 learners. This finding supports that heritage speakers have a particular accent (i.e., heritage accent) that is distinguishable from the speech of monolingual speakers and from foreign accent which is typically found in L2 learners' speech.

Studies in foreign accent rating support that, although heritage speakers excel in the phonology of their heritage languages, compared to L2 learners, their speech is distinguishable from that of monolingual speakers, due to their heritage accent (Benmamoun et al., 2010; Godson, 2004). However, it is unknown to which cues native monolinguals speakers attend when detecting such an accent (Benmamoun et al., 2010). Although Au and colleagues (Au et al., 2002; Au et al., 2008; Knightly et al., 2003) focused on Spanish heritage speakers' production of stop consonants in Spanish and found a strong correlation between accent rating and VOT and degree of lenition (Knightly et al., 2003), it is unclear whether other phonetic features also played a role in the native listeners' judgment. Indeed, Stangen et al. (2015) reported that most raters chose prosody, vowels, and certain consonants as the strongest indicators of foreign accent. Therefore, it is

[1] It is interesting that the heritage speakers in Stangen et al. (2015)'s study, regardless of whether they were first exposed to the majority language before or after age 4, were perceived to have a foreign accent in their majority language relatively more often (around 35% of the time) than the ones in Kupisch et al. (2014a, 2014b)'s study, in which the heritage speakers were rarely judged to have a foreign accent (around 10% of the time). The authors suggest that the differences in their findings may be due to whether the heritage speakers were simultaneous bilinguals from bi-national families (Kupisch et al., 2014a, 2014b) or sequential bilinguals with immigrant parents (Stangen et al., 2015). However, as early sequential bilinguals with immigrant parents, who arrived in the majority society at a very young age, are often found to be indistinguishable from monolingual speakers (Flege et al., 1995), it is possible that other factors, linguistic (e.g., whether the minority language is a European or a non-European language), as well as social (e.g., degree of integration to the mainstream society), also play a role in such differences.

important to examine various phonetic and phonological features at both segmental and suprasegmental levels to understand heritage speakers' phonological system.

2.2 Previous studies on heritage language phonology

2.2.1 Segmental level

Although heritage language phonology is still in its incipient stage, an increasing number of studies have investigated heritage speakers' sound system, focusing mainly on segments, i.e., consonants and vowels. Many of these studies argue that, although heritage speakers have evident advantage over L2 learners in the phonology of their heritage language, their speech also show possible signs of influence from English in certain sounds, and even within the same sound different patterns are observed depending on the linguistic context, task type, heritage speakers' use of the heritage language, and so on.

Among various phonetic and phonological features, the most extensive research has been done on stop consonants in different heritage languages (Arabic: Khattab, 2000; Mandarin Chinese: Chang et al., 2008; Dutch: Simon, 2010; French: Mack, 1990; Sundara et al., 2006; Italian: Nagy & Kochetov, 2013; Japanese: Harada, 2003; Korean: Kang & Guion, 2006; Kang & Nagy, 2012; Oh et al., 2003; Russian: Nagy & Kochetov, 2013; Spanish: Amengual, 2012; Au et al., 2002; Kim, 2011, 2012; Knightly et al., 2003; Magloire & Green, 1999; Rao, 2014, 2015; Ukrainian: Nagy & Kochetov, 2013), mainly on the voice onset time (VOT). According to Jensen (2004), a "voicing language" is a language that has voiced stops with negative VOTs and voiceless stops with short-lag VOTs, such as Spanish, Dutch, and Arabic, while an "aspirating language" is a language that has voiced stops with short-lag or negative VOTs and voiceless stops with long-lag VOTs, such as English, German, and Mandarin. Heritage speakers of a voicing language, who speak an aspirating language as the majority language, tend to produce the voiced stops of their heritage language with short-lag VOTs, following the phonetic norm of the majority language (Khattab, 2000; Kim, 2011, 2012; Knightly et al., 2003; Mack, 1990; Simon, 2010). For instance, Knightly et al. (2003) and Kim (2011, 2012) found that heritage speakers of Spanish (i.e., a voicing language) with English (i.e., an aspirating language) as a majority language produced Spanish word-initial voiced stops /b, d, g/ with full voicing to a lesser degree than native Spanish speakers. Similar results have been observed with child heritage speakers of different languages

(Arabic: Khattab, 2000; French: Mack, 1990; Dutch: Simon, 2010), which indicates that influence from English in the production of voiced stop consonants of the heritage language takes place early on in childhood.

However, this does not imply that heritage speakers are unable to create two separate voiced stop categories. Zampini and Green (2001) argued that once heritage speakers are in a monolingual mode (Grosjean, 2010), through carefully controlling the language of instruction and recruitment process, they are able to successfully distinguish the sound categories of their two languages. Magloire and Green (1999) examined the VOTs of Spanish heritage speakers' bilabial stop consonants, and, in order to control for language mode, they contacted the heritage speakers on two separate occasions, once in English from one lab, and another in Spanish from another lab. Results showed that, with the language mode controlled, heritage speakers' VOTs were comparable to those of monolingual speakers of each language. That is, the heritage speakers produced Spanish voiced stop consonants with negative VOTs, while they produced English voiced stop consonants with short-lag VOTs. In a study examining the production of coronal stops by heritage speakers of Canadian French, Sundara et al. (2006) conducted the English and French sessions with different experimenters, who were native speakers of the language of interest, but had a strong foreign accent in the other language, as a measure to control language mode. As in Magloire and Green (1999), Sundara et al. (2006) found that heritage speakers showed different distributions of Canadian French /d/ and Canadian English /d/, in that Canadian English /d/ was produced with negative VOT less frequently (73.6%) than Canadian French /d/ (100%). However, it is important to take into account that, the heritage speakers in Magloire and Green (1999) and Sundara et al. (2006) were highly proficient bilingual speakers who use the heritage language on a consistent basis, as opposed to the other studies above, in which language proficiency and use were not controlled. In Magloire and Green (1999), the heritage speakers were all recruited in Southern Arizona, which has close linguistic and cultural ties with Mexico, and their average daily use of English and Spanish (i.e., the heritage language) was fairly balanced (English: 56.4%, Spanish: 43.6%). Similarly, based on self-ratings and speech sample rating by monolingual speakers, Sundara et al. (2006) only included heritage speakers whose language proficiency in both Canadian English and Canadian French (i.e., the heritage language) was no less than 6 on a scale from 1 to7, in which 7 indicated native-like proficiency. Therefore, it is likely that heritage speakers' proficiency and use of the heritage language, as well as the majority language, have an effect on their ability to distinguish sound categories in their two languages. However, despite the influence of language proficiency and use, bilingual speakers' voiced stops of one language seem to be easily influenced by the voiced stops of the other language, as the heritage speakers' Canadian English /d/ in Sundara et al. (2006) was produced as prevoiced in the majority of the cases (73.6%), although to a lesser extent than Canadian French /d/, while the monolingual English speakers produced it mostly with short-lag VOTs (95.3%). This, together with the findings of the other studies above, in which the voiced stop consonants of the heritage language were produced with short-lag VOTs, indicate that heritage speakers' voiced stops of the two languages have an effect on each other, whether it is from English to the heritage language, or vice versa.

Unlike voiced stops, heritage speakers are able to successfully distinguish the voiceless stops of their two languages, regardless of their proficiency and use of the heritage language. That is, heritage speakers who are highly proficient in their heritage language and/or who consistently use the heritage language (Amengual, 2012; Magloire & Green, 1999; Italian heritage speakers in Nagy & Kochetov, 2013; Sundara et al., 2006), as well as those who rarely speak the heritage language and/or who are more proficient in English (Au et al., 2002; Khattab, 2000; Kim, 2011, 2012; Knightly et al., 2003), tend to produce word-initial voiceless stops /p, t, k/ of their heritage language (i.e., a voicing language) with short-lag VOTs and /p, t, k/ in English (i.e., an aspirating language) with long-lag VOTs, although the exact VOT values may differ slightly from those of native speakers (Harada, 2003; Kim, 2011, 2012; Mack, 1990; Russian and Ukrainian heritage speakers in Nagy & Kochetov, 2013; Simon, 2010). Successful distinction between the voiceless stops in the heritage language and English (i.e., the majority language) has also been attested among heritage speakers of non-voicing languages, such as Mandarin (Chang et al., 2008, 2011). Mandarin has a laryngeal contrast between unaspirated /p, t, k/ and aspirated /p^h, t^h, k^h/. According to Chang and colleagues (2008, 2011), although the VOT of Mandarin /p^h, t^h, k^h/ and English /p, t, k/ are both within the long-lag range, heritage speakers of Mandarin are able to distinguish English voiceless /p, t, k/ (VOT values around 70 ms.) from Mandarin aspirated /p^h, t^h, k^h/ which are more heavily aspirated (VOT values higher than 100ms.). However, while all speakers successfully distinguished the two voiceless categories, few distinguished Mandarin unaspirated /p, t, k/ and English /b, d, g/, even though the VOT values of Mandarin /p, t, k/ (approximately 10 ms.) were supposed to be slightly, but consistently shorter than those of

English /b, d, g/ when produced with short-lag VOT (approximately 20 ms.). The results in Chang and colleagues (2008, 2011) showed that the heritage speakers produced Mandarin /p, t, k/ with similar VOTs as English /b, d, g/. This finding, together with the findings of the studies with voicing heritage languages, supports that certain sound categories are easily separated in heritage speakers' two languages (e.g., voiceless consonants), while others are not (e.g., voiced consonants).

Studies on heritage speakers' stop consonants have also shown that different patterns may be observed even within the same voicing category. For instance, Au et al. (2002) examined Spanish heritage speakers' production of intervocalic voiced stops /b, d, g/ in Spanish, which are produced as approximants [β, ð, ɣ] in this position, as opposed to stops [b, d, g]. They found that, although the heritage speakers lenited intervocalic /b, d, g/ to a lesser degree than the native speakers, this did not reach statistically significance level. However, in a subsequent study (Knightly et al., 2003), Au and colleagues found that Spanish heritage speakers showed different patterns of lenition when considering each voiced stop separately. That is, the heritage speakers lenited intervocalic /b/ and /g/ significantly less frequently than native speakers, while their percentage of lenited /d/ was similar to that of native speakers. In fact, regarding intervocalic /g/, heritage speakers' percentage of lenited /g/ was noticeably low, to the point that it did not differ from that of L2 learners, who produced this sound with more tension. The authors argued that the asymmetry found in the degree of lenition of the three voiced stops may be due to the difference in Spanish and English sound systems. While lenited /d/ (i.e., [ð]) is very similar to a high-frequency phoneme in English (i.e., /ð/), which is easily found in words such as *the*, *this*, and *other*, lenited /g/ (i.e., [ɣ]) does not exist in the English sound inventory, either phonemically or allophonically. Since the heritage speakers in this study were childhood overhearers (i.e., they were regularly exposed to Spanish during early childhood, but rarely spoke it), the lack of practice of [ɣ], which does not exist in English, may explain why heritage speakers' intervocalic /g/ was less often produced as an approximant, compared to native speakers. Like lenited /g/, lenited /b/ (i.e., [β]) is not an English phoneme, but the heritage speakers showed a clear advantage over L2 learners when producing this sound. The authors claimed that the proximity of [β] to English voiced fricative /v/ may be the reason that heritage speakers were able to recreate a sound that is similar to [β], based on their familiarity with English /v/.

Rao (2014, 2015) examined Spanish heritage speakers' production of intervocalic /b/ more carefully, by taking into account various linguistic and extra-linguistic factors. Following Martínez Celdrán (1984, 1991), Rao (2014, 2015) further divided the realizations of lenited /b/ into pure approximants (PA) [β], which are characterized as vowel-like, with clear formants and lack of turbulence (i.e., more target-like), and tense approximants (TA) [β̞], which have more closure and do not show clear formants (i.e., less target-like). He found that heritage speakers produced less-target-like TA [β̞] and stop allophones [b] more frequently at word boundaries (i.e., at a higher level of the prosodic hierarchy) and in stressed syllables (i.e., in contexts with increased relative prominence). The results also showed an effect of orthography, as heritage speakers produced more tense forms in the grapheme <v> and in the reading task. Rao (2014) posits that increased influence of English phonological rules is likely to occur when orthography is visually available[2]. Lastly, Rao (2014, 2015) found that heritage speakers who had regularly used Spanish since childhood produced intervocalic /b/ with more lenition than those who used Spanish to a lesser extent. The effect of heritage language use on the production heritage language stop consonants has also been attested in other studies (Chang et al., 2008; Nagy & Kochetov, 2013; Oh et al., 2003).

Heritage speakers' stop consonants may also be characterized by differences in cue weighting patterns. For instance, Kang and Nagy (2013) examined Korean heritage speakers' production of Korean stop consonants, which, unlike other languages that have a two-way laryngeal contrast, form a three-way contrast, i.e., lenis /b, d, g/ (slightly aspirated), fortis /p', t', k'/ (unaspirated), and aspirated /p^h, t^h, k^h/ (heavily aspirated). Korean lenis and aspirated stops, in particular, are distinguished mainly by the fundamental frequency (f0) of the following vowel, rather than VOT, due to a "tonogenesis-like sound change" that Korean is undergoing, in which "pitch differences are replacing VOT differences as the key cue to a phonemic contrast between word-initial aspirated and lenis stops". This is most likely to be led by young female speakers (Kang & Nagy, 2012). Kang and Nagy (2012) found a gender effect among first generation immigrant speakers, in that while female speakers used f0 as a primary cue, male speakers used VOT as a primary

[2] Indeed, when compared with intervocalic /d/ and /g/, the Spanish heritage speakers in Rao (2015) produced intervocalic /b/ with lower PA frequency. This is different from what Knightly et al. (2003) found, in which Spanish heritage speakers produced /b/ and /d/ in a more native-like manner than /g/. Rao (2015) suggests that it is likely that this discrepancy is due to methodological differences between the two studies. That is, Knightly et al. (2003) only included /b/ that was spelled with <b>, whereas Rao (2015) used both graphemes <b> and <v>.

cue, reflecting the trend of the Homeland (South Korea). With regard to heritage speakers, Kang and Nagy (2012) found a quite different pattern; male speakers were the ones that used f0 as the primary cue, while female speakers used both f0 and VOT to distinguish the two stop categories. Kang and Nagy (2012) posit that female heritage speakers' use of VOT may be a way to reverse the sound change occurring in the Homeland, by "actively (re)introducing the VOT contrast due to the influence from English". Although gender effect was not examined, Kang and Guion (2006) also found that Korean heritage speakers, the majority of which were female speakers (7 out of 10), used both f0 and VOT to distinguish Korean lenis and aspirated stops.

Although very few compared to stop consonants, there are some studies that have investigated heritage speakers' production of other consonants (Chang et al., 2009, 2011; Henriksen, 2015). For instance, Chang et al. (2009, 2011) examined Mandarin heritage speakers' production of post-alveolar fricatives of English (i.e., palato-alveolar /ʃ/) and Mandarin (i.e., alveolo-palatal /ɕ/ and retroflex /ʂ̻/), which differ in centroid frequency (Mandarin /ɕ/: 5381 Hz > English /ʃ/: 4229 Hz > Mandarin /ʂ̻/: 3583 Hz) (Jongman et al., 2000; Svantesson, 1986). Contrary to the prediction that heritage speakers would produce Mandarin /ʂ̻/ and English /ʃ/ similarly, due their comparable place of articulation, Chang et al. (2009, 2011) found that these two fricative categories were robustly distinguished by the majority of the heritage speakers. In a study on the production of Spanish rhotics by heritage speakers of Mexican Spanish, Henriksen (2015) examined whether heritage speakers are able to distinguish Spanish tap /ɾ/ and trill /r/, by looking at two acoustic correlates of tap-trill contrast, i.e., the number of apical occlusions and overall segmental duration, and compared their behavior to that of first generation immigrant speakers from Mexico. According to Henriksen (2015), no difference was found between the heritage speakers and the first generation immigrant speakers with regard to their use of the acoustic correlates investigated; both groups used duration as a primary cue to the tap-trill contrast, while no consistent difference was found in the number of occlusions[3]. However, as Henriksen (2015) noted, further analysis should be conducted in zero-closure trill variants, as he found that, impressionistically, first generation immigrant speakers and heritage speakers tended to produce these variants with different manner or articulation. That is, while the former produced them

[3] Given that such variability is also observed in monolingual Spanish communities, Henriksen (2015) posits that there may be "a continuity between U.S. Spanish speech norms and those of Latin American speech communities", which supports Otheguy and Zentella (2011), although the possibility of influence from English cannot be completely discarded.

primarily as assibilated rhotics or fricatives, similar to what has been found in the speech of Spanish monolingual speakers (Henriksen & Willis, 2010), the latter group produced them as r-colored approximants, which were more English-like.

Similar to heritage speakers' consonants, which show different patterns depending on consonant types, heritage speakers' vowels also show different patterns based on the type of vowel. With regard to the front-back dimension, which is characterized by second formant (F2) frequencies, Chang et al. (2008, 2011) found that heritage speakers of Mandarin produced Mandarin back rounded vowels /u/ and /o^u/ more fronted (i.e. higher F2) than native speakers of Mandarin, while their production of front rounded vowel /y/ did not differ from the native speakers. They have also found an effect of heritage language use. That is, heritage speakers who used Mandarin more frequently had lower F2 values of /u/ and /o^u/ (i.e., more back), closer to those of the native speakers, than those who used Mandarin less frequently. Similarly, Saadah (2011) found that Arabic heritage speakers' Arabic long high back vowel /u:/ was significantly more fronted (i.e., higher F2) than native Arabic speakers. The fronting of /u/ was also found in Ronquest (2012) in a study on Spanish heritage speakers. Ronquest (2012) found that Spanish heritage speakers' Spanish /u/ was noticeably more fronted than what has been traditionally characterized in Spanish monolingual speech described by Marín Gálvez (1995), to the point that F2 values of their /u/ were significantly higher (i.e., more fronted) than those of their /o/. According to Ronquest (2012), /u/-fronting was attenuated as heritage speakers received more instructions in Spanish. That is, heritage speakers with more experience taking Spanish courses produced /u/ with lower F2 (i.e., more back) than those who had less experience in Spanish instructions. Like Chang et al. (2008, 2011), Ronquest (2012) also found an effect of language use, which was examined based on two factors: the frequency of travel to a Spanish-speaking country and the frequency of Spanish use in a variety of contexts. These effects were found not only on the production of /u/, but also on /i/ and /a/. That is, heritage speakers with more travel experience in a Spanish-speaking country and those who spoke Spanish more frequently and in a larger number of contexts had a more expanded vowel space, due to greater dispersion of point vowels /i/, /u/, and /a/ from the center of the vowel space. Saadah (2011) also found a greater dispersion of long vowels /i:, u:, a:/ from heritage speakers with higher proficiency in Arabic. The fronting of /u/ observed in the speech of heritage speakers of Mandarin, Arabic, and Spanish in the three studies above is likely to be the result of influence from English, as /u/-fronting is a widespread linguistic phenomenon in North America (Labov, 2008), most frequently found in Californian Anglo English (Fought, 1999; Hinton et al., 1987). Resistance to English influence has also been observed in heritage speakers' speech. For instance, Godson (2004) examined Western Armenian vowels produced by heritage speakers of Western Armenian residing in Southern California and found that, even though /u/-fronting is a common linguistic phenomenon in Californian Anglo English (Fought, 1999; Hinton et al., 1987), these speakers did not follow /u/-fronting. Rather, they produced Western Armenian back vowels /o/ and /u/ similarly to those of the first generation immigrant speakers. For these speakers it was their non-back vowels /i/, /ɛ/, and /a/ that were produced differently from the first generation immigrant speakers, closer to the English counterparts; while heritage speakers' /i/ was more fronted (i.e., higher F2), their /ɛ/ and /a/ were lower (i.e., higher F1), compared to those of the first generation immigrant speakers.

Regarding the high-low dimension of vowels, which is characterized by first formant (F1) frequencies, Guion (2003) found that Quichua-Spanish simultaneous and early bilinguals residing in Otavalo, Ecuador[4] showed greater variation in Quichua high vowels, in that the bilinguals' Quichua high front lax vowel /ɪ/ was frequently produced as Spanish /i/, while their Quichua high back lax vowel /ʊ/ was produced similar to either Spanish /u/ or /o/. Bilinguals' Quichua low vowel /a/, on the other hand, was distinguished from Spanish /a/, with significantly lower F1 values (i.e., higher in vowel space) than Spanish /a/.

Different patterns have also been found in heritage speakers' vowels based on whether the vowels are in the periphery in the vowel space or not. For example, in a longitudinal study with a child heritage speaker of Mandarin, Yang et al (2015) found that, although the heritage speaker's Mandarin vowel space remained relatively constant, despite continuous restructuring of his English vowel space, his production of non-peripheral vowels, i.e., high front rounded /y/ and mid back unrounded /ɤ/, continuously changed. That is, while his peripheral vowels (i.e., /i, u, a/) matched those of adult Mandarin monolingual speakers, his non-peripheral vowels did not,

[4] Quichua is the indigenous language of the Otavalo region in Ecuador. Therefore, based on the definition of heritage speakers adopted in the present dissertation (i.e., descendants of immigrants that speak an ethnic minority language in a society where a different language is spoken as the majority language), the bilingual speakers in Guion (2003) cannot be considered as heritage speakers. However, given that Spanish is the "*de facto* official language in Ecuador and is used for virtually all official purposes" (i.e., majority language; L2), while Quichua monolingualism is rare in this region, "limited to older people in isolated communities and children who have not yet reached school age" (i.e., minority language; L1) (Guion et al., 2000), the findings of this study is still relevant to the acquisition of heritage language phonology.

suggesting that the heritage speakers' Mandarin (i.e., L1) vowel system was still undergoing active developmental processes (Yang et al., 2015).

2.2.2 Suprasegmental level

Compared to consonants and vowels, there are few studies that examined heritage speakers' sound system at the suprasegmental level. Robles-Puente (2014) provided an extensive description of the rhythm and intonation of Spanish and English produced by speakers of various language backgrounds residing in Los Angeles, including heritage speakers of Spanish. Looking at normalized pairwise variability index (nPVI) (Low et al., 2000) and voicing ratios (Dellow et al., 2007), Robles-Puente (2014) found that while adult heritage speakers showed an English-like rhythm (i.e., more variability in vowel duration and less voicing in the speech signal) in both Spanish and English, child/adolescent heritage speakers were able to separate the rhythmic patterns of the two languages. With regard to intonation, Robles-Puente (2014) examined heritage speakers' use of pitch accents and boundary tones in various sentence types (i.e., declaratives, imperatives, wh- and yes-no questions, and vocatives). Similar to the case of rhythm, Robles-Puente (2014) found that, in certain properties, such as prenuclear pitch accent, the adult heritage speakers preferred the English pattern (e.g., high H*) over the Spanish one (e.g., late rise L*+H), while the child/adolescent heritage speakers used both patterns independently in the two languages. As the adult heritage speakers differed from the child/adolescent heritage speakers in that they had a longer exposure to English and spoke Spanish less regularly, Robles-Puente (2014) posits that the different patterns found in the two heritage speaker groups may be due to difference in the degree of influence from English. Influence from English at the suprasegmental level has also been found in heritage speakers' use of prosody when marking focus. According to Gries and Miglio (2014) and Harris et al. (2015), heritage speakers of Spanish, as well as English monolingual speakers, use pitch excursion as a strategy to encode new information more frequently than Spanish monolingual speakers.

2.3 Acquisition of phonology

2.3.1 Reorganization of language-general to language-specific speech sounds

There is ample evidence that linguistic experience has a critical role in speech processing (Strange, 1995; Werker & Polka, 1993; Werker & Tees, 2005). A well-known example is the distinction of English /r/ and /l/ by Japanese speakers; these sounds would easily be perceived as different sound categories for native English speakers; but, for native Japanese speakers, who do not have this contrast in their native language, it would seem as if these two sounds pertain to the same sound category (Lively et al., 1993). Many studies in child development have shown that the role of language exposure in speech processing occurs early in life (Kuhl & Iverson, 1995; Werker & Tees, 1984, 2005). According to Kuhl & Iverson (1995), at birth, infants are able to hear differences among all of the sounds that exist in human languages. This language-general ability of speech perception begins to transform into a language-specific one between 6 to 12 months of age, which is also when infants' abilities to discriminate foreign language phonetic units sharply decline (Kuhl et al., 2008). According to Kuhl's Native Language Neural Commitment (NLNC) hypothesis, initial exposure to native language causes physical changes in the neural networks; if infants' native language phonetic perception improves, it reflects that the neural networks are more committed to the phonetic patterns of the native language (Kuhl, 2004; Kuhl et al., 2003; Kuhl et al., 2005; Kuhl et al., 2008). As neural commitment to native language increases, it will be harder for infants to learn new phonetic patterns, causing difficulties in foreign speech perception from then on (Kuhl et al. 2005; Werker & Tees, 1984). In contrast, if infants successfully perceive foreign phonetic patterns (Cheour et al., 1998; Kuhl et al., 2005; Rivera-Gaxiola et al., 2005), it implies that they are still in an immature state of uncommitted circuitry (Kuhl et al., 2008). Thus, learning continues until stability is achieved; neural networks stay flexible and continue to learn until the amount and variability for a particular category reaches stability (Kuhl, 2000, 2004; Kuhl et al., 2005; Kuhl et al., 2008). Therefore, according to Kuhl and colleagues (Kuhl et al., 2003; Kuhl et al., 2008), phonetic perception of native and foreign speech sounds is negatively correlated, i.e., the more the native language is learned, the less the future capacity to learn foreign phonetic patterns. For instance, Werker and Tees (1984) found that English-learning infants at 6 to 8 months of age were successful in distinguishing Hindi dental /t/ from retroflex-dental /T/, a distinction that they had never been exposed to. However, by 10 to 12 months of age, they were no longer able to distinguish these sounds. With regard to the acquisition of prosody, studies have shown that learning the prosodic properties of sound occurs even earlier. Infant studies by Mehler and colleagues (Mehler et al.,1988; Mehler et al., 1996) showed that newborn babies, as young as 4-days-old, had preference for utterances in their mother tongue. Using low-pass filtered sentences (400 Hz), Mehler and colleagues found that newborn babies were able to discriminate utterances in their mother tongue from those in a language that belongs to another rhythmic class. Attenuating speech signal in this fashion leaves the prosody intact, while eliminating most of the distinctive segmental information (Mehler et al., 1988). This implies that there is sufficient information in the prosody to allow infants to distinguish utterances in their native language from those in a foreign language. Specifically in the acquisition of native language stress pattern, Jusczyk et al. (1993) found that, between 7 to 10 months of age, infants learning English had preference for the strong-weak dominant stress pattern of English words.

2.3.2 Optimal period hypothesis and cascade influences

If reorganization of language-general to language-specific ability begins as early as within the first year of life, then when does it end, i.e., how long does this process last? In the second language acquisition (SLA) literature, it is well established that there is a strong relationship between age of acquisition and ultimate proficiency; the earlier one learns a L2, the better. This supports Lennernberg's Critical Period Hypothesis (Lennenberg, 1967), which proposes that language is a system that is deeply constrained by biology and, thus, it can only be acquired during a "critical period" in development which lasts from birth until puberty; before and after the critical period, the system cannot be altered by experiential input (Werker & Tees, 2005). The existence of this opportunity window seems to be especially true in the acquisition of L2 phonology (Bosch et al., 2000; Flege, 1995; Gildersleeve-Neumann et al., 2009; Sebastián-Gallés & Soto-Faraco, 1999; Strange, 1995). It is well documented that adult L2 learners have foreign accents even after using their L2 for decades (Johnson & Newport, 1989; Mayberry & Fisher, 1989; Flege et al., 1995; Flege et al., 1999) and experience difficulties in distinguishing L2 phonemic contrasts that are considered as one phoneme in their native language (e.g., Japanese speakers with English /l/ and /r/). Foreign accents may occur as early as sometime between 4 and 6 years of age (Birdsong, 1999; Dupoux et al., 2010; Flege, 1991; Flege, 1995; Gildersleeve-Neumann et al., 2009; Long, 1990; Strange, 1995; Pallier et al., 1997).

Werker and Tees (2005) also agree that there is a best time period when language can be acquired, but they refute that the onset and offset of language acquisition are absolute and fixed as argued in the Critical Period Hypothesis. Instead, they propose the term “Optimal Period”, in which neither the onset nor the offset is absolute and, thus, language acquisition is open throughout the life span (Figure 1). From this perspective, decline in nonnative speech perception is not loss of perceptual sensitivity, but reorganization, as mentioned earlier.

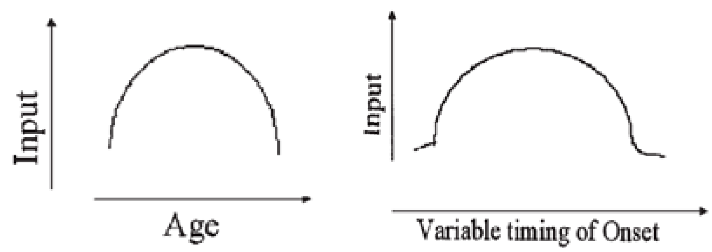


Figure 1 Schematic illustration of critical period (left) and optimal period (right) presented in Werker and Tees (2005)

The notion of reorganization is supported in a number of studies. In the field of L2 instruction, it is argued that L2 learning can be impeded due to difficulties in perceiving non-native speech sounds. In a behavioral training study, Lively et al. (1993) have shown that even in adulthood, learners show improvements in non-native speech perception after training. However, despite improvement, speech training does not necessarily guarantee the levels of accuracy shown by native speakers, presumably due to the influence from the native language (Polka, 1992; Takagi, 2002; Werker & Tees, 2005). For instance, Saalfeld (2009) found that even though Spanish and English are both languages with variable word stress, English L2 learners of Spanish still experienced stress ‘deafness’ (Dupoux et al., 2001) when perceiving Spanish stress minimal pairs of same grammatical categories, possibly due to influence from English in which stress position is not a critical element to distinguish between different lexical meanings or verbal inflections. Similarly, several ERP (Event Related Potential) studies revealed that, although there are neural responses to both native and non-native distinctions, the ERP to the non-native contrast are slower and/or are over different recording sites than is the ERP to native phonetic distinctions (Aaltonen et al., 1987).

Werker and Tees (2005) posit that the reason that adult L2 learners many times fail to achieve native-like competence despite some improvement is due to “cascade influences”. According to Werker and Tees (2005), language involves a number of interrelated, hierarchically organized subsystems, each having different optimal periods; subsystems that require information integration will logically be developed later than those that involve more basic information. Following this line of thinking, tuning one subsystem will lead to the emergence of a new subsystem in a higher level, thus the term “cascade influences”. Experience listening to speech in utero leads to preference for listening to utterances in the native language than those belonging to another rhythmic class (Mehler et al., 1988; Mehler et al., 1996). Early processing of the rhythmical properties of the native language has been attested in American infants’ preference for words that conform to the strong-weak dominant stress pattern of their native language (i.e., English) (Jusczyk et al., 1993; Nazzi et al., 2000). The preference for the rhythmical properties of the native language in turn is used to segment words from continuous speech, and so on (Figure 2). Therefore, Werker and Tees (2005) argued that, once a subsystem is already in place, it will be difficult to learn new properties in that level due to continued reinforcement of higher order linguistic uses.

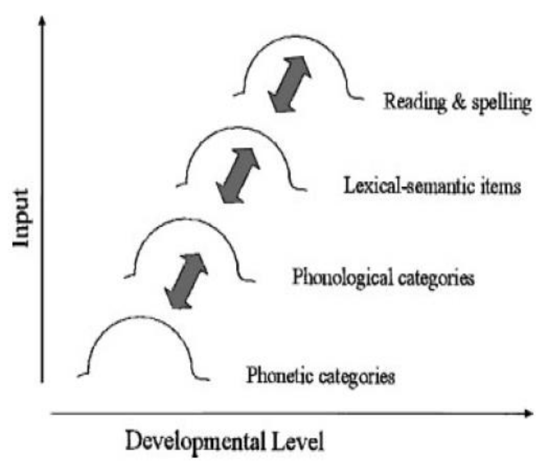


Figure 2 Illustration of the cascade influences involving different subsystems of speech presented in Werker and Tees (2005)

2.3.3 L1-L2 phonetic interference in early bilingual speakers

If learning continues until stability is achieved, bilingual speakers, who acquired both L1 and L2 early in life, such as heritage speakers, should be able to behave like monolingual speakers of each language, mapping two language inputs onto separate perceptual spaces and forming two speech representations (Kuhl & Iverson, 1995, Kuhl et al. 2003; Kuhl et al., 2008). However, even for early bilinguals that are fluent in both of their languages, it is impossible to control two languages exactly the same way as two monolinguals (Mack, 1989; Flege, 1999). Valdés (2001) argues that having absolutely equivalent abilities in two languages is a myth, given that individuals seldom have access to two languages in exactly the same contexts in every domain of interaction or use two languages to carry out the exact same functions to every person with whom they interact. According to the Speech Learning Model (SLM) (Flege, 1995; Flege et al., 2003), L1 and L2 sound systems are not independent, but coexist in one phonetic space. This naturally creates competition between the speech sounds of L1 and L2; bilinguals strive to maintain phonetic contrast between all of the sound categories in their combined L1+L2 phonetic space in the same way that monolinguals strive to maintain phonetic contrast among the categories in their L1-only phonetic space (Flege et al., 2003). The SLM argues that the connection between an L2 speech sound and the closest L1 sound triggers a process called “category assimilation”. When category assimilation occurs, the introduced L2 speech sound fails to create a new category; instead, it merges with the closest L1 sound and, as a result, the production of these sounds will assimilate. If, in contrast, a new category for an L2 sound is established, the closest L1 sound will be deflected away from the phonetic norm in order to maintain phonetic contrast with the L2 sound (i.e., category dissimilation) (Flege et al., 2003). Studies on heritage speakers’ production of stop consonants presented in Section 2.2.1. (Khattab, 2000; Kim, 2011, 2012; Knightly et al., 2003; Mack, 1990; Simon, 2010 among others) provide a good example of category assimilation/dissimilation. For instance, in a case study, Mack (1990) found that her subject, an early bilingual speaker who spoke French at home and English elsewhere (i.e., heritage speaker of French), created three stop categories (i.e., French/English /b, d, g/, French /p, t, k/, and English /p, t, k/) instead of four. Moreover, these categories were realized differently when compared with the monolingual norms; instead of with prevoicing, the child produced French /b, d, g/ with short-lag VOTs, and /p, t, k/ in both French and English were produced with much longer VOTs than the phonetic norm. This result, together with the results of other studies, implies that English /b, d, g/ have gone through category assimilation, while English /p, t, k/ have experienced category dissimilation. When English /b, d, g/ are introduced, they will be treated as variants of French /b, d, g/, thus, the creation of distinct categories for English /b, d, g/ will be blocked and, as a result, French /b, d, g/ and English /b, d, g/ will merge. However, when English /p, t, k/ are introduced, new categories for these sounds will be created through the process of category dissimilation, since the long-lag VOTs of English /p, t, k/ are perceptively different from the short-lag VOTs of the French counterparts.

If L1 and L2 phonetic systems are constantly engaged, mutual influence between L1 and L2 sounds is inevitable. At this point, it is important to note that phonetic interference does not only occur from L1 to L2 (Flege et al., 1995; Flege et al., 1999; Kuhl et al., 2003), but also from L2 to L1 (Chang, 2012; Flege, 1995; Fowler et al., 2008; Grosjean, 1989; Major, 1992; Sancier & Fowler, 1997; Ulbrich & Ordin, 2014) and the direction and strength of the influence depends on factors such as the amount and circumstances of L1 and L2 use, language dominance, and so on (Bosch et al., 2000; Pallier et al., 1997; Yip & Mathews, 2006). Thus, if a child learns two languages when the neural commitment to L1 is not complete, and if L2 is more dominant and used more frequently than L1, there will be greater influence from L2 to L1 than from L1 to L2 (Gildersleeve-Neumann et al., 2009; Simon, 2010; Yip & Mathews, 2006). Such an asymmetrical phonetic transfer is frequently found in heritage speakers in the U.S. Simon (2010) investigated the production of prevoicing in Dutch by a child L1 Dutch speaker who moved to the U.S. and was exposed mostly to English since then (i.e., heritage speaker of Dutch). The results show that the overall majority of Dutch /b, d, g/, and some English /b, d, g/, were produced with prevoicing in the first session (i.e., category assimilation). However, prevoicing in Dutch decreased as the English acquisition process went on; Dutch phonetics moved in the direction of the English target realizations. This implies that influence from L1 to L2 and influence from L2 to L1 do not occur to the same degree; in some cases, the latter could be greater than the former. Flege et al. (2003) proposed that, when a bilingual speaker approximates the phonetic norm of an L2 speech sound, his production of the corresponding L1 speech sound will diverge from L1 phonetic norms. On the contrary, if the influence from L1 to L2 is stronger, the L2 speech sound should follow the L1 phonetic norms.

2.3.4 Perception vs. production

Since heritage speakers usually show an asymmetry in the amount of time that they speak the heritage language and the amount of time that they hear it (i.e., they speak the heritage language less frequently than they hear it), it is of interest whether this asymmetry is reflected in their perception and production of the speech sounds of the heritage language. Numerous studies have shown that being exposed to the heritage language in early childhood alone is beneficial in the perception of the heritage language, compared to L2 learners who learned the same language at a later age (Au et al., 2002; Au et al., 2008; Knightly et al., 2003; Oh et al., 2003). However, as Knightly et al. (2003) pointed out, it is unclear whether exposure to the heritage language in early childhood also leads to a better production of the language. As presented in Section 2.2., studies on heritage language phonology (Chang et al., 2008, 2011; Oh et al., 2003; Rao, 2014, 2015; Robles-Puente, 2014; Ronquest, 2012; Saadah, 2011; Yeni-Komshian et al., 2000, among others) have shown that there is a close relationship between the frequency of use of the heritage language and their production. For instance, Oh et al. (2003) examined the perception and production of three Korean denti-alveolar stops (i.e., aspirated /t^h/, plain /t/, and tense /t'/) by two groups of heritage speakers of Korean: those who regularly heard and spoke Korean during childhood (i.e., childhood speaker) and those who regularly heard Korean, but spoke it minimally (i.e., childhood hearer). Results showed that, the two groups performed similarly in the perception task, while in the production task the childhood speakers outperformed the childhood hearers. Thus, it is likely that, while early exposure to the heritage language is beneficial for later perception of the language, it does not necessarily lead to target-like production. Indeed, in a follow-up study, Oh et al. (2003) further examined a subsample of childhood speakers of Korean who had not spoken Korean at all beyond early childhood or who rarely spoke it. They found that the subsample of childhood speakers did not differ from the childhood hearers, supporting the importance of continued use of the heritage language beyond early childhood. Similarly, in a delayed sentence repetition task with Korean-English bilingual speakers who immigrated to the U.S. at different stages in life, Yeni-Komshian et al. (2000) found that Korean-English bilinguals who immigrated to the U.S. at a very young age (age 1-7) were rated to have a stronger foreign accent in in their L1 Korean than Korean monolingual speakers and other bilinguals who arrived to the U.S. later in life. However, for those that arrived to the U.S. before age 3, which may be considered as heritage speakers of Korean, an inverse relationship was found between their Korean pronunciation and time spent studying Korean beyond early childhood, which suggests that increased use of Korean may have had a positive effect on the heritage speakers' pronunciation in Korean.

In order to understand heritage speakers' perception and production of heritage language speech sounds, it is important to understand the relationship between speech perception and production. Previous studies on infants' perception and production of speech sounds support that language-specific patterns emerge in speech perception prior to speech production, and that there is a strong link between these two domains (deBoysson-Bardies, 1993; Imada et al., 2006; Kuhl & Meltzoff, 1996; Kuhl, et al., 1992; Kuhl et al., 2006; Kuhl et al., 2008; Polka & Werker,1994). According to Native Language Magnet theory, expanded (NLM-e) (Kuhl et al., 2008), infants store perceptual information of speech sounds during the early months of life, when production is still primitive and highly variable (Kuhl et al., 2008). This claim is supported in several neuroimaging studies that show that the brains of bilingual adults do not show a native-like pattern of activity in response to a new language acquired past age 3 (Moore & Guan, 2001; Perani et al., 2003). The NLM-e theory also proposes that the link between speech perception and production is developmental in nature. Infants forge a link between perception and production based on their perceptual experience and learned mapping between the two (Kuhl & Meltzoff, 1982, 1996; Kuhl et al., 2008). That is, by imitating native speech sounds, infants relate auditory results of their own vocalizations to the articulatory movements that caused them, which subsequently creates language-specific mapping between the perception and production (Imada et al., 2006; Kuhl & Iverson, 1995; Kuhl et al., 2008). Imada et al.'s (2006) study on infants supports this claim well. Using Magnetoencephalography (MEG), Imada et al. (2006) found that, when listening to syllables, auditory perceptual brain areas (i.e., superior temporal region) were activated to an equal degree in newborns, 6-month-old, and 12-month-old infants. However, when comparing these areas with the areas responsible for production (i.e., inferior frontal region or Broca's area), results showed that perception of speech syllables alone did not activate the brain areas responsible for production in newborns, but it did so increasingly in 6-month-old and 12-month-old infants. Given that newborns are not able to produce vowels, whereas 6-month-old and 12-month-old infants can, this study provides evidence that the connection between perception and production requires experience with speech production which binds the two.

In the case of heritage speakers, as these speakers regularly hear the heritage language during infancy, mostly likely from their parents and other caregivers (i.e., native speakers of the heritage language), the amount of input they receive in the heritage language during this period may be sufficient to develop mental representations of heritage language speech sounds. Once heritage speakers start producing these sounds, they would compare their own speech to the stored mental representations and adjust their articulators if discrepancy is detected. However, speech production requires complex coordination of articulatory movements and a substantial amount of time and practice are needed to develop a stable perception-production link (Godson, 2004; Knightly et al., 2003). Unlike native-like speech perception which undoubtedly establishes very early in life, native-like speech production does not seem to establish until later. Studies have reported that, around the ages of 5-7, children are able to produce most native speech sounds (Birdsong, 1999; Flege, 1991, 1995; Gildersleeve-Neumann et al., 2009; Long, 1990; Pallier et al., 1997; Strange, 1995). Thus, it is likely that shift to English occurs when heritage speakers are still in the process of developing their articulators and have limited motor control of their articulators (Godson, 2004; Menn & Stoel-Gammon, 1995). Moreover, even if native-like production of heritage language speech sounds is established, whether it maintains throughout life is a question yet to be answered. Several studies have shown that speech production is very malleable and thus even adult speakers experience L1 phonetic drift due to influence from L2. For instance, Major (1992) found that adult speakers of American English who arrived in Brazil after the age of 22 and lived there for more than 12 years produced English voiceless stop consonants with shorter VOTs than monolingual American English speakers. Similarly, Chang (2010) found that over time English late L2 learners of Korean produced English voiceless stop consonants with increased VOTs and the following vowel with higher pitch (f0), similar to Korean aspirated stops /p^h, t^h, k^h/. The findings of these studies suggest that it is crucial to consistently use a language beyond childhood in order to maintain it. Given that heritage speakers tend to speak the heritage language far less frequently than they hear it, it is expected that there would be a discrepancy between their perception and production of the heritage language speech sounds.

2.4 Hypotheses and predictions

Based on previous studies on heritage language phonology and the theories and models regarding the acquisition of phonology, the present study proposes three hypotheses, which are presented below:

- Hypothesis 1: Given that heritage speakers are generally more dominant in English than in the heritage language, heritage speakers' heritage language phonology will be influenced by English.
- Hypothesis 2: A subsystem of speech in the heritage language that is hierarchically lower in the cascade will be less influenced by English than a subsystem that is placed higher in the hierarchy.
- Hypothesis 3: Given that heritage speakers tend to speak the heritage language far more frequently than they hear it, there will be a discrepancy between their perception and production of the heritage language speech sounds.

These three hypotheses will be empirically tested by examining Spanish heritage speakers' perception and production of two types of prominence in Spanish: word-level prominence marking paradigmatic contrast (i.e., lexical stress) and utterance/phrase-level prominence marking information status and focus (i.e., nuclear stress). The acquisition of lexical stress requires the acquisition of phonological word, while the acquisition of nuclear stress requires the acquisition of higher level phonological constituents, such as phonological/intonational phrase and utterance (Nespor & Vogel, 1986; Vogel & Raimy, 2002). Thus, several studies on child language acquisition have shown that lexical stress is acquired early in infancy, while nuclear stress can only be acquired once native semantics/pragmatics is established (around age 12) (Atkinson-King, 1973; Nespor & Vogel, 1986; Vogel & Raimy, 2002; Wieman, 1976). Therefore, with regard to Hypothesis 2, heritage speakers are expected to show less stability in nuclear stress, thus, more influence from English, than lexical stress.

Chapter 3 Word-level prominence: Lexical stress

3.1 Literature review

The present section compares Spanish and English stress systems and reviews previous studies regarding the acoustic correlates of Spanish and English lexical stress. Spanish and English are typologically similar in that both languages have lexical stress. That is, stress is phonologically contrastive in these languages and it is possible to obtain words with different meanings solely by altering the position of stress (Hualde, 2005). For instance, we can find stress minimal pairs, such as *<u>can</u>to* 'I sing' vs. *can<u>tó</u>* 'he/she/you (formal) sang' in Spanish and *<u>con</u>duct* (subject) vs. *con<u>duct</u>* (verb) in English. Stress is characterized as prominence in a syllable of a word resulting from extra muscular energy that acoustically manifests in higher pitch, longer duration and higher intensity (Ladefoged, 2001). Although researchers agree that these three universal parameters are important indices of stress, there are disagreements with regard to the relative strength of these cues, which mostly have their roots in measuring stress in different contexts, resulting in confounding stress with other prosodic factors, such as pitch accent or prosodic boundary adjacency (Kim, 2011). Thus, the effect of stress on these acoustic cues should be understood separately from the effects of other prosodic factors.

3.1.1 Acoustic correlates of stress in Spanish and English

In Spanish, intensity has traditionally been defined as the strongest cue to stress. In his book *Manual de pronunciación* 'Manual of pronunciation' (1914), Navarro Tomás called stress as *acento de intensidad* 'intensity stress', emphasizing the role of intensity as the decisive marker of stress. However, with the advance of technology, researchers have been able to apply objective methods by conducting acoustic analyses on various stress correlates, instead of relying on one's impressionistic judgment, and found that intensity is not the primary acoustic correlate of stress, but rather a secondary one. Some studies found that vowels that bear lexical stress are significantly longer in duration than those that are unstressed, leading to the claim that it is vowel duration that functions as the primary marker of stress in Spanish (Canellada & Madsen, 1987). Other studies argued that it is the fundamental frequency (f0) of the vowel that functions as the primary cue to stress. Quilis (1971, 1981) posits that the f0 is the most important cue for the perception of Spanish stress, while duration is secondary; the role of intensity is so minimal that is could be overlooked. This argument is confirmed by perceptual studies using manipulated stimuli. Enríquez et al. (1989) conducted a perceptual study using synthesized stimuli and concluded that, in Spanish, the f0 is the only parameter that is systematically related to the identification of stressed syllables. Llisterri et al. (2003) also agreed with the central role of the f0, but they claimed that f0 alone is not a sufficient cue to identify stressed syllable; rather, it has to be in combination with duration, intensity, or both duration and intensity.

Nevertheless, it is problematic to conclude that the f0 is the primary cue in identifying Spanish lexical stress, because in most cases the syllable that carries stress also carries a pitch accent (i.e., phrasal prominence). For instance, in Quilis' (1971) study, the target words were produced either in isolation or at the end of short carrier sentences with declarative intonation, such as *Digo la palabra* _____. 'Say the word ______'. In such contexts, stress and accent covary. That is, stressed syllables are always accented and unstressed syllables are always unaccented. However, stressed syllables are not always accented and, even when they are, differences have been found in the alignment of the f0 between stress that carries a nuclear accent, such as the cases above, and stress that carries a prenuclear accent. While in nuclear positions f0 peaks align with stressed syllables, in prenuclear positions they are displaced to a post-tonic syllable, the exact location of which varies depending on several factors such as the number of post-tonic syllables within a word (Estebas-Vilaplana, 2007)[5]. In the latter context, the start of pitch rising (i.e., f0 valley) is consistently anchored to the onset of the stressed syllable (Prieto & Torreira, 2007; Torreira, 2007). Therefore, stress is signaled by the location of the f0 peak in nuclear position, whereas in prenuclear position it is marked by the f0 valley (Estebas-Vilaplana, 2007; Face, 2001; Hualde, 2002; Lleó et al., 2004; Prieto & Torreira, 2007; Torreira, 2007).

Due to such variability of cues, it is important to disentangle stress from pitch accent, in order to understand the primary acoustic cue to stress. Studies done by Ortega-Llebaria and Prieto are the first ones that examined stress correlates of Spanish, while controlling for the effects of pitch accent (Ortega-Llebaria, 2006; Ortega-Llebaria & Prieto, 2007, 2009, 2011). Ortega-Llebaria

[5] Estebas-Vilapalana (2007) found that f0 peaks are most displaced in proparoxytones, followed by paroxytones and oxytones. Thus, she posits that Spanish prenuclear rising accents are characterized by means of an L* pitch accent followed by a loosely aligned H word-edge tone.

and Prieto (2007) examined utterances that were produced in either a declarative sentence, in which the stressed syllable also carried a pitch accent (e.g., *Determino la masa*. 'I determine the mass.'), or a reporting sentence, where the stressed syllable was unaccented (e.g., *La masa del átomo es medible – determino complacida*. 'The atom mass is measurable – I determine pleased.'). Results showed that there was an increase in the f0 in declarative sentences (i.e., accented context), while in reporting sentences (i.e., unaccented context) there was practically no variation in f0. That is, f0 is the primary cue to pitch accent, but when pitch accent is controlled, it no longer leads to the identification of stressed syllables. Thus, in a subsequent study, Ortega-Llebaria and Prieto (2009) investigated which of the two remaining acoustic correlates, i.e., intensity and duration, functions as the primary cue in the identification of stress. Results showed that Spanish listeners detected stress contrasts when the duration was manipulated, but with regard to intensity, they were only sensitive to changes in the overall intensity, while ignoring any changes in spectral tilt. Therefore, Ortega-Llebaria and Prieto maintain that among the three acoustic correlates duration is a most consistent cue to stress in Spanish; stressed syllables are produced with longer duration than unstressed syllables, regardless of whether they are accented or not (Ortega-Llebaria, 2006; Ortega-Llebaria & Prieto, 2007, 2009).

Compared to Spanish, the intent to disentangle the acoustic correlates of stress and pitch accent in English began earlier in the 1990s (Beckman & Edwards, 1994; Campbell & Beckman, 1997; Sluijter & van Heuven, 1996a, 1996b; Sluijter et al., 1997). Traditionally, f0 and duration were considered as the primary and secondary stress correlates in English, respectively, while intensity and vowel quality were claimed to have less importance (Fry, 1955, 1958). However, as in the case of Spanish stress, earlier studies in English stress failed to separate stress and pitch accent, because in those studies stress always covaried with pitch accent. Sluijter and van Heuven (Sluijter & van Heuven, 1996a, 1996b; Sluijter et al., 1997) were among the earliest researchers that controlled for stress-accent covariation. Sluijter and van Heuven (1996a, 1996b) examined the hierarchy of the acoustic correlates of stress in American English without the intervening effect of pitch accent by using target words with and without focal accent. Sluijter and van Heuven (1996a, 1996b) compared multiple acoustic correlates of stress (i.e., f0, duration, overall intensity, spectral tilt, vowel quality), and found that in unaccented contexts, stressed syllables were longer, had greater intensity in higher frequency region (i.e., related to spectral tilt), and had increased distance between F1 and F2 (i.e., related to vowel quality) than their unstressed counterparts; the f0 and overall intensity had little or no cue-value in unaccented contexts. Thus, according to Sluijter and van Heuven (1996a, 1996b), duration, spectral tilt, and vowel quality are respectively the first, second, and third most important acoustic cues to stress in American English, while f0 and overall intensity are the acoustic correlates of pitch accent.

In contrast to the claims of Sluijter and van Heuven, Beckman and colleagues (Beckman & Edwards, 1994; Campbell & Beckman, 1997) posit that the difference between stressed and unstressed syllables is primarily one of vowel quality; other acoustic properties such as duration and intensity are "parasitic" on vowel reduction differences (Ortega-Llebaria & Prieto, 2007). Beckman and Edwards (1994) examined the duration, displacement, and peak velocity for lip-opening movement in accented stressed, unaccented stressed, and reduced (i.e., unaccented unstressed) vowels, and found that the durational differences associated with the contrast between stressed full vowel and reduced vowel in unaccented context were larger and more consistent than those associated with the contrast between stressed full vowels in accented and unaccented contexts. These results are somewhat similar to those of Sluijter and van Heuven. However, instead of concluding that duration is the primary cue to stress, Beckman and Edwards (1994) argued that differences in duration are the consequence of vowel reduction. For instance, a full vowel [ɑ], will be much longer than the reduced counterpart [ə], because it requires a more open vocal tract, lower jaw, and larger lip displacement. In order to factor out the effects of vowel reduction, Campbell and Beckman (1997) controlled for vowel quality by using vowels that reduce to a lesser degree in unaccented contexts, i.e., [æ], [i], [u] (e.g., *Badd-Ellis*, *Beede-Ellis*, *Boode-Ellis*). The results showed that stressed syllables were longer and had greater energy in high frequency region only in accented conditions; in unaccented conditions, speakers varied in their use of duration and spectral tilt. Thus, in the absence of pitch accent, neither duration nor spectral tilt functions as a consistent cue to stress when there is no change in vowel quality. Rather, it is the vowel quality that functions as the decisive marker in identifying lexical stress in English; changes in duration and spectral tilt depend on the presence of vowel reduction.

3.1.2 Differences between Spanish and English lexical stress

From an acoustic/phonetic point of view, Spanish and English are different in the realization of stress (Ortega-Llebaria et al., 2013). One of the most salient acoustic cues in which Spanish and

English differ is whether vowel quality changes in unstressed syllables. In English, full vowels are preserved in stressed syllables, but almost always undergo reduction when unstressed, in most cases resulting in a schwa [ə] (e.g., *a*tom [ˈæ̲təm] vs. *a*tomic [ə̲ˈtɑmɪk]). Unlike in English, vowel reduction is not a phonological phenomenon in Spanish. Spanish vowels maintain the same vowel quality regardless of whether they are within stressed or unstressed syllables (e.g., *átomo* [ˈa̲tomo] vs. *atómico* [a̲ˈtomiko]) (Hualde, 2005; Quilis & Esgueva, 1983). The effect of stress seems to differ in the two languages not only on the realization of vowels, but also on the realization of consonants. In English certain consonants demonstrate allophonic distributions depending on whether the consonant in question is in a stressed or an unstressed syllable. For instance, English voiceless coronal stop /t/ is realized with aspiration [tʰ] in stressed onset or word-initial position (e.g., *attack* [ə.ˈt̲ʰæk]), but with flapping [ɾ] in unstressed and medial position (e.g., *letter* [lɛ.ɾ̲r]). Such allophony does not occur with Spanish consonants (e.g., *ataque* [a.ˈt̲a.que] 'attack' *lata* [ˈla.t̲a] 'can'). Apart from the realization of segments, Spanish and English differ with regard to the duration ratio of stressed to unstressed syllables (Ortega-Llebaria et al., 2013). As mentioned earlier, duration is one of the universal parameters of stress, i.e., stressed syllables are produced with longer duration than their unstressed counterparts. Nevertheless, according to Delattre (1966), there is a cross-linguistic variation, regarding the duration difference between stressed and unstressed syllables. When comparing English and Spanish, the duration difference between stressed and unstressed syllables is much larger in English than in Spanish, which may be attributed to the presence of vowel reduction in English and the absence of it in Spanish (Ortega-Llebaria et al., 2013). Another important difference between Spanish and English is the realization of lexical stress in prenuclear positions. As mentioned above, Spanish prenuclear pitch accents are characterized as a f0 valley aligned with the onset of the stressed syllable and a delayed f0 peak, which is located after the stressed syllable (expressed as L+>H*). However, in English, prenuclear pitch accents are marked as a f0 peak within the stressed syllable (expressed as L+H* or H*) (Estebas-Vilaplana, 2007; Lleó et al., 2004; Robles-Puente, 2014).

From a functional point of view, Spanish and English differ with regard to the role of stress. In English, stress minimal pairs, such as *<u>con</u>duct* and *con<u>duct</u>*, are always related in meaning and in most cases occur across different grammatical categories (i.e., noun vs. verb); while the nouns typically have stress on the left (i.e., *<u>con</u>duct*), the verbs have it on the right (i.e., *con<u>duct</u>*)

(Jensen, 1993). Since words that contrast in stress position pertain to different grammatical categories, listeners would be able to recover the intended message based on syntactic information, even if the stress is placed in the wrong position. For instance, if a listener hears **The CEO must consult with the board of directors prior to taking action*, although it may sound unnatural, he/she will still understand the meaning as *consult* (i.e., verb), not as *consult* (i.e., noun), based on the syntactic cues (Cutler, 1986). In Spanish, however, stress minimal pairs occur both within the same grammatical category (e.g., *papa* ‘pope/potato’ vs. *papá* ‘dad’), as well as across different grammatical categories (e.g., *término* ‘term vs. *termino* ‘I finish’), and they do not necessarily have relationship in meaning (e.g., *plato* ‘plate’ vs. *plató* ‘(scene) set’ (Saalfeld, 2009). For verb minimal pairs, such as *hablo* ‘I speak’ and *habló* ‘he/she/you (formal) spoke’, both forms belong to the same verb (i.e., *hablar* ‘to speak’), but the grammatical properties, such as person (i.e., *yo* ‘I’ vs. *él/ella/Ud.* ‘he/she/you (formal)’) and tense (i.e., present vs. past), are different. In this case, listeners would have to pay attention to the stress position, in order to understand the message, since there might not be cues other than stress that indicate these grammatical properties. For example, if a listener hears **Hable con el director* ‘I spoke with the director’, unless more information is retrievable from the context, he/she will incorrectly understand this sentence as *Hable con el director*. ‘(You (formal)) Talk with the director’, not as *Hablé con el director*. ‘I spoke with the director’ (Saalfeld, 2009). Due to such functional differences between Spanish and English stress, Cutler (1986, 2012) argued that English listeners do not use stress information as much as Spanish listeners for lexical processing, let alone suprasegmental cues.

The difference between Spanish and English listeners in their use of suprasegmental information has been supported by several perception studies (Cooper et al., 2002; Soto-Faraco et al., 2001). Soto-Faraco et al. (2001) investigated whether Spanish listeners attend to suprasegmental cues when accessing words in Spanish, using a cross-modal fragment priming paradigm. In their study, participants listened to auditory primes that were fragments of real word pairs that differed only in the position of lexical stress (e.g., PRINCI… from *principio* ‘beginning’ and PRINCI… from *príncipe* ‘prince’). After listening to a prime, the participants determined as fast as possible whether the visual target word (e.g., either *principio* or *príncipe*) presented on a computer screen was a real word or not. Results showed that, compared to the control condition, in which there was no segmental or suprasegmental overlap between the prime and the target word, the response time was faster when the stress pattern of prime matched with that of the target word (e.g., prime: PRINCI…; target: *príncipe*), whereas the response time slowed down when they did not match (e.g., prime: PRINCI…; target: *príncipe*). The facilitation and inhibition effects found in the matching and mismatching conditions, respectively, suggest that Spanish listeners use both segmental and suprasegmental cues for lexical processing. That is, lexical activation is facilitated when the prime and the target words match both segmentally and suprasegmentally. However, when the prime and the target words only match segmentally, but not suprasegmentally, inhibition of lexical activation occurs, because the mismatching word that was initially activated as a potential candidate would later need to be suppressed as a result of competition with a more favored candidate (i.e., the matching word) (McClelland & Elman, 1986; Norris, 1994).

With regard to English listeners' activation of English words, Cooper et al. (2002) found that, in matching condition (e.g., prime: ADMI…; target: *admiral*), English listeners responded more rapidly than in the control condition, similar to Soto-Faraco et al. (2001)'s study with Spanish listeners' perception of Spanish lexical stress. However, unlike in Soto-Faraco et al. (2001), no inhibition was found in the mismatching condition (e.g., prime: ADMI…; target: *admiration*) in this study. In fact, English listeners had similar response time in both mismatching and control conditions, and even a partial facilitation was observed when the prime was monosyllabic (e.g., prime: MUS…; target: *museum*). Cooper et al. (2002) argued that this may be due to segmental information being used as a more powerful tool in English than suprasegmental information for word recognition. As the authors pointed out, it is extremely difficult to find cases in English that are segmentally ambiguous, but suprasegmentally disambiguating, because in most cases, suprasegmental variation is accompanied by segmental variation (Campbell & Beckman, 1997). Thus, suprasegmental cues only provide redundant information, and because of that English listeners would profit little from such information (Cutler, 2012).

3.1.3 Previous studies on Spanish-English bilingual speakers' perception and production of Spanish lexical stress

Given the cross-linguistic differences between Spanish and English lexical stress, it is of interest whether heritage speakers of Spanish, who are more dominant in English, experience phonetic influence from English when they perceive and produce Spanish lexical stress. According to

Cutler (2012), any acoustic information that is available in the speech signal, whether it is segmental or suprasegmental, has a role in lexical activation. However, this does not signify that all listeners, regardless of their language background, attend to multiple acoustic cues to the same degree. Rather, it is more likely that listeners apply strategies that are efficient for processing their native or more dominant language and these strategies may influence the way they process the speech of a non-native or less dominant language. As the results of Cooper et al.'s (2002) study (presented above) indicate, English listeners do not attend to suprasegmental information to a large degree when identifying stress location, because in English variation in suprasegmental cues in stressed and unstressed contexts almost always covaries with segmental variation. Therefore, when no contextual cue is provided, it would be difficult for English listeners to process lexical stress in a language, such as Spanish, in which segmental cues provide little to no relevant information about stress location.

Indeed, studies on L2 phonology found that English L2 learners of Spanish experience great difficulties in identifying the location of lexical stress in Spanish (Kim, 2014, 2015; Ortega-Llebaria et al., 2013; Romanelli & Menegotto, 2015; Saalfeld, 2009, 2012). Ortega-Llebaria et al. (2013) conducted a detailed analysis on English L2 learners' perception of Spanish lexical stress with the objective to find which suprasegmental cues English listeners attend to when they process Spanish lexical stress, and how their processing strategies differ from those of native Spanish listeners. Ortega-Llebaria et al. (2013) created continua of tokens by acoustically manipulating a natural token *mama* 'mom'[6] and embedded them in two sentences of different prosodic contexts: a declarative sentence (i.e., *Saluda mama contenta*. 'Mom greets happily.'), in which stressed syllables carried a pitch accent, and a reporting clause (i.e., *¡Hola! – saluda <u>mama</u> contenta*. 'Hi! – greets mom happily'), in which the pitch contour was flat and stressed syllables did not carry any pitch accent. Token manipulation was done on pitch alignment and duration for the declarative sentence and on intensity and duration for the reporting clause, creating continua of tokens ranging from the natural range of an oxytone [maˈma] to the natural range of a paroxytone [ˈmama]. The participants were instructed to press a key on the keyboard whenever they heard an oxytone [maˈma]. Results showed that when the tokens were embedded in a declarative sentence, both the native Spanish listeners and the L2 learners responded

---

[6] The participants were instructed that, apart from *mama* [maˈma], which is the word that is mostly taught in L2 classrooms, *mama* [ˈmama] is also used as the colloquial term of *madre* 'mother'.

oxytone [maˈma] the most when the duration of the second syllable was the longest. However, the two groups showed different patterns regarding pitch alignment; while the L2 learners responded oxytone [maˈma] the most when the f0 peak aligned with the stressed syllable, the native listeners did so when the f0 peak was displaced to the syllable after the stressed syllable. Given that in this prosodic context (i.e., prenuclear position in a declarative sentence), English stressed syllables usually carry a H* pitch accent, whereas Spanish stressed syllables carry a L+>*H pitch accent, this finding suggests that the L2 learners were using their native strategies when processing Spanish lexical stress. With regard to the perception of tokens in the reporting clause, there was no effect of intensity, indicating that intensity does not play a critical role in perceiving Spanish lexical stress for any of the two groups. Regarding duration, significant between-group difference was found, specifically in the token that contained the longest second syllable (i.e., oxytone [maˈma]); when listening to this token, the native Spanish listeners responded oxytones [maˈma] significantly more than the L2 leaners. It is important to note that the second syllable of this token was not considerably longer than the first syllable (first syllable: 69 ms.; second syllable: 99 ms.). Given that in English, the duration difference between stressed and unstressed syllables is much larger than that in Spanish (Delattre, 1966), this finding implies that English listeners are not able to attend to fine duration differences, as effectively as Spanish listeners do. Although it is uncertain why the L2 learners attended to durational cues when perceiving tokens in the declarative sentence, while they did not when perceiving tokens in the reporting clause, this study supports the claim that the strategies used for native language processing affects the way L2 leaners perceive foreign language input.

English L2 learners' difficulty in perceiving Spanish lexical stress seems to maintain even after explicit instructions in Spanish stress system. Using a timed ABX stress discrimination task, Saalfeld (2009, 2012) examined L2 learners' ability to distinguish Spanish stress minimal pairs located in different positions within a sentence (i.e., initial, medial, and final). The minimal pairs consisted of Spanish verb pairs of paroxytones and oxytones (e.g., *regreso* 'I return' vs. *regresó* 'he/she/you (formal) returned'). Participants were divided into one of the four groups: L2 leaners that received focused instructions on Spanish stress system for four weeks (i.e., experimental group); L2 learners that received regular instructions on Spanish grammar, but nothing related to Spanish stress system (i.e., control group); naïve English listeners with no experience in Spanish; and native Spanish listeners. A pre-test and a post-test were conducted for both L2 learner groups in order to measure if any improvement was found after the instruction. Results showed that the experimental group did not show any advantage over the control group; both groups improved in accuracy after the four-week instruction[7], but no statistical between-group difference was found in the degree of improvement. Moreover, the improvement found in the two L2 learner groups was not significantly distinguishable from English listeners with no experience in Spanish, suggesting that, despite explicit instruction, English listeners have difficulty processing lexical stress solely based on suprasegmental information. Unlike Saalfeld (2009, 2012), Romanelli and Menegotto (2015) argued that English L2 learners' perception of lexical stress in Spanish may improve after immersed exposure to Spanish, even to a native-like level, but they also acknowledged that exposure to Spanish does not necessarily lead to improvement in the perception of all stress patterns to the same degree. Romanelli and Menegotto (2015) examined the perception of nonce word triplets that differed in the location of stress and/or the final vowel (e.g., SEMAPA vs. SEMAPÁ vs. SEMAPO) by L2 learners enrolled in a three-week Spanish immersion program in Argentina. Results showed that L2 learners' perception of stress position significantly improved after the immersion program, but, the improvement was comparable with native Spanish listeners only when listening to paroxytones; when listening to oxytones the L2 learners performed significantly worse than the native listeners. L2 learners' lower response accuracy in the perception of oxytones also applies to real words. Using stress minimal pairs of Spanish *-ar* verbs (e.g., *paso* 'I pass' vs. *pasó* 'he/she/you (formal) passed'), I have found in a previous study (Kim, 2014, 2015) that L2 learners' response accuracy for paroxytones was similar to that of native Spanish listeners, while their response accuracy for oxytones was lower than chance-level (35.24%). However, rather than being better in perceiving paroxytones than oxytones, it seems more likely that L2 learners were not successful in distinguishing the stress minimal pairs, but instead were having bias toward paroxytones, because when testing L2 learners' sensitivity and bias, results showed that the L2 learners had d-prime scores close to zero (i.e., low sensitivity) and C scores in the direction of bias toward paroxytones. Given that the L2 learners only had two options to choose from (i.e., paroxytones vs. oxytones), their high response accuracy for paroxytones may simply have been a by-product of their preference of paroxytones

[7] As for the improvement that was found in the control group, although this group did not receive explicit instructions regarding Spanish stress system, Saalfeld (2009, 2012) speculates that this may be due to the nature of the design of the experiment. Since ABX judgment tasks were conducted for both pre- and post-tests and no distracter items were included, it is possible that the participants noticed the purpose of the study and thus performed better in the post-test due to learning effect.

over oxytones. Although the exact reason of L2 learners' preference toward paroxytones is still unclear, as Romanelli and Menegotto (2015) pointed out, it is possible that the L2 learners over-generalized Spanish stress rules, according to which stress usually falls on the penultimate syllable if a word ends with a vowel.

Compared to research on perception, there are relatively few studies that examined the use of suprasegmental cues in the production of Spanish lexical stress by English L2 learners of Spanish. Given that vowel reduction is the most salient characteristic of the realization of English lexical stress, Menke and Face (2010) measured the F1 and F2 values of unstressed vowels produced by English L2 learners of different levels of Spanish instructions (i.e., students registered for fourth-semester Spanish course, graduating Spanish majors, and Ph.D. students in Spanish). In an oral reading task, in which participants read a short story, Menke and Face (2010) found that, regardless of the level of Spanish instructions, the L2 learners produced Spanish unstressed vowels with reduction toward the center of the vowel space, /a/ being the vowel that centralized the most. Regarding the other vowels, the reduction occurred more on the front-back dimension (F2) than on the high-low dimension (F1). Therefore, Menke and Face (2010) argued that accurate production of Spanish unstressed vowels appears to be a challenge to English L2 learners, even more than the perception. Although not intended to examine the production of lexical stress *per se*, Stevens (2011) compared the duration of the vowels produced by two groups of English L2 learners: those who participated in a four-week summer study abroad program in Spain and those who learned Spanish "at home" in an American University. Stevens (2011) found that, although overall the study-abroad group produced Spanish vowels shorter than the other group, both groups produced unstressed vowels shorter than stressed vowels. However, since vowel quality was not examined in this study, it is unclear whether the shorter duration of the unstressed vowels found in this study was due to reduction of vowel in the unstressed condition, and not because the L2 learners were using duration as an acoustic correlate to distinguish stressed vowels from unstressed vowels.

The findings in L2 phonology literature are helpful in understanding the influence of English on Spanish heritage speakers' perception and production of lexical stress in Spanish. However, so far, very little research has been done on Spanish heritage speakers' perception and production of Spanish lexical stress and most of them have focused mainly on the effect of stress on vowel quality. Willis (2005) examined the effect of stress on the vowel quality of Spanish /a/ of four female fluent bilingual speakers from the border area between the U.S. and Mexico (e.g., El Paso, TX and Anthony, NM). Semi-spontaneous speech data of the speakers were collected by conducting a picture description task and an oral interview regarding their experience with Spanish and family. Words that contained /a/ in non-final position of the phrase were extracted and the formant values (i.e., F1 and F2) of /a/ in stressed position (e.g., *El niño mi<u>ra</u>ba la <u>ra</u>na.* 'The boy was looking at the frog.') and unstressed position (e.g., *El niño mira<u>ba</u> la ra<u>na</u> saltar.* 'The boy was looking at the frog jump.') were compared. In order to avoid possible confounding effects occurring from different syllable types, only those in open syllables were considered in the study. For three out of the four participants, no significant difference was found between the formant values of the stressed and unstressed syllables, which indicates that these speakers did not use vowel quality to distinguish stressed and unstressed vowels. While one participant showed significant change toward centralization (i.e., lower F1 values and higher F2 values) in unstressed positions, Willis (2005) concluded that this variation was nevertheless non-categorical, because there was a large overlap between the distribution of the vowels in both contexts and those of the other speakers. However, since the effect of stress was examined only on /a/ in this study, it is difficult to generalize the findings to all Spanish vowels without taking into account other vowels as well. Ronquest (2012) conducted an exhaustive investigation on Spanish heritage speakers' production of all five Spanish vowels (i.e., /a, e, i, o, u/) and the effect of lexical stress on them. Using tasks that elicit different speech styles (i.e., narrative retelling task, picture identification task, and carrier phrase task), Ronquest (2012) found that vowels tend to be more centralized when unstressed. That is, unstressed vowels, especially non-high vowels /a, e, o/, had lower F1 values than the stressed vowels, indicating that vowels produced in unstressed condition were generally positioned higher in the vowel space. With regard the front-back dimention (F2), the front vowels /i, e/ showed lower values in unstressed syllables, while the back vowels /o, u/ had higher values. This suggests that all vowels, except for /a/, moved towards the center of the vowel space when unstressed.

Although changes in vowel quality is a good indication of influence from English, it is important to examine heritage speakers' use of suprasegmental cues, especially duration, as this is the primary acoustic cue to lexical stress in Spanish (Ortega-Llebaria, 2006; Ortega-Llebaria & Prieto, 2007, 2009, 2011). Unfortunately, little research has been done on Spanish heritage speakers' use of suprasegmental information. Ronquest (2012) is the only study that I have found so far that investigated the effect of stress on vowel duration, while controlling for the effect of vowel quality. Ronquest (2012) found that heritage speakers produced unstressed vowels shorter than stressed vowels, regardless of vowel type. Moreover, compared to the monolingual Spanish speakers in Marín Gálvez's (1995) study, heritage speakers' duration difference between stressed and unstressed vowels was much larger; heritage speakers' stressed vowels were 45% longer than the unstressed counterparts, while for the monolingual speakers in Marín Gálvez's (1995) study the stressed vowels were only 20.3% longer than the unstressed vowels. As duration difference between stressed and unstressed syllables is much larger in English than in Spanish (Delattre, 1966), this finding implies a possible influence from English. While Ronquest's (2012) study is of significant importance in understanding English influence on Spanish heritage speakers' production of lexical stress in Spanish, this area is severely understudied and more research needs to be done on heritage speakers' use of multiple suprasegmental cues in different prosodic contexts, because, while duration is the primary cue to lexical stress in Spanish, varying suprasegmental cues signal lexical stress, based on the prosodic context in which the stressed syllables are located. Moreover, research on Spanish heritage speakers' perception of Spanish lexical stress is almost non-existent. In a previous study (Kim, 2015), I examined both Spanish heritage speakers' perception and production of stress minimal pairs of Spanish *-ar* verbs (e.g., *paso* 'I pass' vs. *pasó* 'he/she/you (formal) passed') and compared their behavior to that of Spanish monolingual speakers and English L2 learners of Spanish. Results showed that, while heritage speakers successfully identified the location of lexical stress when listening to stress minimal pairs, to the same level as monolingual speakers, they showed a large overlap between the two stress patterns when producing them. That is, unlike the monolingual speakers, who consistently produced the stressed vowels longer than the unstressed vowels, regardless of the stress pattern, the heritage speakers produced the final vowel longer than the penultimate vowel in most cases, even when the target word was a paroxytone. The same pattern was found in L2 learners' speech. I argued that the longer unstressed vowels found in heritage speakers' speech may be due to their preference of past tense verbs, which were oxytones in her study, as heritage speakers tend to be better at narrating an event in the past than talking about current events (Martin et al., 2013). However, further research on heritage speakers' speech pattern, such as conversation topic with Spanish speakers, is necessary to confirm this. Another possible explanation is that heritage speakers' (and L2 learners') longer unstressed vowels may be an artifact of the sentence structure of the stimuli; the target words were all embedded in a carrier sentence with subject-verb inversion (e.g., *Por la plaza* <u>*paso*</u> *yo.* 'Through the square I passed'), which is a strategy used in Spanish to express narrow focus on subject, but not in English. Although no clear boundary cue was found in the speech signal, such as pause and glottalization, I suggested that, due to their unfamiliarity with the sentence structure (Lynch, 2003), it is likely that the heritage speakers and the L2 learners put a phonological phrase boundary [[………paso]$_{PPH}$ [yo]$_{PPH}$]$_{IP}$), which may have led to final vowel lengthening. However, in order to confirm this, it is necessary to examine heritage speakers' production of Spanish lexical stress in various prosodic contexts.

3.2 Research questions and predictions

The present study has its objective in filling the gap in the literature, by examining both Spanish heritage speakers' perception and production of Spanish lexical stress in three different prosodic contexts: (1) nuclear position in a short declarative sentence, (2) prenuclear position in a declarative sentence, and (3) unaccented context, in which the stressed syllable does not bear any pitch accent. Apart from duration, which is the most consistent cue to stress in Spanish (Ortega-Llebaria, 2006; Ortega-Llebaria & Prieto, 2007, 2009), in each prosodic context different acoustic correlates are used that signal lexical stress. In nuclear position, stress covaries with nuclear pitch accent, especially in a short declarative sentence, providing an abundance of suprasegmental information about the stress location (i.e., f0, intensity, and duration). Stress in prenuclear position is characterized as a f0 valley in the stressed syllable followed by a delayed f0 peak (L+>H*). In unaccented context, duration functions as the main acoustic correlate of stress. Therefore, identifying the location of lexical stress in this last context would be very difficult, because, compared to the first two contexts, there would be less acoustic cues available in the speech signal. Indeed, Torreira et al. (2014) found that, in unaccented contexts, stress minimal pairs are produced with overlapping acoustic information, which leads to confusion to Spanish native listeners.

The following are the research questions and the predictions of the present study.

- Research Question 1: Do heritage speakers and L2 learners successfully attend to varying suprasegmental cues when listening to Spanish stress minimal pairs in different prosodic contexts?
    - Prediction 1: If transfer from English occurs, they will show confusion equally across all contexts due to lack of segmental information (i.e., vowel reduction).
    - Prediction 2: If they successfully attend to suprasegmental cues, they will perform better when there are more cues available (N: pitch, intensity, duration; PN: pitch alignment, duration) than when there are less (U: duration).

- Research Question 2: Do heritage speakers and L2 learners successfully use varying suprasegmental cues when producing Spanish stress minimal pairs in different prosodic contexts?
    - Prediction 1: If transfer from English occurs, they will reduce unstressed vowels and show early alignment of pitch peak in prenuclear position.
    - Prediction 2: If they successfully use suprasegmental cues, they will adjust their use of duration and tonal cues based on the prosodic context (N: pitch, intensity, duration; PN: pitch alignment, duration; U: duration), while maintaining the vowel quality in unstressed positions.

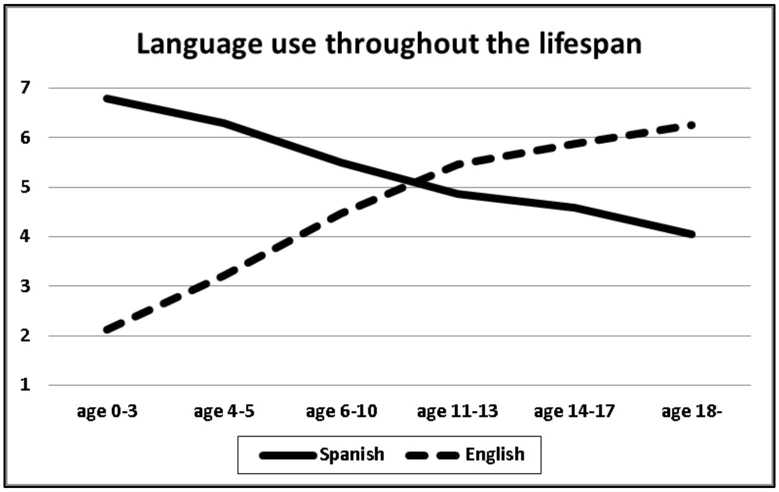
3.3 Experiment 1: Perception of lexical stress

3.3.1 Methods

3.3.1.1 Participants

In total, 68 subjects participated in the study: 24 monolingual native speakers of Spanish (13F, 11M) (avg. age: 22.92 years), 24 heritage speakers of Spanish (18F, 6M) (avg. age: 21.04 years), and 20 English L2 learners of Spanish (14F, 6M) (avg. age: 20.95 years). All the subjects were college students or college-educated. The monolingual native speakers were recruited at the Autonomous University of Querétaro in Santiago de Querétaro, Mexico, and the recruitment of the heritage speakers and the second language learners took place at the University of Illinois at Urbana-Champaign, IL.

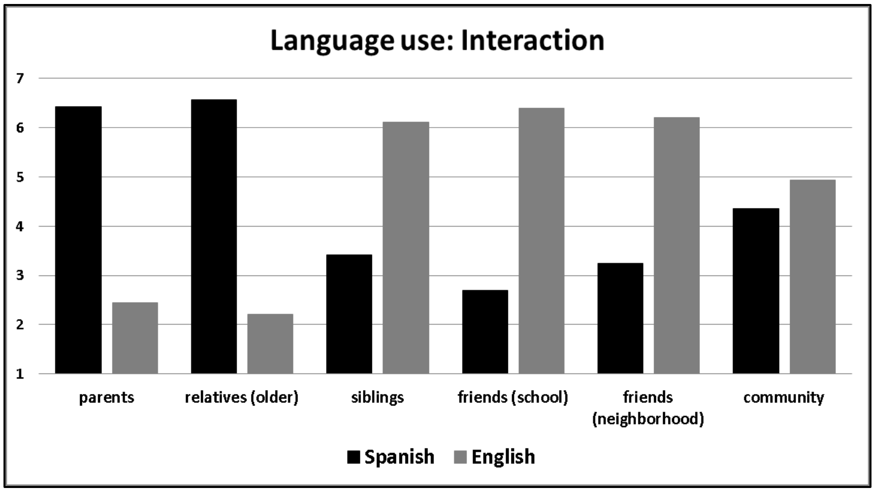
Heritage speakers of Spanish (HS), the focus group of the present study, were first generation U.S.-born Mexican-Americans, whose parents arrived to the U.S. from different areas of Mexico as adults, primarily from the central-west region of the country (i.e., Michoacán, Mexico City, Guerrero, Guanajuato, and Jalisco). All the HSs reported that they learned both Spanish and English from a young age. Twenty one out of 24 HSs reported that they acquired Spanish first at home and English in preschool or elementary school (i.e., early sequential bilinguals), while three HSs reported that they acquired both Spanish and English at home simultaneously (i.e., simultaneous bilinguals). The majority of the HSs (i.e., 19 out of 24 subjects) reported that they grew up in a Hispanic neighborhood in the Chicago metropolitan area, such as Little Village, Pilsen, Brighton Park, and West Lawn, in which the percentage of Hispanic population is 80% or higher, according to the 2010 U.S. Census. Therefore, although the HSs learned both Spanish and English at an early age, it is likely that they were mainly exposed to Spanish in their early childhood until they entered institutions in which the primary language of instruction was English. Indeed, when asked about their use of Spanish and English throughout the lifespan (Figure 3), the HSs responded that their use of Spanish gradually decreased as they grew up, while their use of English increased, although the degree of increase varied by individual, due to the variation in the use of English at early ages (0-5 years).



	age 0-3	age 4-5	age 6-10	age 11-13	age 14-17	age 18-
Spanish	6.79 (0.51)	6.29 (0.75)	5.5 (1.25)	4.88 (1.19)	4.58 (1.06)	4.04 (1.27)
English	2.13 (2)	3.21 (2.01)	4.46 (1.7)	5.46 (1.19)	5.88 (0.96)	6.26 (0.79)

Figure 3 Heritage speakers' average language use throughout the lifespan (1: Never, 7: Always) (values within the parentheses are standard deviation values)

Decrease in Spanish use is also clearly reflected in HSs' current use of Spanish and English. The HSs reported that they used Spanish (avg. 22.42%, s.d. = 9.14%) far less frequently than English (avg. 76.25%, s.d. = 10.42%). Moreover, as shown in Figure 4, their use of Spanish was generally limited to interactions with parents and other family members of older generation (e.g., grandparents, aunt, uncle), while they mainly spoke in English with family members of younger generation (i.e., siblings) and friends from school. With regard to members of their speech community (e.g., friends in the neighborhood, church, grocery stores), the HSs showed varied responses. When interacting with friends in the neighborhood, four HSs (out of 24) responded that they either spoke more Spanish than English or spoke both languages to the same degree, while the rest responded that they spoke more English than Spanish. The number of HSs who use Spanish increased when communicating with other members of the community; eleven HSs reported that they spoke more Spanish than English or spoke both languages to the same degree. HSs' report above suggests that HSs' use of Spanish is usually confined to familial settings and, for some speakers, to their speech community, while English is predominantly used in most settings.



	Parents	Relatives (older)	Siblings	Friends (school)	Friends (neighborhood)	Community
Spanish	6.42 (0.88)	6.56 (0.78)	3.41 (1.51)	2.69 (1.17)	3.25 (1.99)	4.35 (2.15)
English	2.44 (1.3)	2.21 (1.34)	6.11 (0.8)	6.4 (0.77)	6.21 (1.39)	4.94 (1.97)

Figure 4 Heritage speakers' average language use with different interlocutors (1: Never, 7: Always) (values within the parentheses are standard deviation values)

Monolingual native speakers of Spanish (NS) participated in the study as a control group. All the NSs were native speakers of Mexican Spanish and were either born in Santiago de Querétaro, Mexico[8], which is the capital of the state of Querétaro located in the central region of Mexico and grew up there (i.e., 17 out of 24 subjects), or were born in a neighboring state (e.g., Michoacán, Guanajuato, Mexico City) and moved to Santiago de Querétaro during childhood or adolescence. All the NSs were monolingually raised in Spanish and did not learn languages other than Spanish until age 13.29, on average (range: 9-18 years). Although the NSs have learned languages other than Spanish, mostly English, given that they use Spanish most of the time (avg. 82.58%, s.d. = 13.05%) and do not use the other languages functionally, it is considered to be unlikely that these languages would have an effect on the NSs' performance in Spanish in the present study. Therefore, the present study will consider them as an appropriate monolingual control group.

[8] Due to the expansion of industry and close proximity to Mexico City, Santiago de Querétaro has experienced a massive influx from surrounding states such as the State of Mexico, Guanajuato, San Luis Potosí, and Michoacán ("Aumenta migración", 2011). Moreover, because the crime rate in northern states of Mexico has risen significantly, there has been a "silent migration" from northern states as well such as Durango and Chihuahua (Arreola, 2011). Given that the HSs in the present study have their roots in different areas of Mexico, mostly from the central-west region, the NSs from Santiago de Querétaro were considered to be an appropriate control group.

Lastly, English L2 learners of Spanish (L2) who were native speakers of American English and grew up speaking only English participated in the study. Except for one speaker, who was from Central Illinois, all the L2s were born and raised in suburban areas of Chicago, Illinois, in which the percentage of Hispanic population is very low (avg. 7.23%). The L2s also reported that they did not have close contact with Hispanic population while growing up. All the L2s started learning Spanish at the mean age of 13.25 (range: 10-20 years) (i.e., late sequential bilinguals) and were enrolled in an upper-division undergraduate course in Spanish at the time of testing. With regard to current use of Spanish and English, as in the case of the HSs, the L2s reported that they were using Spanish (avg. 10%, s.d. = 7.35%) far less frequently than English (avg. 89.5%, s.d. = 8.04%).

Participants' language dominance was evaluated using two quantitative measures: one subjective and the other objective. For the subjective measure, the participants completed a Bilingual Language Profile (BLP) (Birdsong et al., 2012), which is a questionnaire that produces a continuous score of global language dominance on a scale of 0 to 218, based on participants' self-report on their language history, language use, language proficiency, and language attitudes in Spanish and English. In the case of the NSs, only the Spanish results are reported. HSs' and L2s' BLP scores in Spanish and English were compared using a two-way mixed Analysis of Variance (ANOVA) with group (HS / L2) and language (Spanish / English) as independent variables and BLP scores as the dependent variable. The *aov* function in R (Baayen, 2008) was used for the analysis. For post-hoc pairwise comparisons, Tukey HSD analysis was conducted using the *TukeyHSD* function. Results showed that the BLP scores were higher in English than in Spanish ($F(1, 84) = 705.678$, $p < 0.001$). While no main effect of group was found, there was a significant interaction between group and language ($F(1, 84) = 127.081$, $p < 0.001$). Post-hoc pairwise comparisons revealed that, the HSs' BLP scores in English were significantly lower than those of the L2s ($p < 0.001$), whereas their BLP scores in Spanish were higher than those of the L2s. When comparing HSs' and L2s' BLP scores in Spanish with those of the NSs, a main effect of group was found ($F(2, 65) = 364.1$, $p < 0.001$), which, as the post-hoc pairwise comparisons have revealed, was due to NSs' significantly higher BLP scores than those of the HSs and the L2s ($p < 0.001$).

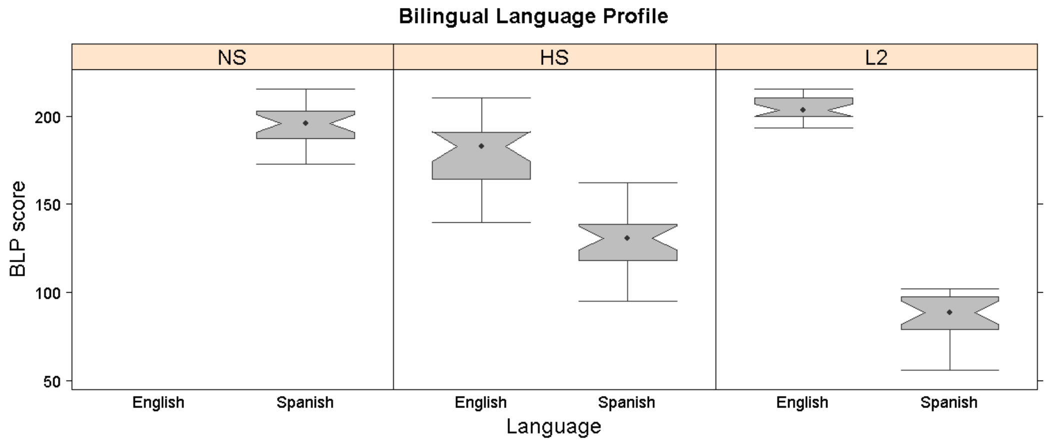


Figure 5 Bilingual language profile scores

For the objective measure, an oral picture-naming task was conducted in each language to evaluate participants' lexical knowledge, as lexical knowledge is considered to be a powerful predictor in determining individuals' general language proficiency (Polinsky & Kagan, 2007; Polinsky, 1997, 2000, 2006). Similar to the methods used in Fairclough and Ramírez Vera (2009), the present study extracted Spanish words from Davies (2006)'s Spanish frequency dictionary, "A frequency dictionary of Spanish: Core vocabulary for learners", which contains a list of the 5,000 most commonly used Spanish words, based on a corpus of 20 million words from various contemporary texts of different registers (i.e., oral and written) and locations (i.e., Spain and Latin America). Only object nouns that are not cognates with English were selected and each object noun was assigned to one of five frequency levels depending on its rank. That is, object nouns among the first 1,000 most frequent Spanish words were assigned to level 1, those among the second 1,000 most frequent words were assigned to level 2, and so on. The object pictures for the picture-naming task were then selected from the International Picture Naming Project (IPNP) database (Szekely et al., 2004)[9], which is a corpus of 520 black-and-white drawings of common objects that are normed for picture naming in seven different languages, including Mexican Spanish and American English. In total, 60 picture items (i.e., 12 items per frequency level) with the highest name agreement rates and the lowest number of alternative names were chosen to avoid possible confusion in identifying the object. For the Spanish version, the average name agreement rate was 98.3% (s.d. = 3.16%) and the average number of alternative names was 1.47 (s.d. = 0.65). For the English version, the average name agreement

[9] The database is available online at: http://www.crl.ucsd.edu/~aszekely/ipnp/

rate was 97.13% (s.d. = 4.9%) and the average number of alternative names was 1.63 (s.d. = 0.76). Special care was taken to match the word frequency[10] between the items in the two language versions as much as possible. The average absolute difference in word frequency (calculated through log natural transformation) was 0.71.

The oral picture-naming task was conducted using *PsychoPy2* (Peirce, 2007). After a beep, the object pictures were presented on a computer screen in a randomized order and the participants were instructed to name each object out loud in the language of interest as quickly as possible. The HSs and the L2s completed the task in both Spanish and English. To avoid possible priming effect, the Spanish task was conducted at the beginning of the study and the English task was conducted at the end with an interval of approximately 90 minutes in between. The NSs completed only the Spanish task at the beginning of the study. Prior to the task, the participants completed a practice trial with five items to familiarize themselves with the task format. The accuracy and response time were coded by two trained research assistants. Correct responses were coded as "1" and incorrect responses were coded as "0". Only complete words were considered correct and errors in any segments of the word were considered incorrect (e.g., **flora* instead of *flor* 'flower', **corbota* instead of *corbata* 'tie'). In the case of responses with a definite or an indefinite article in the Spanish task (e.g., *el sol* 'the sun', *un libro* 'a book'), errors in gender (e.g., **un mano* instead of *una mano* 'a hand') were disregarded and the responses were considered as correct as long as the word was produced correctly. The response times were measured only for the correct responses and they were calculated as the time elapsed between the end of the beep until the beginning of the response. Responses with a filler word (e.g., *um*, *uh*, *eh*), coughing, sigh, elongated syllable (e.g., *caaaaaaja* 'box'), stuttering (e.g., un cor… corazón 'a har... heart'), or self-correction (e.g., *¿Paloma? Pluma.* 'Dove? Pen.') were excluded from the analysis, although they were considered as correct responses. For the responses with an article, the response time was calculated until the beginning of the article as long as no noticeable pause between the article and the word was detected. Response times that were more than two standard deviations away from the average response time were excluded from the analysis. Among 5,380

[10] The IPNP database also contains word frequency, which is calculated as the log natural frequency from various corpora. For Spanish, the corpus of Alameda and Cuetos (1995) was used to measure word frequency and, for English, the CELEX database was used.

correct responses, 257 tokens were excluded due to lengthy response times and the reasons above, resulting in a total number of 5,123 tokens for the analysis.

HSs' and L2s' dominance in Spanish and English was compare by examining the effects of group (HS / L2), language (Spanish / English), and the interaction between the two fixed factors on accuracy and response time with subject and item as random effects. For the analysis of accuracy, logit mixed effects modeling was conducted using the *glmer* function in the *lme4* package in R (Baayen, 2008) and, for the analysis of response time, linear mixed effects modeling was performed using the *lmer* function. The best fitting model selected through backward elimination included random intercepts for subject and item with by-item random slope for group, in the case of the analysis of accuracy, and random intercepts for subject and item with by-item random slope for group and by-subject random slope for language. All the fixed factors (i.e., group and language) were centered using contrast-coding. Further pairwise analyses were conducted using the *lsmeans* function in the *lsmeans* package.

Results showed that the accuracy rates were higher and the response times were shorter in the English task than in the Spanish task (accuracy: $\beta = -4.993$, SE = 0.68, $z = -7.337$, $p < 0.001$; response time: $\beta = 0.315$, SE = 0.034, $t = 9.14$). Moreover, there was a significant interaction between language and group (accuracy: $\beta = -3.655$, SE = 0.684, $z = -5.346$, $p < 0.001$; response time: $\beta = 0.206$, SE = 0.052, $t = 3.94$), which indicates that the difference in accuracy and in response time between the two languages was significantly larger for the L2s than for the HSs. Pairwise comparisons with language and group of the accuracy data confirmed that, while the HSs and the L2s received similar scores in the English task, the HSs performed significantly better than the L2s in the Spanish task ($p < 0.001$). However, no significant difference was found between HSs' and L2s' response time in any of the two languages. When comparing HSs' and L2s' accuracy and response time in Spanish with those of the NSs, results showed that the NSs had higher accuracy rates and responded faster than both the HSs (accuracy: $\beta = -4.027$, SE = 0.681, $z = -5.913$, $p < 0.001$; response time: $\beta = 0.276$, SE = 0.056, $t = 4.907$) and the L2s (accuracy: $\beta = -2.285$, SE = 0.46, $z = -4.968$, $p < 0.001$; response time: $\beta = 0.144$, SE = 0.067, $t = 2.164$).

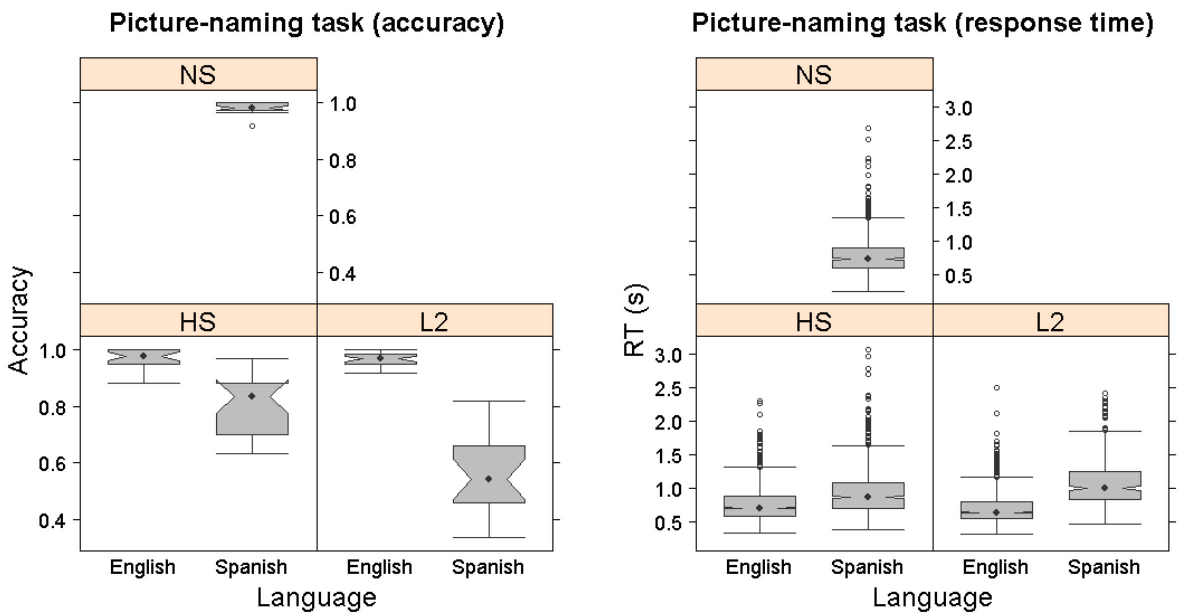


Figure 6 Oral picture-naming task (left: accuracy, right: response time)

The findings of the two language dominance measures (i.e., Bilingual Language Profile and oral picture-naming task) suggest that the HSs and the L2s considered that they are more dominant in English than in Spanish and this was confirmed when objectively testing their lexical knowledge, which is considered to be a strong predictor in determining general language proficiency (Polinsky & Kagan, 2007; Polinsky, 1997, 2000, 2006). Moreover, when compared with the NSs, HSs' and L2s' dominance in Spanish was lower than that of the NSs in both the subjective and objective measures, although HSs' Spanish dominance was significantly higher than that of the L2s. Interestingly, while the HSs self-rated their English dominance lower than the L2s, no significant difference was found in their lexical knowledge from those of the L2s. These findings indicate that HSs' early and immersed exposure to Spanish not only puts them in an advantageous position over L2s with regard to their dominance in Spanish, but also it does not interfere with their dominance in English. However, the findings also imply that HSs' early and immersed exposure to Spanish alone does not guarantee Spanish dominance that is comparable to monolingual native speakers of Spanish and continued use of Spanish past childhood may be necessary.

3.3.1.2 Materials

Sixty minimal pairs (i.e., 120 words) that only differed in the position of lexical stress (e.g., *Canto*. 'I sing'. vs. *Cantó*. 'he/she sang'.) were used in Experiment 1. The minimal pairs consisted of di- and tri-syllabic Spanish regular *-ar* verbs in the first person singular of the present indicative form (e.g., *Canto*. 'I sing'.) and the same verb in the third person singular of the preterit or (simple) past perfective tense (e.g., *Cantó*. 'he/she/you (formal) sang'.). The former case always had stress in the penultimate syllable (i.e., paroxytone) and the latter case always had it on the last syllable (i.e., oxytone). The word frequency of the minimal pairs was matched, based on whether they were high or low frequency words, using *Corpus del Español* (Davies, 2002). The cut-off point between high and low frequency words (i.e., 49 tokens) were determined as the median of all first person singular of the present indicative verbs and all third person singular of the preterit verbs found in the corpus. The target items were equally distributed with regard to the number of syllables (i.e., di- and tri-syllabic), vowel types in the verb stem (i.e., the penultimate vowel) (i.e., /a, e, i, o, u/), stress pattern (i.e., paroxytone and oxytone), and tense (i.e., present indicative and preterit). Apart from the target items, 40 filler items of inflected Spanish verbs that do not form stress minimal pairs (e.g., *Sufrió*. 'He/she/you (formal) suffered.', *Aprendes*. 'You learn.') were included. The filler items were included to confirm that the participants were able to successfully conjugate Spanish verbs. If participants correctly identify the subject of the filler items, this would indicate that any non-target-like behavior found in the target words would not be confounded with possible problems in verb conjugation. The filler items matched in the number of syllables (i.e., di- and tri-syllabic), stress pattern (i.e., paroxytone and oxytone), and tense (i.e., present indicative and preterit) with the target items. The items were embedded in meaningful sentences in three prosodic contexts: nuclear position (N), prenuclear position (PN), and unaccented context (U), as shown in examples (1)-(3) below.

(1) Nuclear position (N)

e.g., *Canto*. 'I sing'.

(2) Prenuclear position (PN)

e.g., *Canto la balada*. 'I sing the ballad'.

(3) Unaccented context (U)

e.g., *¿Dónde canto con micrófono*? ‘Where do I sing with microphone?’

With the goal to facilitate the extraction of the target items for further acoustic analyses, the final vowels were adjacent to consonants which are characterized to mark a clear acoustic boundaries in Spanish, such as voiceless obstruents (e.g., /p/, /k/, /t/, /s/), non-palatal nasals (e.g., /m/, /n/), and liquids (e.g., /l/). Thus, for PN and U, the following word always began with one of these consonants, creating a clear boundary from the final vowel of the target word. With regard to the target words in PN, the following word always began with a voiced consonant (i.e., non-palatal nasal or liquid) in order to identify the location of the pitch (f0) peak (H).

A total number of 480 items (i.e., 120 target items and 40 filler items in three prosodic contexts) were produced by a male native speaker of Mexican Spanish and the recording took place in a sound-attenuated booth at the University of Illinois Phonetics and Phonology Lab using an AKG C520 head-mounted microphone, which was positioned approximately 2 inches away from the participants’ lips, and a Marantz PMD570 solid state recorder with a sampling rate of 48 kHz and a sample size of 16 bits. The target items were divided into six lists, each consisting of 180 items (i.e., 20 target items and 40 filler items in three prosodic contexts). This was a measure to avoid priming effect, so that the minimal pairs of the same verb did not appear in the same list. That is, an inflected verb in one of the three prosodic context (e.g., paroxytone in PN: *Canto la balada* ‘I sing the ballad.’) only appeared in one list. Acoustic analyses of the stimuli confirmed that, among various acoustic correlates, duration was the only cue that systematically distinguished stressed from unstressed vowels regardless of prosodic context. Moreover, it was revealed that items in U had the least correlates available (i.e., duration), while more acoustic information signaled stress when the items were in N (i.e., duration, intensity, pitch) or in PN (i.e., duration, pitch alignment). The list of stimuli and the acoustic information of the stimuli are presented in Appendix E.1. and Appendix F.1.

3.3.1.3 Procedures

Each participant was assigned to one of the six lists. A forced-choice identification task was conducted using *PsychoPy2*, in which the participants had to identify which of the two subjects that appeared on each side of the computer screen was the subject of the sentence that they heard, by pressing either the left key (i.e., “a”) or the right key (i.e., “l”) on the keyboard as quickly as possible. The stimuli were presented in a randomized order and the order of the options was counter-balanced. In both testing sites (i.e., Mexico and U.S.), the experiment was conducted in a phonetics laboratory equipped with a sound-attenuated booth. The stimuli were presented through a Lenovo T430s laptop computer with Sennheiser HD 558 headphones. Before the initiation of the experiment, the participants were informed that they would listen to Spanish sentences, in which the subjects were absent, and that their task was to choose which of the two options on the computer screen was the subject of the sentence. A practice trial with five items, which were not the target items, was conducted for the familiarization with the task format.

3.3.1.4 Coding and analysis

All the responses were collected through *PsychoPy2*. The correct responses were automatically coded as “1” and the incorrect responses were coded as “0”. The response times were measured only for the correct responses and they were calculated as the time elapsed between the onset of the target word until the moment the participants pressed the key. Response times that were more than two standard deviations away from the average response time were excluded from the analysis. The effects of group (NS / HS / L2), prosodic context (N / PN / U), stress pattern (paroxytone / oxytone), and the interaction among the fixed factors on participants’ accuracy and response time were analyzed with subject and item as random effects. For the analysis of accuracy, logit mixed effects modeling was conducted using the *glmer* function in the *lme4* package in R (Baayen, 2008) and, for the analysis of response time, linear mixed effects modeling was performed using the *lmer* function. The best fitting model selected through backward elimination included random intercepts for subject and item, in the case of the analysis of accuracy, and random intercepts for subject and item with by-item random slope for group and by-subject random slope for stress pattern, for the analysis of response time. All the fixed factors (i.e., group, prosodic context, and stress pattern) were centered using contrast-coding. Further pairwise analyses were conducted using the *lsmeans* function in the *lsmeans* package.

3.3.2 Results

3.3.2.1 Accuracy

The accuracy of a total number of 4,080 target tokens (i.e., 20 items * 3* prosodic contexts * 68 participants) were analyzed. Figure 7 shows the participants' accuracy rates by group, prosodic context, and stress pattern. Results showed that there was a main effect of group for both the HS and the L2, which indicates that overall the NSs (i.e., the baseline group) performed better than both the HSs (β = -1.332, SE = 1.889, z = -7.051, $p < 0.001$) and the L2s (β = -1.988, SE = 0.206, z = -9.633, $p < 0.001$). A main effect of prosodic context was also found for both PN and U. That is, the overall accuracy in N (i.e., the baseline prosodic context) was higher than that in PN (β = -0.588, SE = 0.153, z = -3.819, $p < 0.001$) and U (β = -0.879, SE = 0.165, z = -5.317, $p < 0.001$). While no main effect was found for stress pattern, there was a significant interaction between group (L2) and stress pattern. This suggests that the accuracy difference between the NSs and the L2s was larger for oxytones than for paroxytones (i.e., the baseline stress pattern) (β = -2.176, SE = 0.25, z = -8.708, $p < 0.001$). A significant interaction was also found between group (L2) and prosodic context (U), indicating that the accuracy difference between the NSs and the L2s was larger in U than in N (β = 0.784, SE = 0.285, z = 2.747, $p < 0.01$). Indeed, Figure 7 clearly shows that while L2s' accuracy rates for paroxytones were similar to those of the NSs and the HSs, their accuracy rates for oxytones were considerably lower than those of the NSs and the HSs. Pairwise comparisons of group and stress pattern confirmed that, while L2s' accuracy rates were significantly lower than those of the NSs for both stress patterns ($p < 0.001$ for both), the accuracy difference between the two stress patterns were found to be significant only for L2s. That is, L2s' accuracy rates for oxytones were significantly lower than those for paroxytones, while this was not the case for the NSs, who had similar accuracy rates for the two stress patterns. Unlike the L2s, the HSs had significantly higher accuracy rates for oxytones than for paroxytones ($p < 0.05$). However, their accuracy rates did not differ from the NSs in any of the two stress patterns. Moreover, like the NSs, the HSs had significantly higher accuracy rates than the L2s for paroxytones ($p < 0.01$), as well as for oxytones ($p < 0.001$).

With regard to the interaction between group and prosodic context, Figure 7 shows that L2s' overall accuracy was noticeably lower in U than in N, mainly due to the decrease in the accuracy rates for paroxytones from N (average 86%) to U (average 56%). The drop was not observed to such a degree for either the NSs (from average 93.75% in N to average 88.75% in U) or the HSs (from average 91.67% in N to average 77.92% in U). Pairwise comparisons of group and prosodic context showed that, while all three groups had significantly lower accuracy rates in U than in N (NS & L2: $p < 0.01$, HS: $p < 0.001$), the L2s had significantly lower accuracy rates than both the NSs and the HSs in all three prosodic contexts ($p < 0.001$ for all). As for the accuracy difference between PN and U, both the NSs and the HSs had significantly higher accuracy rates in PN than in U ($p < 0.001$ for both), while the L2s did not have significantly different accuracy rates in the two prosodic contexts. This indicates that while for the NSs and the HSs the accuracy rates dropped abruptly in U, for the L2s, they dropped steadily from N to PN and from PN to U.

Lastly, three-way interactions were found (1) among group (HS), prosodic context (PN), and stress pattern ($\beta = 1.235$, $SE = 0.57$, $z = 2.166$, $p < 0.05$), (2) among group (HS), prosodic context (U), and stress pattern ($\beta = 1.55$, $SE = 0.591$, $z = 2.621$, $p < 0.01$), and (3) among group (L2), prosodic context (U), and stress pattern ($\beta = 1.405$, $SE = 0.571$, $z = 2.459$, $p < 0.05$). These results suggest that (1) the effect of group (HS) * prosodic context (U) is different for paroxytones compared to oxytones, (2) the effect of group (L2) * prosodic context (U) is different for paroxytones compared to oxytones, and (3) the effect of group (HS) * prosodic context (PN) is different for paroxytones compared to oxytones. For a better understanding of the complex nature of three-way interactions, pairwise comparisons of group, prosodic context, and stress pattern were conducted. Regarding (1), the results showed that the HSs had similar accuracy rates as the NSs regardless of the prosodic context and the stress patterns. However, the two groups differed in that, while the NSs had significantly lower accuracy rates in U than in N when the target word was an oxytone ($p < 0.01$), for the HSs, such a pattern was found when the target word was a paroxytone ($p < 0.05$). The effect of (2) can be explained in a similar way: higher accuracy rates were found in N than in U in L2s' perception of paroxytones ($p < 0.001$), while this was the case for NSs' perception of oxytones ($p < 0.01$). Moreover, for paroxytones, the NSs and the L2s did not differ in their accuracy when the target words were in N, while they did when the target words were in U ($p < 0.001$). For oxytones, on the other hand, the NSs had significantly higher accuracy rates than the L2s both in N and in U ($p < 0.001$ for both). Indeed, as seen in Figure 7, the accuracy difference between L2s' perception of the two stress patterns was significant in N ($p < 0.001$), while in U, it was not significant. Although not entered in the model, similar patterns were found between NSs' and L2s' accuracy rates of the two stress patterns in PN and in U. That is, higher accuracy rates were found in PN than in U in L2s' perception of paroxytones ($p < 0.05$), while this was the case for NSs' perception of oxytones ($p < 0.01$). Like the effect of (2), the NSs had higher accuracy rates than the L2s in PN (as in N and U) when the target word was an oxytone ($p < 0.001$) and the accuracy difference between L2s' perception of the two stress patterns was significant in PN (as in N) ($p < 0.001$). However, unlike (2), the NSs had significantly higher accurate rates than the L2s in PN when the target word was a paroxytone (as in U, but unlike in N) ($p < 0.001$). Another important difference that was not entered in the model, but highly relevant in this study, is the comparison between the HSs and the L2s. Pairwise comparisons showed that the HSs behaved similarly to the NSs in that, except for paroxytones in N, the HSs performed significantly better than the L2s in all conditions. Lastly, for (3), as mentioned above, the pairwise comparisons did not show any significant group difference (NS vs. HS) in N as opposed to PN, regardless of whether the target word was a paroxytone or an oxytone. However, based on the examination of Figure 7, it is possible that this effect may derive from the group difference being larger in PN compared to N, when the target word was a paroxytone, as the group difference seems minute in the two prosodic contexts when the target word was an oxytone.

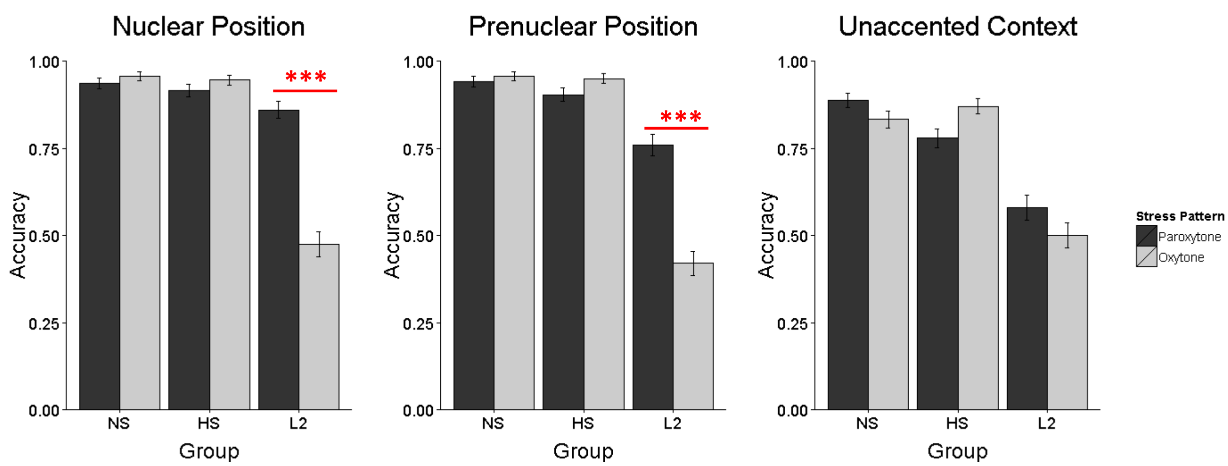


Figure 7 Participants' accuracy rates
(***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

To summarize, despite a slight difference from the NSs with regard to the stress pattern which showed a drop in accuracy from N to U (i.e., oxytones for the NSs and paroxytones for the HSs), the HSs behaved very similarly to the NSs in their perception of Spanish lexical stress; their performance was significantly worse when the target words were in U than when they were in N or in PN. Since there were less acoustic cues available that signaled lexical stress in U (i.e., longer duration) than in N (i.e., f0 peak, longer duration) and in PN (i.e., f0 valley with displaced f0 peak, longer duration), this finding shows that the NSs and the HSs attended to such acoustic cues when perceiving Spanish stress minimal pairs (e.g., *Canto*. 'I sing'. vs. *Cantó*. 'He/she/you (formal) sang'.), the distinction of which could otherwise be ambiguous. Moreover, like the NSs, the HSs performed significantly better than the L2s in their perception of oxytones in all three prosodic contexts (i.e, N, PN, and U) and in their perception of paroxytones in PN and in U. Also, unlike the L2s, who had higher accuracy rates for paroxytones than for oxytones in N and PN, the HSs and the NSs showed similar accuracy rates for the two stress patterns.

It is important to note that, in spite of the drop in accuracy rates from N and PN to U, NSs' and HSs' accuracy rates were well above chance level, which indicates that the two groups were able to successfully identify the location of lexical stress by attending to varying suprasegmental cues. On the other hand, the L2s' performance became significantly worse as the acoustic cues to lexical stress became less straightforward (N: f0 peak on the stressed syllable, PN: f0 valley on the stress syllable with f0 peak displaced to a following syllable, U: no tonal cues). Particularly, when perceiving oxytones, L2s' accuracy rates were extremely low compared to those of the NSs and the HSs. In fact, in N and in PN, they were below chance level (N: average 47.5%, PN: average 42%). As seen in Figure 7, this is strikingly different from the accuracy rates for paroxytones (N: average 86%, PN: average 76%), which were comparable to those of the NSs and the HSs. Since the participants were given two options (i.e., paroxytone or oxytone), this finding suggests that it is very likely that the L2s responded paroxytone far more frequently when the stimuli were oxytones than vice versa, showing preference toward paroxytones.

In order to test participants' sensitivity to distinguish the two stress patterns and their response bias toward paroxytones, their d-prime scores and response criterion (i.e., C scores) were calculated, respectively, based on their hit (HIT) and false alarm (FA) rates. HIT was considered as instances in which the participants selected a paroxytone when the stimulus was a paroxytone (i.e., correctly chose paroxytones), while FA was considered as instances in which the participants selected a paroxytone when the stimulus was an oxytone (i.e., incorrectly chose paroxytones). D-prime scores were calculated as the difference between the z-score of the HIT rates and the z-score of the FA rates (i.e., z(HIT) - z(FA)) and C scores were calculated as the sum of the z-score of the HIT rates and the z-score of the FA rates multiplied by -0.5 (i.e., -0.5*(z(HIT) + z(FA)). Figure 8 and 9 respectively show the d-prime scores and the C scores by group and prosodic context. As shown in Figure 8, the HSs clearly patterned with the NSs across the three prosodic contexts, while the L2s had noticeably low d-prime scores in all three prosodic contexts. A two-way mixed ANOVA with group and prosodic context as independent variables were conducted on participants' d-prime scores using the *aov* function in R (Baayen, 2008). Results showed that there was a main effect of group ($F(2, 195) = 133.05$, $p < 0.001$) and a main effect of prosodic context ($F(2, 195) = 21.89$, $p < 0.001$). No interaction between group and prosodic context was found. Tukey HSD post-hoc analysis was conducted for pairwise comparisons within group and prosodic contexts. Results confirmed that the L2s had significantly lower d-prime scores than both the NSs and the HSs in all three prosodic contexts ($p < 0.001$ for all). Moreover, while both the NSs and the HSs had significantly higher d-prime scores in N and in PN than in U (for NSs, $p < 0.05$ for both N vs. U and PN vs. U; for HSs, N vs. U: $p < 0.01$, PN vs. U: 0.001). With respect to the L2s, their d-prime scores did not differ based on the prosodic context. This indicates that the NSs and the HSs were successful in distinguishing the paroxytones from the oxytones in all three prosodic contexts, although their sensitivity reduced when there were less acoustic cues available in the speech signal (i.e., in U). The L2s, on the other hand, had difficulty distinguishing the two stress patterns, regardless of the prosodic context.

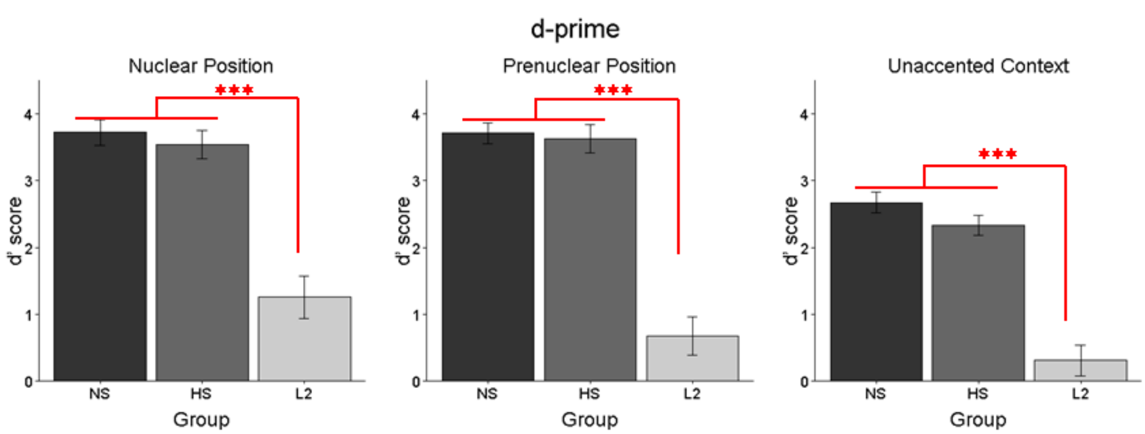


Figure 8 Participants' d-prime scores by group and prosodic context
(***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

As in the case of the d-prime scores, the HSs' C scores also patterned similarly with those of the NSs. Figure 9 shows participants' C scores by group and prosodic context. Results of a two-way mixed ANOVA with group and prosodic context as independent variables showed that there was a main effect of group ($F(2, 195) = 29.77$, $p < 0.001$), a marginally significant main effect of prosodic context ($F(2, 195) = 2.87$, $p = 0.059$), and a significant interaction between group and prosodic context ($F(4, 195) = 3.09$, $p < 0.05$). Tukey HSD post-hoc analysis revealed that the L2s had significantly lower C scores than both the NSs and the HSs in N ($p < 0.001$ for both) and in PN (NS vs. L2: $p < 0.01$, NS vs. HS: $p < 0.001$), while in U, the L2s' C scores did not differ from those of the NSs and the HSs. The results also showed that the L2s had significantly higher C scores in U than in N ($p < 0.05$). These findings suggest that the L2s had a clear bias toward responding paroxytones (as shown in the negative C scores) when the target words were in N or in PN, while the NSs and the HSs did not show such a bias (as shown in their C scores close to zero). When the target words were in U, the three groups did not differ in their degree of bias, as all of them had similar C scores, which supports L2s' significantly higher C scores in this prosodic context compared to N. As seen in Figure 9, although the HSs had a higher degree of positive C scores compared to the other two groups, which indicates a bias toward oxytones, these values did not differ from them at a significant level.

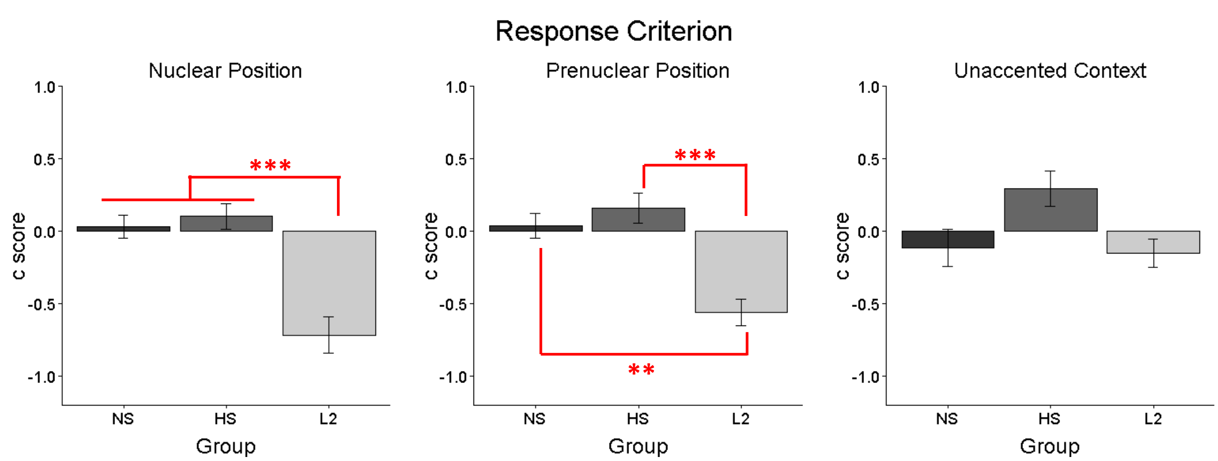


Figure 9 Participants' C scores by group and prosodic context
(***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

3.3.2.2 Response time

Among 3,331 correct responses of the target tokens, 169 were excluded due to short/lengthy response times, resulting in a total number of 3,162 tokens for the analysis. Figure 10 shows the participants' accuracy rates by group, prosodic context, and stress pattern. Results showed that there was a main effect of group for both HS and L2, which indicates that overall the NSs (i.e., the baseline group) responded faster than both the HSs ($\beta = 0.284$, $SE = 0.07$, $t = 4.08$) and the L2s ($\beta = 0.417$, $SE = 0.085$, $t = 4.92$). A main effect of stress pattern was also found ($\beta = 0.079$, $SE = 0.023$, $t = 3.45$), indicating that the participants were faster in responding when listening to paroxytones (i.e., the baseline stress pattern) than when listening to oxytones. With respect to prosodic context, main effects were found for both PN ($\beta = 0.391$, $SE = 0.024$, $t = 16.44$) and U ($\beta = 0.117$, $SE = 0.028$, $t = 4.15$). That is, the overall response time in N (i.e., the baseline prosodic context) was faster than that in PN and U. Recall that in N the target words were presented in isolation or at the end of the utterance after a clitic, while in PN and in U, they were at the beginning and in the middle of the utterance, respectively. Thus, it is reasonable that the response time was longer in the latter two contexts, because in these contexts the participants tended to withhold their responses until they finished listening to the entire sentence. Among various interactions, the interaction between L2 and PN was found to be significant ($\beta = -0.134$, $SE = 0.06$, $t = -2.21$), which indicates that the difference in response time between the NSs and the L2s was larger in N than in PN. Pairwise comparisons with group, stress pattern, and prosodic context showed that, despite the interaction, the NSs responded significantly faster than the L2s in all contexts, except for the oxytones in U. When comparing the response time between the HSs and the L2s, the HSs responded significantly faster than the L2s when the target words were in N; in PN, their response time differed only when the target word was an oxytone, and, in U, no significant difference was found between their response time. No significant difference was found in NSs' and HSs' response time. Moreover, although there was a general tendency that the response time for oxytones were longer than that for paroxytones, this difference was not statistically significant regardless of group or prosodic context, as seen in the large overlap between the two stress patterns across groups and prosodic contexts in Figure 10.

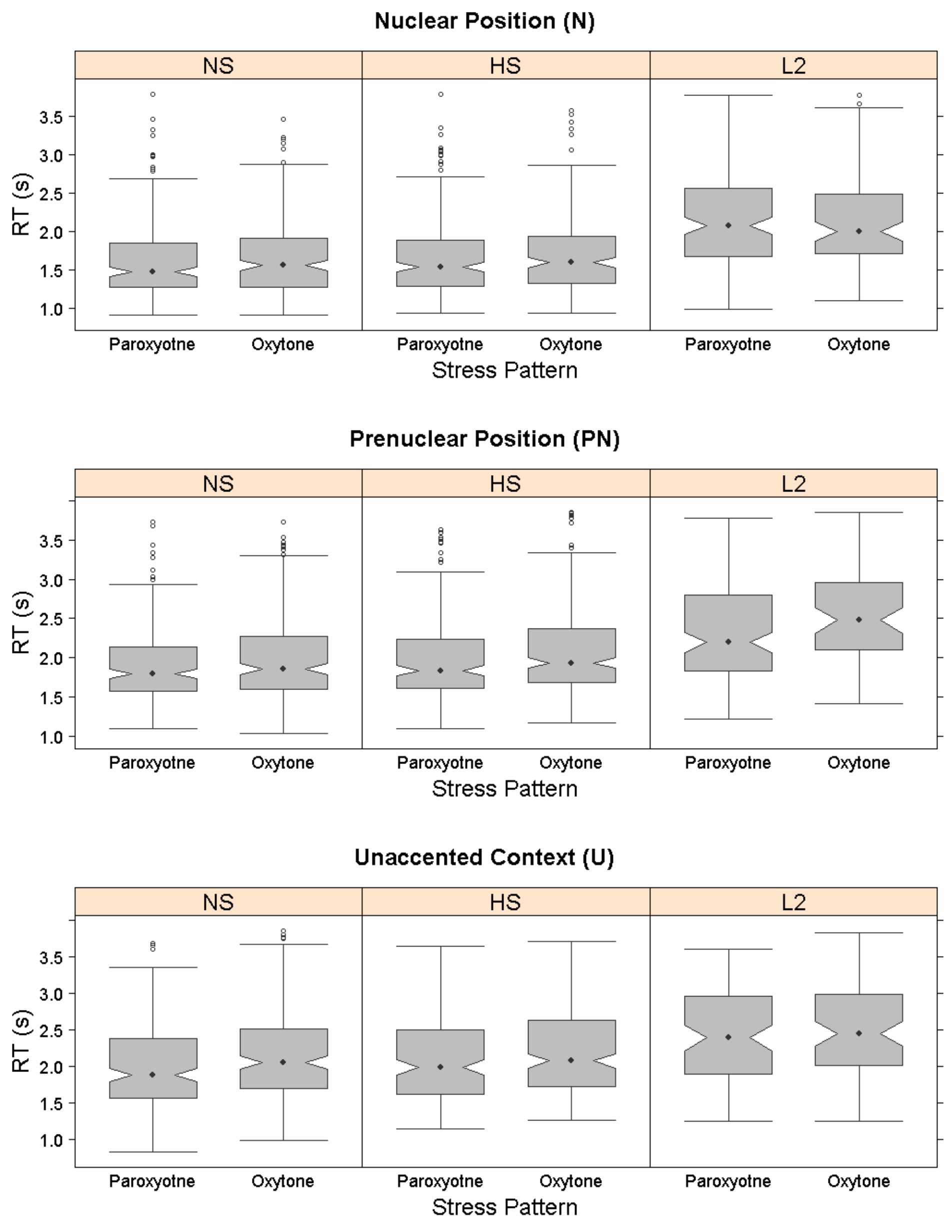


Figure 10 Participants' response rate by group (NS / HS / L2), prosodic context (N / PN / U) and stress pattern (paroxytone / oxytone)

3.4 Experiment 2: Production of lexical stress

3.4.1 Methods

3.4.1.1 Participants

The same speakers that participated in Experiment 1, participated in Experiment 2.

3.4.1.2 Materials

Similar to Experiment 1, stress minimal pairs of Spanish regular *-ar* verbs (i.e, first person singular of the present indicative form vs. third person singular of the preterit form) were used in Experiment 2. Among the 60 pairs used in Experiment 1, 20 pairs (i.e., 40 items) that contain consonants that are characterized to mark a clear acoustic boundaries in Spanish, such as voiceless obstruents (e.g., /p/, /k/, /t/, /s/), non-palatal nasals (e.g., /m/, /n/), and liquids (e.g., /l/, /r/) were chosen for the facilitation of segmentation. The target items were equally distributed with regard to the number of syllables (i.e., di- and tri-syllabic), vowel types in the verb stem (i.e., /a, e, i, o, u/), stress pattern (i.e., paroxytone and oxytone), and tense (i.e., present indicative and preterit). Apart from the target items, 20 pairs of Spanish verbs that do not form stress minimal pairs were included as filler items. The pairs consisted of a first person singular of the preterit tense (e.g., *Salí*. 'I left.') and the same verb in the third person singular of the present indicative form (e.g., *Sale*. 'He/She leaves. / You (formal) leave.'). The former case always had stress in the last syllable (i.e., oxytone) and the latter case always had it on the penultimate syllable (i.e., paroxytone). The filler items were included to confirm whether HSs' tendency to produce paroxytones as if they were oxytones found in Kim (2015) was due to a possible preference for past tense verbs over present tense verbs or due to final vowel lengthening. That is, if HSs prefer past tense verbs, then they would mistakenly produce filler items of third person singular of the present indicative form (e.g., *Sale*. 'He/She leaves. / You (formal) leave.') as first person singulars of the preterit tense (e.g., *Salí*. 'I left.'). The filler items matched in the number of syllables, vowel types, stress pattern, and tense with the target items. Same with Experiment 1, all the items were embedded in meaningful sentences in three prosodic contexts: nuclear position (N) (e.g., *La <u>saco</u>.* 'I take it out.'), prenuclear position (PN) (e.g., *<u>Saco</u> la basura.* 'I take out the trash.'), and unaccented context (U) (e.g., *¿De dónde saco caramelos?* 'From where do I take out candies?'), in order to examine whether the participants were able to adjust their use of different suprasegmental cues when they produced Spanish lexical stress in various prosodic contexts. The use of various prosodic contexts was also intended to investigate whether final vowel lengthening, which is considered to be one of the possible reasons for HSs' preference for oxytones (Kim, 2015), occurs even in contexts in which it is not expected (i.e., PN and U). In order to maintain the consistency of the carrier phrase in which the target stress pairs were located, the stimuli were presented without any subject. The list of stimuli used for Experiment 2 is presented in Appendix E.2.

3.4.1.3 Procedures

After completing Experiment 1, the participants continued with Experiment 2. The participants read out loud the stimuli, which were presented as PowerPoint slides on a Lenovo T430s laptop computer. All the items were presented in a randomized order and special care was taken to make sure that the minimal pairs of the same verb did not appear in a consecutive order. Given that the stimuli consisted of sentences of various prosodic contexts, subjects, and tenses, reading them in a random order may lead to confusion in the subject and the tense of the verb. To avoid this possible confound, prior to seeing each stimulus, the participants were presented with the subject (e.g., *yo* 'I') and the time (e.g., *pasado* 'past') of the sentence, in order to make sure that they understood the context of the sentence. Apart from the context, the participants were also informed that the sentences that they had to read were either declarative sentences or questions. The participants were instructed that they had to use different intonation patterns when reading these two sentences types: a falling contour for declarative sentences and a rising contour for questions. This was to avoid any confounding effect coming from various intonation patterns. A practice trial with five items, which were not the target items, was conducted for the familiarization with the task format. In both the U.S. and in Mexico, the productions were recorded in a sound-attenuated booth. In the U.S., the recordings were collected using an AKG C520 head-mounted microphone and a Marantz PMD570 solid state recorder with a sampling rate of 48 kHz and a sample size of 16 bits. In Mexico, the recordings were collected using an AKG C520 head-mounted microphone and a Zoom H4n handy portable digital recorder with a sampling rate of 44.1 kHz and a sample size of 16 bits. During the productions in both locations, the microphone was positioned approximately 2 inches away from the participants' lips. The sound files collected in the U.S. were resampled to 44.1 kHz to match with the ones collected in Mexico.

3.4.1.4 Coding and Analysis

Acoustic analyses were conducted on the penultimate (V1) and last vowels (V2) of the target items (e.g., /a/ and /o/ in *Canto* 'I sing'), as the stress minimal pairs used in the present study are distinguished by the relative prominence of these two vowels. Segmentation of the vowels was first performed using *EasyAlign* (Goldman, 2011), which is an automatic phonetic alignment tool developed as a plug-in of *Praat*. Later, the results were individually checked and manually corrected when needed. Both the formant structure in the spectrogram and the periodicity of the waveform were used for manual correction. That is, the beginning and end of a vowel were identified as the zero-crossing points of the regular periodic signal (in the waveform) closest to the onset and the offset of a continuous F2 (in the spectrogram) (Baker, 2006; Recasens, 1999). Suprasegmental information, such as the duration, average intensity, and average pitch, of the V1 and the V2 were extracted using scripts in *Praat* (Boersma & Weenink, 2015). The duration of the vowels were calculated as the time of the onset the vowel subtracted from the time of the offset of the vowel. Average pitch (Hz) was extracted using the *To Pitch* function in *Praat*, with a time step window of 0.01 second, pitch floor of 75 Hz, and pitch ceiling of 600 Hz, which are the default values. As for average intensity (dB), the values were extracted using the *Get Intensity (dB)* function. In order to control for individual differences, such as speech rate, gender (female / male), and test location (Mexico / U.S.), the raw duration, pitch, and intensity values were normalized using z-score normalization, which is calculated as the distance between the raw value and the mean value of each speaker divided by the standard deviation. Then, the difference between the normalized duration, pitch, and intensity of the V1 and the V2 were calculated to examine whether these values vary depending on the location of lexical stress. For instance, it is expected that the difference in duration, pitch, and intensity between /e/ and /o/ in *ceno* 'I eat dinner.' would be larger than between /e/ to /o/ in *cenó* 'He/She/You (formal) ate dinner.' With respect to PN, in particular, apart from the suprasegmental information of the vowels mentioned above, the degree of pitch peak (H) displacement was analyzed, which was defined as the distance from the onset of the stressed syllable to the location of H. The location of H was determined semi-automatically by selecting a region in which a peak was detected and extracting the point of the maximum pitch in that region. As the comparisons were done across different items, the H displacement values were normalized by dividing them into the duration of the stressed syllable. Although Spanish lexical stress is expressed mainly at the suprasegmental level, vowel quality was also examined to see whether unstressed vowels undergo vowel reduction, as in the case of English. F1 and F2 frequencies of the same vowel (i.e., V1) across minimal pairs (e.g., stressed /a/ in *paso* vs. unstressed /a/ in *pasó*) were extracted using the *To Formant (burg)* function with a time step window of 0.01, maximum number of 5 formants, maximum formant of 5,500 Hz for female speakers and 5,000 Hz for male speakers, window length of 0.025 second, and pre-emphasis from 50 Hz, which are the default values. The values were extracted at the middle of each vowel in order to avoid any possible transition effect from the adjacent segments. Similar to duration, pitch, and intensity, the raw formant values were normalized using Lobanov's (1971) z-score procedure to control for physiological differences resulting from individual differences, such as vocal tract length and gender. The idea is that, if vowel reduction occurs in unstressed position, the F1 and the F2 of unstressed vowels would have more centralized values than those of their stressed counterparts.

Tokens that were missing, produced before/after a pause, with creaky voice, devoicing, unexpected intonation pattern (i.e., falling contour for questions, rising contour for declarative sentences), or unclear articulation were excluded from the analyses. Additionally, for H alignment, tokens with H within a voiceless segment (e.g., within /t/ in *canto* 'I sing') and with flat pitch contour were excluded from the analysis. Moreover, with regard to stress pattern, comparisons between paroxytones and oxytones were done only with complete minimal pairs. That is, if a token was excluded due to any of the reasons above, the other token from the same minimal pair was also excluded from the analyses. The effects of group (NS / HS / L2), stress pattern (paroxytone / oxytone), prosodic context (N / PN / U), and the interactions among the fixed factors on the relative difference between the acoustic information of the stressed and those of the unstressed vowels mentioned above (i.e., normalized duration, pitch, intensity, and F1 and F2 frequency) were analyzed using linear mixed effects modeling with participants and items as random effects. The *lmer* function in the *lme4* package in R (Baayen, 2008) was used for the analyses. The best fitting models were selected through backward elimination, which will be explained in more detail below. All the fixed factors (i.e., group, prosodic context, and stress pattern) were centered using contrast-coding. With regard to pitch alignment in PN, the effects of group (NS / HS / L2), stress pattern (paroxytone / oxytone), and the interaction among the fixed factors on the normalized H displacement values were analyzed using linear mixed effects modeling with subject and item as random effects. Similarly, the best fitting model was selected through backward elimination. All the fixed factors (i.e., group and stress pattern) were centered using contrast-coding. Further pairwise analyses were conducted using the *lsmeans* function in the *lsmeans* package.

3.4.2 Results

Among the 8,160 target tokens (i.e., 40 items * 3 prosodic contexts * 68 participants), 1,809 tokens were excluded due to a pause before/after the token, creaky voice, devoicing, unexpected intonation pattern (i.e., falling contour for questions, rising contour for declarative sentences), and unclear articulation, and 745 tokens were excluded due to incomplete stress minimal pairs, leaving a total number of 5,606 tokens (i.e., 2,803 stress minimal pairs) to analyze.

3.4.2.1 Duration

The best fitting model for the difference between the normalized duration of the V1 and the V2 included random intercepts for subject and item with by-item random slope for group and by-subject random slope for stress pattern. Figure 11 shows the duration difference of paroxytones and oxytones produced by the three groups across prosodic contexts. Values higher than 0 (marked with red dotted lines) indicate that V1 was longer than V2 and values lower than 0 indicate that V1 was shorter than V2. Results showed that there was a main effect of group for both HSs ($\beta = -0.583$, $SE = 0.085$, $t = -6.834$) and L2s ($\beta = -0.258$, $SE = 0.099$, $t = -2.605$), indicating that duration difference of the NSs (i.e., the baseline group) was overall larger than those of the other two groups. There were also main effects of condition for PN ($\beta = 0.56$, $SE = 0.028$, $t = 20.253$) and U ($\beta = 0.14$, $SE = 0.027$, $t = 3.299$), which suggests that the duration difference of the items in N (i.e., the baseline prosodic context) was larger than that of the items in PN and in U. Moreover, significant interactions were found between group (HS) and prosodic context for both PN ($\beta = 0.212$, $SE = 0.057$, $t = 3.72$) and U ($\beta = 0.177$, $SE = 0.055$, $t = 3.242$) and between group (L2) and prosodic context for PN ($\beta = -0.462$, $SE = 0.087$, $t = -5.301$). That is, the difference between the NSs and the HSs was larger in N than in PN and in U, and the difference between the NSs and the L2s was larger in PN than in N. Regarding stress pattern, a main effect of stress pattern ($\beta = -1.273$, $SE = 0.15$, $t = -8.498$) and significant interactions between stress pattern and group (HS) ($\beta = 0.555$, $SE = 0.168$, $t = 3.299$) were found, suggesting that the duration difference of the paroxytones (i.e., the baseline stress pattern) was higher than that of the oxytones and the difference between the two stress patterns was larger for the NSs than for the HSs. Also, there were significant interactions between stress pattern and prosodic context for both PN ($\beta = 0.862$, $SE = 0.055$, $t = 15.569$) and U ($\beta = 0.378$, $SE = 0.053$, $t = 7.078$), which indicates that the difference between the two stress patterns was larger in N compared to that in PN and in U. Lastly, three-way interactions were found among group (HS), stress pattern, and prosodic context for both PN ($\beta = -0.904$, $SE = 0.113$, $t = -7.95$) and U ($\beta = -0.238$, $SE = 0.109$, $t = -2.18$), as well as among group (L2), stress pattern, and prosodic context (PN) ($\beta = -0.407$, $SE = 0.174$, $t = -2.335$). This suggests that the interaction between group (HS) and stress pattern was different in PN and in U, compared to N, and the interaction between group (L2) and stress pattern was different in PN, compared to N.

For a better understanding of the data, pairwise comparisons of group, stress pattern, and prosodic context were conducted. Results showed that all the groups had significantly higher duration difference for paroxytones than for oxytones regardless of the prosodic context, which indicates that all the three groups used duration to distinguish Spanish stress minimal pairs. However, among the three groups, only the NSs produced the stressed vowels consistently longer than the unstressed vowels, as seen in Figure 11, in which the majority of NSs' duration difference for paroxytones was above zero (i.e., V1 was longer than V2) and those for oxytones were below zero (i.e., V2 was longer than V1). The HSs and the L2s, on the other hand showed such a tendency only for the oxytones; with regard to the paroxytones, in 50% or more of the cases, the duration difference was below zero, which indicates that even when the lexical stress was in the penultimate vowel (V1), they produced the last vowel (V2) longer in many of the cases. Indeed, when comparing with the NSs, both the HSs and the L2s had a significantly lower duration difference when producing paroxytones, regarless of the prosodic contexts ($p < 0.001$

for all, except for NS vs. HS in PN and for NS vs. L2 U in which $p < 0.01$). With regard to the oxytones, while no difference was found among the three groups in N, in the other prosodic contexts, the L2s had a lower duration difference than the other two groups: in PN, L2s' duration difference was significantly lower than the NSs ($p < 0.01$) and, in U, it was significantly lower than both the NSs ($p < 0.01$) and the HSs ($p < 0.05$). This indicates that, when producing oxytones, the L2s produced the final vowel longer than the penultimate vowel to a larger degree than the NSs and the HSs.

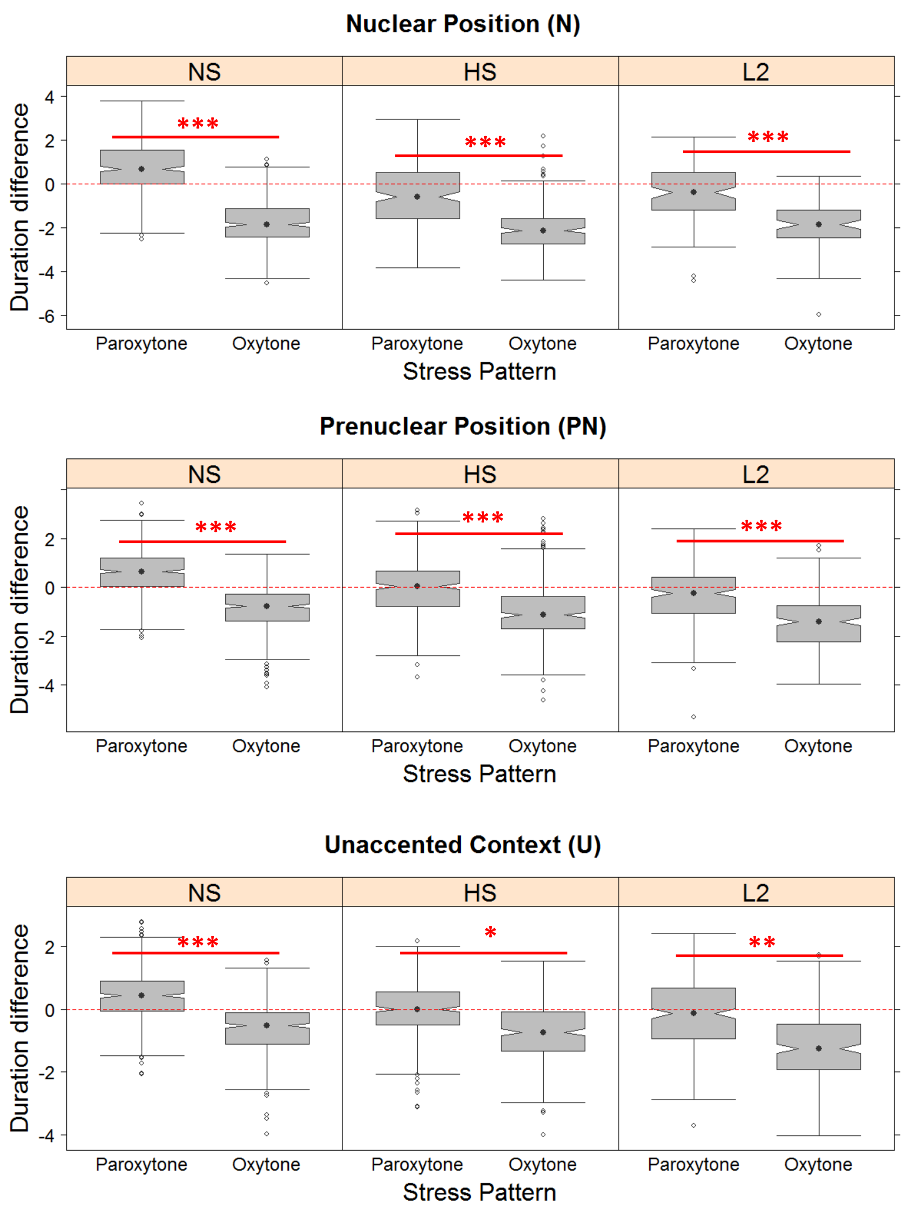


Figure 11 Duration difference between penultimate vowel and final vowel
(***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$)

3.4.2.2 Pitch

The best fitting model for the difference between the normalized pitch of the V1 and the V2 included random intercepts for subject and item with by-item random slope for group and by-subject random slope for prosodic context. Figure 12 shows the pitch difference of paroxytones and oxytones produced by the three groups across prosodic contexts. Values higher than 0 (marked with red dotted lines) indicate that V1 was produced with higher pitch than V2 and values lower than 0 indicate that V1 was produced with lower pitch than V2. Results showed that there was a main effect of condition for both PN (β = -0.837, SE = 0.07, t = -11.94) and U (β = 0.849, SE = 0.066, t = 12.88), which suggests that the pitch difference in N (i.e., the baseline prosodic context) was overall higher than that in PN, while compared to the pitch difference in U, it was lower. There was also a main effect of group (L2) (β = -0.298, SE = 0.093, t = -3.19) and a significant interaction between group (L2) and prosodic context (U) (β = -0.4, SE = 0.164, t = -2.44). That is, the NSs (i.e., the baseline group) in general had higher pitch difference than the L2s, and the difference between the two groups was larger in U than in N. With regard to stress pattern, although no main effect was found, there were significant interactions between stress pattern and group for both HSs (β = 0.174, SE = 0.042, t = 4.1) and L2s (β = -0.247, SE = 0.066, t = -3.73), and between stress pattern and prosodic context for both PN (β = 1.484, SE = 0.042, t = 34.89) and U (β = 0.431, SE = 0.042, t = 10.19). This indicates that the pitch difference of paroxytones (i.e., the baseline stress pattern) was higher than that of oxytones, and the difference between these two stress patterns was larger for the NSs than for the HSs and the L2s, and in N than in PN and in U. Lastly, there was a three-way interaction between group (HS), stress pattern, and prosodic context for both PN (β = -1.124, SE = 0.087, t = -12.91) and U (β= -0.215, SE = 0.087, t = -2.48), indicating that the interaction between group (HS) and stress pattern was different in N, compared to PN and U.

Pairwise comparisons with group, stress pattern, and prosodic context revealed that for all three groups the pitch difference was significantly higher in paroxytones than in oxytones only in N. However, as seen in Figure 12, only the NSs consistently produced the stressed vowels with higher pitch than the unstressed vowels, regardless of stress pattern (i.e., pitch difference higher than zero in paroxytones and pitch difference lower than zero in oxytones), while for the HSs and the L2s, this was the case only for paroxytones. That is, when producing oxytones, the HSs and the L2s tended to produce the unstressed vowels with higher pitch than the stressed vowels, as seen in Figure 12, in which the pitch difference values were higher than zero in 52.97% of the cases for the HSs and 36.36% of the cases for the L2s. For the NSs, these values were higher than zero in only 11.75% of the cases. Indeed, when comparing across groups, the NSs had significantly lower pitch difference in this stress pattern than the HSs ($p < 0.001$). Although the NSs also had lower pitch difference than the L2s, this did not reach significance. With regard to PN, no significant difference was found between the two stress patterns in any of the three groups. In fact, all three groups had pitch difference values lower than zero (i.e., V2 was produced with higher pitch than V1) in both paroxytones and oxytones. It is suspected that this is due to the displacement of H to a following syllable, which will be analyzed further below. Lastly, in U, the NSs and the HSs had significantly lower pitch difference in paroxytones than in oxytones, which is opposite from the trend shown in N. However, it is unlikely that these speakers used this opposite trend to signal lexical stress in U, because, in both stress patterns, there was a large degree of variation of the pitch difference values that centered on zero, and the values of the two stress patterns overlapped greatly. No significant difference was found between L2s' pitch difference of paroxytones and oxytones in this prosodic context.

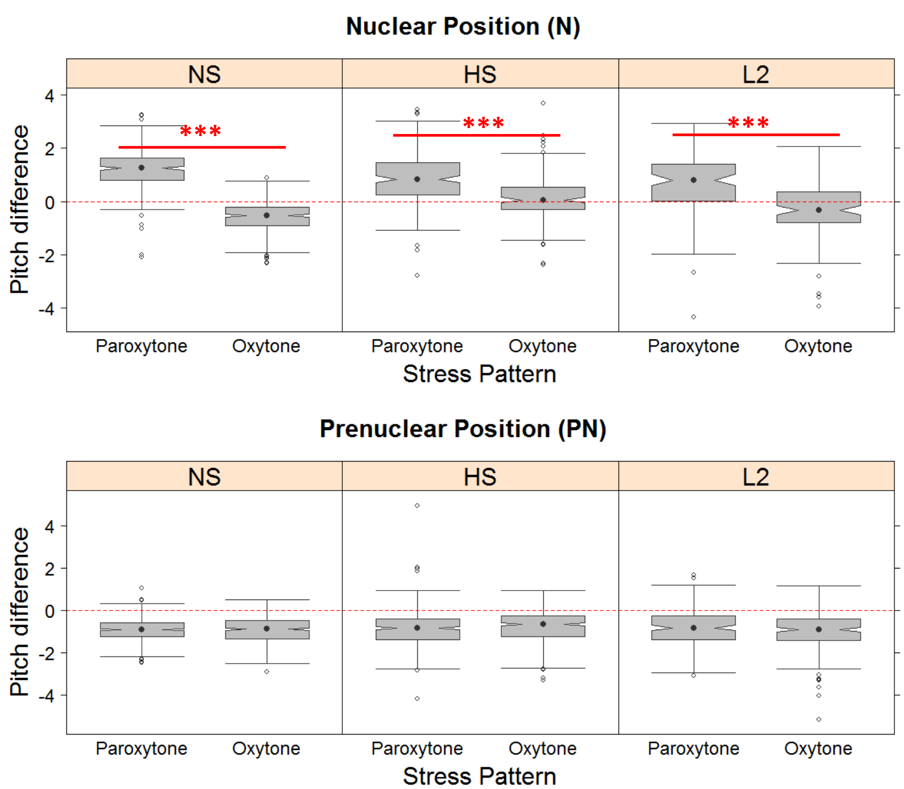


Figure 12 Pitch difference between penultimate vowel and final vowel
(***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$)

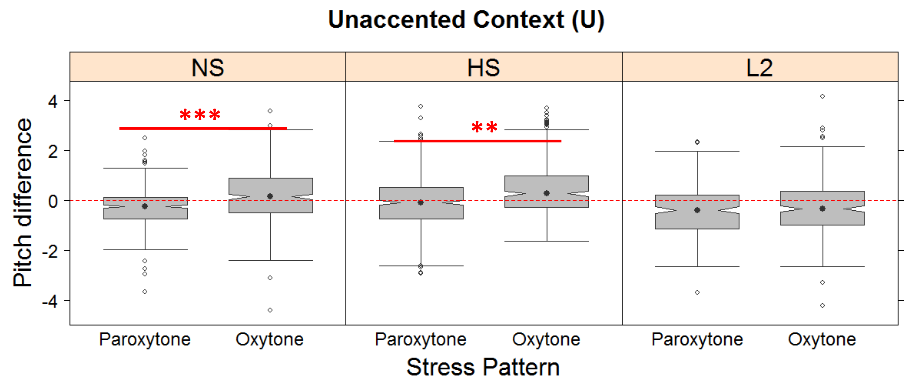


Figure 12 (cont.)

3.4.2.3 Intensity

The best fitting model for the difference between the normalized intensity of the V1 and the V2 included random intercepts for subject and item with by-item random slope for group and by-subject random slope for stress pattern. Figure 13 shows the intensity difference of paroxytones and oxytones produced by the three groups across prosodic contexts. Values higher than 0 (marked with red dotted lines) indicate that V1 was produced with higher intensity than V2 and values lower than 0 indicate that V1 was produced with lower intensity than V2. Results showed that there was a main effect of stress pattern (β = -0.634, SE = 0.109, t = -5.84), indicating that the intensity difference was overall higher for paroxytones (i.e., baseline stress pattern) than for oxytones. Main effects were also found in prosodic context for both PN (β = -0.909, SE = 0.024, t = -37.18) and U (β = 0.424, SE = 0.024, t = 18.02), and an interaction between stress pattern and prosodic context for both PN (β = 1.368, SE = 0.049, t = 28.1) and U (β = 0.324, SE = 0.047, t = 6.89). This suggests that the intensity difference in N (i.e., baseline prosodic context) was in general higher than that in PN, while it was lower than that in U. Moreover, the difference between the two stress patterns was larger in N than in PN and in U. Although no group effect was found, there were significant interactions between group (HS) and stress pattern (β = 0.307, SE = 0.122, t = 2.52) and between group (L2) and prosodic context (PN) (β = 0.185, SE = 0.077, t = 2.41). That is, while the three groups did not differ in their use of intensity, the intensity difference between two stress patterns was larger for the NSs (i.e., the baseline group) than for the HSs, and the difference between the NSs and the L2s was larger in N than in PN. As for three-way interactions, there was a significant interaction among group (HS), stress pattern, and prosodic context for both PN ($\beta = -0.907$, SE = 0.1, $t = -9.08$) and U ($\beta = -0.22$, SE = 0.096, $t = -2.29$), suggesting that the interaction between group (HS) and stress pattern was different in N, compared to that in PN and in U.

Pairwise comparisons with group, stress pattern, and prosodic context revealed that only in N the three groups had higher intensity difference for paroxytones than for oxytones. However, as seen in Figure 13, the stressed vowels were produced with higher intensity than the unstressed vowels only for paroxytones (i.e., intensity difference values higher than zero), while for oxytones, there were many cases, in which the unstressed vowels were produced with higher intensity. This was the case even for the NSs, who have shown clear distinctions between stress and unstressed vowels regarding duration and pitch for both stress patterns. In 48%, 72.6%, and 47.47% of the oxytones produced by the NSs, the HSs, and the L2s, respectively, the intensity difference values were higher than zero. Despite the similarity among the three groups, it seems that the NSs distinguished the two stress patterns more clearly, as the results showed that they had significantly a higher intensity difference than the HSs and the L2s ($p < 0.05$ for both) for paroxytones. In the case of oxytones, the NSs had significantly lower intensity difference values than the HSs ($p < 0.01$). With regard to PN, only the NSs showed a significant difference between the paroxytones and the oxytones. However, the large variation of the intensity difference values centered on zero in both stress patterns indicates that it is not likely that these speakers used intensity in a consistent manner to distinguish the two stress patterns. Lastly, in U, none of the three groups showed any significant difference between the two stress patterns. These findings are in line with those of pitch difference, in that the three groups used intensity as an acoustic correlate of lexical stress when the word was located in N, and the NSs used this cue to a larger degree than the HSs and the L2s, although not consistently when producing oxytones.

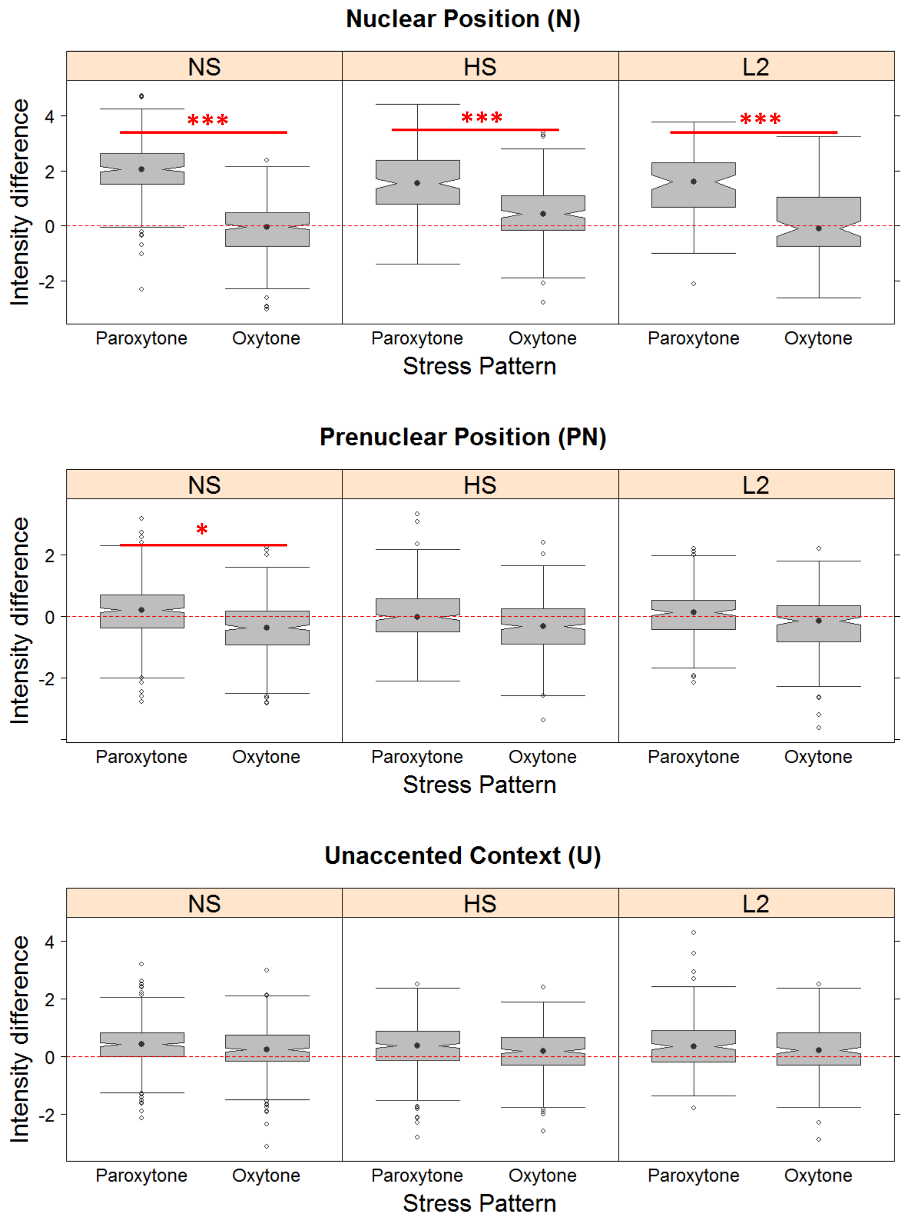


Figure 13 Intensity difference between penultimate vowel and final vowel
(***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$)

3.4.2.4 Pitch peak displacement (for PN)

Among the 2,720 target tokens in PN (i.e., 40 items * 68 participants), 685 tokens were excluded due to H within a voiceless segment and flat pitch contour, apart from the reasons mentioned above (i.e., pause before/after the token, creaky voice, devoicing, unexpected intonation pattern, and unclear articulation), and 383 tokens were excluded due to incomplete stress minimal pairs, leaving a total number of 1,652 tokens (i.e., 826 stress minimal pairs) to analyze. Figure 14 shows the normalized degree of H displacement for paroxytones and oxytones produced by the three groups of speakers, which was calculated as the distance from the onset of the stressed syllable to the location of H divided by the duration of the stressed syllable. The two red dotted lines indicate the onset and the offset of stressed syllables. Thus, if the value is within the two lines, it suggests that the H is aligned within the stressed syllable and if the value is outside of the two lines, it means that the H is aligned with a previous or a following syllable, depending on whether it is on the left or the right side of the region between the two lines.

The best fitting model for the normalized H displacement included random intercepts for subject and item with by-item random slope for group and by-subject random slope for stress pattern. Results showed that there was a main effect of group (L2) ($\beta = -0.275$, $SE = 0.08$, $t = -3.43$) and stress pattern ($\beta = -0.577$, $SE = 0.069$, $t = -8.34$). That is, overall the NSs (i.e., baseline group) produced with a larger degree of H displacement than the L2s and the H was displaced to a larger degree for paroxytones (i.e., the baseline stress pattern) than for oxytones. Moreover, a significant interaction between group (HS) and stress pattern ($\beta = -0.577$, $SE = 0.094$, $t = -6.14$) suggests that the difference in H displacement between the two stress patterns was larger for the HSs than for the NSs.

Pairwise comparisons with group and stress pattern revealed that, although H displacement occurred to a larger degree for paroxytones than for oxytones in all three groups, this difference was found to be statistically significant only for the HSs and the L2s. This can be seen in Figure 14, in which HSs' and L2s' H displacement values were further right for the paroxytones, compared to the oxytones. In the case of the HSs, the H for paroxytones was displaced to the point that their H displacement values were significantly higher than those of the NSs ($p < 0.01$). Also, it is interesting to note that the HSs and the L2s tended to produce the H within the stressed syllable more frequently than that the NSs, particularly in the case of oxytones. Indeed, in the case of the L2s, the H displacement values for oxytones were significantly lower than those of both the NSs ($p < 0.001$) and the HSs ($p < 0.05$). HSs' H displacement values were also lower than those of the NSs, but the difference did not reach significance ($p = 0.094$).

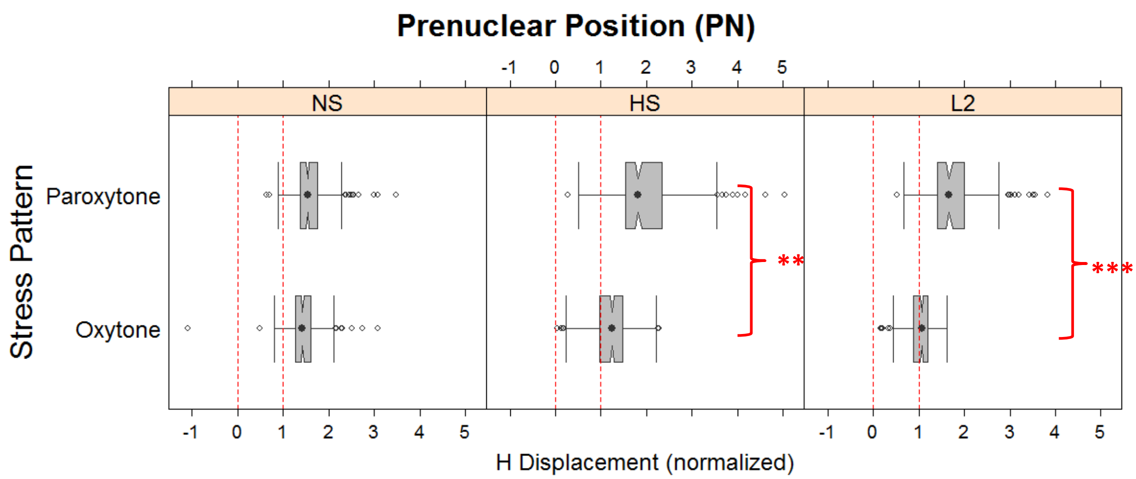


Figure 14 Normalized distance between stressed syllable onset and pitch peak (H) (***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$)

The frequency of cases in which H was displaced to a following syllable (i.e., displaced H) was further analyzed (Figure 15). Tokens with displaced H were coded as “1” and those with H aligned within the stressed syllable were coded as “0”. Logit mixed effects modeling was performed with group (NS / HS / L2) and stress pattern (paroxytone / oxytone) as fixed factors and subject and item as random effects. The best fitting model selected through backward elimination included random intercepts for subject and item with by-item random slope for group. All the fixed factors (i.e., group and stress pattern) were centered using contrast-coding. Results showed that there was a main effect of group for the HSs ($\beta = -2.747$, $SE = 0.691$, $z = -3.977$, $p < 0.001$) and the L2s, although the effect for the latter group was marginally significant ($\beta = -1.04$, $SE = 0.549$, $z = -1.896$, $p = 0.058$). Moreover, a main effect of stress pattern ($\beta = -2.581$, $SE = 0.385$, $z = -6.7$, $p < 0.001$) was found. That is, the overall rate of displaced H was higher for the NSs (i.e., the baseline group) than the HSs and the L2s, and higher for paroxytones (i.e., the baseline stress pattern) than for oxytones.

Pairwise comparisons with group and stress pattern revealed that, while, for the NSs, the rate of displaced H was similar for the two stress patterns, for the HSs and the L2s, it was significantly lower for oxytones than for paroxytones. Comparisons across the groups also confirmed that HSs’ and L2s’ rate of displaced H for oxytones was significantly lower than that of the NSs ($p < 0.001$ for both). Thus, while all the three groups displaced the H to a following syllable in PN, which explains why their final vowel was produced with higher pitch than the penultimate vowel in both stress patterns (see Section 3.4.2.2.), the HSs and the L2s tended to do it to a lesser degree when producing oxytones and instead aligned the H within the stressed syllable. The NSs, on the other hand displaced the H to a following syllable in the majority of the cases, regardless of the stress pattern.

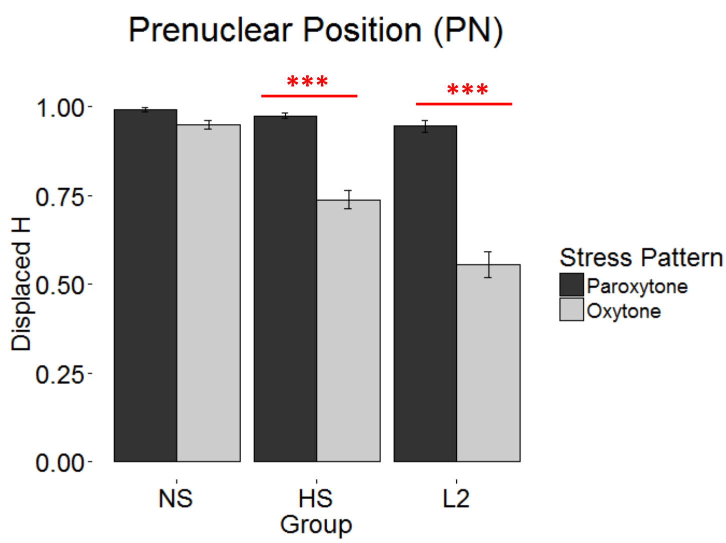


Figure 15 H Displacement rate for paroxytones and oxytones
(***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$)

3.4.2.5 Vowel Quality

Apart from the suprasegmental information presented above, participants' normalized F1 and F2 frequencies of stressed and unstressed vowels were also analyzed to examine whether stress had an effect on their vowel quality. The effects of group (NS / HS / L2), stress condition (stressed / unstressed), vowel type (a / e / i / o / u), and the interactions among the three fixed factors on the normalized F1 and F2 values of the vowels were examined for each prosodic context using linear mixed effects modeling with subject and item as random effects (i.e., 6 models). The best fitting models selected through backward elimination included random intercepts for subject and item with by-subject random slope for vowel for all measures, except for the normalized F1 in N, of which the model included random intercepts for subject and item with by-subject random slope for stress condition.

For F1, results showed that main effects were found for all vowel types, regardless of the prosodic context: /e/ (N: $\beta = -1.436$, SE = 0.063, $t = -22.71$; PN: $\beta = -1.427$, SE = 0.075, t = -

19.13; U: β = -1.233, SE = 0.081, t = -15.174), /i/ (N: β = -2.526, SE = 0.063, t = -40.02; PN: β = -2.481, SE = 0.06, t = -41.45; U: β = -2.272, SE = 0.072, t = -31.567), /o/ (N: β = -1.5, SE = 0.063, t = -23.68; PN: β = -1.444, SE = 0.075, t = -19.28; U: β = -1.27, SE = 0.086, t = -14.766), /u/ (N: β = -2.434, SE = 0.063, t = -38.67; PN: β = -2.403, SE = 0.063, t = -37.9; U: β = -2.31, SE = 0.076, t = -30.317). This indicates that overall /a/ (i.e., the baseline vowel type) was produced with higher F1 than the other vowels. There was also a main effect of stress condition (N: β = -0.369, SE = 0.093, t = -3.99; PN: β= -0.236, SE = 0.084, t = -2.8; U: β = -0.225, SE = 0.101, t = -2.223), suggesting that stressed vowels (i.e., baseline stress condition) were in general produced with higher F1 than the unstressed counterparts. In the case of N, there were significant interactions between stress condition and vowel type for /i/ (β = 0.36, SE = 0.126, t = 2.86) and for /u/ (β = 0.302, SE = 0.126, t = 2.4), which indicates that the difference between stressed and unstressed vowels was larger in /a/ than in /i/ and /u/. No significant interaction was found in the other two prosodic contexts. Although no effect of group was found in any of the three prosodic contexts, there were significant interactions between group and vowel type. In N, significant interactions were found between /i/ and both HSs (β = 0.264, SE = 0.075, t = 3.53) and L2s (β = 0.353, SE = 0.114, t = 3.1), and between /e/ and HSs (β = 0.21, SE = 0.075, t = 2.81). This indicates that the difference between /a/ and /i/ was larger for the NSs (i.e., the baseline group) than for the HSs and the L2s, and the difference between /a/ and /e/ was larger for the NSs (i.e., the baseline group) than for the HSs. Significant interactions were also found in PN between /i/ and both HSs (β = 0.162, SE = 0.057, t = 2.83) and L2s (β = 0.187, SE = 0.077, t = 2.42). Lastly, in U, there were significant interactions between /o/ and HSs (β = -0.265, SE = 0.119, t = -2.227) and between /u/ and L2s (β = -0.232, SE = 0.107, t = -2.171), and a marginally significant interaction between /e/ and L2s (β = -0.245, SE = 0.128, t = -1.912), indicating that the difference between /a/ and the mid vowels /e, o/ was larger for the NSs than for the L2s, and the difference between /a/ and /u/ was larger for the NSs than for the HSs. Particularly in PN, three-way interactions were found among group (HS), stress condition, and vowel type for both /i/ (β = 0.228, SE = 0.11, t = 2.06) and /o/ (β = 0.254, SE = 0.11, t = 2.3), and among group (L2), stress condition, and vowel type (/o/) (β = 0.389, SE = 0.149, t = 2.62). That is, the interaction between group (HS) and stress condition was different in /i/ and /o/, compared to /a/, and the interaction between group (L2) and stress condition was different in /o/, compared to /a/. Pairwise comparisons with group, stress condition, and vowel type were performed, but the results showed that there was no significant difference between stressed and unstressed vowels, regardless of the vowel type, in any of the three groups. This is clearly shown in Figure 16, in which there is a great degree of overlap between the distribution of the stressed vowels and their unstressed counterparts.

With regard to F2, results showed that main effects were found for all vowel types, regardless of the prosodic context: /e/ (N: $\beta = 0.917$, $SE = 0.107$, $t = 8.603$; PN: $\beta = 0.909$, $SE = 0.115$, $t = 7.881$; U: $\beta = 0.872$, $SE = 0.114$, $t = 7.676$), /i/ (N: $\beta = 1.293$, $SE = 0.12$, $t = 10.758$; PN: $\beta = 1.49$, $SE = 0.12$, $t = 12.424$; U: $\beta = 1.385$, $SE = 0.121$, $t = 11.401$), /o/ (N: $\beta = -0.586$, $SE = 0.106$, $t = -5.51$; PN: $\beta = -0.468$, $SE = 0.113$, $t = -4.147$; U: $\beta = -0.486$, $SE = 0.113$, $t = 04.286$), and /u/ (N: $\beta = -0.893$, $SE = 0.107$, $t = -8.316$; PN: $\beta = -0.762$, $SE = 0.114$, $t = -6.676$; U: $\beta = -0.817$, $SE = 0.114$, $t = -7.17$). This indicates that /a/ (i.e., baseline vowel type) was produced with lower F2 than /e/ and /i/, while it was produced with higher F2 than /o/ and /u/. There was also a main effect of group in N and in PN. In N, the group effect was found for the L2s ($\beta = -0.357$, $SE = 0.096$, $t = -3.711$), and in PN, the group effect was found for both the HSs ($\beta = -0.123$, $SE = 0.051$, $t = -2.416$) and the L2s ($\beta = -0.122$, $SE = 0.067$, $t = -1.829$). That is, in these two prosodic contexts the NSs (i.e., the baseline group) overall produced with higher F2 than the HSs and/or the L2s. In U, only a marginally significant effect was found for L2s ($\beta = -0.119$, $SE = 0.066$; $t = -1.792$). Apart from the main effects, significant interactions were found between vowel type and group. In N and PN, there were interactions between /u/ and both HSs (N: $\beta = 0.167$, $SE = 0.082$, $t = 2.04$; PN: $\beta = 0.293$, $SE = 0.073$, $t = 3.994$) and L2s (N: $\beta = 0.465$, $SE = 0.122$, $t = 3.808$; PN: $\beta = 0.28$, $SE = 0.097$, $t = 2.887$), suggesting that the difference between /a/ and /u/ was larger for the NSs than for the HSs and the L2s. In U, these interactions were marginally significant (HS: $\beta = 0.136$, $SE = 0.072$, $t = 1.884$; L2: $\beta = 0.16$, $SE = 0.092$, $t = 1.731$). Additionally, in this prosodic context, significant interactions were found between L2 and both /e/ ($\beta = 0.287$, $SE = 0.089$, $t = 3.243$) and /i/ ($\beta = 0.316$, $SE = 0.14$, $t = 2.25$). That is, the difference between /a/ and /e/ and between /a/ and /i/ was larger for the L2s than the NSs. Lastly, a three-way interaction was found among group (L2), stress condition, and vowel type (/i/) in N ($\beta = -0.559$, $SE = 0.214$, $t = -2.61$) and U ($\beta = 0.498$, $SE = 0.174$, $t = 2.867$), indicating that the interaction between group (L2) and stress condition was different in /i/, compared to /a/. However, as in the case of F1, pairwise comparisons with group, stress condition, and vowel type did not show any significant difference between stressed and unstressed vowels, regardless of the vowel type, in any of the three groups, as shown in the large overlap between the distribution of the stressed vowels and their unstressed counterparts in Figure 16.

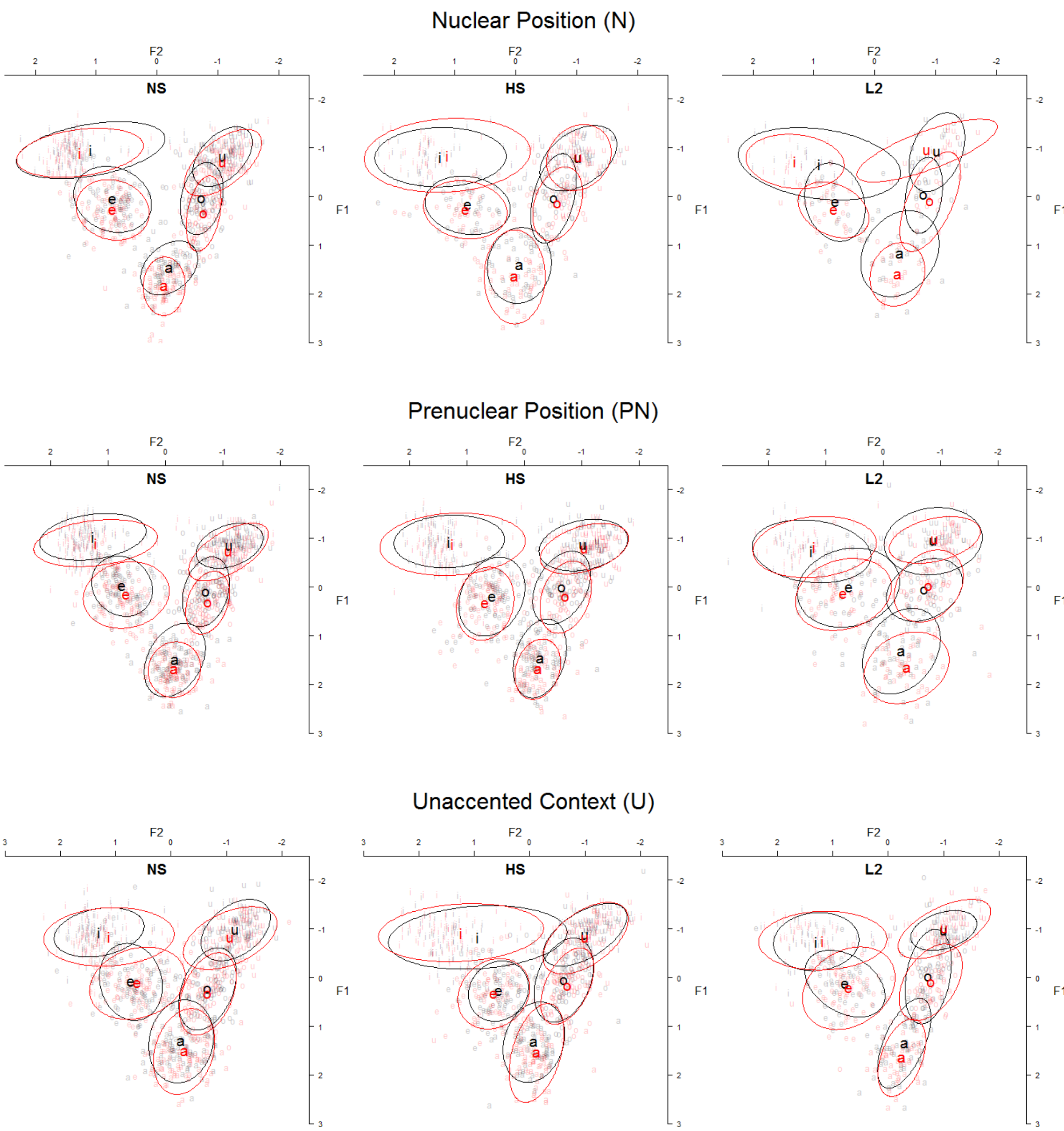


Figure 16 Distributions of normalized F1 and F2 frequencies of stressed (in red) and unstressed vowels (in black)

To summarize, even though a trend of vowel reduction was found in the three groups with regard to vowel height (F1), particularly for /a/ in N, when comparing the individual vowels in stressed and in unstressed conditions, no significant difference was found in their formant frequencies in any of the three groups. This finding implies that the three groups behaved similarly regarding vowel quality, in that they produced the stressed and unstressed vowels with similar formant frequencies, without reducing the unstressed vowels. Instead of vowel reduction, the three groups seem to differ in the distance between the vowels in the vowel space. For instance, as seen in Figure 17, the L2s produced /a/ further back (i.e., lower F2) than the NSs, while their /u/ was produced more fronted (i.e., higher F2), leading to a shorter horizontal distance between /a/ and /u/, compared to the one of the NSs. This may be due to the influence from English, which has a low back vowel /ɑ/ and in which /u/-fronting often occurs. Moreover, L2s' /i/ was produced lower than that of the NSs, leading to a shorter vertical distance between /a/ and /i/. As orthographic <i> in English is often pronounced as a lax vowel /ɪ/ (e.g., *bit*, *sit*, *pick*), L2s' lower /i/ suggests that these speakers may have produced this vowel similarly to English /ɪ/. The HSs also showed a similar pattern as the L2s (i.e., more fronted /u/, lower /i/ than the NSs), despite to a much smaller degree. Therefore, although the quality of each vowel may differ among the three groups, vowel quality *per se* does not seem to be a reliable cue to distinguish stress minimal pairs for these speakers.

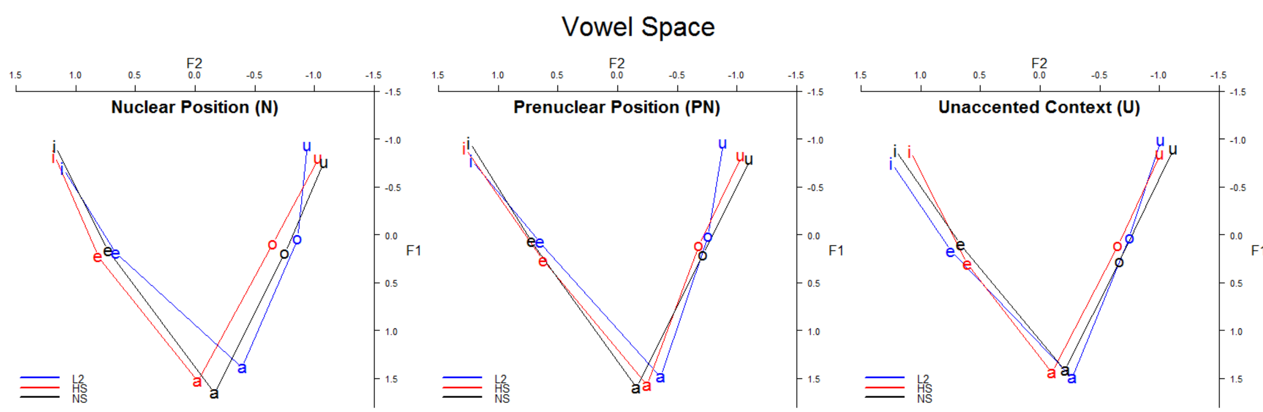


Figure 17 Vowel space by group (NS: black, HS: red, L2: blue)

3.5 Discussion

3.5.1 Perception of lexical stress

As seen in the findings of the perception task, both the NSs and the HSs were able to successfully distinguish Spanish stress minimal pairs in all three prosodic contexts, despite the varying stress correlates. Given that in Spanish different suprasegmental cues play a critical role in the identification of lexical stress, depending on the prosodic context in which the target word is located, this finding shows that both the NSs and the HSs were sensitive to these varying suprasegmental cues. The finding that the NSs and the HSs were slightly, but significantly less successful in distinguishing the Spanish stress minimal pairs in U than in the other two contexts, also supports their attention to suprasegmental information, because, among the three prosodic contexts tested in the present study, U was the context in which there were the least suprasegmental cues available (i.e., duration). This finding is consistent with Torreira et al. (2014), which found that, in unaccented contexts, stress minimal pairs are produced with overlapping acoustic information, which leads to confusion to Spanish native listeners. Indeed, when comparing with the results of the filler items, only in U, NSs' and HSs' accuracy rates of the target items (i.e., oxytones for the NSs and paroxytones for the HSs) were significantly lower than those of the filler items (NS: $p < 0.001$, HS: $p < 0.01$), while no significant difference was found between the accuracy rates of the target items and those of the filler items in other prosodic contexts. In the case of the NSs, the difficulty in identifying the stress location in U was also reflected in the response time data, which showed that the NSs' response time was significantly longer for the oxytones in this prosodic context than for the filler items ($p < 0.001$). Given that the subject of the filler items are easily identifiable without suprasemgnetal information, due to their verbal suffix (e.g., *-es* in *Aprendes*. 'You learn.') that disambiguates the word from other possible competitors, the lower accuracy rates found in the target items in U, compared to those of the filler items, confirms that for the NSs and the HSs suprasemgnetal information plays a critical role in the identification of stress location, when other contextual cues are not available.

The L2s, on the other hand, did not seem to successfully attend to the suprasegmental cues when perceiving Spanish lexical stress, because they were not only unable to correctly identify the location of stress, but also showed a clear bias toward paroxytones, even when the stimuli were oxytones. L2s' high accuracy rates for the filler items in all the three prosodic contexts (N: average 92.63%, PN: average 88.5%, U: average 79.25%) confirm that they have successfully acquired Spanish verb conjugation rules. Thus, it is unlikely that their non-target-like behavior in the perception of Spanish minimal pairs was due to difficulties in conjugating Spanish verbs. Moreover, the response time data showed that, except for paroxytones in U, the L2s spent a significantly longer time in responding when the stimuli were target items than when they were filler items. Regarding the NSs and the HSs, no significant difference was found between the response time for the target items and the filler items, except for the oxytones in U in the case of the NSs. This finding suggests that, for the L2s, morphological cues (i.e., the verbal suffix) are more important when identifying verb subject than the position of lexical stress, while, for the NSs and the HSs, both morphological and prosodic cues are important.

There are several possibilities that may explain L2s' bias toward paroxytones. Firstly, the paroxytones that were used in the present study were all verbs of present indicative form and the oxytones were all preterit tense verbs. Given that the present tense is the verbal form that is first learned in the classroom, it is possible that the L2s were having a bias toward the verbal form that they were more familiar with. However, this does not seem to be the definitive reason, as lower response accuracy in oxytones has also been found with nonce words (Romanelli & Menegotto, 2015). Moreover, when examining the errors of the filler items, although L2s' error rates were slightly higher for the past tense verbs (i.e., 172 out of 317 incorrect responses) than for present tense verbs (i.e., 145 out of 317 incorrect responses), this difference was not large enough to conclude that they had a bias toward present tense verbs. Another possible explanation to L2s' bias toward paroxytones is that paroxytone is the most common stress pattern in Spanish, which is approximately four times more frequent than oxytones (Quilis, 1993). Therefore, it is possible that the L2s picked up this pattern during the process of learning Spanish and applied it while completing the task, since the context was ambiguous. Also, as Romanelli and Menegotto (2015) suggested, the L2s may have over-simplified the Spanish stress rules, which generally assigns stress on the penultimate syllable if a word ends with a vowel, /n/, or /s/, and on the final syllable if a word ends with other consonants. Lord (2007) also argued that L2 learners are able to acquire Spanish stress knowledge, even though it is usually not taught in great detail in L2 classrooms. Therefore, since stress minimal pairs cannot be disambiguated based on segmental information, it is possible that the L2 learners in the present study had confusion when listening to these words and applied the general Spanish stress rules. However, it is premature to conclude that the L2s were not attending to suprasegmental information at all based on the present data, because such a bias should be observed equally in all three prosodic contexts, if suprasegmental information did not have any effect on their perception. However, in the present study, the bias was found only in N and in PN, whereas in U, the L2s had difficulty identifying the location of lexical stress equally for paroxytones, as well as for oxytones. Given that U is the context in which the least suprasemgnetal cues were available, it is possible that the L2s were attending to suprasegmental information to some extent, but were not successful in selectively paying attention to the critical cue, i.e., duration, in the absence of tonal cues.

3.5.2 Production of lexical stress

With regard to the production of lexical stress, the HSs, together with the L2s, seemed to use suprasegmental cues to a lesser extent than the NSs. In N, while all the groups used duration, pitch, and intensity to distinguish paroxytones from oxytones, only the NSs consistently used these cues to produce the stressed vowels with more prominence than the unstressed vowels, regardless of the stress pattern. Athough intensity was used less consistently when producing oxytones than duration and pitch, the NSs still made a clearer distinction between paroxytones and oxytones using this cue, compared to the other two groups. The HSs and the L2s showed deviation from the NSs, especially in their use of duration when producing paroxytones. That is, in more than half of the cases, the HSs and the L2s produced the final vowel longer than the penultimate vowel, even when the stress was located in the penultimate vowel. In PN, the three groups used duration and pitch alignment to distinguish the stress minimal pairs, while overall pitch and intensity did not play a significant role in the distinction. However, the HSs and the L2s behaved differently from the NSs in that, they displaced the H farther from the stressed syllable when producing paroxytones than when producing oxytones. Moreover, they aligned the H in the following syllable more frequently when producing paroxytones. The NSs, on the other hand, did not differ significantly in the degree of H displacement when producing the two stress patterns and aligned the H in the following syllable with similar frequency. With regard to duration, the HSs and the L2s produced the unstressed vowels of the oxytones longer than the stressed vowels in half of the cases. This tendency was also found in N. Lastly in U, the only cue that was used consistently across the three groups was duration. However, like in N and in PN, the HSs and the L2s produced almost half of their oxytones with longer unstressed syllables.

Thus, the HSs and the L2s showed a deviating pattern in all three prosodic contexts from the NSs, mainly in their use of duration cues. That is, they produced the final vowel (V2) longer than the penultimate vowel (V1), even when producing paroxytones, which created a large overlap in duration between the two stress patterns. It is unclear whether this behavior is due to influence from English, because if English had an effect on HSs' and L2s' use of duration when producing

Spanish lexical stress, the duration difference between the stressed and unstressed vowels should be larger than the NSs, as a result of vowel reduction (Delattre, 1966). However, the HSs and the L2s of the study did not reduce the unstressed vowels, but, rather, produced them longer than the stressed vowels in the case of the paroxytones. This finding is consistent with what I have found in a previous study (Kim, 2015), even though the present study considered various prosodic contexts. In Kim (2015), I suggested two possibilities to explain the longer unstressed vowels found in HSs' and L2s' paroxytones: preference of past tense verbs and final vowel lengthening. Firstly, it is possible that the HSs and the L2s incorrectly produced the paroxytones as oxytones, because they tend to prefer past tense verbs, which were all oxytones in the present study. This may be the case for the HSs, as Martin et al. (2013) found that, when conversing about topics beyond personal experience, intermediate-level Spanish heritage speakers tend to be more successful in talking about an event in the past than narrating a current event. Given that the HSs in the present study were mostly of intermediate-to-advanced-level (i.e., subjective language dominance score: 59.58%, objective language dominance score: 80.07%) and the stimuli of the present study were sentences that the HSs would not usually say in real life, it is possible that they unknowingly produced the past tense (i.e., oxytones) when reading the present tense, because this is the verb tense that they are more familiar with. Although special care was taken to avoid any literacy effects (i.e., presenting the subject of the sentence and the time in which the event occurred prior to each stimulus), it is possible that the HSs of the study were not familiar with reading texts in Spanish (Campbell & Rosenthal, 2000), which have promoted them to rely more on the past tense.

Preference of past tense may explain HSs' deviance from the NSs, but the same explanation does not apply to the L2s, because, usually in L2 classrooms, past tense verbs are taught after present tense verbs. Moreover, the perception results showed that the L2s had a bias toward paroxytones. Thus, if the L2s had a preference toward a particular tense, it should the present tense (i.e., paroxytones), not the past tense (i.e., oxytones). Rather, the second possibility (i.e., final vowel lengthening) seems to be a more plausible explanation to L2s' behavior. Recall that the L2s had lower duration difference (i.e., the last syllables were produced longer than the penultimate syllables to a larger degree) than the NSs and the HSs when producing oxytones, even when the target word was not at a phrase-boundary. That is, when the target word was at the end of the phrase (i.e., in N), the three groups produced the target words with a relatively low duration difference (see Figure 11), which suggests that they may have lengthened the final vowel in this prosodic context. However, when the target word was not located at a phrase-boundary (i.e., in PN and in U), the NSs and the HSs no longer seemed to show final vowel lengthening, as evidenced in the significantly higher duration difference in these contexts, compared to that in N. The L2s, on the other hand, consistently had a low duration difference across the prosodic contexts. This implies that the L2s lengthened the final vowels, even in the prosodic contexts, in which the target words were not located at a phrase-boundary. Although tokens with any signs of a phrasal-boundary in the acoustic signal (e.g., pause, glottalization) were excluded in the analyses, it is possible that the L2s, who were less dominant in Spanish (i.e., subjective language dominance score: 40%, objective language dominance score: 55.83%) than the NSs (i.e., subjective language dominance score: 90.04%, objective language dominance score: 98.33%) and the HSs (i.e., subjective language dominance score: 59.58%, objective language dominance score: 80.07%), positioned an “invisible” boundary (i.e., a phonological phrase boundary) after the target words, due to their lack of fluency in Spanish, resulting in final vowel lengthening.

A closer inspection of the errors in the production of the filler items corroborates the possibilities above. Fillers were considered incorrect if they were produced with incorrect tense (e.g., *Salí* ‘I left.’ instead of *Sale*. ‘He/She/You (formal) leave(s).’) or with incorrect position of lexical stress (e.g., **Sale instead of Sale*. ‘He/She/You (formal) leave(s).’). Table 2 shows the frequency of the incorrect productions of fillers by error type. As Table 2 demonstrates, the majority of the errors were related to producing oxytones instead of paroxytones (i.e., past tense verbs instead of present tense verbs or lexical stress of the paroxytones on the final syllable instead of the penultimate syllable). Out of the 136 of HSs’ errors (i.e., 6.46% of the entire fillers), 120 tokens were produced as oxytones instead of paroxytones (i.e. 88.24% of the errors), and out of the 126 of L2s’ errors (i.e., 9.9% of the entire fillers), 80 tokens were produced this way (i.e., 63.49% of the errors). Interestingly, while these errors were mainly due to the production of incorrect tense (i.e., past tense instead of present tense) for the HSs (58.82% of the errors), for the L2s they were mainly due to the production of incorrect stress location (i.e., stress on the final syllable for paroxytones instead of the penultimate syllable) (47.62% of the errors). This finding suggests that both the HSs and the L2s produced the paroxytones with longer unstressed vowels, but for different reasons: while the HSs produced these words as oxytones, due to their preference toward past tense verbs, the L2s produced them with lengthened final vowels, possibly as a result of positioning a phrase boundary after the words. However, this does not mean that verb tense had no effect on L2s' behavior or final vowel lengthening did not occur in HSs' speech, as there were also noticeable amount of cases, in which the L2s produced the present tense as past tense (15.87%) and the HSs incorrectly put the stress in the final vowel for paroxytones (29.41%).

Error type	Incorrect tense		Incorrect stress position		Total
	present → past	past → present	penultimate syll → final syll	final syll → penultimate syll	
HS	80 (58.82%)	8 (5.88%)	40 (29.41%)	8 (5.88%)	136
L2	20 (16.13%)	18 (14.52%)	60 (48.39%)	26 (20.97%)	124

Table 2 HSs' and L2s' incorrect production of fillers by error type

Therefore, based on the errors found in the filler data, HSs' and L2s' longer unstressed vowels of many paroxytones can be explained through a combination of preference toward past tense verbs and final vowel lengthening. This, however, does not seem to be the case for the NSs, as no error was found in their production of fillers. Stress misplacement in reading tasks has also been found in other studies, both in the speech of L2s (Adams, 1979; Knightly et al., 2003; Lord, 2007) and in the speech of HSs (Knightly et al., 2003; Robles-Puente, 2014). For instance, Robles-Puente (2014) also found cases, in which HSs in California produced verbs with incorrect stress position (e.g., **porfia<u>ban</u>* instead of *por<u>fia</u>ban* 'they strived') and verbs with incorrect tense or mood (e.g., *logra<u>rá</u>* (future tense, indicative mood) instead of *lo<u>gra</u>ra* (present tense, subjective mood)) when they were reading a passage in Spanish. HSs' and L2s' longer unstressed vowels of paroxytones may also explain why the H was displaced to a larger degree in this stress pattern, compared to the oxytones. If it is the case that the HSs and the L2s produced the paroxytones with longer unstressed vowels, because they were producing this stress pattern as if they were oxytones, it would be reasonable that the H was aligned farther toward the end of the word. Also, it is interesting to note that, when producing oxytones, the L2s aligned the H within the stressed syllable in 44.44% of the cases, instead of displacing it to a following syllable. This is likely to be due to influence from English, in which prenuclear pitch accents are generally marked as a H within the stressed syllable (i.e., L+H* or H*) (Estebas-Vilaplana, 2007; Lleó et al., 2004; Robles-Puente, 2014). The HSs also showed some cases of H within the stressed syllable when producing oxytones, which occurred significantly more frequently than when producing paroxytones. However, it is difficult to conclude that English had an influence in HSs' pitch alignment, to the same degree as the L2s, because the rate of H within the stressed syllable was relatively low (25.56%), compared to that of the L2s (44.44%).

Chapter 4 Utterance-level prominence: Nuclear stress

4.1 Literature review

Another type of prosodic prominence that was investigated in the present study is phrase- or utterance-level prominence, namely, nuclear stress. Nuclear stress conveys semantic and pragmatic information about a constituent in a discourse. It may indicate the information status (i.e., new vs. given information) of a constituent or whether the constituent receives contrastive or narrow focus. Focus is defined as the non-presupposed or new information in the utterance (Zubizarreta, 1998). As focused constituents convey new information in the discourse, they always contain a word that bears a nuclear stress, which is aligned with the nuclear pitch accent, i.e., the intonational nucleus of the utterance (Reinhart, 2006; Zubizarreta, 1998; Zubizarreta & Nava, 2011). Thus, the nuclear stress of an utterance plays a critical role in identifying the focused constituent (Zubizarreta, to appear). The sentences below taken from Breen et al. (2010) show some examples of nuclear stress with varying locations, depending on the focused constituent of the utterance:

(1a) Broad focus

What happened? – Damon fried an OMELET.

(1b) Narrow focus on subject

Who fried an omelet? – DAMON fried an omelet.

(1c) Narrow focus on verb

What did Damon do to an omelet? – Damon FRIED an omelet.

(1d) Narrow focus on object

What did Damon fry? – Damon fried an OMELET.

In broad focus contexts, nuclear stress is generally assigned to the rightmost prominent constituent of the utterance (Chomsky & Halle, 1968), although variability has been found in the location of nuclear stress, depending on several factors, such as the semantics of the verb (e.g., unaccusatives vs. unergatives) and the predictability and noteworthiness of the predicate-subject relation (Chafe, 1974; Lozano, 2006; Nava & Zubizarreta 2010; Sasse, 1987; Selkirk, 1984, 1995; Zubizarreta 1998, to appear).

4.1.1 Focus marking in Spanish and English

Although many languages are similar in that focused constituents contain a word that receives a nuclear stress (Zubizarreta, 1998; Zubizarreta & Nava, 2011; Reinhart, 2006), the way focus is marked varies across different languages. In the case of English, word order is relatively rigid, similar to other Germanic languages. Thus, focus and information status are usually expressed prosodically by stressing the focused constituent *in situ* (2a and 2b). With regard to Spanish, like many Romance languages, word order is very flexible, thus, prosody does not necessarily play a significant role in expressing focus and information status. Rather, this information is more likely to be expressed syntactically by placing the focused constituent utterance-finally (3a) (Cole, 2014; Contreras, 1976; Donati & Nespor, 2003; Steedman, 2014; Zubizarreta, 1998). Focus can also be realized prosodically in Spanish, but this is considered to be a marked form and it is used mainly to express contrastive focus (3b). Thus, Germanic languages are claimed to have a flexible nuclear stress, while Romance languages are claimed to have a rigid nuclear stress (Zubizarreta & Nava, 2011). Similarly, Vallduví (1990) used the terms "plastic languages" for Germanic languages and "nonplastic languages" for Romance languages to distinguish them.

(2a) Focus in English: Narrow focus on subject

Who ate the chocolate? - EDUARDO ate the chocolate.

(2b) Focus in English: Contrastive focus on subject

Did Julia ate the chocolate? - No, EDUARDO ate the chocolate.

(3a) Focus in Spanish: Narrow focus on subject

¿Quién comió el chocolate? - Comió el chocolate EDUARDO.

(3b) Focus in Spanish: Contrastive focus on subject

¿Comió el chocolate Julia? - No, EDUARDO comió el chocolate.

However, recently this argument has been challenged by several experimental studies that showed that Spanish native speakers in fact frequently stress the focused constituent *in situ*, as in English, rather than moving it to the utterance-final position (Gabriel, 2007, 2010; Gupton & Leal Méndez, 2013; Hoot, 2016, to appear). For instance, in an acceptability judgement task, Hoot (2016, to appear) found that, when there was a narrow focus on the subject, Spanish native listeners rated the S(ubject)V(erb)O(bject) word order more acceptable than the supposedly felicitous VOS word order. Similarly, in an elicited production task, Gupton and Leal Méndez (2013) found that Spanish native speakers maintained the canonical SVO word order in the majority of the cases, instead of moving the subject to the utterance-final position (i.e., VOS). The findings of these experimental studies call for a reevaluation of theories that argue a categorical use of word order to mark focus in Spanish.

If focus can be realized prosodically in both Spanish and English, what are the phonetic and phonological properties that speakers use to mark focus? There are several factors that contribute to the prosodic marking of focus, including the type and location of pitch accent, pitch range, duration, and intensity, which are shared across many languages (Burdin et al., 2015; Féry, 2013; Jun, 2005; Ladd, 2008). Focused constituents are generally produced with higher pitch, longer duration, higher intensity, and larger pitch range, compared to the non-focused constituents of the same utterance. While several aspects are shared cross-linguistically to express focus, there are also others that are encoded differently across different languages and dialects (Gussenhoven, 2002). With regard to English and Spanish, studies have shown that these two languages show language-specific patterns, particularly in pitch information, such as the type of pitch accent and pitch excursion. In English, lexical items in a non-final position of a declarative sentence usually bear a high pitch accent (H*). However, when these words are focused, they tend to be expressed with a rising pitch accent (L+H*) (Burdin et al., 2015). Moreover, non-focused words tend to be deaccented when they are located after a focused word (i.e., post-focal deaccenting), lending prosodic prominence to this word (Breen et al., 2010; Burdin et al., 2015; Cole, 2014; Ladd, 2008; Ito & Speer, 2006). In Spanish, lexical items are usually produced with a rising pitch movement that continues throughout the stressed syllable until the syllable(s) that follow(s) (L+>H*) when they are located in a non-final position in a declarative sentence. Rising pitch movement also occurs when words in this position are focused, but the rise usually ends within the stressed syllable (L+H*) (de la Mota, 1995, 1997; Face, 2000, 2001, 2002, Face & D'Imperio, 2005; Hualde, 1999; Nibert, 2000). With respect to pitch excursion, Harris et al. (2015) conducted a cross-linguistic study to compare monolingual speakers of Spanish and English in their use of pitch excursion in the production of new and given information. Results showed that English monolinguals demonstrated significantly more instances of pitch movement when expressing new information, compared to given information, while Spanish monolinguals did not differ in their use of pitch movement in the two contexts. This indicates that pitch excursion is not used as an important cue to mark focus in Spanish as it is in English.

4.1.2 Previous studies on Spanish-English bilingual speakers' perception and production of Spanish nuclear stress

With regard to focus marking in Spanish by Spanish-English bilingual speakers, research has been done mostly on bilinguals' use of syntactic cues, particularly word order (Gupton & Leal Méndez, 2013; Hertel, 2003; Hoot, 2016; Lozano, 2006). For instance, studies have shown that English L2 learners of Spanish are able to acquire SV/VS alternation in Spanish, but they tend to show optionality, accepting both SV and VS word order (Hertel, 2003; Lozano, 2006). That is, although it is possible that L2 learners acquire the target option over time, they are not able to completely expunge the dispreferred non-target option, if it does not necessarily lead to ungrammaticality, but rather to pragmatic anomaly (Montrul, 2005; Sorace, 2000). Optionality was also found with heritage speakers of Spanish. In a contextualized aural acceptability judgment task, Hoot (2016) found that Spanish monolingual speakers showed a clear preference toward the SVO word order (e.g., *Mi TÍO compró un carro*. 'My uncle bought a car.') when the context called for a narrow focus on subject (e.g., *¿Quién compró un carro?* 'Who bought a car?'), while disfavoring the VOS word order (e.g., *Compró un carro mi TÍO*.), even more than the mismatch condition (e.g., *Mi tío compró un CARRO*.). Although Spanish heritage speakers also showed a preference toward the SVO word order, unlike the monolingual speakers, they accepted the VOS word order with higher ratings than the mismatch condition. Such optionality occurred to a larger degree among heritage speakers of lower proficiency level than those of higher proficiency level, suggesting that, similar to L2 learners, heritage speakers show optionality in their use of syntactic cues when processing focus in Spanish.

Compared to bilinguals' use of syntactic cues, little research has been done on their use of phonetic and phonological properties when they mark focus. Gries and Miglio (2014) and Harris et al. (2015), for instance, compared Spanish heritage speakers with monolingual speakers of Spanish and English, and found that the three groups differed in their use of pitch excursion. That is, unlike Spanish monolinguals, who did not use pitch excursion to a large degree to mark focus, heritage speakers, as well as English monolinguals showed a clear distinction between new and given information by using this cue. This finding indicates that English prosody has an effect on heritage speakers' production of focus in Spanish (see also van Maastricht et al., 2015 for L1Spanish-L2Dutch and L1Dutch-L2Spanish bilingual speakers).

4.2 Research questions and predictions

The present study centers upon how focused constituents are identified and realized by heritage speakers of Spanish, and whether they show similar/different patterns from Spanish monolingual speakers and English L2 learners of Spanish. As found in several experimental studies (Gabriel, 2010; Gupton & Leal Méndez, 2013; Hoot, 2016, to appear), syntactic marking of focus using word order may not be as frequent in Spanish as it has been previously claimed. Rather, it is likely that prosody plays an important role in focus marking in Spanish. The results of several studies have suggested that, although both Spanish and English use prosody to express focus, the phonetic realization of focus may be different in the two languages (Gries & Miglio, 2014; Harris et al., 2015). Based on these contradictive findings, the present study intends to answer the following questions:

- Research question 1: Do heritage speakers and L2 learners attend to prosody or word order when they listen to sentences with conflicting cues?
    - Prediction 1: If they mainly process focus prosodically, they will identify the focused constituents as the lexical items that bear the nuclear stress.
    - Prediction 2: If they mainly process focus syntactically, they will identify the focused constituents as the lexical items that are located utterance-finally.

- Research question 2: Do heritage speakers and L2 learners use prosody or word order to distinguish focused and non-focused constituents in their production?
    - Prediction 1: If they mainly use prosody, they will place the nuclear stress on the focused constituents *in situ* and produce them with higher prosodic prominence compared to the non-focused constituents.
    - Prediction 2: If they mainly use word order, they will place the focused constituent at the end of the utterance.

4.3 Experiment 3: Perception of nuclear stress

4.3.1 Methods

4.3.1.1 Participants

The same subjects in Experiments 1 and 2 participated in Experiment 3.

4.3.1.2 Materials

Twenty minimal pairs (i.e., 40 sentences) with focus on the subject were used in Experiment 3. The pairs only differed in the position of the subject: pre-verbal position (e.g., <u>*LILIANA*</u> *lo preparó.* 'Liliana prepared it.') and post-verbal position (e.g., *Lo preparó* <u>*LILIANA*</u>. 'Liliana prepared it.'). Each sentence consisted of a subject, a transitive verb, and a direct object clitic. The subjects (S) were tri- or quadri-syllabic Spanish names and were all paroxytones (e.g., *LiLIAna*). The verbs (V) were di- or tri-syllabic transitive verbs in the third person singular of the preterit or (simple) past perfective tense (e.g., *preparó* 'he/she/you (formal) prepared') and were accompanied by a direct object clitic of third person singular *lo* 'it-masculine'. Transitive verbs were used, due to their large number of examples, compared to intransitive verbs. Moreover, unlike intransitive verbs, of which the position relative to the subject in neutral contexts (i.e., broad focus contexts) tends to differ based on the semantics of the verb (i.e., unaccusative vs. unergative)[11], transitive verbs are generally positioned after the subject in neutral contexts, resulting in a SVO word order. Clitic (Cl) *lo* 'it-masculine' was used as the direct object (O) instead of a full nominal phrase (e.g., *el pastel de chocolate* 'the chocolate cake'), because focused subjects are strongly inclined to be produced with prominence *in situ* in sentences with a full nominal phrase (i.e., $[_F$S]VO), while in sentences with clitics, they tend to be placed post-

---

[11] Preferences in the location of the nuclear stress have been attested on different verb types, specifically in two types of intransitive verbs, i.e., unaccusatives (e.g., *arrive, come, leave*) and unergatives (e.g., *cry, yell, sleep*) (Chafe, 1974; Nava & Zubizarreta 2010; Sasse, 1987; Selkirk, 1984, 1995; Zubizarreta 1998, to appear). It has been shown that the nuclear stress tends to fall on the subject with unaccusative verbs, while with unergative verbs, it varies depending on pragmatic factors such as predictability and noteworthiness of the predicate-subject relation (Nava & Zubizarreta, 2010; Zubizarreta & Nava, 2011). As Spanish and English are considered to differ in the nature of nuclear stress (i.e., Spanish: rigid nuclear stress, English: flexible nuclear stress) (Zubizarreta & Nava, 2011), in English sentences with these verbs are expressed by positioning the nuclear stress *in situ* (e.g., *MY FRIEND arrived*.), whereas in Spanish it is expressed by locating the stressed word to the end of the utterance (e.g., *Llegó MI AMIGO*.)

verbally (i.e., Cl+V[$_F$S]) (Gabriel, 2010). As the target sentences either had a canonical word order, but with focus in a non-final position (i.e., [$_F$S]Cl+V), which is possible in Spanish, but considered to be a marked form (Vallduví 1990), or had focus in the final position, which is considered to be the unmarked form, but with a non-canonical word order (i.e., Cl+V[$_F$S]), it would be possible to examine how participants attend to word order and prominence when they process focus. That is, if word order plays a larger role than prominence, participants would perceive the Cl+V[$_F$S] structure as focus on the subject and the [$_F$S]Cl+V structure as focus on the verb. If prosody plays a larger role than word order, participants would perceive both sentence structures as focus on the subject. Apart from the target sentences above, 20 pairs of filler sentences (i.e., 40 sentences) with focus on the verb were used. The filler sentences had the same structure as the target sentences (i.e., subject, transitive verb, and direct object clitic) and, as in the case of the target sentences, the only difference between the pairs was whether the subjects were located pre-verbally and post-verbally (e.g., *Emilio lo <u>TIRÓ</u>.* vs. *Lo <u>TIRÓ</u> Emilio.* 'Emilio threw it.'). Thus, the filler sentences either had a canonical word order with focus on the final position (i.e., S+Cl[$_F$V]) or had a non-canonical word order with focus on a non-final position (i.e., Cl[$_F$V]S). The filler sentences were included in order to examine whether different patterns are found depending on the type of the focused constituent (Gupton & Leal Méndez, 2013; Hoot, 2016, to appear), as well as to confirm that participants attend to the cues of interest (i.e., prominence and word order). That is, if participants attend to either of these cues, they would show a clear categorization of the S+Cl[$_F$V] structure as focus on the verb, while showing a confusion when listening to the Cl[$_F$V]S structure. Filler items of other sentence structures were not included, because, as Ito and Speer (2006) pointed out, people "fail to disambiguate prosodically when other linguistic information is available to express linguistic meaning".

The sentences were produced by a male native speaker of Mexican Spanish. The speaker was instructed to listen to prompt questions, which were produced by another male native speaker of Mexican Spanish, and read out loud the sentences as if he was responding to the questions. Different prompt questions were provided to elicit natural productions of the two sentence structures. In order to elicit target sentences with the S+Cl+V order (e.g., *LILIANA lo preparó.* 'LILIANA prepared it.'), questions with contrastive focus on the subject (CF_S) (e.g., *¿Lo preparó Leonardo?* 'Did Leonardo prepare it?') were asked and, to elicit target sentences with the Cl+V+S order (e.g., *Lo preparó LILIANA.* 'LILIANA prepared it.'), questions with narrow focus on the subject as new information (NF_S) (e.g., *¿Quién lo preparó?* 'Who prepared it?') were asked. Regarding the filler sentences, to elicit sentences with the S+Cl+V order (e.g., *Emilio lo TIRÓ.* 'Emilio THREW it.'), questions with narrow focus on the verb as new information (NF_V) (e.g., *¿Qué hizo Emilio?* 'What did Emilio do?') were asked and, to elicit sentences with the Cl+V+S order (e.g., *Lo TIRÓ Emilio.* 'Emilio THREW it.'), questions with contrastive focus on the verb (CF_V) (e.g., *¿Lo sacó Emilio?* 'Did Emilio take it out?') were asked. For the sentences with nuclear stress on the non-final position (i.e., $[_F S]$Cl+V and Cl$[_F V]$S structures), a negative particle *no* 'no' was included in the beginning of the sentence to provide a natural context in which the speaker places focus on a non-canonical position (e.g., *No, LILIANA lo preparó.* 'No, LILIANA prepared it.'). The negative particles were later removed, leaving only the subject, the verb, and the clitic. The recording of the sentences took place in a sound-attenuated booth at the University of Illinois Phonetics and Phonology Lab using an AKG C520 head-mounted microphone, which was positioned approximately 2 inches away from the participants' lips, and a Marantz PMD570 solid state recorder with a sampling rate of 48 kHz and a sample size of 16 bits. The items were divided into two lists, each consisting of 20 target sentences and 20 filler sentences, as a measure to avoid priming effect, so that the minimal pairs of the same subject and verb did not appear in the same list. The list of the stimuli and the acoustic information of them are presented in Appendix E.3 and Appendix F.3.

4.3.1.3 Procedures

After completion of Experiments 1 and 2, the participant took a 10 minute break and continued with Experiment 3. Each participant was assigned to one of the two lists. Similar to Swerts et al. (2002), a forced-choice identification task was conducted using *PsychoPy2*, in which the participants listened to Spanish sentences and reconstructed the preceding utterance based on the sentences that they heard. The four question types presented above (i.e., NF_S, NF_V, CF_S, CF_V) were shown on the four quadrants of the computer screen and the participants selected the question that they thought that could have been the question of the aural stimuli by clicking on it using a mouse (Figure 18).

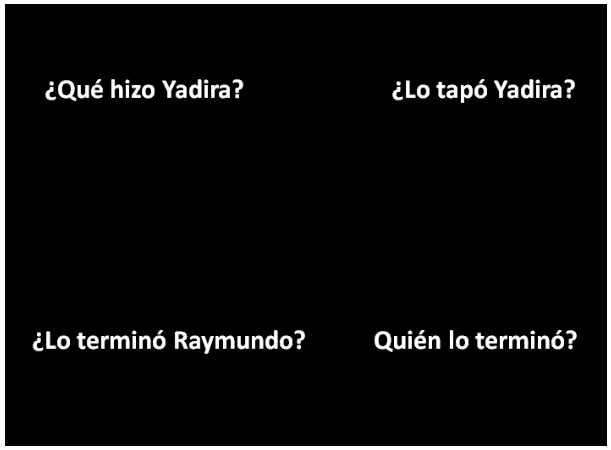


Figure 18 Example of the options shown to the participants

In the case that there were more than one question that matched well with the stimuli, the participants were instructed to choose the best option. The participants were also instructed that they spend enough time to read all four questions. The sentences were presented in a randomized order and the order of the options was counter-balanced. In both testing sites (i.e., Mexico and U.S.), the experiment was conducted in a phonetics laboratory equipped with a sound-attenuated booth. The stimuli were presented through a Lenovo T430s laptop computer with Sennheiser HD 558 headphones and a Logitech LS1 Laser Mouse. Before the initiation of the main task, a practice test with five items, which were not the target items, was conducted for the familiarization with the task format.

4.3.1.4 Coding and analysis

The location of the mouse clicks on the four quadrants of the computer screen was automatically collected through *PsychoPy2*. Responses with a negative value on the x-axis and a positive value on the y-axis were coded as the questions in the first quadrant, responses with positive values on both x-axis and y-axis were coded as the questions in the second quadrant, responses with negative values on both x-axis and y-axis were coded as the questions in the third quadrant, and responses with a positive value on the x-axis and a negative value on the y-axis were coded as the questions in the fourth quadrant.

Participants' response rate of the target stimuli was analyzed, which was measured as the rate of instances in which the participants selected the question types eliciting focus on the subject (i.e., NF_S and CF_S). If participants' responses were either NF_S or CF_S, they were coded as "1" and, if they were either NF_V or CF_V, they were coded as "0". If word order plays a critical role in participants' perception of focus, they would respond NF_S or CF_S for the Cl+V[$_F$S] structure, leading to high response rates for this word order, and respond NF_V or CF_V for the [$_F$S]Cl+V structure, leading to low response rates for this word order. On the other hand, if prominence plays a larger role, they would respond NF_S or CF_S for both Cl+V[$_F$S] and [$_F$S]Cl+V structures, leading to high response rates for both word orders. Similarly, for the filler sentences, participants' response rate of questions eliciting focus on the verb (i.e., NF_V and CF_V) was analyzed. If participants' responses were either NF_V or CF_V, they were coded as "1" and, if they were either NF_S or CF_S, they were coded as "0". If word order plays a critical role in participants' perception of focus, they would respond NF_S or CF_S for the Cl[$_F$V]S structure, leading to low response rates for this word order, and respond NF_V or CF_V for the S+Cl[$_F$V] structure, leading to high response rates for this word order. However, if prominence plays a larger role, they would respond NF_V or CF_V for both S+Cl[$_F$V] and Cl[$_F$V]S structures, leading to high response rates for both word orders.

The effects of group (NS / HS / L2), word order (S+Cl+V / Cl+V+S), and the interaction between the two fixed factors on participants' response rate were analyzed separately for the target sentences and the filler sentences, using logit mixed effects modeling with subject and item as random factors. For both models, *glmer* function in the *lme4* package in R (Baayen, 2008) was used. The best fitting models selected through backward elimination included random intercepts for subject and item with by-subject random slope for word order. All the fixed factors (i.e., group and word order) were centered using contrast-coding. Further pairwise analyses were conducted using the *lsmeans* function in the *lsmeans* package.

4.3.2 Results

Figure 19 shows participants' response rates by group and word order. As seen in Figure 19, regardless of the word order, participants' response rate of focus on the subject (response rate[S]) was higher than chance-level for the target sentences (i.e., focus on the subject), while their response rate of focus on the verb (response rate[V]) was lower than chance-level for the filler sentences (i.e., focus on the verb). With regard to the target sentences, results showed that there was a main effect of group for L2 ($\beta = -1.576$, $SE = 0.497$, $z = -3.172$, $p < 0.01$), suggesting that overall the response rate[S] was higher for the NSs (i.e., the baseline group) than the L2s. Pairwise comparisons of group and word order revealed that this is mainly due to L2s' lower response rate[S] when listening to the Cl+V[$_F$S] structure, compared to when listening to the [$_F$S]Cl+V structure. Although the difference between the two word orders was only marginally significant ($p = 0.079$), L2s' response rate[S] for the Cl+V[$_F$S] structure was significantly lower than that of the NSs ($p < 0.05$), as well as that of the HSs ($p < 0.01$), while the response rate[S] for the [$_F$S]Cl+V structure did not differ among the three groups.

Based on the high response rate[S] for both word orders, one can suggest that, regardless of language background, listeners tend to pay more attention to acoustic prominence than to word order when they process focus in Spanish. This is contrary to what has been predicted for the NSs, who were expected to attend to the word order cue more than the L2s. In order to confirm that prominence plays a critical role in the perception of Spanish focus, analyses on the filler sentences, which had focus on the verb, were carried out. As mentioned earlier, if participants mainly attend to the word order cue, they would show a high response rate[V] for the S+Cl[$_F$V] structure and a low response rate[V] for the Cl[$_F$V]S structure; if prominence plays a larger role, their response rate[V] would be high for both word orders. Results showed that there were main effects of group for both the HSs ($\beta = 0.817$, $SE = 0.396$, $z = 2.066$, $p < 0.05$) and the L2s ($\beta = 1.936$, $SE = 0.465$, $z = 4.164$, $p < 0.001$), which suggests that overall the response rate[V] of the NSs (i.e., the baseline group) was lower than that of the HSs and the L2s. Pairwise comparisons of group and condition showed that L2s' response rate[V] for the Cl[$_F$V]S structure was significantly higher than the other two groups ($p < 0.01$ for both). Similar trend was found between L2s' response rate[V] for the S+Cl[$_F$V] structure and that of the NSs and the HSs, although significant difference was found only between the L2s and the HSs ($p < 0.05$). The results of the filler sentences reject the possibility that prominence had an important role in participants' perception of focus, because the response rate[V] was found to be very low across the two word orders.

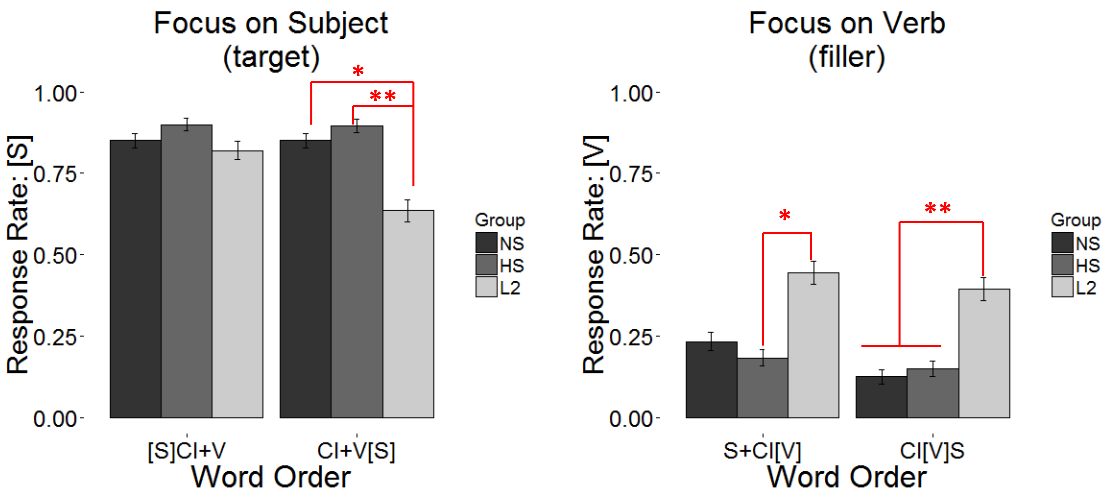


Figure 19 Participants' response rate by group and word order
(***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

The findings of the filler sentences suggest that none of the two cues (i.e., word order and prominence) played a critical role in participants' perception of focus in Spanish. Thus, participants' response rates of the four question types were analyzed in order to further examine their response pattern. As shown in Figure 20, the participants selected the questions eliciting narrow focus on the subject (NF_S) (e.g., ¿*Quién lo preparó*? 'Who prepared it?') with higher frequency than the other question types, regardless of the word order and the position of the focused constituent. However, compared to the other two groups, the L2s had higher response rates of NF_V (i.e., narrow focus on the verb). As seen in Figure 20, when listening to the Cl+V[$_F$S] structure, the L2s identified it as NF_V on average 34.5% of the time, while the NSs and the HSs did so on average 15% and 10.42% of the time, respectively. Regarding the S+Cl[$_F$V] structure, the L2s responded NF_V on average 43.5% of the time, while the NSs and the HSs responded NF_V on average 23.33% and 17.92% of the time, respectively. Similarly, for the Cl[$_F$V]S structure, L2s' average response rate of NF_V (39.5%) was more than twice as high as that of the NSs (12.5%) and the HSs (15%). Only for the [$_F$S]Cl+V structure was L2s' average response rate of NF_V (17.5%) was comparable to that of the NSs (15%) and the HSs (10%). It is also interesting to note that there were few cases in which the participants selected questions that elicited contrastive focus (i.e., CF_S and CF_V), the rates of which were expected to be higher in sentences with focus in a non-final position (i.e., the [$_F$S]Cl+V structure and the Cl[$_F$V]S structure). However, the results showed that the response rates for these question types were only 5% or less. Thus, L2s' higher NF_V rates, compared to the other two groups, explains their relatively low response rate[S] observed in the Cl+V[$_F$S] structure and their relatively high response rate[V] observed in S+Cl[$_F$V] and Cl[$_F$V]S structures.

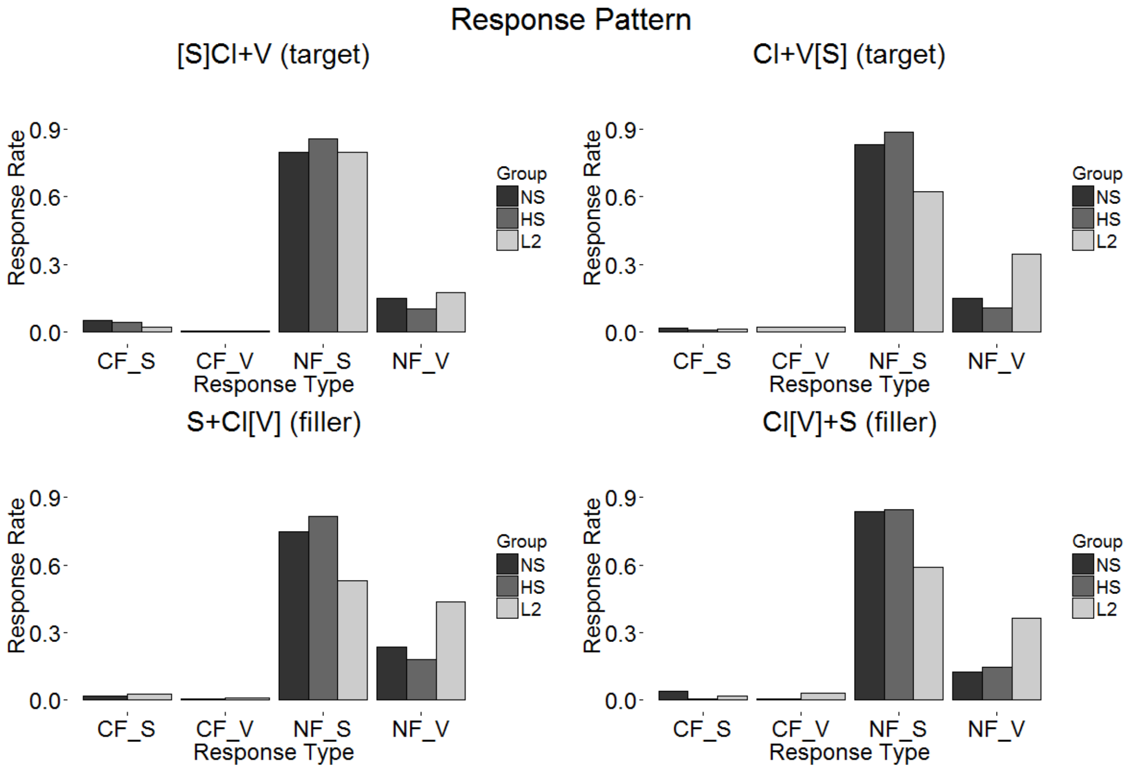


Figure 20 Paticipants' response pattern by response type

To summarize, the participants seemed to prefer choosing preceding contexts that elicit focus on the subject (i.e., NF_S and CF_S), regardless of the focused constituent and word order of the sentence. Such a tendency was shown to a lesser degree for the L2s, whose response rate for preceding contexts eliciting focus on the verb (i.e., NF_V and CF_V) was higher when listening to the filler sentences, in which the focused constituent was the verb, as well as the Cl+V[$_F$S] structure, in which the focused subject was located post-verbally.

In order to test participants' sensitivity to the location of focus and their response bias toward focus on the subject, their d-prime scores and response criterion (i.e., C scores) were calculated, respectively, based on their hit (HIT) and false alarm (FA) rates. HIT was considered as instances in which the participants selected questions eliciting focus on the subject (i.e., NF_S or CF_S) when the focused constituent of the stimulus was the subject (i.e., [$_F$S]Cl+V and Cl+V[$_F$S] structures), while FA was considered as instances in which the participants selected questions eliciting focus on the subject (i.e., NF_S or CF_S) when the focused constituent of the stimulus was the verb (i.e., S+Cl[$_F$V] and Cl[$_F$V]S structures). Figure 21 shows the d-prime scores by group and word order. As shown in Figure 21, all the three groups had noticeably low d-prime scores in the two word orders, indicating that participants' sensitivity to distinguish focus on the subject and focus on the verb was very low, regardless of word order. A two-way mixed ANOVA with group and word order as independent variables and d-prime scores as the dependent variable was conducted using the *aov* function in R (Baayen, 2008). Results showed that there was a main effect of group ($F(2, 130) = 3.279$, $p < 0.05$), which indicates that the d-prime scores were different among the three groups. There was also a main effect of word order ($F(1, 130) = 8.808$, $p < 0.01$) and a marginally significant interaction between group and word order ($F(2, 130) = 2.505$, $p < 0.086$), suggesting that the d-prime scores were different between the two word orders and this effect was not the same across all groups. Tukey HSD post-hoc analysis revealed that this was due to the L2s having significantly higher d-prime scores than the HSs ($p < 0.05$) and, marginally, than the NSs ($p = 0.078$) when the subject was positioned pre-verbally, while the d-prime scores did not differ among the three groups when the subject was positioned post-verbally. Indeed, the L2s had significantly higher d-prime scores in the former context than in the latter context ($p < 0.05$). This indicates that, although the L2s were slightly more sensitive in distinguishing the location of focus than the NSs and the HSs when the sentences are structured in a canonical word order (i.e., pre-verbal subject), overall the three groups had great difficulty distinguishing whether the focus was on the subject or on the verb.

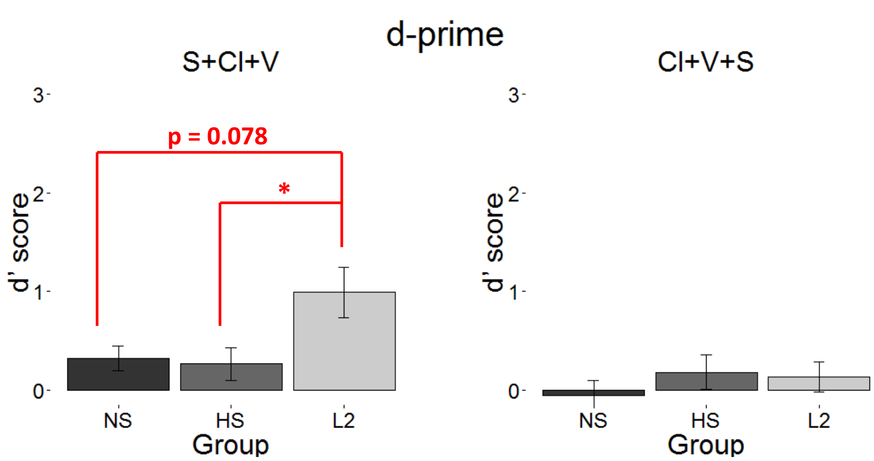


Figure 21 Participants' d-prime scores by group and word order
(***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

In addition to the d-prime scores, participants' C scores were analyzed in order to examine whether they had bias toward focus on the subject. As shown in Figure 22, the participants had negative C scores in both word orders, suggesting that there was a stronger tendency of selecting questions eliciting focus on the subject than those eliciting focus on the verb. A two-way mixed ANOVA with group and word order as independent variables and C scores as the dependent variable showed that there was a main effect of group ($F(2, 130) = 12.644$, $p < 0.001$), which indicates that the C scores were different among the three groups. No main effect of word order or significant interaction between group and word order was found. Tukey HSD post-hoc analysis confirmed that the L2s had significantly higher C scores than the NSs the HSs ($p < 0.01$) when the subject was positioned after the verb. Although the same trend was found when the subject was located pre-verbally, the group difference did not reach significance level; only a marginally significant difference was found between L2s' C scores and those of the HSs ($p = 0.062$). These findings imply that neither prominence nor word order plays a critical role in participants' perception of focus in Spanish. Rather, in the case of the NSs and the HSs, a clear bias was found toward preceding contexts that elicit focus on the subject. The L2s, on the other hand, appeared to be slightly more sensitive to the location of focus, which may be due to their attention to prominence.

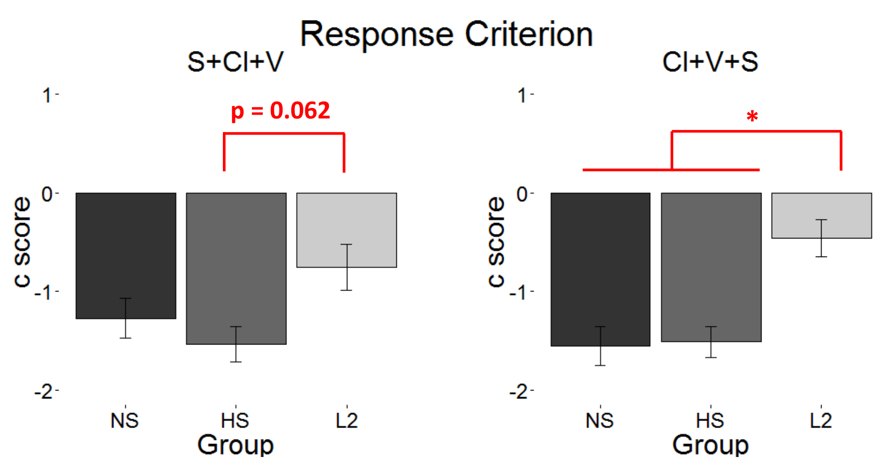


Figure 22 Participants' C scores by group and word order
(***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

4.4 Experiment 4: Production of nuclear stress

4.4.1 Methods

4.4.1.1 Participants

The same subjects in Experiments 1, 2, and 3 participated in Experiment 4.

4.4.1.2 Materials

Among the sentences used in Experiment 3, 12 sentences with the canonical S+Cl+V word order (e.g., *Liliana lo preparó*. 'Liliana prepared it.') were chosen as the target items for Experiment 4. Only sentences, in which the stressed syllables of the subjects and the verbs containing consonants that are characterized to mark a clear acoustic boundaries in Spanish, such as voiceless obstruents (e.g., /p/, /k/, /t/, /s/), non-palatal nasals (e.g., /m/, /n/), and liquids (e.g., /l/, /r/), were selected. This criterion was used for the facilitation of segmentation. These sentences were paired with three prompt questions, each eliciting different scope and location of focus: broad focus (BF) (e.g., *¿Qué pasó?* 'What happened?'), narrow focus on subject (S) (e.g., *¿Quién lo preparó?* 'Who prepared it?'), and narrow focus on verb (V) (e.g., *¿Qué hizo Liliana?* 'What did Liliana do?'). Apart from the 12 triplet target items, 36 sentences that had different structures from the target items (e.g., sentences with direct objects expressed in full nominal phrases, sentences with prepositional phrases) were used as filler items. The filler items were paired with a question eliciting broad focus (9 items), narrow focus on the verb (9 items), or narrow focus on the prepositional phrase (18 items). All the questions were followed by a sentence expressing failure of communication, i.e., *Perdón, no te oí.* 'Sorry, I did not hear you.', which were produced by a male native speaker of Mexican Spanish. In both the U.S. and in Mexico, the productions were recorded in a sound-attenuated booth. In the U.S., the recordings were collected using an AKG C520 head-mounted microphone and a Marantz PMD570 solid state recorder with a sampling rate of 48 kHz and a sample size of 16 bits. In Mexico, the recordings were collected using an AKG C520 head-mounted microphone and a Zoom H4n handy portable digital recorder with a sampling rate of 44.1 kHz and a sample size of 16 bits. During the productions in both locations, the microphone was positioned approximately 2 inches away from the participants' lips. The sound files collected in the U.S. were resampled to 44.1 kHz to match with the ones collected in Mexico. The list of the stimuli used for Experiment 4 is presented in Appendix E.4.

4.4.1.3 Procedures

After completion of Experiment 3, the participants continued with Experiment 4. A simulated interactive elicitation task was conducted, using *PsychoPy 2*. The participants were instructed to imagine as if they were having a conversation with someone on the phone. They first read out loud the sentences presented on the computer screen, as if they were initiating a conversation by informing the imaginary interlocutor on the other end with the message of the sentences. Soon after completing the sentences, the participants pressed a key on the keyboard and listened to the pre-recorded prompt questions described above. Their task was to answer the questions as freely as possible with one condition; the action (i.e., the verb) and the person who performed the action (i.e., the subject) had to be included in their response. Apart from the phase in which the participants had to read the sentence to initiate the conversation, nothing was shown on the computer screen. All the items were presented in a randomized order and special care was taken to make sure that the triplets of the same item did not appear in a consecutive order. An example of the format of the simulated interaction is presented below.

Example: Participant: *Liliana lo preparó.* 'Liliana prepared it.'
Interlocutor: *Perdón no te oí. ¿Quién lo preparó?* 'Sorry, I did not hear you. Who prepared it?'
Participant: ________________________________

Although it is not as natural as collecting spontaneous speech data, this method was used with the expectation that, compared to reading out loud tasks, it would allow participants have more freedom to use various prosodic and syntactic cues to express different types of focus, while eliciting the target lexical items of interest (i.e., the subject and the verb) and controlling for interlocutor effect. Before the initiation of the main task, a practice test with 10 items, which were not the target items, was conducted for the familiarization with the task format.

4.4.1.4 Coding and analysis

Participants' responses were first coded based on the word order of the response (i.e., S+Cl+V or Cl+V+S). Acoustic analyses were conducted on the stressed syllable of the focused and the non-focused constituents to examine whether focus is expressed prosodically through the relative prominence between them. Suprasegmental information, such as the duration, intensity, pitch, and pitch range of the stressed syllables of the subjects and the verbs, was extracted using scripts in *Praat* (Boersma & Weenink, 2015). Speech segmentation was first performed using *EasyAlign* (Goldman, 2011). Later, the results were individually checked and manually corrected when needed. The formant structure in the spectrogram and the periodicity of the waveform were used as criteria for the manual correction. That is, the beginning and end of a vowel were identified as the zero-crossing points of the regular periodic signal (in the waveform) closest to the onset and the offset of a continuous second formant (F2) (in the spectrogram) (Baker, 2006; Recasens, 1999). The duration of the stressed syllables was calculated as the time of the onset of the stressed syllable subtracted from the time of the offset of the stressed syllable. Average pitch (Hz) was extracted using the *To Pitch* function in *Praat*, with a time step window of 0.01 second, pitch floor of 75 Hz, and pitch ceiling of 600 Hz, which are the default values. In addition to the average pitch, the pitch range of the stressed syllable of the first content word (Syll_W1) and the stressed syllable of the second content word (Syll_W2) were compared. Pitch range was calculated as the difference between the maximum and the minimum pitch of the stressed syllables. As for average intensity (dB), the values were extracted using the *Get Intensity (dB)* function. The comparison between the focused and the non-focused constituents was carried out based on the relative prominence (i.e., duration, intensity, pitch, and pitch range) of the Syll_W1 and the Syll_W2 of each sentence (e.g., *-lia-* in *Liliana* and *-ró* in *preparó* in the sentence *Liliana lo preparó*. 'Liliana prepared it'). The relative prominence was compared across the three contexts of different scope and location of focus (i.e., BF, S, and V). The idea is that, if the stressed syllable of the focused word is produced with longer duration, higher intensity, higher pitch, and larger pitch range than the stressed syllable of the unfocused word, then the difference between Syll_W1 and Syll_W2 would be the largest when the focus is on W1, followed by cases in which the scope of the focus encompasses both Syll_W1 and Syll_W2 (i.e., BF) and cases in which the focus is on W2. In order to control for individual differences, such as speech rate, gender (female / male), and test location (Mexico / U.S.), the raw duration, intensity, pitch, and pitch range values were normalized using z-score normalization, which is calculated as the distance between the raw value and the mean value of each speaker divided by the standard deviation. Then, the difference between the normalized duration, intensity, pitch, and pitch range of the Syll_W1 and the Syll_W2 were calculated to examine whether these values vary depending on the focus type. For instance, it is expected that the difference between *-lia-* in *Liliana* and *-ró* in *preparó* in the sentence *Liliana lo preparó.* 'Liliana prepared it' would be larger if the focused constituent is *Liliana* than when the focused constituent is *preparó*. Apart from the suprasegmental information of the stressed syllables mentioned above, the degree of the displacement of pitch peak (H) of W1, which was defined as the distance from the onset of the stressed syllable to the location of H, was analyzed across the three focus types (i.e., BF, S, and V). As H in a non-final position is generally displaced to a following syllable when the word receives a prenuclear pitch accent (L+>H*), it was of interest to see whether early alignment of H (L+H*) is observed when W1 (i.e., non-final position) is focused, compared to when it is not. The location of H was determined semi-automatically by selecting a region in which a peak was detected and extracting the point of the maximum pitch in that region. As the comparisons were made across tokens with varying stressed syllable duration, the H displacement values were normalized by dividing them into the duration of the stressed syllable. Tokens that were missing, produced with hesitation, rising contour, incorrect subject or verb, missing subject or verb, unclear articulation, creaky voice throughout the utterance, or radically different sentence structures (e.g., cleft construction) were excluded from the analyses. Moreover, the comparisons between Syll_W1 and Syll_W2 were done only with complete minimal triplets. That is, if a token was excluded due to any of the reasons above, the other tokens from the same minimal triplet were also excluded from the analyses. The effects of group (NS / HS / L2), focus type (BF / S / V), and the interactions between the fixed factors on the relative difference between the acoustic information of Syll_W1 and W2 mentioned above (i.e., normalized duration, intensity, pitch, and pitch range) were analyzed using linear mixed effects modeling with participants and items as random effects. The *lmer* function in the *lme4* package in R (Baayen, 2008) was used for the analyses. The best fitting models selected through backward elimination included random intercepts for subject and item with by-subject random slope for focus type for all measures, except for the normalized f0 range, of which the model included random intercepts for subject and item without any slope terms. All the fixed factors (i.e., group and focus type) were centered using contrast-coding. With regard to pitch alignment of Syll_W1, the effects of group (NS / HS / L2), focus type (BF / S / V), and the interaction between the fixed factors on the normalized H displacement values were analyzed using linear mixed effects modeling with participants and items as random effects. The best fitting model selected through backward elimination included random intercepts for subject and item with no slope terms. All the fixed factors (i.e., group and focus type) were centered using contrast-coding. Further pairwise analyses were conducted using the *lsmeans* function in the *lsmeans* package.

4.4.2 Results

Among the total number of 2376 tokens (i.e., 66 participants * 3 focus types * 12 items), 1164 tokens were excluded due to the reasons mentioned above, resulting in 1212 tokens for the analyses. Before presenting the results of the acoustic analyses, it is important to note that the majority of participants' responses were constructed in the S+Cl+V word order. As seen in Figure 23, there were few cases in which the participants responded in the Cl+V+S word (141 out of the 1212 tokens). The effects of group (NS / HS / L2), focus type (BF / S / V) and the interaction between the two fixed factors on the word order of the participants' responses were examined using logit mixed effects modeling with subject and item as random effects. The best fitting model selected through backward elimination included random intercepts for subject and item with no slope terms. Results showed that there was a main effect of focus type for S ($\beta = 0.761$, $SE = 0.28$, $z = 2.713$, $p < 0.01$), suggesting that there were significantly more cases of Cl+V+S word order in S than in BF (i.e., baseline focus type). Although no significant group effect was found, there was a marginal effect for HS ($\beta = -1.865$, $SE = 1.082$, $z = -1.722$, $p = 0.085$) and a significant interaction between group (HS) and focus type (S) ($\beta = -1.921$, $SE = 0.599$, $z = -3.205$, $p < 0.01$), which indicates that the NSs produced Cl+V+S word order slightly more frequently than the HSs and the difference between the HSs and the NSs (i.e., baseline group) was significantly larger in V than in BF. Pairwise comparisons with group and focus type revealed that the HSs produced the Cl+V+S word order more frequently in S than in BF ($p < 0.01$) and V ($p < 0.05$). However, apart from the HSs, no such difference was found in the production of Cl+V+S of the NSs and the L2s.

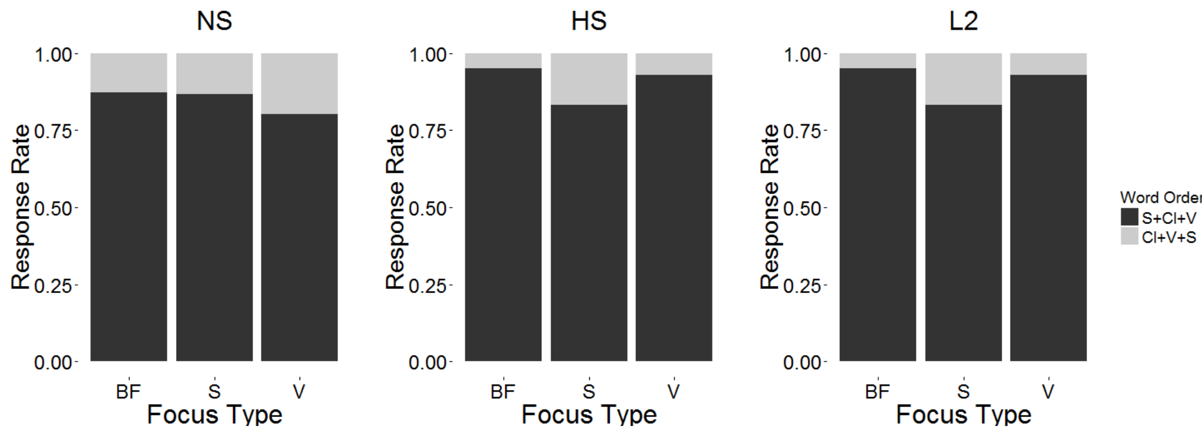


Figure 23 Distribution of responses with preverbal (S+Cl+V) and postverbal (Cl+V+S) subjects by group and focus type

Due to the small number of triplets produced with the Cl+V+S word order, acoustic analyses were conducted only with tokens of the S+Cl+V word order that formed triplets (i.e., 927 tokens). It is important to note that a large amount of tokens of the S+Cl+V word order (233 out of the 927 tokens) were produced with a prosodic boundary, which was identified as a short pause or glottalization at the word edge. As prosodic boundary has a large effect on the suprasegmental information of the syllables located at the boundary (Cole, 2015), these tokens were not included in the acoustic analyses. This study also excluded 66 tokens in which the W2 (i.e., the verb) was produced with sustained creaky voice, most likely due to deaccenting. The excluded tokens are further discussed in Section 4.5.2. Here I report the results of the acoustic analyses conducted on the remaining 465 tokens with complete minimal triplets.

4.4.2.1 Difference in duration, intensity, pitch, and pitch range

Figure 24 shows the difference in the normalized duration, intensity, pitch, and pitch range between Syll_W1 and Syll_W2 of the three groups across the three focus types. Values higher than 0 (marked with red dotted lines) indicate that Syll_W1 was produced with longer duration, higher intensity, higher pitch, and larger pitch range than Syll_W2 and values lower than 0 indicate that Syll_W1 was produced with shorter duration, lower intensity, lower pitch, and smaller pitch range than Syll_W2. As seen in Figure 24, although the exact values differed, a trend was observed across the types of measure, in which the values in S were higher, compared to those in BF and V. This suggests that, even though the relative prominence between W1 (i.e., the subject) and W2 (i.e., the verb) differed depending on the type of measure, in general the degree of the relative prominence was larger in S than in BF and in V.

Results showed that for all the measures, except for pitch range, there was a (marginally) significant main effect of group for HS (Duration: $\beta = -0.607$, SE = 0.308, t = -1.968; Intensity: $\beta = 0.34$, SE = 0.189, t = 1.801; Pitch: $\beta = 1.484$, SE = 0.362, t = 4.1), suggesting that the difference between Syll_W1 and Syll_W2 was larger in the NSs (i.e., the baseline group) than the HSs. There was also a main effect of focus type for V in all measures (Duration: $\beta = -1.104$, SE = 0.384, t = -2.874; Intensity: $\beta = -0.833$, SE = 0.168, t = -4.944; Pitch: $\beta = -0.61$, SE = 0.185, t = -3.296; Pitch range: $\beta = -0.628$, SE = 0.166, t = -3.774), which indicates that the difference between Syll_W1 and Syll_W2 was larger in BF (i.e., the baseline focus type) than in V. Additionally, in the cases of intensity and pitch, a main effect of focus type for S was found (Intensity: $\beta = 0.29$, SE =0.122, t = 2.372; Pitch: $\beta = 0.258$, SE = 0.102, t = 2.529). This indicates that the difference between Syll_W1 and Syll_W2 was larger in S, compared to that in BF. However, despite these effects, pairwise comparisons with group and focus type revealed that the differences among the three focus types did not reach significance level in most cases. Significant or marginally significant difference was found only between HSs' duration difference in S and V ($p = 0.07$), between HSs' intensity difference in S and V ($p < 0.01$), between HSs' intensity difference in S and BF ($p = 0.07$), and between NSs' pitch range difference in S and V ($p < 0.05$). When comparing across groups, (marginally) significant difference was found only in the pitch difference in S between the NSs and the L2s ($p = 0.069$) and the pitch difference in V between the same groups ($p < 0.05$).

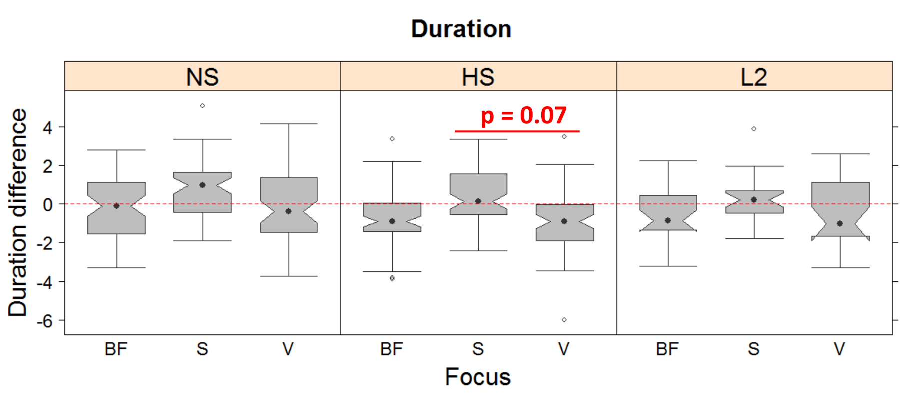


Figure 24 Difference in normalized duration, intensity, pitch, and pitch range between stressed syllables of the first and second content words
(***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$)

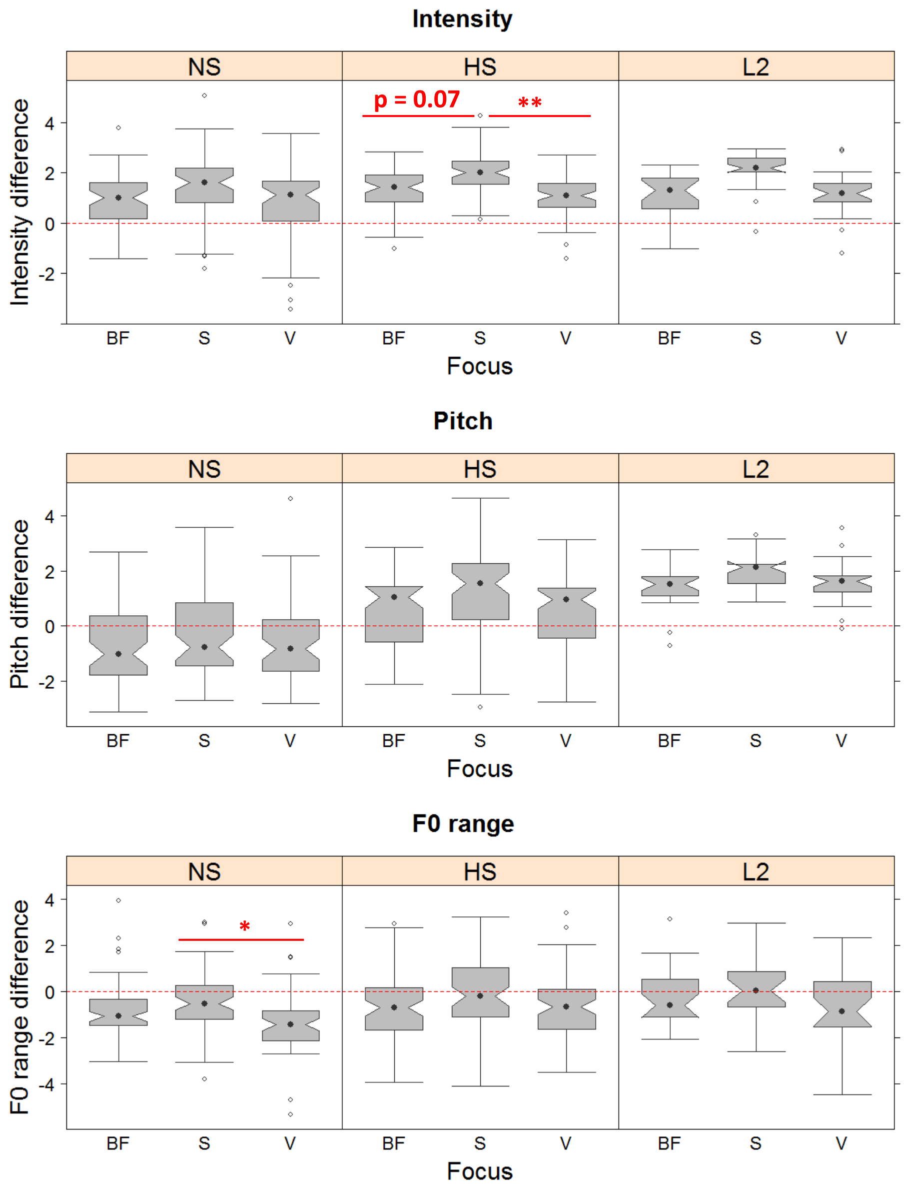


Figure 24 (cont.)

4.4.2.2 Pitch alignment

Figure 25 shows the normalized degree of H displacement for the three focus types produced by the NSs, the HSs, and the L2s, which was calculated as the distance from the onset of Syll_W1 to the location of H divided by the duration of Syll_W1. The two red dotted lines indicate the onset and the offset of Syll_W1. Thus, if a value is within the two lines, it suggests that the H was aligned within Syll_W1 and if the value is outside of the two lines, it means that the H was aligned a following syllable. In Figure 25, a trend is observed in which the H in S was distributed more toward the left, compared to the H in BF and in V. This suggests that the H was displaced to a lesser degree when W1, in this case the subject, was focused.

Results showed that there was a main effect of group for the HSs (β = -0.328, SE = 0.109, t = -3.003) and main effects of focus type of both S (β = -0.111, SE = 0.051, t = -2.169) and V (β = 0.298, SE = 0.072, t = 4.109). This indicates that overall the H was displaced to a lesser degree in the L2s, compared to the NSs (i.e., the baseline group), and, regarding the focus type, the H was displaced to a lesser degree in S and to a larger degree in V, compared to BF (i.e., the baseline focus type). Also, significant interactions were found between group (HS) and focus type (S) (β = -0.211, SE = 0.074, t = -2.861), between group (HS) and focus type (V) (β = 0.429, SE = 0.124, t = 3.462), and between group (L2) and focus type (S) (β = -0.212, SE = 0.089, t = -2.39). That is, the difference in the degree of H displacement between the NSs and the HSs was larger in S and in V than in BF. Also, the difference in the degree of H displacement between the NSs and the L2s was larger in S. Indeed, pairwise comparisons with group and focus type revealed that, in S, L2s' H displacement was significantly lower than that of the NSs and the HSs ($p < 0.05$ for both), while no group difference was found in BF and in V. The pairwise comparisons also showed that HSs' H displacement was higher than that of the NSs, although this did not reach significance level ($p = 0.086$). When comparing across the three focus types, results showed that H displacement occurred to a significantly lesser degree in S than in BF and in V for both the HSs and the L2s, while for the NSs the degree of H displacement did not differ among the three focus types.

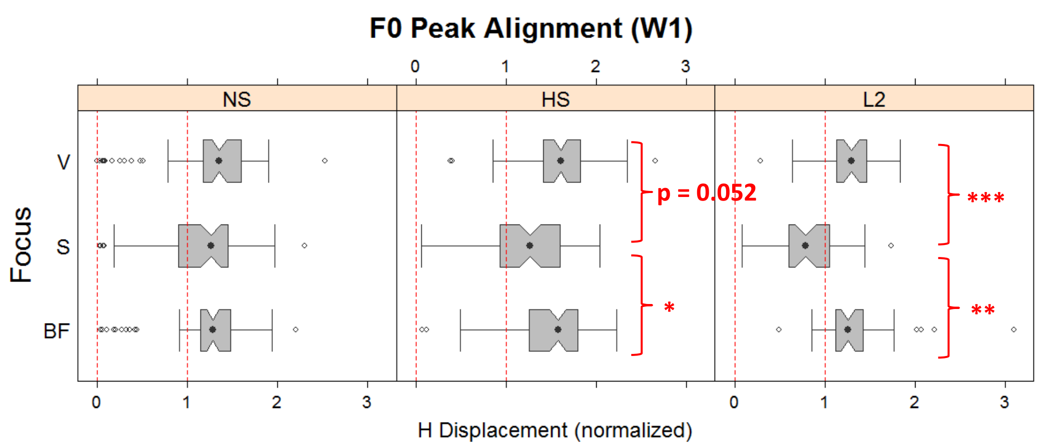


Figure 25 Normalized distance between stressed syllable onset and pitch peak (H)
(***: $p < 0.001$, **: $p < 0.01$; *: $p < 0.05$)

4.5 Discussion

4.5.1 Perception of nuclear stress

The findings of the perception task suggest that word order may not play a critical role in identifying focus and information status in Spanish as one might expect. If word order was an important cue, the participants would have chosen the word that was located at the end of the utterance. That is, a high response rate[S] for the Cl+V[$_{F}$S] structure and a low response rate[S] for the [$_{F}$S]Cl+V structure would have been observed. Similarly, for the filler items, a low response rate[V] for the Cl[$_{F}$V]S structure and a high response rate[V] for the S+Cl[$_{F}$V] structure would have been observed. Instead, the results showed that the NSs and the HSS had a high response rate[S] equally for both Cl+V[$_{F}$S] and [$_{F}$S]Cl+V structures and a low response rate[V] equally for both Cl[$_{F}$V]S and S+Cl[$_{F}$V] structures. This outcome also rejects the possibility that prominence had an effect on their responses, because, if that was the case, not only that a high response rate[S] should have been observed for the Cl+V[$_{F}$S] and the [$_{F}$S]Cl+V structures, but also that a high response rate[V] should have been observed for the Cl[$_{F}$V]S and the S+Cl[$_{F}$V] structures. NSs' and the HSs' high response rate[S] when the focus was on the subject and low response rate[V] when the focus was on the verb indicate that, regardless of word order and prominence, they selected questions eliciting focus on the subject (i.e., NF_S or CF_S) in the majority of the trials. Indeed, this was confirmed when testing for participants' sensitivity (i.e., d-prime scores) and response bias (i.e., C scores); the NSs and the HSs did not distinguish the stimuli that had focus on the subject from those that had focus on the verb, and this was due to the NSs and the HSs having a bias toward focus on the subject. The L2s also considered that the focus was on the subject in many cases, but they did not have such a strong bias toward focus on the subject. Unlike the NSs and the HSs, who showed a categorical pattern preferring NF_S across sentence structures, the L2s had higher response rates of NF_V (i.e., narrow focus on the verb) than the other two groups, mostly when listening to the filler items which had focus on the verb. This suggests that L2s are more sensitive to prosody than NSs and HSs, although this may not be the most important cue to which they attend when identifying focus.

NSs' and HSs' strong bias toward focus on the subject may be an indication that listeners do not pay close attention to syntactic (i.e., word order) or prosodic (i.e., prominence) cues when processing information status and focus. Rather, according to a follow-up question to some of the participants of the study, it is likely that they put more weight on other factors, such as the content and the structure of the message. When asked how they would respond in real life to the questions that appeared in the experiment (e.g., *¿Quién lo preparó?* 'Who prepared it?'), the participants responded that they would usually answer in one word/phrase (e.g., *Liliana*) or use a completely different sentence structure (e.g., *Liliana fue la que lo preparó.* 'Liliana was the one who prepared it.), rather than repeating given information, which they considered unnatural or "forced". The participants also commented that if they had to maintain the constituents used in the study (i.e., S, Cl, and V), it would be more natural if the focus was on the subject than if it was on the verb. For instance, the complete sentence *Liliana lo preparó.* 'Liliana prepared it.' would be somewhat natural if it was responding to the question *¿Quién lo preparó?* 'Who prepared it?' (i.e., focus on the subject), while it would sound odd if it was a response to the question *¿Qué hizo Liliana?* 'What did Liliana do? (i.e., focus on the verb). In this context, the most natural answer would be *Lo preparó.* without an overt subject. Given that the subjects were expressed in all the stimuli, this may have led the NSs and the HSs think that the questions must have been asking about the subject of the sentence, which explains why the NSs and the HSs were having a bias toward focus on the subject. Asymmetry in the realization of focus on different constituents is not a new finding. For instance, Hoot (2016, to appear) found that, while Spanish listeners considered sentences with non-final focus as highly acceptable, regardless of whether the focused constituent was a subject (e.g., *Mi TÍO compró un carro.* 'My UNCLE bought a car.') or an object (e.g., *Mi mamá le dio un CHOCOLATE a mi sobrino*. 'My mom gave a CHOCOLATE to my nephew.'), they showed a different pattern when the focused constituents were located utterance-finally. That is, the listeners rated higher when the utterance-final focus was an object (e.g., *Mi mamá le dio a mi sobrino un CHOCOLATE.*) than when it was a subject (e.g., *Compró un carro mi TÍO.*). Further research needs to be carried out to understand why such asymmetries occur in sentences with focus on different constituents.

4.5.2 Production of nuclear stress

With regard to the production of nuclear stress, the participants did not seem to use syntactic cues (i.e., word order) to express focus. Responses with inverted word order (i.e., Cl+V+S) only occurred in 18.04% of the time for the NSs, 10.57% of the time for the HSs, and 7.62% of the time for the L2s. When examining the use of inverted word order across the three focus types, the HSs and the L2s produced this sentence structure more frequently in S than in BF or in V, although the difference reached significance level only for the HSs. Despite such a pattern, given that the use of inverted word order was so infrequent, it is difficult to conclude that word order played a significant role in the realization of focus for these speakers. With regard to the NSs, although they produced inverted word order more frequently than the HSs and the L2s, focus type did not have any effect on word order. This finding suggests that, contrary to what has been expected, focus in Spanish is not mainly expressed by moving the focused constituents utterance-finally. Rather, stressing the focused constituents *in situ* seems to be a more common strategy that speakers use when expressing focus, supporting the claims of recent experimental studies (Gabriel, 2010; Gupton & Leal Méndez, 2013; Hoot, 2016, to appear). Results of the acoustic analyses confirmed that, when the subject was focused in sentences with the S+Cl+V word order, its duration, intensity, pitch, and pitch range, relative to those of the verb, were longer, higher, and larger than when the focus was on the verb or in broad focus contexts.

Although the three groups showed a similar pattern regarding the relative prominence between the subject and the verb, a clear distinction was found among them with respect to the pitch and the alignment of pitch peak of the subject (i.e., W1). Firstly, as seen in Figure 24, the L2s produced the subjects with higher pitch than the verbs in a categorical way (96% of the cases). Although to a lesser degree than the L2s, the HSs also produced the subjects with higher pitch in the majority of the cases (71.5% of the cases). On the other hand, NSs' subjects were mostly produced with lower pitch than the verbs (71.04% of the cases). This is likely to be due to NSs' flat or falling contour during the stressed syllable of the subject (Figure 26). This contour is similar to L*+H pitch contour, characterized as a pitch valley on the stressed syllable and a delayed pitch rise at the onset of the post-stressed syllable, and its variants, which are often found in Mexican Spanish, as described in Robles-Puente (2014). However, it is unclear whether the NSs used this contour to mark focus, since the same pattern was also found in other focus types. This is similar to what has been found in Muntendam and Torreira (2016) with both native speakers of Peninsular Spanish and Spanish-Quechua bilingual speakers. Therefore, it may be the case that there are additional pragmatic values, in conjunction with focus, that are associated with the flat contour. For instance, Escandell-Vidal (2011) reported that a low pitch contour with elongated duration is used to reiterate and emphasize information that has been formed as part of the interlocutors' common ground. Although the existence of a common ground cannot be assumed in the present study, it is possible that the NSs used this pitch contour to express that they were reiterating what has been told already. However, this pattern was mainly found in the speech of the NSs, while HSs and the L2s showed a larger tonal movement (Figure 26), supporting the findings of Gries and Miglio (2014) and Harris et al. (2015).

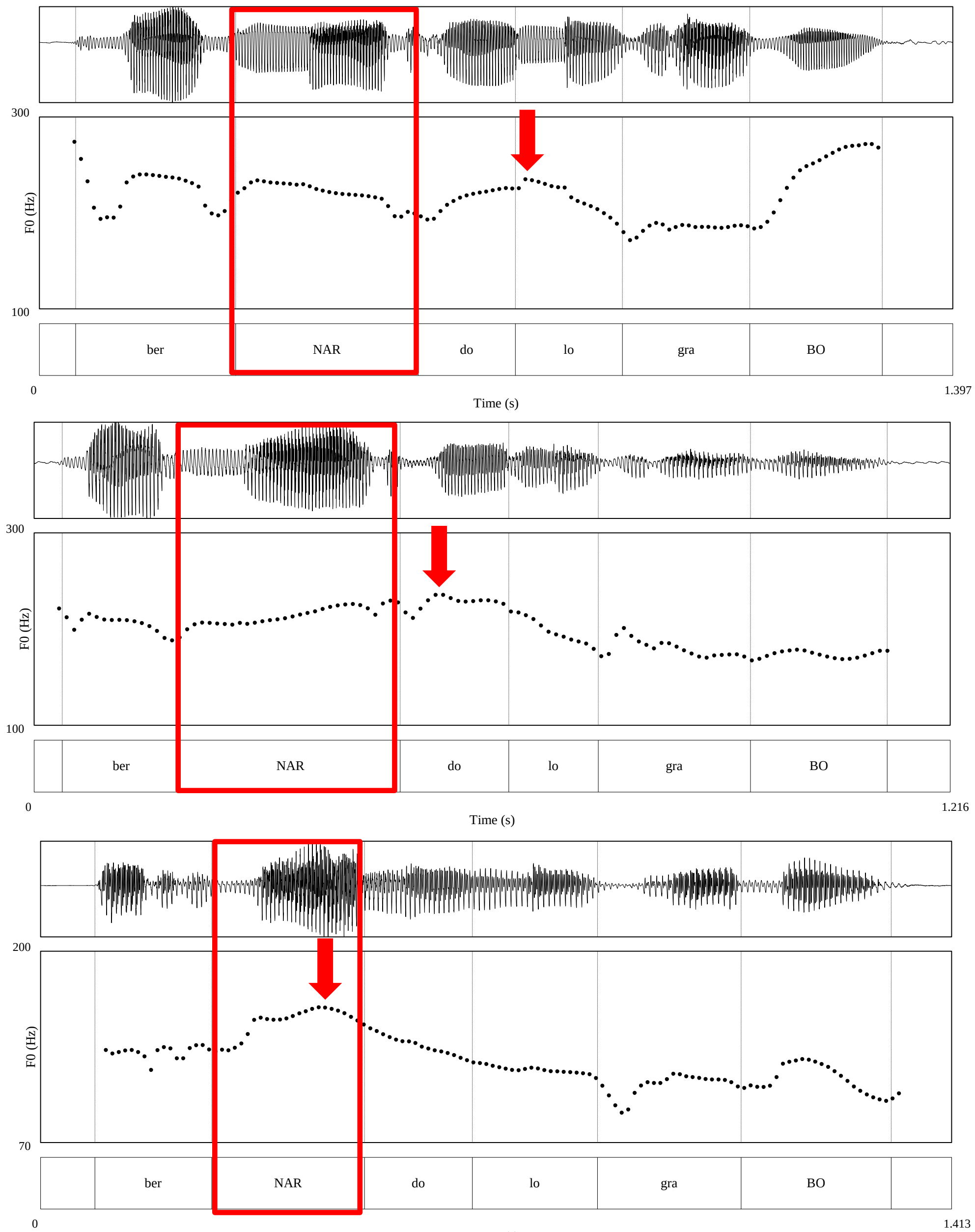


Figure 26 Pitch contour of [$_F$*Bernardo*] *lo grabó*. 'Bernardo recorded it.' produced by a monolingual speaker (top), a heritage speaker (middle), and a second language learner (bottom)

As for the alignment of pitch peak, the NSs displaced the pitch peak to a following syllable in the majority of the cases and the frequency of displaced pitch peaks did not differ across the three focus types (BF: 78.69%, S: 72.13%, V: 77.05%), whereas the L2s displaced the pitch peak mostly in BF (79.49%) and in V (84.62%); in S, the pitch peak was displaced in only 28.21% of the cases. That is, in this focus type, the L2s tended to align the pitch peak within the stress syllables. With regard to the HSs, they seemed to have the characteristics of both the NSs and the L2s. Similar to the NSs, the HSs displaced the pitch peak most of the time in all three focus types. However, unlike the NSs, HSs' pitch peak was displaced less frequently in S (74.39%), compared to BF (92.68%) and V (95.12%), which is more similar to what has been found with the L2s. This trend is also schematized in Figure 26 (see red arrows). Given that early alignment of pitch peak in focused constituents is commonly found in English, the alignment pattern found in HSs' speech may be a result of a fusion between Spanish (L+>H*) and English pitch contour of nuclear stress (L+H*).

It is important to note that the results presented in this study represent less than a fifth of the entire production data (i.e., 465 out of 2376 tokens). As explained earlier, the majority of the tokens were excluded in the acoustic analyses, due to multiple reasons, including hesitation, rising contour, incorrect subject or verb, missing subject or verb, unclear articulation, creaky voice, radically different sentence structures (e.g., cleft construction), prosodic boundary, and incomplete triplets. This implies that speakers use various strategies to express focus in Spanish, apart from stressing the focused constituent *in situ* and inverting word order (Face & D'Imperio, 2005). Therefore, the excluded data were examined further. The most common strategies that the participants of the present study used were the insertion of a prosodic boundary at the edge of the focused word (Figure 28) and post-focal deaccenting (Figure 30). Out of the 927 tokens that formed complete triplets of the S+Cl+V word order (NS: 381 tokens, HS: 339 tokens, L2: 207 tokens), after excluding irreparable cases, such as hesitation, incorrect subject or verb, and unclear articulation, 215 tokens (NS: 119 tokens, HS: 42 tokens, L2: 55 tokens) were produced with a prosodic boundary at the right-edge of the subject, the majority of which occurred when the focus was on the subject (see Figure 27 and Table 3). This pattern was consistent across the groups, which indicates that, although the three groups differed in the overall frequency of prosodic boundaries, they used this cue to mark focus in a systematic manner. There were also a few cases (18 tokens), mostly in the L2 speech, in which a prosodic boundary was inserted at the left-edge of the verb. However, since the number was very small and the tokens were distributed relatively evenly across the three focus types (BF: 4 tokens, S: 8 tokens, V: 5 tokens), it is unlikely that this type of prosodic boundary was used to mark focus.

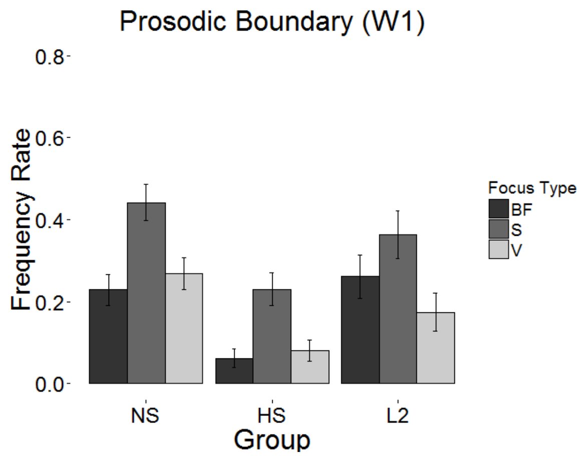


Figure 27 Frequency rate of prosodic boundary at the right-edge of the subject (W1)

Prosodic Boundary	BF	S	V	Total # of tokens
NS	29 (24.58%)	56 (46.61%)	34 (28.81%)	119
HS	7 (16.67%)	26 (61.9%)	9 (21.42%)	42
L2	18 (32.73%)	25 (45.45%)	12 (21.82%)	55

Table 3 Distribution of prosodic boundary across focus type

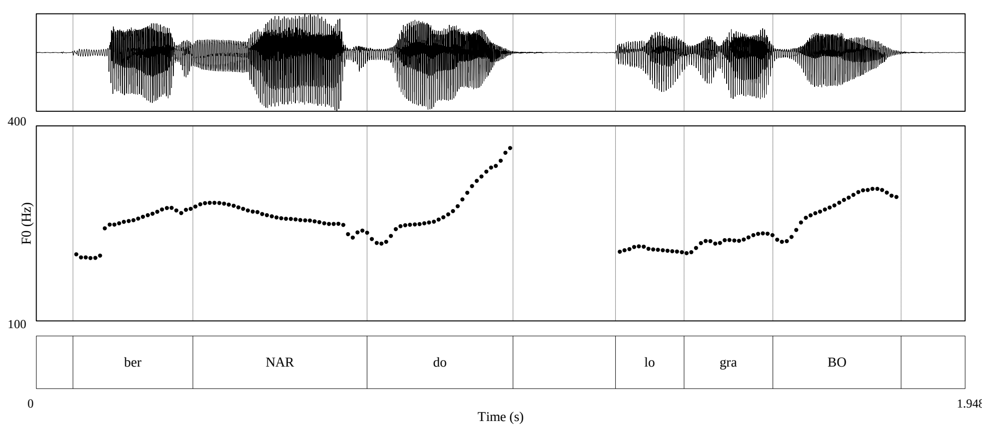


Figure 28 An example of prosodic boundary in [$_F$ *Bernardo*] *lo grabó.* ‘Bernardo recorded it.’ produced by a monolingual speaker

With regard to post-focal deaccenting, which was characterized as a sustained glottalization after a focused constituent (Figure 30), there were 66 tokens, mostly in the L2 speech (NS: 2 tokens, HS: 15 tokens, L2: 49 tokens), in which the verb (W2) was produced this way and this occurred mainly when the focus was on the subject (see Figure 29 and Table 4). As post-focal deaccenting is a strategy that is frequently used in English (Breen et al., 2010; Burdin et al., 2015; Cole, 2014; Ladd, 2008; Ito and Speer, 2006), the deaccenting of the verb after a focused subject found in the L2 speech are likely to be due to influence from English.

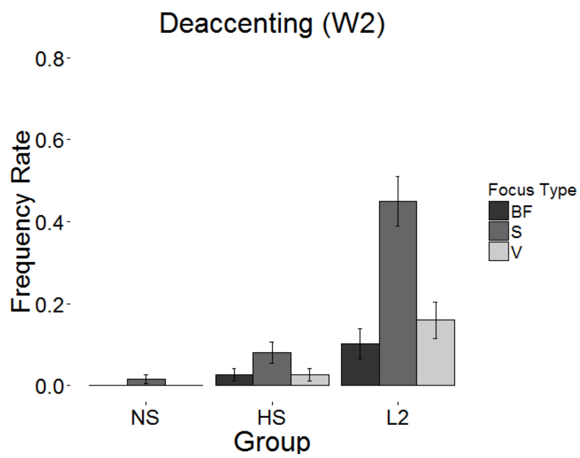


Figure 29 Frequency rate of deaccenting of the verb (W2)

Deaccenting	BF	S	V	Total # of tokens
NS	0 (0%)	2 (100%)	0 (0%)	2
HS	3 (20%)	9 (60%)	3 (20%)	15
L2	7 (14.29%)	31 (63.27%)	11 (22.44%)	49

Table 4 Distribution of prosodic boundary across focus type

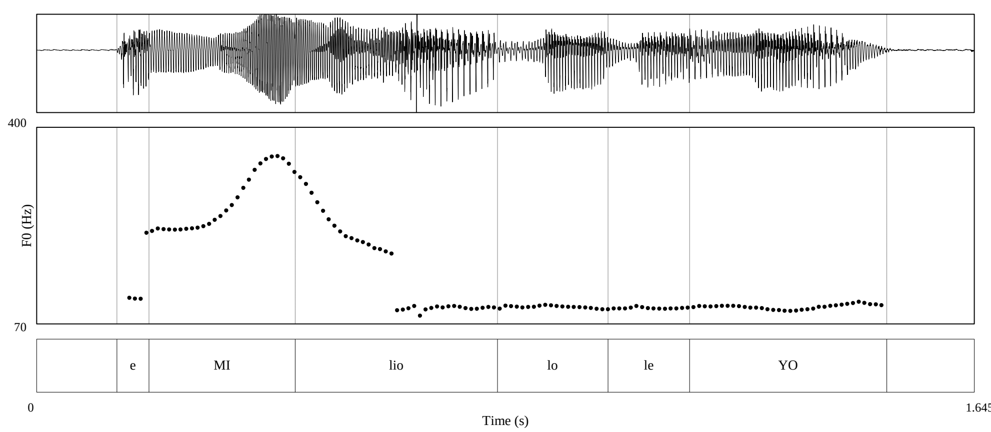


Figure 30 An example of post-focal deaccenting in [$_F$ *Emilio*] *lo leyó.* ‘Emilio read it.’ produced by a second language learner

Although not considered in the 927 tokens that formed complete triplets, other strategies were observed, apart from prosodic boundary and post-focal deaccenting, which are presented below:

- The use of *que* as a possible strategy to mark BF

  e.g., *¿Qué pasó?* 'What happened?
  - *Que Leonardo lo diseño.* '(I told you that) Leonardo designed it.'

Utterance initial *que*	BF	S	V	Total # of tokens
NS	40 (85.11%)	3 (6.38%)	4 (8.51%)	47
HS	1 (100%)	0 (0%)	0 (0%)	1
L2	N/A	N/A	N/A	N/A

Table 5 Distribution of utterance initial *que* across focus type

- Cleft construction as a possible strategy to mark S

  e.g., *¿Quién lo diseño?* 'Who designed it?'
  - *Leonardo fue el que lo diseño.* 'Leonardo was the one who designed it.'

Cleft construction	BF	S	V	Total # of tokens
NS	9 (22.5%)	20 (50%)	11 (27.5%)	40
HS	4 (18.18%)	14 (63.64%)	4 (18.18%)	22
L2	N/A	N/A	N/A	N/A

Table 6 Distribution of cleft constructions across focus type

- Word/Phrase in isolation as a possible strategy to mark narrow focus (S and V)

  e.g., *¿Quién lo diseño?* 'Who designed it?'
  - *Leonardo.* 'Leonardo.'

Word in isolation	BF		S		V		Total # of tokens	
	S	Cl+V	S	Cl+V	S	Cl+V	S	Cl+V
NS	0 (0%)	1 (4.76%)	4 (100%)	0 (0%)	0 (0%)	20 (95.24%)	4	21
HS	0 (0%)	0 (0%)	6 (100%)	1 (20%)	0 (0%)	4 (80%)	6	5
L2	0 (0%)	0 (0%)	2 (100%)	0 (0%)	0 (0%)	6 (100%)	2	6

Table 7 Distribution of word in isolation across focus type

These response patterns are indeed very interesting, as they provide possible explanations to why the participants had difficulty in distinguishing sentences that only differed in word order or the location of nuclear stress in the perception task. That is, it is possible that there are multiple strategies, other than stressing the focused constituents and positioning the focused constituent utterance-finally, that may be used more frequently to mark focus. Therefore, it is not necessary to use certain cues in a categorical manner to signal the focused element and, at the same time, the same cues may be used to mark different focus types (Burdin et al., 2015; Muntendam & Torreira, 2016). Thus, it is likely that the participants in the present study were having problem attending to these cues, because they are rarely used in real life. Moreover, the higher frequency of cases in which the subject was omitted when the focused word was the verb, compared to cases in which the verb was omitted when the focused word was the subject, provides a possible explanation to the bias toward focus on the subject found in the perception data. That is, as some participants pointed out, it is more natural to repeat the verb (e.g., *¿Quién lo preparó?* 'Who prepared it?' - *Liliana lo preparó.* 'Liliana prepared it.') than to repeat the subject (e.g. *¿Qué hizo Liliana?* 'What did Liliana do?' - #*Liliana lo preparó.* 'Liliana prepared it.'). Further research should be conducted to examine the (in)felicity of reiteration of given information of different types of constituents.

Chapter 5 Conclusion

5.1 Summary of the findings

The present study examined the perception and production of lexical stress (i.e., word-level prominence) and nuclear stress (i.e., utterance-level prominence) by heritage speakers of Spanish (HS) and compared their behavior to that of English L2 learners of Spanish (L2) and Spanish monolingual native speakers (NS). Four experimental studies were conducted: two forced-choice identification tasks for the perception of lexical stress (Experiment 1) and nuclear stress (Experiment 3), a reading aloud task for the production of lexical stress (Experiment 2), and a simulated interactive elicitation task for the production of nuclear stress (Experiment 4). The results of each task are summarized below.

5.1.1 Experiment 1: Perception of lexical stress

In Spanish, suprasegmental information, such as duration, pitch, and intensity, plays a critical role in distinguishing stress minimal pairs when no additional contextual cues are provided. Moreover, varying suprasegmental cues signal lexical stress in Spanish, depending on the prosodic context in which it is located. Thus, listeners need to attend to these varying cues in order to successfully identify the location of lexical stress. However, in English, stress minimal pairs are rare and they can be distinguished through cues other than suprasegmental information, such as syntactic (i.e., noun vs. verb) and/or segmental cues (i.e., vowel quality). Thus, suprasegmental information does not play a critical role in identifying the location of lexical stress in English (Cutler, 1986, 2012).

The goal of Experiment 1 was to examine whether heritage speakers of Spanish are able to adjust their attention to varying suprasegmental cues when identifying Spanish lexical stress in different prosodic contexts. Stress minimal pairs of paroxytones and oxytones composed of Spanish regular *-ar* verbs (e.g., *<u>can</u>to* ‘I sing’ vs. *can<u>tó</u>* ‘he/she/you (formal) sang’) were aurally presented in three prosodic contexts with different stress correlates: nuclear position (N) (pitch, intensity, duration), prenuclear position (PN) (pitch alignment, duration), and unaccented context (U) (duration). As U, among the three prosodic contexts, contains the least suprasegmental information that signal lexical stress, I predicted that if heritage speakers successfully attended to suprasegmental cues, they would identify the location of lexical stress less successfully in U, compared to the other contexts. However, if English had an influence in their perception, they would not attend to suprasegmental information to a large degree, thus, confusion will be observed across the prosodic contexts.

Results showed that both the monolingual speakers and the heritage speakers successfully distinguished Spanish minimal pairs in all three prosodic contexts with slightly, but significantly lower accuracy rates in U. This indicates that the monolingual speakers and the heritage speakers were attending to these varying cues when processing Spanish lexical stress. The L2 learners, on the other hand, did not seem to pay attention to these cues, because not only that they were unable to distinguish the paroxytones (*canto*) from the oxytones (*cantó*), but also they showed a clear bias toward the paroxytones. Although the reason to this bias needs to be investigated further, it is likely that these speakers were more sensitive to morphological cues than suprasegmental ones, given that they showed high accuracy rates when listening to the filler items, which, due to their verbal suffix (e.g., *-es* in *Aprendes*. 'You learn.'), were easily identifiable without attention to suprasemgnetal information. As English listeners do not attend to suprasegmental cues to a large degree for lexical processing (Cutler, 1986, 2012), the confusion found in the L2 learners of the present study is likely to be a result of transfer from English. Such an effect was not found in the heritage speakers.

5.1.2 Experiment 2: Production of lexical stress

Although in general lexical stress is manifested acoustically as higher pitch, longer duration and higher intensity of the stressed syllables, resulting from extra muscular energy when producing them (Ladefoged, 2001), there are cross-linguistic differences between Spanish and English in several acoustic properties. One of the most salient differences in the realization of Spanish and English lexical stress is whether vowel quality changes in unstressed positions. In English unstressed vowels undergo reduction (e.g., *atom* [ˈætəm] vs. *atomic* [əˈtɑmɪk]), while vowel quality does not change in such a systematic manner in Spanish (e.g., *átomo* [ˈatomo] vs. *atómico* [aˈtomiko]) (Hualde, 2005). Another important difference between Spanish and English is the alignment of pitch (f0) peak in prenuclear position in declarative sentences. In Spanish f0 peaks of stressed syllables are usually displaced to a following syllable (L+>H*), whereas in English they are often aligned within the stressed syllables (L+H* or H*) (Estebas-Vilaplana, 2007).

The goal of Experiment 2 was to examine whether heritage speakers are able to distinguish Spanish minimal pairs when they produce these pairs in various prosodic contexts. Special attention was paid to vowel quality and pitch alignment in prenuclear position. Stress minimal pairs in three different prosodic contexts (i.e., N, PN, and U) that differed only in the location of lexical stress, such as the ones in Experiment 1, were used in Experiment 2. It was predicted that, if heritage speakers successfully use varying suprasegmental cues, they would adjust their use of duration and tonal cues based on the prosodic context (N: pitch, intensity, duration; PN: pitch alignment, duration; U: duration), while maintaining the vowel quality in unstressed positions. However, if transfer from English occurs, heritage speakers would reduce unstressed vowels and show early alignment of pitch peak in prenuclear position.

Results showed that, while all the groups used duration to distinguish the paroxytones from the oxytones across the three prosodic contexts, the monolingual speakers used this cues in a more consistent manner than the heritage speakers and the L2 learners. That is, the monolingual speakers produced the stressed vowels longer than the unstressed vowels, regardless of the stress pattern, whereas the heritage speakers and the L2 learners did so only when producing the oxytones (*cantó*). When producing the paroxytones (*canto*), they produced the unstressed final vowels longer than the stressed penultimate vowels in more than half of the cases. Although it is unclear why the heritage speakers and the L2 learners produced the unstressed final vowels longer in this stress pattern, based on the types of error found in the filler data, it is possible that the heritage speakers and the L2 learners showed such patterns for different reasons. That is, while the most common error that the heritage speakers made when producing the filler times was the production of incorrect tense (i.e., past tense instead of present tense) (e.g., *Sal<u>í</u>* ‘I left.’ instead of *<u>Sa</u>le*. ‘He/She/You (formal) leave(s).’), for the L2 learners, the most common error was the production of incorrect position of lexical stress (i.e., stress on the final syllable for paroxytones instead of the penultimate syllable) (e.g., **Sa<u>le</u> instead of <u>Sa</u>le*. ‘He/She/You (formal) leave(s).’).

With regard to vowel quality, although some differences were observed among the three groups in the height (F1) and anteriority (F2) of certain vowels, no difference was found between the quality of stressed and unstressed vowels in any of the three groups. That is, based on this finding, it is unlikely that there was an influence from English in the vowel quality of heritage speakers' and L2 learners' speech. Lastly, as for the pitch alignment in PN, both the heritage speakers and the L2 learners behaved differently from the monolingual speakers in that, they displaced the f0 peak farther from the stressed syllable when producing paroxytones than when producing oxytones. Moreover, both groups aligned the f0 peak within the following syllable more frequently when producing paroxytones, although this occurred less frequently for the heritage speakers (25.56%) than the L2 learners (44.44%). The monolingual speakers, on the other hand, did not show any difference between the degree of f0 peak displacement of the two stress patterns and the number of cases in which the f0 peak aligned within the stressed syllable. Given that f0 peak in English often aligns within the stressed syllable in prenuclear position (L+H* or H*) (Estebas-Vilaplana, 2007), the L+H*/H* frequently found in L2 learners' speech may be an indication of transfer from English. With regard to the heritage speakers, although there were fewer cases of L+H*/H* in their speech, compared to that of the L2 learners, the earlier alignment of f0 peak in paroxytones, similar to the L2s, may be interpreted as an influence from English prosody.

5.1.3 Experiment 3: Perception of nuclear stress

Many languages are similar in that nuclear stress is used to encode semantic and pragmatic information about a constituent in a discourse, such as whether it is new or given information or whether it receives contrastive or narrow focus. Word order is flexible in Spanish, while it is relatively rigid in English. Thus, in Spanish information status and focus are mainly marked syntactically, by locating the focused constituent utterance-finally, and sometimes prosodically, by stressing the focused constituent *in situ*, while they are mostly marked prosodically in English (Cole, 2015; Contreras, 1976; Donati & Nespor, 2003; Steedman, 2014; Zubizarreta, 1998).

The goal of Experiment 3 was to examine whether heritage speakers attend to prosody or word order when they listen to stimuli with conflicting cues. Sentences with focus on subject that varied in word order, $[_F S]$Cl+V and Cl+V$[_F S]$, were used in Experiment 3. The prediction was that if heritage speakers process focus prosodically, like in English (and sometimes in Spanish), they would attend to the relative prosodic prominence of the constituents and, thus, identify the subject as the focused constituent. However, if they process focus syntactically, like what has been claimed in Spanish, they would attend to word order and, thus, identify the last word (i.e., the verb in the [$_F$S]Cl+V structure and the subject in the Cl+V[$_F$S] structure) as the focused constituent.

Results showed that all the three groups identified the subject as the focused constituent when listening to the two sentence structures, which may indicate that the participants were attending to the prosodic cues more than the syntactic cues. However, the results of the filler items, which had similar sentence structures as the target items, but with focus on the verb (i.e., S+Cl[$_F$V] and sCl[$_F$V]S), showed that the participants, particularly the monolingual speakers and the heritage speakers, identified the subject as the focused constituent in these sentence structures as well. This finding refutes the possibility that the participants were paying attention to prosodic cues. Further analysis confirmed that, rather than attending to prosodic cues, the participants were having a strong preference for focus on the subject, regardless of the relative prominence of the constituents and the word order. This effect was observed to a significantly lesser degree with the L2 learners, who perceived sentences structures with focus in the verb (S+Cl[$_F$V] and sCl[$_F$V]S) as such more frequently than the other two groups. This suggests that, although prosody may not function as the most important cue, the L2 learners were more sensitive to this cue than the monolingual speakers and the heritage speakers. Despite the findings of recent experimental studies, which support the importance of prosody in focus marking in Spanish, prosody undoubtedly plays a larger role in English than in Spanish. Thus, L2 learners' sensitivity to prosodic prominence observed in the present study indicates that English had an effect in their perception of nuclear stress in Spanish. Such an effect was not found in the heritage speakers.

5.1.4 Experiment 4: Production of nuclear stress

Apart from the possibility of using syntactic cues to express focus, cross-linguistic differences have also been found between Spanish and English with regard to the acoustic properties that speakers of these two languages use. Although prosody can be used in both languages for focus marking, it seems to have a larger effect in English than in Spanish. That is, focus in English tends to be expressed with a larger pitch excursion in the focused constituents than Spanish (Gries & Miglio, 2014; Harris et al., 2015) and post-focal deaccenting (Breen et al., 2010; Burdin et al., 2015; Cole, 2015; Ladd, 2008; Ito & Speer, 2006).

The goal of Experiment 4 was to examine whether heritage speakers use prosody or word order to distinguish focused and non-focused constituents in their production. The participants were instructed to express three types of focus using the same types of constituents as Experiment 3 (i.e., subject, transitive verb, direct object clitic): broad focus (BF), narrow focus on subject (S), and narrow focus on verb (V). It was predicted that if the heritage speakers mainly use prosody, they would place the nuclear stress on the focused constituents *in situ*, producing them with higher prosodic prominence compared to the non-focused constituents. However, if they mainly use word order, they would place the focused constituent in utterance-final position.

Results showed that the three groups did not use syntactic cues (i.e., word order) in a systematic manner to express focus, but rather maintained the canonical word order (S+Cl+V) and stressed the focused constituent *in situ*, supporting the findings of recent experimental studies (Gabriel, 2010; Gupton & Leal Méndez, 2013; Hoot, 2016, to appear). Moreover, among the sentences with the S+Cl+V word order, the duration, intensity, pitch, and pitch range of the subject relative to those of the verb, were longer, higher, and larger in S, compared to BF and V. Although the three groups showed a similar pattern regarding the relative prominence between the subject and the verb, a clear distinction was found among them with respect to the alignment of the f0 peak of the subject (i.e., non-final constituent). While the monolingual speakers displaced the pitch peak to a following syllable in the majority of the cases and the frequency of displaced pitch peaks did not differ across the three focus types, the L2 learners displaced the f0 peak mostly in BF and in V; in S, the L2 learners tended to align the f0 peak within the stress syllables. With regard to the heritage speakers, they seemed to have the characteristics of both the monolingual speakers and the L2 learners. Similar to the monolingual speakers, the heritage speakers displaced the f0 peak most of the time in all three focus types. However, unlike the monolingual speakers, heritage speakers' f0 peak was displaced less frequently in S, compared to BF and V, which is more similar to what has been found with the L2 learners.

The results also showed that the participants used various strategies to mark focus, apart from prosodic prominence and pitch alignment. The strategies that were commonly used across the three groups were prosodic boundary after a focused constituent (e.g., *¿Quién lo diseño?* 'Who designed it?' - *Leonardo, lo diseño*. 'Leonardo, designed it.') and omission of given information (e.g., *¿Quién lo diseño?* 'Who designed it?' - *Leonardo (lo diseño)*. 'Leonardo (designed it).'). There were also strategies that were observed only in certain groups. For instance, the monolingual speakers used *que* to express BF (e.g., *¿Qué pasó?* 'What happened? - *Que Leonardo lo diseño.* '(I told you that) Leonardo designed it.'), and cleft construction to mark S (e.g., *¿Quién lo diseño?* 'Who designed it?' - *Leonardo fue el que lo diseño*. 'Leonardo was the one who designed it.'). These strategies were used to a lesser extent in the heritage speakers and were never observed in L2 learners' speech. Rather, the L2 learners mainly used post-focal deaccenting, possibly due to influence from English prosody. The monolingual speakers and the heritage speakers also used post-focal deaccenting, but to a much lesser degree.

5.2 General discussion

5.2.1 Perception and production of lexical stress

The findings of Experiments 1 and 2 suggest that, while the heritage speakers performed similarly to the monolingual speakers in the perception of lexical stress in Spanish, they showed a deviant pattern in their production. Unlike the monolingual speakers, who used various suprasegmental cues (i.e., difference in duration, pitch, and intensity between penultimate and final vowels, and pitch alignment in PN) in a systematic manner to distinguish paroxytones and oxytones, the heritage speakers demonstrated a preference toward past tense verbs (i.e., oxytones), which is the verb tense that they are more familiar with, when dealing with topics that are unrelated to their personal experience (Martin et al., 2013). Therefore, HSs' deviance from the NSs in their production of Spanish lexical stress is likely to be due to their limited opportunities to talk in Spanish on topics beyond their personal experience, which usually requires the use of various verb forms. The L2 learners, on the other hand, showed influence from English in both the perception and production of Spanish lexical stress. As English listeners do not use stress information, let alone suprasemgnetal information, for lexical processing as much as Spanish listeners do (Cutler, 1986, 2012), when exposed to minimal pairs that differed only in stress position, the L2 learners were not successful in distinguishing them. Instead, they had a bias toward paroxytones, possibly as a result of over-generalizing Spanish stress rules.

With respect to the production of lexical stress, the L2 learners aligned the H within the stressed syllable when producing oxytones in PN, which is how pitch accents are generally expressed in English in this prosodic context. With regard to the production paroxytones, the L2 learners showed a similar pattern as the heritage speakers, in that they produced the unstressed vowels longer than the stressed vowels. Although it is possible that the L2 learners also had preference toward past tense verbs, like the heritage speakers, it is more likely that they positioned a phrase boundary after the target words, due to lack of fluency in Spanish, which may have led to final vowel lengthening. Thus, although both the heritage speakers and the L2 learners were more dominant in English than in Spanish, influence from English seemed to occur to a larger degree to the L2 learners.

5.2.2 Perception and production of nuclear stress

With regard to the perception and production of nuclear stress, both Experiment 3 and 4 showed unexpected outcomes; neither prosodic prominence, nor word order seemed to play a critical role in participants' identification and realization of focus in Spanish. When presented with sentences that differed only in the relative prominence of the constituents and word order, the three groups of participants were not able to distinguish the sentences based on these cues. However, when examining the response patterns of the three groups, the heritage speakers performed similarly as the monolingual speakers, in that they showed a strong bias to focus on the subject, whereas the L2 learners did not show such a bias. In fact, the L2 learners seemed to be more sensitive to the relative prominence of the two constituents of the stimuli (i.e., the subject and the verb) than the heritage speakers and the monolingual speakers. With regard to the production task, in which the participants had to orally express focus of different types (i.e., BF, S, and V), the results showed that the three groups maintained the canonical S+Cl+V word order, even when the focused constituent was the subject. In this context, it was expected that the participants, especially the monolingual speakers, move the subject utterance-finally (Cl+V+S), because theories in focus claim that Spanish has a rigid nuclear stress that falls in the utterance-final position (Zubizarreta & Nava, 2011). Therefore, in order to mark focus, the focused constituent should be placed in this position. However, the participants of the present study rarely changed the word order and, instead, distinguished the three focus types by using the relative prominence (i.e., duration, intensity, pitch, pitch range) between the subject (W1) and the verb (W2). Although this behavior was equally found in the three groups, group difference was observed with respect to the alignment of the f0 peak of the subject. That is, while the heritage speakers and the L2 learners aligned the pitch peak earlier when the subject was focused, the monolingual speakers did not show such an early alignment. Thus, based on the present data, it seems that both the heritage speakers and the L2 learners deviated from the monolingual speakers regarding pitch alignment. It is important to remember that the acoustic analyses conducted in the present study were restricted to sentences of S+Cl+V word order without any prosodic boundaries or creaky voice resulting from post-focal deaccenting. Thus, this finding may be helpful in understanding participants' production of nuclear stress in Spanish, but not in gaining a comprehensive understanding of their production of focus *per se*, as the participants used multiple strategies, apart from relative prominence and pitch alignment, to express different types of focus. Indeed, these additional cues may help us understand why the participants in the present study had difficulty in identifying focus solely based on prosodic prominence and word order. Although this study has only provided a description of such strategies, it is important to conduct further research on these cues to see whether heritage speakers show any patterns that differ from or similar to the monolingual speakers and the L2 learners.

5.2.3 Examination of the hypotheses

The present research was carried out with the goal to empirically test the following three hypotheses:

- Hypothesis 1: Given that heritage speakers are generally more dominant in English than in the heritage language, heritage speakers' heritage language phonology will be influenced by English.
- Hypothesis 2: A subsystem of speech in the heritage language that is hierarchically lower in the cascade will be less influenced by English than a subsystem that is placed higher in the hierarchy.
- Hypothesis 3: Given that heritage speakers tend to speak the heritage language far more frequently than they hear it, there will be a discrepancy between their perception and production of the heritage language speech sounds.

The results of the four experiments showed several patterns in the heritage speakers that diverged from those of the monolingual speakers, such as the early alignment of f0 peak and the elongation of unstressed final vowels in the production of paroxytones (lexical stress), and the early alignment of f0 peak in the production of focused constituents (nuclear stress). The same patterns occurred in the L2 learners, but to a larger degree. Since in English f0 peak in prenuclear position is usually aligned within the stressed syllable (H* or L+H*), unlike in Spanish, in which f0 peak tends to be displaced to a following syllable (L+>H*), the early f0 peak alignment found in the heritage speakers and the L2 learners is likely to be due to influence from English phonology (Hypothesis 1). Also, note that these divergent patterns only occurred in heritage speakers' production, whereas in the perception tasks the heritage speakers performed similarly to the monolingual speakers. Thus, the present study supports that influence from English occurs to a larger degree in heritage speakers' production than in their perception (Hypothesis 3). The L2 learners, on the other hand, showed divergent patterns from the monolingual speakers in both their perception and production. With regard to Hypothesis 2, heritage speakers' discrepancy between their perception and production was found in both lexical stress and nuclear stress. However, as the nuclear stress studies bore unexpected results that are not necessarily phonological in nature (e.g., bias toward focus on subject, use of cleft constructions to mark focus), it was not possible to directly compare the degree of English influence and see whether more influence was found in one linguistic feature over the other.

5.3 Limitations of the study and future direction

Although the findings of the present study shed some light on heritage language phonology, especially on the discrepancy between heritage speakers' perception and production, there are some limitations that need to be adjusted and taken into account in future research. First of all, the present study used a reading aloud method for Experiment 2 (production of lexical stress) and a part of Experiment 4 (production of nuclear stress). As heritage speakers usually use the heritage language orally, they tend to have limited literacy skills in the heritage language (Campbell & Rosenthal, 2000; Montrul, 2011). This may explain why the heritage speakers in Experiment 2 produced paroxytones with long unstressed final vowels. Although further research needs to be conducted to confirm whether this is due to stress misplacement resulting from their preference for past tense verbs or final vowel lengthening, this kind of behavior has also been observed in other studies that involved reading tasks (Knightly et al., 2003; Robles-Puente, 2014). The reading task format in the beginning of Experiment 4 may have also affected the word order of the heritage speakers, as well as other participants, who responded with the S+Cl+V word order in the majority of the cases. Although the main section in this task did not involve any reading (i.e., respond to a question that was presented aurally), since the participants had to read a sentence in the beginning of each trial to "initiate a conversation", which was in the S+Cl+V word order, it is possible that this has primed the participants to respond in this word order. Although recent experimental studies have also found that Spanish speakers do not change word order to mark focus as frequently as previously argued (Gabriel, 2010; Gupton & Leal Méndez, 2013; Hoot, 2016, to appear), heritage speakers' spontaneous speech should be examined whether the same patterns are shown when they speak naturally. As Ito and Speer (2006) pointed out, prosody may differ in read and spontaneous speech (Ayers, 1994; Blaauw, 1994; Howell & Kadi-Hanifi, 1991), because "[r]eaders and talkers have different pragmatic goals and different processing demands".

In the present study, the divergent patterns found in heritage speakers' behaviors, compared to those of the monolingual speakers, were considered as possible influence from English. There are two assumptions related to this view that needs further investigation. The first assumption is that the Spanish that heritage speakers are exposed to is of the same variety that is spoken by the monolingual speakers in the country of origin. It has been attested among late second language learners that phonetic drift in the first language may occur toward the second language after immersion in that language (Chang, 2010; Major, 1992). As the parents of the heritage speakers, who are the major sources of Spanish input that heritage speakers receive while growing up, had resided in the U.S. for a considerable amount of time, it is possible that the Spanish spoken by the parents shows divergent patterns from the Spanish spoken in the monolingual communities of the country of origin, due to contact with English, and the heritage speakers learned these patterns from them (Otheguy, 2016; Potowski, 2013). Thus, it is important to examine the speech of heritage speakers' parents and other caregivers to have a better understanding of heritage speakers' input in Spanish. Montrul (2014) examined the structural changes with Differential Object Marking (DOM) often observed in Spanish heritage speakers, which is characterized as the omission of the obligatory use of *a* with animate, specific direct objects. By comparing

Mexican-Americans of different generations (i.e., first and second generation immigrants), Montrul (2014) found that the omission of *a* was found, not only in the second generation immigrants (i.e., heritage speakers), but also in their parents' generation. Interestingly, among the first generation immigrants, those who omitted obligatory *a* were older, acquired English later in life, and had resided in the U.S. for a longer period of time, which implies that this change does not occur until long after the first generation immigrants arrive to the U.S. Thus, with regard to this linguistic feature, it is unlikely that the quality of input that the heritage speakers received while growing up was different from the monolingual variety in Mexico. A similar line of research should be conducted on heritage speakers' phonology.

The second assumption that needs confirmation is that the heritage speakers use these divergent patterns in English as well. Although the heritage speakers in the present study were more dominant in English than Spanish, this does not necessarily mean that these patterns are present in their English. Research in foreign accent rating showed conflicting results with regard to heritage speakers' global accent in the majority language. That is, some studies, such as Flege et al. (1995) and Kupisch et al. (2014a), reported that heritage speakers do not differ from monolingual speakers in their overall accent in the majority language, while others found that heritage speakers' speech in the majority language is perceived as more accented, compared to that of monolingual speakers (Flege et al. 1999; Stangen et al., 2015; Yeni-Komshian et al., 2000). As these different findings imply, it is important to take into account individual variations when examining heritage speakers' phonology. Studies in heritage language phonology have shown that heritage speakers' phonology may vary depending on several extra-linguistic factors, such as minority language use, formal education in the minority language, and the frequency of travel to countries in which the minority language is used (Chang et al., 2008, 2011; Oh et al., 2003; Rao, 2014, 2015; Ronquest, 2012; Saadah, 2011, among others). Since presentation of group tendencies can obscure individual differences among heritage speakers (Henriksen, 2015; Rothman, 2007, 2009), it is important to consider various extra-linguistic factors in order to have a better understanding of heritage language phonology.

Appendix A Short questionnaire for screening

A.1 English version (for Spanish heritage speakers and English L2 learners of Spanish)

Name*

E-mail*

Age*

Sex*

- Male
- Female

Where were you born (country and city)?*

If you were not born in the US, at what age did you arrive to the US?

Where was your FATHER born (country and city)? *

If your FATHER was not born in the US, at what age did he arrive to the US?

What is your FATHER's native language?*

Where was your MOTHER born (country and city)? *

If your MOTHER was not born in the US, at what age did she arrive to the US?

What is your MOTHER's native language?*

Did you learn both Spanish and English before age 5?*

- Yes
- No

Which language did you learn first?*

- Spanish
- English
- Both Spanish and English at the same time

At what age did you learn English?*

At what age did you learn Spanish?*

Have you studied in a bilingual school or immersion/dual language program (a school where both Spanish and English were the languages of instruction)?*

- Yes
- No

If you answered "Yes" above, specify (at what age, for how long).

What language are you more dominant in NOW?*

- Spanish
- English
- Both Spanish and English equally

Have you lived abroad?*

- Yes
- No

If you answered "Yes" above, specify (where, how old were you, for how long).

Are you proficient in, or have you had prolonged exposure to, a language other than Spanish and English?*

- Yes
- No

If you answered "Yes" above, please specify.

Have you ever been identified as having a vision problem, hearing impairment, language disability, or learning disability?*

- Yes
- No

If you answered "Yes" above, please specify.

Comment below anything else you would like to add about your language background.

A.2 Spanish version (for Spanish monolingual native speakers)

Nombre y apellido*

Correo electrónico*

Edad*

Sexo*

- Hombre
- Mujer

Lugar de nacimiento (país y ciudad)*

¿Qué estudia usted? (si no es estudiante, ¿a qué se dedica?)*

Lengua(s) materna(s)*

Lugar de nacimiento de su padre (país y ciudad)*

Lengua(s) materna(s) de su padre*

Lugar de nacimiento de su madre (país y ciudad)*

Lengua(s) materna(s) de su madre*

¿A qué edad aprendió usted español?*

¿A qué edad aprendió usted inglés?*

¿Ha vivido usted en un país extranjero?*

- Sí
- No

Si respondió "Sí" arriba, explique con más detalle (¿a qué edad? ¿dónde? ¿por cuánto tiempo?)

¿Ha aprendido lenguas además de español e inglés?*

- Sí
- No

Si respondió "Sí" arriba, explique con más detalle (¿qué lengua? ¿por cuánto tiempo?)

¿Cuándo está disponible (fecha y hora)? Días posibles: el 2 y 3 de junio

¿Tiene un problema de visión, deficiencia o discapacidad del lenguaje, problema de audición o de aprendizaje?*

- Sí
- No

Si respondió "Sí" arriba, explique con más detalle.

Comente abajo si hay algo más que quiere añadir de su experiencia de aprendizaje de lenguas.

Appendix B Bilingual Language Profile (Birdsong et al., 2012)

B.1 English version (for Spanish heritage speakers and English L2 learners of Spanish)

I. Biographical Information

Participant number*

Age*

Sex*

- M
- F

Where did you grow up in? (City/State)*

Highest level of formal education*

- Less than high school
- High school
- Some college
- College (B.A, B.S.)
- Some graduate school
- Masters
- PhD / MD / JD
- Other:

II. Language History

In this section, we would like you to answer some factual questions about your language history. Please answer each question by selecting the appropriate answer from the drop-down menu.

1. At what age did you start learning the following languages?

At what age did you start learning ENGLISH?

At what age did you start learning SPANISH?

2. At what age did you start to feel comfortable using the following languages?

At what age did you start to feel comfortable using ENGLISH?

At what age did you start to feel comfortable using SPANISH?

3. How many years of classes (grammar, history, math, etc.) have you had in the following languages (primary school through university)?

How many years of classes (grammar, history, math, etc.) have you had in ENGLISH (primary school through university)?

How many years of classes (grammar, history, math, etc.) have you had in SPANISH (primary school through university)?

4. How many years have you spent in a country/region where the following languages are spoken?

How many years have you spent in a country/region where ENGLISH is spoken?

How many years have you spent in a country/region where SPANISH is spoken?

5. How many years have you spent in a family where the following languages are spoken?

How many years have you spent in a family where ENGLISH is spoken?

How many years have you spent in a family where SPANISH is spoken?

6. How many years have you spent in a work environment where the following languages are spoken?

How many years have you spent in a work environment where ENGLISH is spoken?

How many years have you spent in a work environment where SPANISH is spoken?

III. Language use

In this section, we would like you to answer some questions about your language use. Please answer each question by selecting the appropriate answer from the drop-down menu.

7. In an average week, what percentage of the time do you use the following languages with friends?

Total use for all languages should equal 100%.

In an average week, what percentage of the time do you use ENGLISH with friends?

In an average week, what percentage of the time do you use SPANISH with friends?

In an average week, what percentage of the time do you use OTHER LANGUAGES with friends?

8. In an average week, what percentage of the time do you use the following languages with family?

Total use for all languages should equal 100%.

In an average week, what percentage of the time do you use ENGLISH with family?

In an average week, what percentage of the time do you use SPANISH with family?

In an average week, what percentage of the time do you use OTHER LANGUAGES with family?

9. In an average week, what percentage of the time do you use the following languages at school/work?

Total use for all languages should equal 100%.

In an average week, what percentage of the time do you use ENGLISH at school/work?

In an average week, what percentage of the time do you use SPANISH at school/work?

In an average week, what percentage of the time do you use OTHER LANGUAGES at school/work?

10. When you talk to yourself, how often do you talk to yourself in the following languages?

Total use for all languages should equal 100%.

When you talk to yourself, how often do you talk to yourself in ENGLISH?

When you talk to yourself, how often do you talk to yourself in SPANISH?

When you talk to yourself, how often do you talk to yourself in OTHER LANGUAGES?

11. When you count, how often do you count in the following languages?

Total use for all languages should equal 100%.

When you count, how often do you count in ENGLISH?

When you count, how often do you count in SPANISH?

When you count, how often do you count in OTHER LANGUAGES?

IV. Language proficiency

In this section, we would like you to rate your language proficiency by giving marks from 0 to 6. Please answer each question by clicking on the appropriate button.

12. How well do you speak the following languages?

How well do you speak ENGLISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

How well do you speak SPANISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

13. How well do you understand the following languages?

How well do you understand ENGLISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

How well do you understand SPANISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

14. How well do you read the following languages?

How well do you read ENGLISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

How well do you read SPANISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

15. How well do you write the following languages?

How well do you write ENGLISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

How well do you write SPANISH?

	0	1	2	3	4	5	6	
not well at all	○	○	○	○	○	○	○	very well

V. Language attitudes

In this section, we would like you to respond to statements about language attitudes by giving marks from 0-6. Please respond to each statement by clicking on the appropriate button.

16. I feel like myself when I speak the following languages.

I feel like myself when I speak ENGLISH.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

I feel like myself when I speak SPANISH.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

17. I identify with the following cultures.

I identify with an ENGLISH-speaking culture.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

I identify with a SPANISH-speaking culture.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

18. It is important to me to use (or eventually use) the following languages like a native speaker.

It is important to me to use (or eventually use) ENGLISH like a native speaker.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

It is important to me to use (or eventually use) SPANISH like a native speaker.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

19. I want others to think I am a native speaker of the following languages.

I want others to think I am a native speaker of ENGLISH.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

I want others to think I am a native speaker of SPANISH.

	0	1	2	3	4	5	6	
disagree	○	○	○	○	○	○	○	agree

B.2 Spanish version (for Spanish monolingual native speakers)

I. Información biográfica

Número de participante*

Edad*

Género*

- M
- F

La ciudad donde creció.*

País*

Nivel más alto de formación académica completada*

- Menos de la escuela secundaria
- Escuela secundaria
- Un poco de universidad
- Universidad (diplomatura/licenciatura)
- Un poco de escuela graduada
- Máster
- Doctorado
- Other:

II. Historial lingüístico

En esta sección, nos gustaría que contestara algunas preguntas sobre su historial lingüístico. Por favor conteste a cada pregunta seleccionando la respuesta apropiada en el menú desplegable.

1. ¿A qué edad empezó a aprender las siguientes lenguas?

¿A qué edad empezó a aprender ESPAÑOL?

¿A qué edad empezó a aprender INGLÉS?

2. ¿A qué edad empezó a sentirse cómodo usando las siguientes lenguas?

¿A qué edad empezó a sentirse cómodo usando ESPAÑOL?

¿A qué edad empezó a sentirse cómodo usando INGLÉS?

3. Cuántos años de clases (gramática, historia, matemáticas, etc.) ha tenido en las siguientes lenguas (desde la escuela primaria a la universidad)?

¿Cuántos años de clases (gramática, historia, matemáticas, etc.) ha tenido en ESPAÑOL?

¿Cuántos años de clases (gramática, historia, matemáticas, etc.) ha tenido en INGLÉS?

4. ¿Cuántos años ha pasado en un país/región donde se hablan las siguientes lenguas?

¿Cuántos años ha pasado en un país/región donde se habla ESPAÑOL?

¿Cuántos años ha pasado en un país/región donde se habla INGLÉS?

5. ¿Cuántos años ha pasado en familia hablando las siguientes lenguas?

¿Cuántos años ha pasado en familia hablando ESPAÑOL?

¿Cuántos años ha pasado en familia hablando INGLÉS?

6. ¿Cuántos años ha pasado en un ambiente de trabajo donde se hablan las siguientes lenguas?

¿Cuántos años ha pasado en un ambiente de trabajo donde se habla ESPAÑOL?

¿Cuántos años ha pasado en un ambiente de trabajo donde se habla INGLÉS?

III. Uso de lenguas

En esta sección, nos gustaría que contestara algunas preguntas sobre su uso de lenguas marcando la casilla apropiada. Por favor conteste a cada pregunta seleccionando la respuesta apropiada en el menú desplegable.

7. En una semana normal, ¿qué porcentaje del tiempo usa las siguientes lenguas con sus amigos?

El total sumando las respuestas del uso de ESPAÑOL, INGLÉS y OTRAS LENGUAS debe llegar al 100%.

En una semana normal, ¿qué porcentaje del tiempo usa ESPAÑOL con sus amigos?

En una semana normal, ¿qué porcentaje del tiempo usa INGLÉS con sus amigos?

En una semana normal, ¿qué porcentaje del tiempo usa OTRAS LENGUAS con sus amigos?

8. En una semana normal, ¿qué porcentaje del tiempo usa las siguientes lenguas con su familia?

El total sumando las respuestas del uso de ESPAÑOL, INGLÉS y OTRAS LENGUAS debe llegar al 100%.

En una semana normal, ¿qué porcentaje del tiempo usa ESPAÑOL con su familia?

En una semana normal, ¿qué porcentaje del tiempo usa INGLÉS con su familia?

En una semana normal, ¿qué porcentaje del tiempo usa OTRAS LENGUAS con su familia?

9. En una semana normal, ¿qué porcentaje del tiempo usa las siguientes lenguas en la escuela/el trabajo?

El total sumando las respuestas del uso de ESPAÑOL, INGLÉS y OTRAS LENGUAS debe llegar al 100%.

En una semana normal, ¿qué porcentaje del tiempo usa ESPAÑOL en la escuela/el trabajo?

En una semana normal, ¿qué porcentaje del tiempo usa INGLÉS en la escuela/el trabajo?

En una semana normal, ¿qué porcentaje del tiempo usa OTRAS LENGUAS en la escuela/el trabajo?

10. Cuando se habla a usted mismo, ¿con qué frecuencia se habla a sí mismo en las siguientes lenguas?

El total sumando las respuestas del uso de ESPAÑOL, INGLÉS y OTRAS LENGUAS debe llegar al 100%.

Cuando se habla a usted mismo, ¿con qué frecuencia se habla a sí mismo en ESPAÑOL?

Cuando se habla a usted mismo, ¿con qué frecuencia se habla a sí mismo en INGLÉS?

Cuando se habla a usted mismo, ¿con qué frecuencia se habla a sí mismo en OTRAS LENGUAS?

11. Cuando hace cálculos contando, ¿con qué frecuencia cuenta en las siguientes lenguas?

El total sumando las respuestas del uso de ESPAÑOL, INGLÉS y OTRAS LENGUAS debe llegar al 100%.

Cuando hace cálculos contando, ¿con qué frecuencia cuenta en ESPAÑOL?

Cuando hace cálculos contando, ¿con qué frecuencia cuenta en INGLÉS?

Cuando hace cálculos contando, ¿con qué frecuencia cuenta en OTRAS LENGUAS?

IV. Competencia lingüística

En esta sección, nos gustaría que considerara su competencia de lengua marcando la casilla de 0 a 6. Por favor conteste a cada pregunta seleccionando el botón apropiado.

12. ¿Cómo habla en las siguientes lenguas?

¿Cómo habla en ESPAÑOL?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

¿Cómo habla en INGLÉS?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

13. ¿Cómo entiende en las siguientes lenguas?

¿Cómo entiende en ESPAÑOL?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

¿Cómo entiende en INGLÉS?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

14. ¿Cómo lee en las siguientes lenguas?

¿Cómo lee en ESPAÑOL?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

¿Cómo lee en INGLÉS?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

15. ¿Cómo escribe en las siguientes lenguas?

¿Cómo escribe en ESPAÑOL?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

¿Cómo escribe en INGLÉS?

	0	1	2	3	4	5	6	
no muy bien	○	○	○	○	○	○	○	muy bien

V. Actitudes lingüísticas

En esta sección, nos gustaría que contestara a las siguientes afirmaciones sobre actitudes lingüísticas. Por favor responda a cada frase seleccionando el botón apropiado.

16. Me siento "yo mismo" cuando hablo en las siguientes lenguas

Me siento “yo mismo” cuando hablo en ESPAÑOL.

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy de acuerdo

Me siento “yo mismo” cuando hablo en INGLÉS.

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy de acuerdo

17. Yo me identifico con las siguientes culturas.

Me identifico con una cultura HISPANOHABLANTE

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy de acuerdo

Me identifico con una cultura ANGLOHABLANTE

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy de acuerdo

18.Es importante para mi usar (o llegar a usar) las siguientes lenguas como un hablante nativo.

Es importante para mi usar (o llegar a usar) ESPAÑOL como un hablante nativo.

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy de acuerdo

Es importante para mi usar (o llegar a usar) INGLÉS como un hablante nativo.

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy acuerdo

19. Quiero que los demás piensen que soy un hablante nativo de las siguientes lenguas

Quiero que los demás piensen que soy un hablante nativo de ESPAÑOL

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy de acuerdo

Quiero que los demás piensen que soy un hablante nativo de INGLÉS.

	0	1	2	3	4	5	6	
no estoy de acuerdo	○	○	○	○	○	○	○	estoy de acuerdo

Appendix C Additional questionnaire for Spanish heritage speakers

Participant number*

How often did you use SPANISH? (choose one in each row)

	Never (=1)	2	3	4	5	6	Always (=7)
age 0-3	○	○	○	○	○	○	○
age 4-5 (preschool)	○	○	○	○	○	○	○
age 6-10 (elementary school)	○	○	○	○	○	○	○
age 11-13 (middle school)	○	○	○	○	○	○	○
age 14-17 (high school)	○	○	○	○	○	○	○
age 18-	○	○	○	○	○	○	○

How often did you use ENGLISH? (choose one in each row)

	Never (=1)	2	3	4	5	6	Always (=7)
age 0-3	○	○	○	○	○	○	○
age 4-5 (preschool)	○	○	○	○	○	○	○
age 6-10 (elementary school)	○	○	○	○	○	○	○
age 11-13 (middle school)	○	○	○	○	○	○	○
age 14-17 (high school)	○	○	○	○	○	○	○
age 18-	○	○	○	○	○	○	○

Rate your father's proficiency in SPANISH.

	1	2	3	4	5	6	7	
Very poor	○	○	○	○	○	○	○	Native speaker command

Rate your father's proficiency in ENGLISH.

	1	2	3	4	5	6	7	
Very poor	○	○	○	○	○	○	○	Native speaker command

Rate your mother's proficiency in SPANISH.

	1	2	3	4	5	6	7	
Very poor	○	○	○	○	○	○	○	Native speaker command

Rate your mother's proficiency in ENGLISH.

	1	2	3	4	5	6	7	
Very poor	○	○	○	○	○	○	○	Native speaker command

What is your father's highest level of formal education?

- ○ Lower than high school
- ○ High school
- ○ Some college
- ○ College (B.A., B.S.)
- ○ Graduate school

What is your mother's highest level of formal education?

- ○ Lower than high school
- ○ High school
- ○ Some college
- ○ College (B.A., B.S.)
- ○ Graduate school

What does your father do for living?

What does your mother do for living?

I speak to my parents in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I speak to my parents in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My parents speak to me in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My parents speak to me in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How often do you interact with your parents?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Very frequently

Do you have siblings?

- ○ Yes
- ○ No

If you answered "Yes" above, specify (how many, older or younger?).

I speak to my siblings (brothers and sisters) in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I speak to my siblings (brothers and sisters) in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My siblings (brothers and sisters) speak to me in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My siblings (brothers and sisters) speak to me in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How often do you interact with your siblings (brothers and sisters)?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Very frequently

I speak to my relatives (uncles/aunts, grandparents) in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I speak to my relatives (uncles/aunts, grandparents) in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My relatives (uncles/aunts, grandparents) speak to me in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My relatives (uncles/aunts, grandparents) speak to me in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How often do you interact with your relatives (uncles/aunts, grandparents)?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Very frequently

I speak to my friends at school in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I speak to my friends at school in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My friends at school speak to me in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My friends at school speak to me in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How often do you interact with your friends at school?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Very frequently

I speak to my friends in my neighborhood in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I speak to my friends in my neighborhood in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My friends in my neighborhood speak to me in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

My friends in my neighborhood speak to me in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How often do you interact with your friends in your neighborhood?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Very frequently

I speak to people in my community (grocery stores, church) in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I speak to people in my community (grocery stores, church) in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

People in my community (grocery stores, church) speak to me in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

People in my community (grocery stores, church) speak to me in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How often do you interact with the people in your community (grocery stores, church)?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Very frequently

I watch TV/movies in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I watch TV/movies in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I listen to radio/music in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I listen to radio/music in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I read books/newspapers in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I read books/newspapers in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I write e-mails/text messages in SPANISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I write e-mails/text messages in ENGLISH.

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

In your perception, how much foreign accent do you have in SPANISH?

	1	2	3	4	5	6	7	
None	○	○	○	○	○	○	○	Pervasive

In your perception, how much foreign accent do you have in ENGLISH?

	1	2	3	4	5	6	7	
None	○	○	○	○	○	○	○	Pervasive

How frequently, do others identify you as a non-native speaker of Spanish based on your accent in SPANISH?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

How frequently, do others identify you as a non-native speaker of English based on your accent in ENGLISH?

	1	2	3	4	5	6	7	
Never	○	○	○	○	○	○	○	Always

I am simply a Mexican who lives in North America.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I keep Mexican and American cultures separate.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I feel Mexican-American.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I feel part of a combined culture.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I am conflicted between the American and Mexican ways of doing things.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I feel like someone moving between two cultures.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I feel caught between the Mexican and American cultures.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

I don't feel trapped between the Mexican and American cultures.

	1	2	3	4	5	6	7	
Strongly disagree	○	○	○	○	○	○	○	Strongly agree

If there were questions that were not clear, please comment below.

Appendix D List of items used for picture-naming task

item (Sp)	rank[a]	freq. level[a]	picture[b]	etype[b]	elex1[b]	efreq[b]	item (Eng)	etype[b]	elex1[b]	efreq[b]
casa	116	1	obj213	1	1	7.482	house	2	0.98	6.409
mano	150	1	obj195	1	1	7.003	hand	2	0.98	6.586
ojo	247	1	obj148	1	1	5.263	eye	2	0.98	6.261
libro	253	1	obj051	1	1	6.351	book	1	1	6.075
puerta	354	1	obj132	1	1	6.711	door	1	1	5.958
pie	386	1	obj166	1	1	5.958	foot	2	0.98	5.79
corazón	649	1	obj201	1	1	5.94	heart	1	1	5.106
sol	686	1	obj431	1	1	6.36	sun	1	1	5.03
árbol	833	1	obj469	1	1	4.71	tree	1	1	5.257
perro	939	1	obj128	1	1	5.416	dog	1	1	4.754
flor	950	1	obj163	3	0.96	4.779	flower	1	1	4.543
hoja	1000	1	obj236	1	1	4.277	leaf	1	1	4.407
tren	1220	2	obj467	2	0.92	4.71	train	1	1	4.407
dedo	1248	2	obj153	2	0.96	4.883	finger	2	0.98	4.82
caja	1258	2	obj056	1	1	4.796	box	1	1	4.635
silla	1307	2	obj089	1	1	4.89	chair	1	1	4.92
rueda	1352	2	obj503	2	0.88	3.912	wheel	1	1	3.807
puente	1353	2	obj062	2	0.98	4.419	bridge	2	0.98	4.205
pan	1392	2	obj060	2	0.96	4.934	bread	2	0.98	4.317
gato	1412	2	obj086	1	1	4.949	cat	3	0.96	4.22
reloj	1685	2	obj099	1	1	4.963	clock	2	0.98	3.689
hueso	1695	2	obj050	1	1	3.807	bone	1	1	4.248
huevo	1900	2	obj144	1	1	4.007	egg	2	0.98	4.466
zapato	1932	2	obj393	2	0.98	3.526	shoe	1	1	4.382
bandera	2021	3	obj159	2	0.96	4.174	flag	1	1	3.296
vela	2287	3	obj075	2	0.98	4.007	candle	1	1	2.833
pluma	2444	3	obj151	2	0.96	3.989	feather	2	0.98	3.091
pantalón	2489	3	obj298	2	0.82	4.127	pants	3	0.9	2.833
cajón	2620	3	obj134	3	0.96	3.97	drawer	1	1	3.219
bota	2735	3	obj052	3	0.96	2.303	boot	3	0.9	3.689
león	2756	3	obj246	1	1	4.025	lion	1	1	3.258
corona	2772	3	obj117	1	1	3.761	crown	3	0.94	3.219
manzana	2853	3	obj009	1	1	3.178	apple	1	1	3.434
cuchillo	2945	3	obj228	2	0.96	3.871	knife	1	1	3.807
mosca	2956	3	obj165	3	0.94	3.466	fly	3	0.9	3.611
campana	3000	3	obj040	1	1	3.555	bell	1	1	3.332
casco	3033	4	obj204	1	1	3.367	helmet	2	0.96	2.639
enfermera	3122	4	obj284	2	0.98	3.434	nurse	3	0.96	3.912
queso	3182	4	obj090	1	1	3.219	cheese	1	1	3.466

item (Sp)	rank[a]	freq. level[a]	picture[b]	etype[b]	elex1[b]	efreq[b]	item (Eng)	etype[b]	elex1[b]	efreq[b]
corbata	3238	4	obj453	1	1	3.951	tie	2	0.98	3.555
lápiz	3296	4	obj311	1	1	3.045	pencil	1	1	2.996
sierra	3390	4	obj376	1	1	1.609	saw	1	1	0.693
guante	3392	4	obj183	1	1	0	glove	1	1	2.996
globo	3408	4	obj025	1	1	3.367	balloon	1	1	1.946
máscara	3449	4	obj258	1	1	3.401	mask	2	0.98	3.045
uva	3682	4	obj186	2	0.96	2.89	grapes	2	0.9	0
cuna	3726	4	obj115	2	0.98	3.401	crib	3	0.84	0.693
mariposa	3883	4	obj067	1	1	2.773	butterfly	1	1	2.398
tambor	4057	5	obj138	1	1	3.135	drum	2	0.8	2.833
conejo	4316	5	obj343	1	1	2.89	rabbit	2	0.84	2.996
clavo	4363	5	obj277	2	0.98	2.89	nail	1	1	3.258
bastón	4454	5	obj076	1	1	3.367	cane	3	0.96	2.398
oso	4555	5	obj034	2	0.98	3.85	bear	2	0.82	2.833
pala	4601	5	obj395	1	1	2.485	shovel	1	1	1.609
martillo	4688	5	obj193	1	1	2.773	hammer	1	1	2.485
araña	4717	5	obj416	1	1	2.944	spider	1	1	2.079
flecha	4781	5	obj012	2	0.98	2.773	arrow	2	0.98	2.773
pato	4786	5	obj139	3	0.96	2.398	duck	3	0.96	0
cometa	4857	5	obj227	1	1	1.099	kite	1	1	1.792
cohete	4937	5	obj359	2	0.98	3.045	rocket	3	0.9	2.708

[a]: data taken from Davies (2006)
[b]: data taken from International Picture Naming Project (IPNP) database (Szekely et al., 2004)
freq.level: frequency level based on rank
picture: picture item number
etype: number of alternative names
elex1: percent name agreement
efreq: log frequency

Appendix E List of stimuli used for experiments

E.1 Experiment 1 (Perception of lexical stress)

Target items

Nuclear position (60 pairs)	Pre-nuclear position (60 pairs)	Unaccented context (60 pairs)
Canto/ó.	*Canto/ó la balada.*	*¿Dónde canto/ó con micrófono?*
Las marco/ó.	*Marco/ó la diferencia.*	*¿Cómo las marco/ó correctamente?*
Paso/ó.	*Paso/ó la aduana.*	*¿Cuándo paso/ó por la plaza?*
La saco/ó.	*Saco/ó la basura.*	*¿De dónde saco/ó caramelos?*
La tapo/ó.	*Tapo/ó la nariz.*	*¿Por qué la tapo/ó con un corcho?*
Los mando/ó.	*Mando/ó la dirección.*	*¿Por qué les mando/ó tres cuadernos?*
Los aclaro/ó.	*Aclaro/ó la vajilla.*	*¿Cómo aclaro/ó tus respuestas?*
Los espanto/ó.	*Espanto/ó la abeja.*	*¿Por qué lo espanto/ó con las manos?*
Descanso/ó.	*Descanso/ó la muñeca.*	*¿Dónde descanso/ó por la tarde?*
Los levanto/ó.	*Levanto/ó la bebida.*	*¿Cómo levanto/ó tu mochila?*
Los reparo/ó.	*Reparo/ó la batidora.*	*¿Dónde reparo/ó computadoras?*
Trabajo/ó.	*Trabajo/ó muy motivado.*	*¿Cuándo trabajo/ó por Laurita?*
Los dejo/ó.	*Dejo/ó la mochila.*	*¿Dónde dejo/ó sus cuadernos?*
Llego/ó.	*Llego/ó muy despacio.*	*¿Adónde llego/ó por la tarde?*
Las lleno/ó.	*Lleno/ó la nevera.*	*¿Por qué la lleno/ó con manzanas?*
Las llevo/ó.	*Llevo/ó la maleta.*	*¿Cuándo llevo/ó tus maletas?*
Ceno/ó.	*Ceno/ó muy ligero.*	*¿Cuándo ceno/ó con Manuela?*
La seco/ó.	*Seco/ó la bufanda.*	*¿Cómo seco/ó sus cazuelas?*
Contesto/ó.	*Contesto/ó la demanda.*	*¿Cómo contesto/ó con confianza?*
Enseño/ó.	*Enseño/ó la gramática.*	*¿Cuándo enseño/ó con pasión?*
Espero/ó.	*Espero/ó la noticia.*	*¿Cómo espero/ó con paciencia?*
Lo Intento/ó.	*Intento/ó manejar bien.*	*¿Cómo intento/ó manejar bien?*
Manejo/ó.	*Manejo/ó la motocicleta.*	*¿Cómo manejo/ó con cuidado?*
Regreso/ó.	*Regreso/ó meditando.*	*¿Cuándo regreso/ó con Manuela?*
Grito/ó.	*Grito/ó muy molesto.*	*¿Dónde grito/ó por la noche?*
Las miro/ó.	*Miro/ó la ventana.*	*¿Cuándo miro/ó telenovelas?*
La pico/ó.	*Pico/ó la batata.*	*¿Cómo pico/ó como gallinas?*
Pinto/ó.	*Pinto/ó la habitación.*	*¿Cuándo pinto/ó rascacielos?*
Silbo/ó.	*Silbo/ó la melodía.*	*¿Cómo silbo/ó con los labios?*
La tiro/ó.	*Tiro/ó la moneda.*	*¿Dónde tiro/ó los calcetines?*
Los animo/ó.	*Animo/ó la navidad.*	*¿Cómo animo/ó tu esperanza?*
Los alquilo/ó.	*Alquilo/ó la habitación.*	*¿Cuándo alquilo/ó la caravana?*
Los aviso/ó.	*Aviso/ó la llegada.*	*¿Cuándo aviso/ó con precaución?*
Camino/ó.	*Camino/ó muy lentamente.*	*¿Adónde camino/ó con cuidado?*
Les explico/ó.	*Explico/ó la metáfora.*	*¿Cómo explico/ó con certeza?*
Los termino/ó.	*Termino/ó la entrevista.*	*¿Cuándo termino/ó sus tareas?*
Las compro/ó.	*Compro/ó la lotería.*	*¿Dónde compro/ó palomitas?*
Las corto/ó.	*Corto/ó la madera.*	*¿Cómo corto/ó tangerinas?*
Los logro/ó.	*Logro/ó modificarlo.*	*¿Cómo logro/ó superarlo?*
Los noto/ó.	*Noto/ó la envidia.*	*¿Cómo noto/ó tu defecto?*

Nuclear position (60 pairs)	Pre-nuclear position (60 pairs)	Unaccented context (60 pairs)
La toco/ó.	*Toco/ó la guitarra.*	*¿Por qué la toco/ó con cuidado?*
La tomo/ó.	*Tomo/ó la decisión.*	*¿Por qué la tomo/ó por la noche?*
Las apoyo/ó.	*Apoyo/ó la decisión.*	*¿Cómo apoyo/ó tu decisión?*
Las arrojo/ó.	*Arrojo/ó la manzana.*	*¿Por qué la arrojo/ó por el puente?*
Los devoro/ó.	*Devoro/ó la manzana.*	*¿Cómo lo devoro/ó con rapidez?*
Lo perdono/ó.	*Perdono/ó la mentira.*	*¿Por qué lo perdono/ó sin problema?*
Las añoro/ó.	*Añoro/ó la libertad.*	*¿Cuándo añoro/ó tu presencia?*
Los adorno/ó.	*Adorno/ó la galleta.*	*¿Cómo adoro/ó tu hogar?*
Juzgo/ó.	*Juzgo/ó normalmente.*	*¿Por qué lo juzgo/ó sin evidencia?*
Los busco/ó.	*Busco/ó la leyenda.*	*¿Por qué lo busco/ó por teléfono?*
Lo dudo/ó.	*Dudo/ó la verdad.*	*¿Por qué lo dudo/ó sin pensar?*
Sudo/ó.	*Sudo/ó muy obviamente.*	*¿Dónde sudo/ó por el calor?*
Juro/ó.	*Juro/ó liberarlo.*	*¿Por qué lo juro/ó sin motivo?*
Fumo/ó.	*Fumo/ó normalmente.*	*¿Cuándo fumo/ó sin comer?*
Las anudo/ó.	*Anudo/ó la bufanda.*	*¿Cómo anudo/ó tu corbata?*
Dibujo/ó.	*Dibujo/ó la bandera.*	*¿Dónde dibujo/ó por la calle?*
Las escucho/ó.	*Escucho/ó la bachata.*	*¿Cuándo la escucho/ó por internet?*
Mascullo/ó.	*Mascullo/ó la oración.*	*¿Cuándo mascullo/ó sin parar?*
Pregunto/ó.	*Pregunto/ó la identidad.*	*¿Por qué lo pregunto/ó con curiosidad?*
Circulo/ó.	*Circulo/ó la noticia.*	*¿Cuándo circulo/ó por la ciudad?*

Filler items

Nuclear position (40 items)	Pre-nuclear position (40 items)	Unaccented context (40 items)
Los haces.	*Bebes el batido.*	*¿Cuándo pides un favor?*
Creces.	*Sigues el manual.*	*¿Cuándo temes por el futuro?*
Los metes.	*Subes el volumen.*	*¿Dónde pones las pilas?*
La comes.	*Vives en Noruega.*	*¿Cómo rompes las reglas?*
La coses.	*Sabes el número.*	*¿Cuándo corres en el gimnasio?*
Aprendes.	*Supones el peligro.*	*¿Cómo respondes las preguntas?*
Lo cometes.	*Decides ayudarlo.*	*¿Cómo ofendes a tus padres?*
Lo comprendes.	*Recibes una llamada.*	*¿Adónde insistes en caminar?*
Lo conoces.	*Propones un objetivo.*	*¿Dónde escondes las evidencias?*
Los escoges.	*Escribes un ensayo.*	*¿Cómo prometes ser bueno?*
Los muestras.	*Juegas al baloncesto.*	*¿Cuándo cierras la puerta?*
Lo cueces.	*Vuelves a Holanda.*	*¿Cuándo quieres una manzana?*
Lo sientes.	*Puedes olvidarlo.*	*¿Cómo cuentas los números?*
Los viertes.	*Mueves el armario.*	*¿Cuándo piensas en tu niñez?*
Duermes.	*Vienes a ayudarme.*	*¿Cuándo tienes una inspiración?*
Las aprietas.	*Devuelves la llamada.*	*¿Cómo obtienes lo que quieres?*
Entiendes.	*Disuelves la bebida.*	*¿Cómo sostienes la esperanza?*
Los defiendes.	*Envuelves el regalo.*	*¿Cuándo extiendes el contrato?*
Empiezas.	*Mantienes el legado.*	*¿Cuándo detienes el servicio?*
Las enciendes.	*Te refieres a Manuela.*	*¿Cómo encuentras una excusa?*
Nací.	*Abrí la nevera.*	*¿Por qué corrí tan despacio?*
Lo sentí.	*Bebí la limonada.*	*¿Por qué lo cubrí con maquillaje?*

Nuclear position (40 items)	Pre-nuclear position (40 items)	Unaccented context (40 items)
Lo comí.	*Viví en Granada.*	*¿Dónde dormí profundamente?*
Lo cumplí.	*Moví la lavadora.*	*¿Adónde salí con mucha prisa?*
Los rompí.	*Volví a llamarlo.*	*¿Cuándo perdí tu confianza?*
Aprendí.	*Conseguí el dinero.*	*¿Cuándo lo conocí por primera vez?*
Entendí	*Decidí abandonarlo.*	*¿Cuándo preferí no hacerlo?*
Comprendí	*Le devolví la mirada.*	*¿Cómo encendí tu cigarrillo?*
Lo prometí	*Escribí una novela.*	*¿Dónde escondí mi celular?*
Respondí	*Recibí un abrigo.*	*¿Cuándo insistí en el tema?*
Lo perdió.	*Murió en Madrid.*	*¿Cómo mordió tus zapatos?*
Sufrió.	*Olió a limón.*	*¿Cuándo mostró su riqueza?*
Pensó.	*Subió al mirador.*	*¿Cuándo se sintió traicionado?*
Lo cubrió.	*Durmió en el avión.*	*¿Cómo vertió la basura?*
Nació.	*Mordió al ladrón.*	*¿Dónde vendió cacahuates?*
Atendió.	*Devolvió la ayuda.*	*¿Cuándo apretó sus dientes?*
Descendió.	*Envolvió la muñeca.*	*¿Cuándo encendió su teléfono?*
Empezó.	*Recordó la melodía.*	*¿Cómo encontró su camino?*
Entendió.	*Resolvió el enigma.*	*¿Cuándo extendió su contrato?*
Los defendió.	*Consiguió un empleo.*	*¿Cuándo prefirió caminar allí?*

E.2 Experiment 2 (Production of lexical stress)

Target items

Nuclear position (20 pairs)	Pre-nuclear position (20 pairs)	Unaccented context (20 pairs)
La saco/ó.	*Saco/ó la basura.*	*¿De dónde saco/ó caramelos?*
La tapo/ó.	*Tapo/ó la nariz.*	*¿Por qué la tapo/ó con un corcho?*
Los reparo/ó.	*Reparo/ó la batidora.*	*¿Dónde reparo/ó computadoras?*
Trabajo/ó.	*Trabajo/ó muy motivado.*	*¿Cuándo trabajo/ó por Laurita?*
Los dejo/ó.	*Dejo/ó la mochila.*	*¿Dónde dejo/ó sus cuadernos?*
Ceno/ó.	*Ceno/ó muy ligero.*	*¿Cuándo ceno/ó con Manuela?*
Espero/ó.	*Espero/ó la noticia.*	*¿Cómo espero/ó con paciencia?*
Manejo/ó.	*Manejo/ó la motocicleta.*	*¿Cómo manejo/ó con cuidado?*
Las miro/ó.	*Miro/ó la ventana.*	*¿Cuándo miro/ó telenovelas?*
La tiro/ó.	*Tiro/ó la moneda.*	*¿Dónde tiro/ó los calcetines?*
Los animo/ó.	*Animo/ó la navidad.*	*¿Cómo animo/ó tu esperanza?*
Camino/ó.	*Camino/ó muy lentamente.*	*¿Adónde camino/ó con cuidado?*
Los noto/ó.	*Noto/ó la envidia.*	*¿Cómo noto/ó tu defecto?*
La tomo/ó.	*Tomo/ó la decisión.*	*¿Por qué la tomo/ó por la noche?*
Los devoro/ó.	*Devoro/ó la manzana.*	*¿Cómo lo devoro/ó con rapidez?*
Lo perdono/ó.	*Perdono/ó la mentira.*	*¿Por qué lo perdono/ó sin problema?*
Juro/ó.	*Juro/ó liberarlo.*	*¿Por qué lo juro/ó sin motivo?*
Fumo/ó.	*Fumo/ó normalmente.*	*¿Cuándo fumo/ó sin comer?*
Las anudo/ó.	*Anudo/ó la bufanda.*	*¿Cómo anudo/ó tu corbata?*
Dibujo/ó.	*Dibujo/ó la bandera.*	*¿Dónde dibujo/ó por la calle?*

Filler items

Nuclear position (20 items)	Pre-nuclear position (20 items)	Unaccented context (20 items)
Sale.	*Sale corriendo.*	*¿Adónde sale con much prisa?*
Salí.	*Salí corriendo.*	*¿Adónde salí con mucha prisa?*
Lo reparte.	*Reparte los regalos.*	*¿Cuándo reparte los exámenes?*
Lo repartí.	*Repartí los regalos.*	*¿Cuándo repartí los exámenes?*
Lo teme.	*Teme a la muerte.*	*¿Cuándo teme por el futuro?*
Lo temí.	*Temí a la muerte.*	*¿Cuándo temí por el futuro?*
Lo comete.	*Comete un error.*	*¿Cuándo comete errores?*
Lo cometí.	*Cometí un error.*	*¿Cuándo cometí errores?*
Vive.	*Vive en Noruega.*	*¿Cómo vive en la montaña?*
Viví.	*Viví en Noruega.*	*¿Cómo viví en la montaña?*
Lo decide.	*Decide ayudarlo.*	*¿Dónde decide comprarlo?*
Lo decidí.	*Decidí ayudarlo.*	*¿Dónde decidí compararlo?*
La Come.	*Come empanadas.*	*¿Cuándo come chocolate?*
La Comí.	*Comí empanadas.*	*¿Cuándo comí chocolate?*
Lo conoce.	*Conoce a Susana.*	*¿Cómo conoce a Guillermo?*
Lo conocí.	*Conocí a Susana.*	*¿Cómo conocí a Guillermo?*
Sube.	*Sube el volumen.*	*¿Cuándo sube a la montaña?*
Subí.	*Subí el volumen.*	*¿Cuándo subí a la montaña?*
Discute.	*Discute sobre el tema.*	*¿Cuándo discute con su padre?*
Discutí.	*Discutí sobre el tema.*	*¿Cuándo discutí con mi padre?*

E.3 Experiment 3 (Perception of nuclear stress)

Target items: Focus on subject (20 pairs)

$[_F$S]Cl+V	Cl+V$[_F$S]
ADRIANA lo adornó.	*Lo adornó ADRIANA.*
AURELIANO lo apagó.	*Lo apagó AURELIANO.*
GUILLERMO lo ayudó.	*Lo ayudó GUILLERMO.*
RONALDO lo borró.	*Lo borró RONALDO.*
AMANDA lo cerró.	*Lo cerró AMANDA.*
ESMERALDA lo construyó.	*Lo construyó ESMERALDA.*
LEONARDO lo diseñó.	*Lo diseñó LEONARDO.*
MAGDALENA lo esperó.	*Lo esperó MAGDALENA.*
BERNARDO lo grabó.	*Lo grabó BERNARDO.*
DOLORES lo lavó.	*Lo lavó DOLORES.*
EMILIO lo leyó.	*Lo leyó EMILIO.*
EMILIA lo llamó.	*Lo llamó EMILIA.*
EDGARDO lo miró.	*Lo miró EDGARDO.*
MANOLO lo observó.	*Lo observó MANOLO.*
OLIVIA lo ofendió.	*Lo ofendió OLIVIA.*

[$_F$S]Cl+V	Cl+V[$_F$S]
DANIELA lo pagó.	*Lo pagó DANIELA.*
LILIANA lo preparó.	*Lo preparó LILIANA.*
RAYMUNDO lo reparó.	*Lo reparó RAYMUNDO.*
VIVIANA lo saludó.	*Lo saludó VIVIANA.*
YADIRA lo terminó.	*Lo terminó YADIRA.*

Filler items: Focus on verb (20 pairs)

S+Cl[$_F$V]	Cl[$_F$V]S
Guillermo lo AGARRÓ.	*Lo AGARRÓ Guillermo.*
Esmeralda lo ANUDÓ.	*Lo ANUDÓ Esmeralda.*
Magdalena lo APOYÓ.	*Lo APOYÓ Magdalena.*
Leonardo lo CAPTURÓ.	*Lo CAPTURÓ Leonardo.*
Aureliano lo CARGÓ.	*Lo CARGÓ Aureliano.*
Liliana lo CASTIGÓ.	*Lo CASTIGÓ Liliana.*
Manolo lo CULTIVÓ.	*Lo CULTIVÓ Manolo.*
Manuela lo ENSEÑÓ.	*Lo ENSEÑÓ Manuela.*
Osvaldo lo EXAMINÓ.	*Lo EXAMINÓ Osvaldo.*
Mariana lo EXTRAÑÓ.	*Lo EXTRAÑÓ Mariana.*
Raymundo lo INSTALÓ.	*Lo INSTALÓ Raymundo.*
Adriana lo LLEVÓ.	*Lo LLEVÓ Adriana.*
Rodrigo lo MANDÓ.	*Lo MANDÓ Rodrigo.*
Bernardo lo PEGÓ.	*Lo PEGÓ Bernardo.*
Daniela lo QUEBRÓ.	*Lo QUEBRÓ Daniela.*
Edgardo lo QUEMÓ.	*Lo QUEMÓ Edgardo.*
Noelia lo RECORDÓ.	*Lo RECORDÓ Noelia.*
Mariana lo SALVÓ.	*Lo SALVÓ Mariana.*
Emilio lo TIRÓ.	*Lo TIRÓ Emilio.*
Osvaldo lo TRAGÓ.	*Lo TRAGÓ Osvaldo.*

E.4 Experiment 4 (Production of nuclear stress)

Target items (12 triplets)

Context	Question (by Focus Type)	
Adriana lo adornó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo adornó?*
	V	*Perdón no te oí. ¿Qué hizo Adriana?*
Aureliano lo apagó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo apagó?*
	V	*Perdón no te oí. ¿Qué hizo Aureliano?*

Context		Question (by Focus Type)
Guillermo lo ayudó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo ayudó?*
	V	*Perdón no te oí. ¿Qué hizo Guillermo?*
Esmeralda lo construyó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo construyó?*
	V	*Perdón no te oí. ¿Qué hizo Esmeralda?*
Leonardo lo diseñó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo diseñó?*
	V	*Perdón no te oí. ¿Qué hizo Leonardo?*
Magdalena lo esperó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo esperó?*
	V	*Perdón no te oí. ¿Qué hizo Magdalena?*
Bernardo lo grabó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo grabó?*
	V	*Perdón no te oí. ¿Qué hizo Bernardo?*
Emilio lo leyó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo leyó?*
	V	*Perdón no te oí. ¿Qué hizo Emilio?*
Edgardo lo miró.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo miró?*
	V	*Perdón no te oí. ¿Qué hizo Edgardo?*
Daniela lo pagó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo pagó?*
	V	*Perdón no te oí. ¿Qué hizo Daniela?*
Liliana lo preparó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo preparó?*
	V	*Perdón no te oí. ¿Qué hizo Liliana?*
Raymundo lo reparó.	BF	*Perdón no te oí. ¿Qué pasó?*
	S	*Perdón no te oí. ¿Quién lo reparó?*
	V	*Perdón no te oí. ¿Qué hizo Raymundo?*

(BF: broad focus, S: focus on subject, V: focus on verb)

Filler items (36 items)

Context		Question (by Focus Type)
Adriana bebió leche.	BF	*Perdón no te oí. ¿Qué pasó?*
Osvaldo movió el armario.	BF	*Perdón no te oí. ¿Qué pasó?*
Esmeralda devolvió el libro.	BF	*Perdón no te oí. ¿Qué pasó?*
Rodrigo recibió una beca.	BF	*Perdón no te oí. ¿Qué pasó?*
Manuela escribió una carta.	BF	*Perdón no te oí. ¿Qué pasó?*
Guillermo descubrió un secreto.	BF	*Perdón no te oí. ¿Qué pasó?*
Adriana borró la pizarra.	BF	*Perdón no te oí. ¿Qué pasó?*
Esmeralda ofendió a sus padres.	BF	*Perdón no te oí. ¿Qué pasó?*
Rodrigo obersvó la clase.	BF	*Perdón no te oí. ¿Qué pasó?*

Context		Question (by Focus Type)
Osvaldo terminó sus tareas.	V	*Perdón no te oí. ¿Qué hizo Osvaldo?*
Guillermo lavó los platos.	V	*Perdón no te oí. ¿Qué hizo Guillermo?*
Manuela saludó a Liliana.	V	*Perdón no te oí. ¿Qué hizo Manuela?*
Adriana llevó una maleta.	V	*Perdón no te oí. ¿Qué hizo Adriana?*
Manuela enseñó griego.	V	*Perdón no te oí. ¿Qué hizo Manuela?*
Osvaldo examinó el agua.	V	*Perdón no te oí. ¿Qué hizo Osvaldo?*
Guillermo agarró un abrigo.	V	*Perdón no te oí. ¿Qué hizo Guillermo?*
Rodrigo mandó una carta.	V	*Perdón no te oí. ¿Qué hizo Rodrigo?*
Esmeralda anudó su corbata.	V	*Perdón no te oí. ¿Qué hizo Esmeralda?*
Edgardo regresó en la noche con Manuela.	PP	*Perdón no te oí. ¿Cuándo regresó con Manuela?*
En la fiesta Leonardo cantó con micrófono.	PP	*Perdón no te oí. ¿Dónde cantó con micrófono?*
En la mañana Emilio llevó mis maletas.	PP	*Perdón no te oí. ¿Cuándo llevó tus maletas?*
En el mercado Leonardo compró palomitas.	PP	*Perdón no te oí. ¿Dónde compró palomitas?*
Adriana caminó a su habitación con cuidado.	PP	*Perdón no te oí. ¿Adónde caminó con cuidado?*
En la plaza Guillermo vendió cacahuates.	PP	*Perdón no te oí. ¿Dónde vendió cacahuates?*
Con un cuchillo Esmeralda cortó tangerinas.	PP	*Perdón no te oí. ¿Cómo cortó tangerinas?*
Anoche Liliana terminó sus tareas.	PP	*Perdón no te oí. ¿Cuándo terminó sus tareas?*
Raymundo sudó en el aula por el calor.	PP	*Perdón no te oí. ¿Dónde sudó por el calor?*
A las dos Bernardo cenó con Manuela.	PP	*Perdón no te oí. ¿Cuándo cenó con Manuela?*
De su mochila Raymundo sacó caramelos.	PP	*Perdón no te oí. ¿De dónde sacó caramelos?*
El pasado verano Aureliano trabajó por Laurita.	PP	*Perdón no te oí. ¿Cuándo trabajó por Laurita?*
El lunes Edgardo extendió su contrato.	PP	*Perdón no te oí. ¿Cuándo extendió su contrato?*
Después de la película Bernardo encendió su teléfono.	PP	*Perdón no te oí. ¿Cuándo encendió su teléfono?*
En la mañana Adriana pasó por la plaza.	PP	*Perdón no te oí. ¿Cuándo pasó por la plaza?*
Ayer Guillermo dejó sus cuadernos en casa.	PP	*Perdón no te oí. ¿Cuándo dejó sus cuadernos?*
Magdalena descansó en un café por la tarde.	PP	*Perdón no te oí. ¿Dónde descansó por la tarde?*
Muy fácilmente Liliana encontró su camino.	PP	*Perdón no te oí. ¿Cómo encontró su camino?*

BF: broad focus
V: focus on verb
PP: focus on prepositional phrase

Appendix F Acoustic analyses of stimuli used in perception experiments

F.1 Experiment 1 (Perception of lexical stress)

Acoustic analyses of the target stimuli used in Experiment 1 were conducted. Suprasegmental information, such as the duration, average intensity, and average pitch, of the stressed and unstressed vowels of the target words in the three prosodic contexts, were extracted using scripts in *Praat* (Boersma & Weenink, 2015). Speech segmentation was first performed using *EasyAlign* (Goldman, 2011), which is an automatic phonetic alignment tool developed as a plug-in of *Praat*. Later, the results were individually checked and manually corrected when needed. The formant structure in the spectrogram and the periodicity of the waveform were used as criteria for the manual correction. That is, the beginning and end of a vowel were identified as the zero-crossing points of the regular periodic signal (in the waveform) closest to the onset and the offset of a continuous second formant (F2) (in the spectrogram) (Baker, 2006; Recasens, 1999). The comparison between stressed and unstressed vowels of paroxytones and oxytones was carried out based on the relative prominence of the penultimate (V1) and last vowels (V2) of the target items (e.g., /a/ and /o/ in *Canto.* 'I sing'). The idea is that, if stressed vowels (i.e., V1 for paroxytones and V2 for oxytones) are produced with longer duration, higher intensity, and higher pitch than unstressed vowels (i.e., V2 for paroxytones and V1 for oxytones), then the difference between V1 and V2 would be higher than zero for paroxytones and lower than zero for oxytones. The duration of the vowels were calculated as the time of the onset the vowel subtracted from the time of the offset of the vowel. Average pitch (Hz) was extracted using the *To Pitch* function in *Praat*, with a time step window of 0.01 second, pitch floor of 75 Hz, and pitch ceiling of 600 Hz, which are the default values. As for average intensity (dB), the values were extracted using the *Get Intensity (dB)* function. Then, the difference between the duration, pitch, and intensity of the V1 and the V2 was calculated to examine whether these values vary depending on the location of lexical stress. In total, 360 tokens (i.e., 2 stress patterns * 3 prosodic contexts * 5 vowel types * 12 verbs) were analyzed.

The effects of stress pattern (paroxytones / oxytones), prosodic context (N / PN / U), and the interaction between the two fixed factors on difference in duration, average intensity, and average pitch between V1 and V2 were analyzed using linear mixed effects modeling with item as a random effect. The *lmer* function in the *lme4* package in R (Baayen, 2008) was used for the analysis. Further pairwise analyses were conducted using the *lsmeans* function in the *lsmeans* package. For duration, a main effect of stress condition (β = -0.053, SE = 0.003, t = -19.834) was found, which indicates that overall the duration difference for paroxytones (i.e., the baseline stress pattern) was higher than that of the oxytones. There was also a main effect of prosodic context for PN (β = 0.014, SE = 0.002, t = 8.444), suggesting that overall the duration difference in N (i.e., the baseline prosodic context) was higher than that in PN. Apart from the main effects, significant interactions between stress condition and PN (β = 0.045, SE = 0.003, t = 13.305) and between stress condition and U (β = 0.012, SE = 0.004, t = 3.145) were found. This suggests that the difference between the two stressed patterns was significantly smaller in PN and U than in N. Indeed, Figure 31 shows that the duration difference values between the two stress patterns are farther apart from each other in N, compared to PN and U. Pairwise comparisons with stress pattern and prosodic context confirmed that the duration difference for oxytones in N was significantly lower than that in both PN and U ($p < 0.001$ for both), while the duration difference for paroxytones in this prosodic context was significantly higher than that in U ($p < 0.01$). The duration difference for paroxytones in PN was between that in N and in U, but significant difference was found only between the values in PN and those in U ($p < 0.05$). That is, the gap between paroxytones and oxytones with regard to the duration difference between the V1 and the V2 was largest in N, followed by that in PN and in U. The pairwise comparison results also showed that the duration difference for paroxtyones was consistently higher than that for oxytones in all three prosodic contexts. This, together with the tendency that the duration difference values were mostly above zero for paroxytones (i.e., V1 was produced longer than V2) and below zero for oxytones (i.e., V1 was produced shorter than V2) suggest that duration was used as a strong acoustic correlate of lexical stress.

With regard to the difference between the intensity of the V1 and the V2, there was a main effect of stress pattern (β = -4.705, SE = 0.347, t = -13.575), suggesting that the overall intensity difference for paroxytones (i.e., baseline stress pattern) was higher than that for oxytones. There were also main effects of prosodic context for both PN (β = -4.061, SE = 0.218, t = -18.658) and U (β = 0.947, SE = 0.252, t = 3.762). That is, the intensity difference in N (i.e., baseline prosodic context) was higher than that in PN and lower than the one in U. The results also showed that there was a significant interaction between stress pattern and prosodic context (PN) (β = 9.235, SE = 0.435, t = 21.212), which indicates that the difference between the two stress patterns was larger in N than in PN. This can be clearly seen in Figure 31, in which the values for paroxytones and oxytones are much farther apart in N than in PN and in U. Pairwise comparisons with stress pattern and prosodic context revealed that the large difference between the two stress patterns in N is mainly due to the high intensity difference values of the paroxytones in this prosodic context, which were significantly higher than those in PN and in U ($p < 0.001$ for both). As for the oxytones, the values in both N and PN were found to be significantly lower than that those in U ($p < 0.05$). That is, the gap between paroxytones and oxytones was the largest in N among the three prosodic contexts. In PN and in U, the intensity difference values were close to zero and the values of the two stress patterns overlapped greatly, as seen in Figure 31. Therefore, although significant difference was found between the two stress patterns regardless of the prosodic context, it seems that intensity is used as an acoustic correlate of lexical stress in N, while in other prosodic contexts, it is used less consistently.

As for pitch difference, results showed that there were main effects of stress pattern ($\beta = -32.634$, SE = 1.07, t = -30.5), which indicates that the pitch difference was higher for paroxytones (i.e., baseline stress pattern) than for oxytones. Main effects of prosodic context were also found for both PN ($\beta = -21.378$, SE = 0.926, t = -22.44) and U ($\beta = 20.686$, SE = 1.101, t = 18.79), suggesting that and the pitch difference in N (i.e., baseline prosodic context) in general higher than that in PN and lower than that in U. Moreover, there was a significant interaction between stress pattern and prosodic context (PN) ($\beta = 61.928$, SE = 1.905, t = 32.51). That is, the difference between paroxytones and oxytones was larger in N than in PN. Pairwise comparisons with stress pattern and prosodic context showed very similar results as the ones found in the intensity data. That is, the difference between paroxytones and oxytones was much larger in N than in the other two prosodic contexts, mainly due to the significantly higher pitch difference values of paroxytones in this context, compared to those in PN and U ($p < 0.001$ for both). Also, the values of the oxytones in both N and PN were found to be significantly lower than that those in U ($p < 0.05$), which indicates that the gap between the pitch difference values of the paroxytones and the oxytones was much larger in N, compared to the gap in PN and in U. Particularly, in PN, the V2 was produced with higher pitch than the V1 regardless of the stress pattern (i.e., pitch difference values lower than zero), the reason of which will be discussed below in the pitch alignment section. Thus, like in the case of intensity, use of pitch in distinguishing paroxytones from oxytones is usually limited to N.

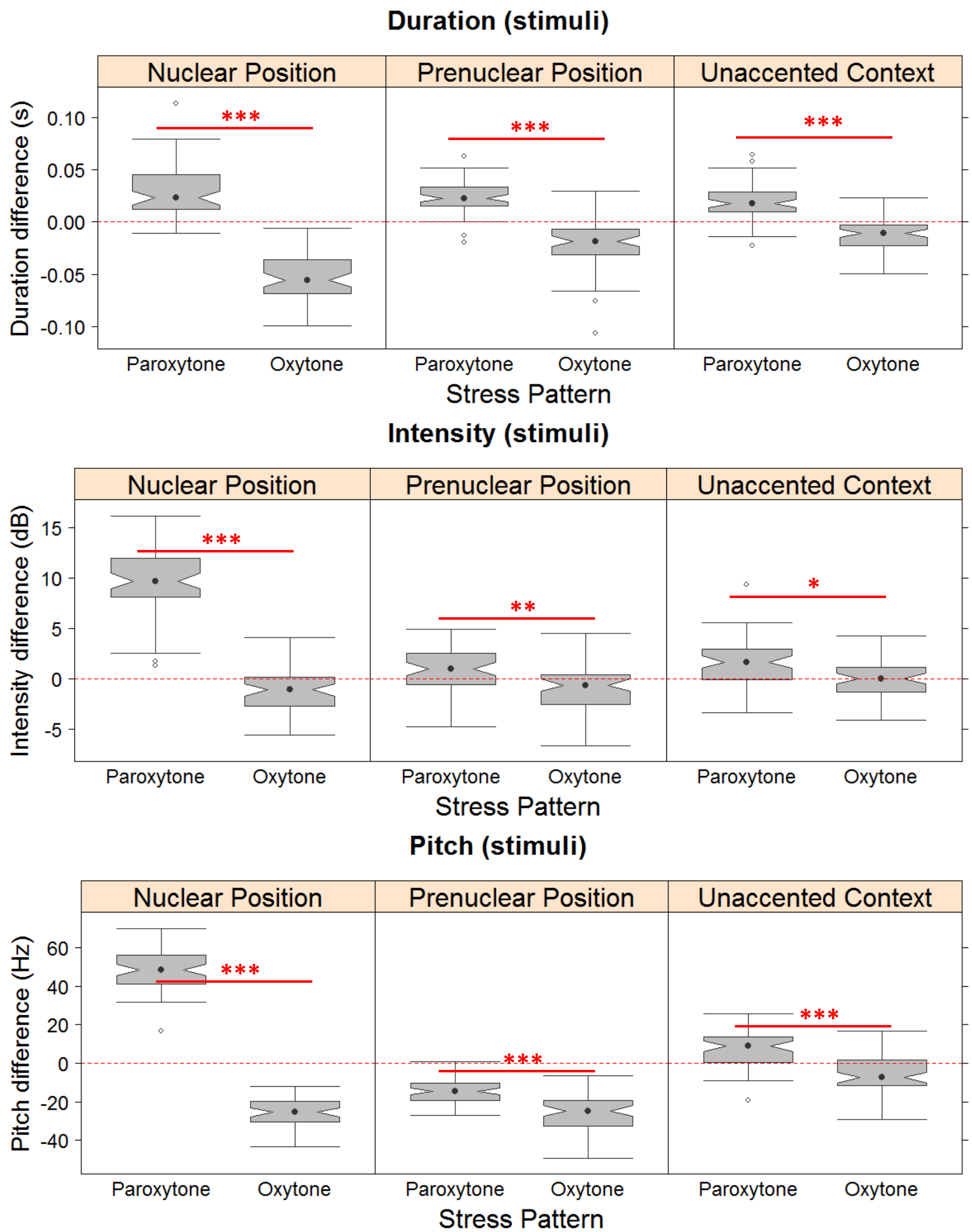


Figure 31 Difference in duration, intensity, and pitch of the last two vowels (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

With respect to PN, in particular, apart from the suprasegmental information of the vowels mentioned above, the degree of the displacement of pitch peak (H), which was defined as the distance from the onset of the stressed syllable to the location of H was analyzed. The location of

H was determined semi-automatically by selecting a region in which a peak was detected and extracting the point of the maximum pitch in that region. As the comparisons were done across different items, the H displacement values were normalized by dividing them into the duration of the stressed syllable. Tokens with H within a voiceless segment (e.g., within /t/ in *canto* 'I sing') were excluded from the analysis and comparisons between paroxytones and oxytones were done only with complete minimal pairs. Among the 120 tokens in PN (i.e., 2 stress patterns * 5 vowel types * 12 verbs), 23 tokens were excluded due to this reason and 23 tokens were excluded due to incomplete stress minimal pairs, leaving a total number of 74 tokens (i.e., 37 stress minimal pairs) to analyze. For 72 out of the 74 tokens (i.e., 97.3%), the H was displaced to a following syllable. H was aligned within the stressed syllable only for 2 tokens (i.e., 2.7%), all of which were oxytones. The effects of stress pattern (paroxytone / oxytone) on the normalized degree of H displacement was analyzed using linear mixed effects modeling with item as a random effect. Result showed that there was a main effect of stress pattern (β = -0.25, SE = 0.045, t = -5.58). This suggests that while H was displaced to a following syllable in both stress patterns, the degree to which it was displaced was significantly larger for paroxytones (i.e., the baseline stress pattern) than for oxytones.

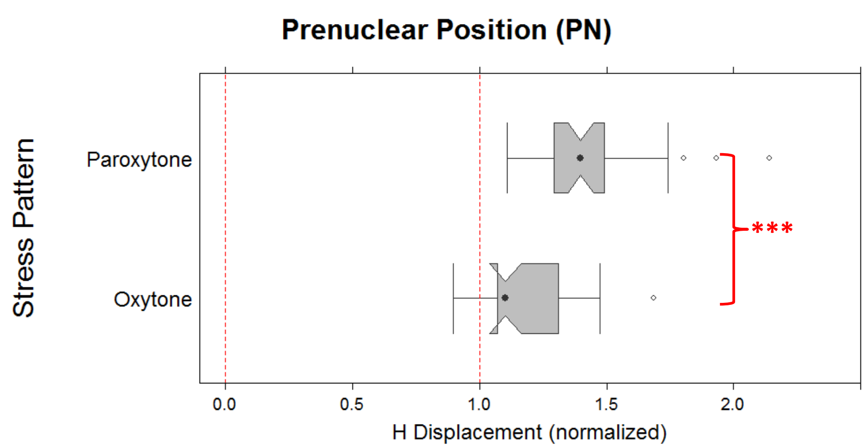


Figure 32 Normalized H displacement (dotted lines: onset and offset of stressed syllable) (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

Although in Spanish segmental information, such as vowel quality, does not vary in a systematic manner based on lexical stress, the present study examined the vowel quality of the stimuli to ensure that this segmental cue was not available in the speech signal. F1 and F2 values of the V1 of the stress minimal pairs (e.g., /a/ in *Canto.* 'I sing.' and /a/ in *Cantó* 'He/She/You (formal) sang.') were extracted, because this way it was possible to compare the same vowels in stressed and unstressed positions. The *To Formant (burg)* function was used with a time step window of 0.01, maximum number of 5 formants, maximum formant of 5,000 Hz, as the tokens were produced by a male speaker, window length of 0.025 second, and pre-emphasis from 50 Hz, which are the default values. The values were extracted at the middle of each vowel in order to avoid any transition effect from the adjacent segments. The effects of stress condition (stressed / unstressed) and vowel type (/a, e, i, o, u/), and the interaction between the two fixed factors were examined on F1 and F2 values separately for each prosodic context using linear mixed effects modeling with item as a random effect. As expected, for both F1 and F2, a main effect of vowel type was found for all vowels in all three prosodic contexts. That is, regardless of the prosodic context, /a/ (i.e., the baseline vowel type) overall had higher F1 values than the other vowels. With regard to F2, /a/ had lower F2 values than /e/ and /i/, while it had higher F2 values than /o/ and /u/. The statistical results of the main effect of vowel type on the F1 and F2 values across prosodic contexts are presented in Table 8.

F1: intercept: /a/			estimate	standard error	t-value
N	vowel	/e/	-254.87	16.88	-15.1
		/i/	-380.25	16.88	-22.52
		/o/	-225.75	16.88	-13.37
		/u/	-349.5	16.88	-20.7
PN	vowel	/e/	-266.083	10.556	-25.21
		/i/	-374.958	10.556	-35.52
		/o/	-214.667	10.556	-20.34
		/u/	-347.792	10.556	-32.95
U	vowel	/e/	-214.46	10.92	-19.64
		/i/	-326.75	10.92	-29.93
		/o/	-181.04	10.92	-16.58
		/u/	-301.67	10.92	-27.63

F2: intercept: /a/			estimate	standard error	t-value
N	vowel	/e/	626.04	16.88	-15.1
		/i/	789.46	16.88	-22.52
		/o/	-458.42	16.88	-13.37
		/u/	-548.54	16.88	-20.7
PN	vowel	/e/	605.37	10.556	-25.21
		/i/	801.29	10.556	-35.52
		/o/	-385.58	10.556	-20.34
		/u/	-506.38	10.556	-32.95

Table 8 Statistical results of the main effect of vowel type on F1 (above) and F2 (below) across prosodic contexts

F2: intercept: /a/			estimate	standard error	t-value
U	vowel	/e/	537.83	85.44	6.295
		/i/	710.71	85.44	8.319
		/o/	-399.71	85.44	-4.678
		/u/	-465.75	85.44	-5.451

Table 8 (cont.)

Apart from vowel type, there was a main effect of stress condition on the F1 values of the vowels in N ($\beta = -111.00$, SE: 16.16, $t = -6.87$) and in PN ($\beta = -26.333$, SE: 12.399, $t = -2.12$), which indicates that in these two prosodic contexts, the stressed vowels (i.e., the baseline stress condition) overall had higher F1 values than the unstressed vowels. For F1 in N, significant interactions were found between stress condition and all the vowel types (/e/: $\beta = 77.08$, SE = 22.85, $t = 3.37$; /i/: $\beta = 91.83$, SE = 22.85, $t = 4.02$; /o/: $\beta = 98.5$, SE = 22.85, $t = 4.31$; /u/: $\beta = 95$, SE = 22.85, $t = 4.16$). This suggests that, in N, the difference in the F1 between the stressed and unstressed vowels was significantly larger when the vowel was /a/ than the vowel was one of the other vowel types. This may be due to the low F1 values of unstressed /a/ in this prosodic context, as seen in Figure 33. Pairwise comparisons with stress condition and vowel type revealed that the unstressed /a/ had significantly lower F1 values than the stressed /a/ in N ($p < 0.001$), while this was not the case for the other vowel types. In the case of the F2, a significant interaction was found between stress condition and /i/ in PN ($\beta = -122.08$, SE = 42.55, $t = -2.869$), suggesting that the difference in the F2 between the stressed and unstressed vowels was significantly smaller when the vowel was /a/ than when the vowel was /i/. Indeed, pairwise comparison results confirmed that while the F2 of stressed /i/ was significantly higher than that of unstressed /i/, no significant difference was found between the F2 of the stressed and unstressed /a/. A significant interaction was also found between stress condition and /u/ in N ($\beta = -248.08$, SE = 109.12, $t = 2.273$), but no significant difference was found between stressed and unstressed /a/ and between stressed and unstressed /u/ in this prosodic context, according to pairwise comparison results. Apart from the F1 of /a/ in N and the F2 of /i/ in PN, none of the vowel types showed a significant difference between the formant frequencies of the stressed and unstressed vowels in any of the three prosodic contexts. Thus, the stimuli data of the present study confirmed that vowel quality is not a systematic acoustic correlate of lexical stress in Spanish.

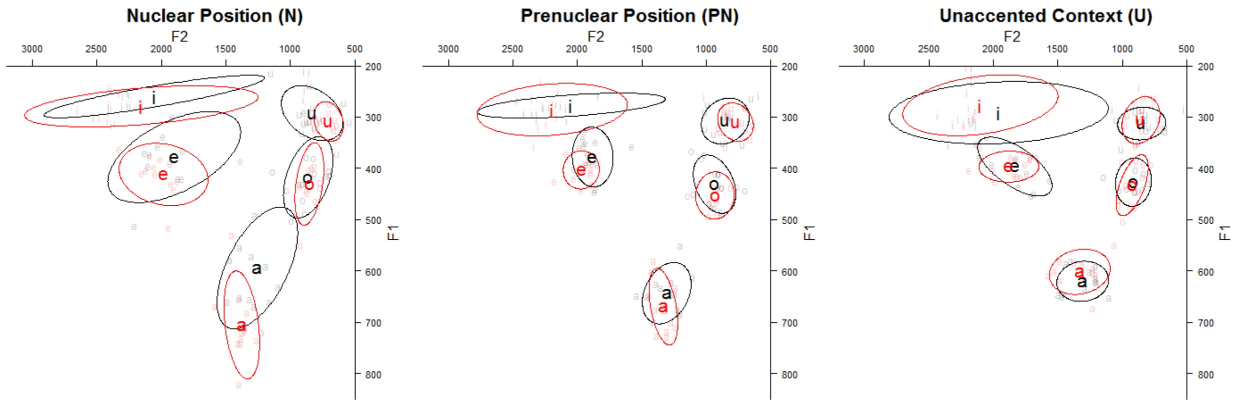


Figure 33 Vowel quality of stressed (red dots) and unstressed vowels (black dots) used in stimuli (dotted circle: significantly different F1 and/or F2 values)

F.2 Acoustic analyses of stimuli in Experiment 3 (Perception of nuclear stress)

Acoustic analyses of the stimuli used in Experiment 3 were conducted for both the target sentences and the filler sentences. Suprasegmental information, such as the duration, average intensity, average pitch of the stressed syllables of the subjects and the verbs, as well as the pitch peaks associated with them, was extracted using scripts in *Praat* (Boersma & Weenink, 2015). Speech segmentation was first performed using *EasyAlign* (Goldman, 2011), which is an automatic phonetic alignment tool developed as a plug-in of *Praat*. Later, the results were individually checked and manually corrected when needed. The formant structure in the spectrogram and the periodicity of the waveform were used as criteria for the manual correction. That is, the beginning and end of a vowel were identified as the zero-crossing points of the regular periodic signal (in the waveform) closest to the onset and the offset of a continuous second formant (F2) (in the spectrogram) (Baker, 2006; Recasens, 1999).

The comparison between the focused and the non-focused words was carried out based on the relative prominence of the stressed syllables of the first content word (W1) and the second content word (W2) of each sentence (e.g., *-lia-* in *Liliana* and *-ró* in *preparó* in the sentence *Liliana lo preparó.* ‘Liliana prepared it’). The idea is that, if the stressed syllable of the focused word is produced with longer duration, higher intensity, and higher pitch than the stressed syllable of the unfocused word, then the difference between W1 and W2 would be higher than zero for sentences in which the focus is on the first content word and lower than zero for those in which the focus is on the second content word. The duration of the stressed syllables was calculated as the time of the onset the stressed syllable subtracted from the time of the offset of the stressed syllable. Average pitch (Hz) was extracted using the *To Pitch* function in *Praat*, with a time step window of 0.01 second, pitch floor of 75 Hz, and pitch ceiling of 600 Hz, which are the default values. In addition to the average pitch, the pitch peaks (H) (Hz) of W1 and W2 were compared. The location of H was determined semi-automatically by selecting a region in which a peak was detected and extracting the point of the maximum pitch in that region. As for average intensity (dB), the values were extracted using the *Get Intensity (dB)* function. Then, the difference between the duration, pitch, and intensity of the W1 and the W2 was calculated to examine whether these values vary depending on the position of focus and word order. In total, 80 tokens (i.e., 2 focus positions * 2 word orders * 20 verbs) were analyzed.

As distinct items were used for the target sentences and the filler sentences, which differed in the focus location (i.e., subject vs. verb), the effect of word order (S+Cl+V / Cl+V+S) on the difference in duration, average intensity, average pitch, and pitch peak between W1 and W2 were analyzed separately for the target sentences and filler sentences. Linear mixed effects modeling was conducted with item as a random effect, using the *lmer* function in the *lme4* package in R (Baayen, 2008). Additionally, pairwise analyses were conducted using the *lsmeans* function in the *lsmeans* package, in order to extract the p-values. Apart from intensity for focus on the verb, main effect of word order was found in all measures: duration for both focus on the subject (β = -0.178, SE = 0.015, t = -11.714) and focus on the verb (β = 0.062, SE = 0.017, t = 3.657), intensity for focus on the subject (β = -9.264, SE = 0.501, t = -18.49), pitch for both focus on the subject (β = -38.557, SE = 1.868, t = -20.64) and on the verb (β = 22.386, SE = 2.249, t = 9.954), and pitch peak for both focus on the subject (β = -57.563, SE = 3.671, t = -15.68) and on the verb (β = 23.101, SE = 2.973, t = 7.77). This indicates that, except for intensity, the difference between W1 and W2 was significantly higher when W1, regardless of whether it was a subject or a verb, received focus than when it did not. Among the four measures, duration seems to be the acoustic cue that varies according to the position of focus in the most systematic manner. As shown in Figure 34, while W1-W2 difference in intensity, pitch, and peak was mostly above zero (i.e., the red dotted lines) regardless of focus location and the word order, the differences in duration varied depending on whether W1 was focused or not; if it was focused, then the values were above zero, if W2 was focused instead, the values were below zero. That is, unlike intensity and pitch, which generally decreased throughout the sentence (i.e., higher in W1 than in W2), duration was longer in focused constituents than in non-focused constituents. Given that the stressed syllables of the subjects generally contained a diphthong (e.g., *Liliana*) or a coda (e.g., *Leonardo*), while those of the verbs were always open syllables without any diphthongs (e.g., *preparó*), it is possible that the inherent duration resulting from the weight of the stressed syllable led to longer W1 duration for the S+Cl+V structure (W1 = S) and shorter W1 duration for the Cl+V+S (W1 = V) in the target sentences, in which the focused constituents were the subjects. However, syllable weight does not seem to be the main reason for the longer stressed syllable of subjects in the target sentences, because the opposite pattern was found when the focus was on the verb (i.e., shorter W1 for the S+Cl+V structure and longer W1 for the Cl+V+S structure). Rather, it is likely that duration is a systematic cue in the realization of focus in Spanish.

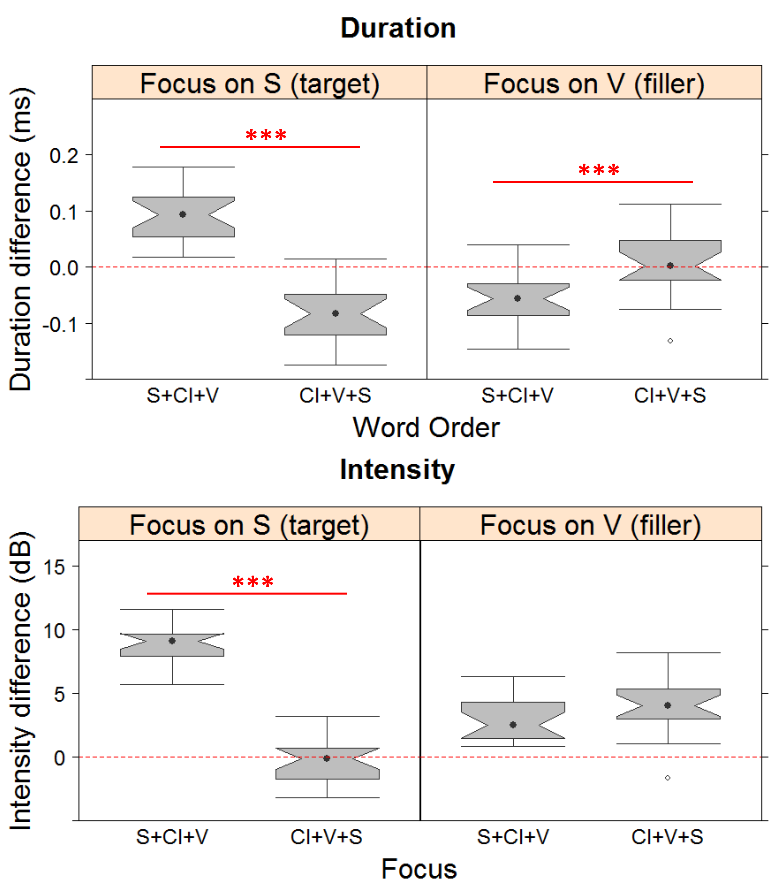


Figure 34 Difference in duration, intensity, pitch, and pitch peak of the stressed syllables of the first and the second content words
(***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$)

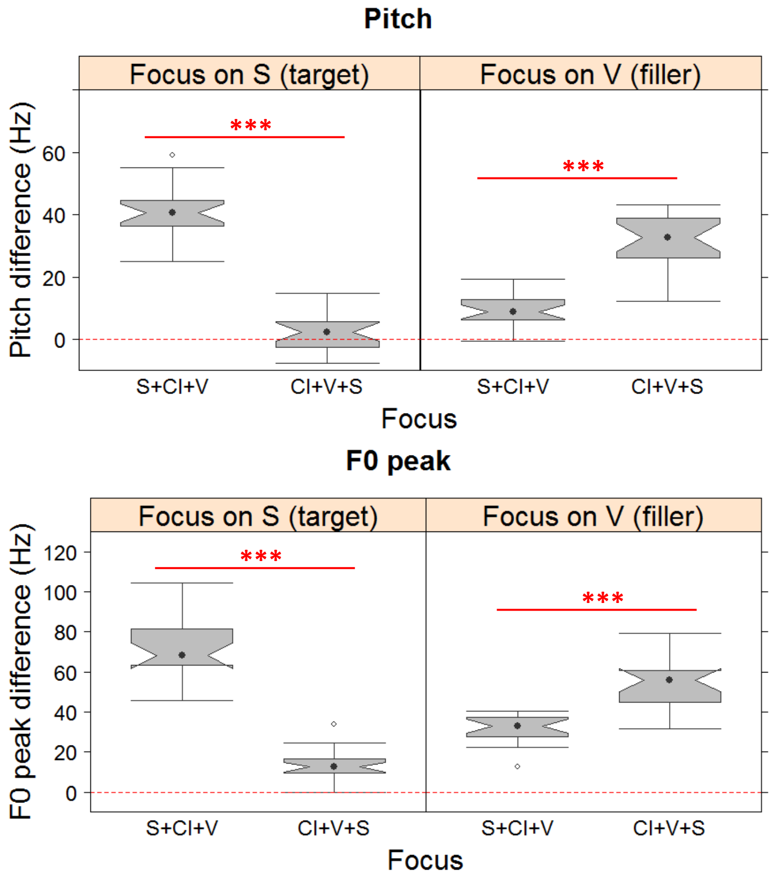


Figure 34 (cont.)

Apart from the suprasegmental information of the stressed syllables mentioned above, the degree of the displacement of pitch peak (H) of the content word in the prenuclear position (i.e., W1), which was defined as the distance from the onset of the stressed syllable to the location of H was analyzed. The location of H was determined semi-automatically by selecting a region in which a peak was detected and extracting the point of the maximum pitch in that region. As the comparisons were done across different items, the H displacement values were normalized by dividing them into the duration of the stressed syllable. As seen in Figure 35, the H was displaced to a following syllable in most of the cases when W1 did not receive focus (i.e., the V in the Cl+V[$_F$S] structure and the S in the S+Cl[$_F$V] structure). On the other hand, H was mostly aligned within the stressed syllable when W1 was focused (i.e., the S in the [$_F$S]Cl+V structure and the V in the Cl[$_F$V]S structure). The effects of word order on the normalized degree of H displacement was analyzed separately for the target sentences (i.e., focus on the subject) and filler sentences (i.e., focus on the verb), using linear mixed effects modeling with item as a random effect. Result showed that there was a main effect of word order both when the focus was on the subject ($\beta = 0.492$, SE = 0.067, t = 7.31) and when the focus was on the verb ($\beta = 0.4$-0.621, SE = 0.084, t = -7.388). This suggests that H was displaced to a following syllable to a significantly larger degree when W1 was not focused than when it was.

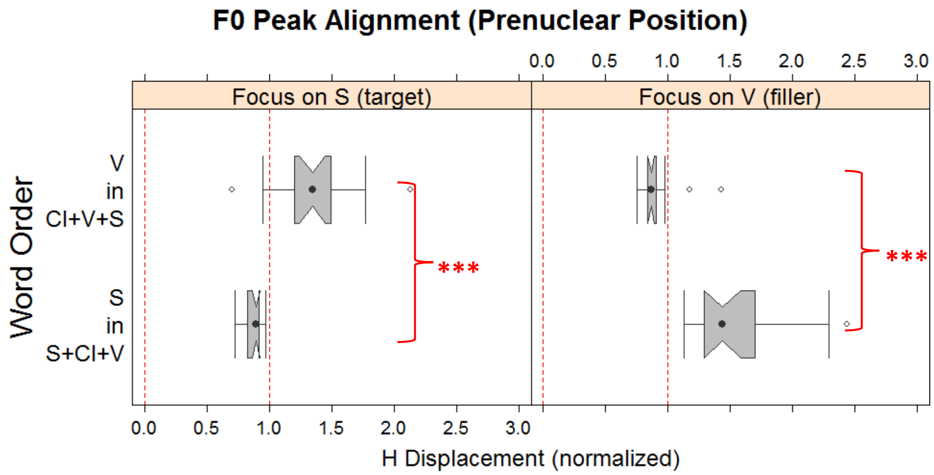


Figure 35 Normalized H displacement (dotted lines: onset and offset of stressed syllable)
(***: p < 0.001, **: p <0.01, *: p < 0.05)

REFERENCES

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology, 24*, 197-207.

Acosta, Y. D., & de la Cruz, G.P. (2011). *The foreign born from latin america and the caribbean: 2010.* US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Adams, C. (1979). *English speech rhythm and the foreign learner*. The Hague: Mouton.

Alameda, J. R., & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, *15*(3), 517-530.

Arreola, J. J. (2011), Aumenta migración a Querétaro por violencia. *El Universal,* 149. Retrieved from http://www.eluniversal.com.mx/estados/79353.html

Atkinson-King, K. (1973). *Children's acquisition of phonological stress contrasts* (Unpublished doctoral dissertation). University of California, Los Angeles.

Au, T. K., Knightly, L. M., Jun, S. A., & Oh, J. S. (2002). Overhearing a language during childhood. *Psychological Science, 13*(3), 238-243.

Au, T. K., Oh, J. S., Knightly, L. M., Jun, S. A., & Romo, L. F. (2008). Salvaging a childhood language. *Journal of Memory and Language, 58*(4), 998-1011.

Au, T. K., & Romo, L. F. (1997). Does childhood language experience help adult learners? In H.-C. Chen (Ed.), *The cognitive processing of chinese and related asian languages* (pp. 417-441). Hong Kong: Chinese University Press.

Aumenta migración a Querétaro por temor a la inseguridad. (2011). *Notisistema.Com.* Retrieved from http://www.notisistema.com/noticias/?p=418411

Ayers, G. M. (1994). Discourse functions of pitch range in spontaneous and read speech. *OSU Working Papers in Linguistics, 44*, 1-49.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baker, G. K. (2006). An acoustic basis for palatal geminate behavior in spanish. In J. P. Montreuil (Ed.), *New perspectives on romance linguistics: Vol. II: Phonetics, phonology and dialectology. selected papers from the 35th linguistic symposium on romance languages* (pp. 1-14). John Benjamins Publishing.

Beaudrie, S., & Ducar, C. (2005). Beginning level university heritage programs: Creating a space for all heritage language learners. *Heritage Language Journal, 3*(1), 1-26.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form. papers in laboratory phonology III* (pp. 7-33). Cambridge: Cambridge University Press.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics, 39*(3/4), 129-181.

Birdsong, D. (1999). Whys and why nots of the critical period hypothesis for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and critical period hypothesis* (pp. 1-22). Mahwah, NJ: Lawrence Erlbaum.

Birdsong, D., Gertken, L.M., & Amengual, M. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. Web. 20 Jan. 2012. <https://sites.la.utexas.edu/bilingual/>.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication, 14*, 359-375.

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.14, retrieved 27 July 2015 from http://www.praat.org/

Boomershine, A. (2013). The perception of english vowels by monolingual, bilingual, and heritage speakers of spanish and english. In C. Howe, S. E. Blackwell & M. L. Quesada (Eds.), *Selected proceedings of the 15th hispanic linguistics symposium* (pp. 103-118). Somerville, MA: Cascadilla Proceedings Project.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology, 12*, 189-221.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25*(7), 1044-1098.

Brown, A., & Patten, E. (2014, April 29). *Statistical portrait of Hispanics in the United States, 2012*. Washington, DC: Pew Hispanic Center. Retrieved from http://www.pewhispanic.org/2014/04/29/statistical-Portrait-of-Hispanics-in-the-Unitedstates-2012

Brown, C. L. (2003). The making of a bilingual/bicultural identity: A case study of a Japanese-American child. *Japan Journal of Multilingualism and Multiculturalism, 9*(1), 52-68.

Brown, C. L. (2009). Heritage language and ethnic identity: A case study of Korean-American college students. *International Journal of Multicultural Education, 11*, 1-16.

Burdin, R. S., Phillips-Bourass, S., Turnbull, R., Yasavul, M., Clopper, C. G., & Tonhauser, J. (2015). Variation in the prosody of focus in head- and head/edge-prominence languages. *Lingua, 165*, 254-276.

Campbell, N., & Beckman, M. E. (1997). Stress, prominence and spectral tilt. In A. Botinis, G. Kouroupetroglou & G. Carayiannis (Eds.), *Intonation: Theory, models and applications: Proceedings of an ESCA workshop* (pp. 67-70). Athens: ESCA and University of Athens.

Campbell, R. N., & Rosenthal, J. W. (2000). Heritage languages. In J. W. Rosenthal (Ed.), *Handbook of Undergraduate Second Language Education* (pp.165-184). Mahwah, NJ: Lawrence Erlbaum.

Canellada, M. J., & Madsen, J. K. (1987). *Pronunciación del español. Lengua hablada y literaria*. Madrid, Spain: Editorial Castalia.

Carreira, M. (2013). The vitality of Spanish in the united states. *Heritage Language Journal, 10*(3), 396-413.

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 20-51). Amsterdam: John Benjamins.

Chang, C. B. (2010). *First language phonetic drift during second language acquisition* (Doctoral disssertaion). Retrieved from http://linguistics.berkeley.edu/phonlab/annual_report/documents/2010/Chang_diss.pdf

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*(2), 249-268.

Chang, C. B., Haynes, E. F., Yao, Y., & Rhodes, R. (2008). The phonetic space of phonological categories in heritage speakers of Mandarin. *Chicago Linguistic Society, 44*(1), 31-45.

Chang, C. B., Haynes, E. F., Yao, Y., & Rhodes, R. (2009). A tale of five fricatives: Consonantal contrast in heritage speakers of Mandarin. *University of Pennsylvania Working Papers in Linguistics, 15*(1), 37-43.

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America, 129*(6), 3964-3980.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351-353.

Cho, G. (2000). The role of heritage language in social interactions and relationships: Reflections from a language minority group. *Bilingual Research Journal, 24*(4), 369-384.

Chomsky, N., & Halle, M. (1969). *The sound pattern of English*. New York: Harper Row.

Cole, J. (2015). *Prosody in context: A review*. *Language, Cognition and Neuroscience, 30*, 1-31.

Contreras, H. (1976). *A theory of word order with special reference to Spanish*. Amsterdam: North-Holland.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in Snglish: Evidence from native and non-native listeners. *Language and Speech, 45*(Pt 3), 207-228.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*, 201-220.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

Davies, M. (2002). *Corpus del español: 100 million words, 1200s-1900s.* Retrieved from http://www.corpusdelespanol.org

Davies, M. (2006). *A frequency dictionary of Spanish: Core vocabulary for learners*. New York/London: Routledge.

de Boysson-Bardies, B. (1993). Ontogeny of language-specific syllabic productions. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 353-363). Dordrecht, The Netherlands: Kluwer Academic Publishers.

de la Mota, C. (1997). Prosody of sentences with contrastive new information in Spanish. In A. Botinis, G. Kouroupetroglou & G. Carayiannis (Eds.), *Intonation: Theory, models and applications: Proceedings of an ESCA workshop* (pp. 75-78).

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching, 4*(14), 183-198.

Dellwo, V., Fourcin, A., & Abberton, E. (2007). Rhythmical classification based on voice parameters. In J. Trouvan, W.J. Barry (Eds.), *International Congress of Phonetic Sciences*, Saarbruecken, Germany (pp. 1129-1132).

Donati, C., & Nespor, M. (2003). From focus to syntax. *Lingua, 113*, 1119-1142.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress "deafness" . *Journal of the Acoustical Society of America, 110,* 1606-1618.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress ‘deafness’ in simultaneous French–Spanish bilinguals. *Cognition, 114*(2), 266-275.

Enríquez, E. V., Casado, C., & Santos, A. (1989). La percepción del acento en español. *Lingüística Española Actual, 11,* 241-269.

Escandell-Vidal, M. V. (2011). *Verum focus y prosodia: Cuando la duración (sí que) importa*. *Oralia*, *14,* 181-201.

Estebas-Vilaplana, E. (2007). The phonological status of English and Spanish prenuclear f0 peaks. *Atlantis, 29*(2), 39-57.

Face, T. (2000). Prosodic manifestations of focus in Spanish. *Southwest Journal of Linguistics, 19*, 45-62.

Face, T. (2001). Focus and early peak alignment in Spanish intonation. *Probus, 13*, 223-246.

Face, T. (2002). Local intonational marking of Spanish contrastive focus. *Probus, 14,* 71-92.

Face, T., & D'Imperio, M. (2005). Reconsidering a focal typology: Evidence from Spanish and Italian. *Rivista Di Linguistica, 17*(2), 271-289.

Féry, C. (2013). Focus as prosodic alignment. *Natural Language and Linguistic Theory, 31*(3), 683-734.

Fishman, J. A. (2001). Can threatened languages be saved? Clevedon, UK: Multilingual Matters.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America, 89*(1), 395-411.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timodium, MD: York Press.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). The effect of age of second language learning on the production of English consonants. *Speech Communication, 16,* 1-26.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between native and second language phonetic subsystems. *Speech Communication*, *40*, 467-491.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second language learning. *Journal of Memory and Language, 41*, 78-104.

Fought, C. (1999). A majority sound change in a minority community: /u/-fronting in Chicano English. *Journal of Sociolinguistics, 3*(1), 5-23.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Halle, P. (2008). Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics, 36*(4), 649-663.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27,* 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*(2), 126-152.

Fuller, J. M. (2012). *Bilingual pre-teens: Competing ideologies and multiple identities in the US and Germany.* New York: Routledge.

Gabriel, C. (2010). On focus, prosody, and word order in Argentinean Spanish: A minimalist OT account. *Revista Virtual De Estudos Da Linguagem, 4*, 183-222.

Gildersleeve-Neumann, C. E., Peña, E. D., Davis, B. L., & Kester, E. S. (2009). Effects on L1 during early acquisition of L2: Speech changes in Spanish at first English contact. *Bilingualism: Language and Cognition, 12*(2), 259-272.

Godson, L. (2004). Vowel production in the speech of western Armenian heritage speakers. *Heritage Language Journal, 2*(1), 44-69.

Goldman, J. (2011). EasyAlign: An automatic phonetic alignment tool under Praat. *Proceedings of the 12th Annual Conference of the International Speech Communication Association,* Retrieved from http://archive-ouverte.unige.ch/unige:18188

González-Barrera, A., & Lopez, M. (2013). Spanish is the most spoken non-English language in US homes, even among non-Hispanics. Washington, DC: Pew Hispanic Center. Retrieved from http://www.pewresearch.org/fact-tank/2013/08/13/spanish-is-the-most-spoken-non-english-language-in-u-s-homes-even-among-non-hispanics/

Grieco, E. M., Trevelyan, E., Larsen, L., Acosta, Y. D., Gambino, C., de la Cruz, G Patricia, et al. (2012). The size, place of birth, and geographic distribution of the foreign-born population in the United States: 1960 to 2010. *Population Division Working Paper, 96.* Retrieved from https://www.census.gov/population/foreign/files/WorkingPaper96.pdf

Gries, S. T., & Miglio, V. G. (2014). New information in naturalistic data is also signaled by pitch movement: An analysis from monolingual English/Spanish and bilingual Spanish speakers. *Complutense Journal of English Studies, 22*, 11-33.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language, 36*, 3-15.

Grosjean, F. (2010). *Bilingual: Life and reality*. Cambridge: Harvard University Press.

Guardado, M. (2014). The discourses of heritage language development: Engaging ideologies in Canadian Hispanic communities. *Heritage Language Journal, 11*(1), 1-28.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica, 60*(2), 98-128.

Guion, S. G., Flege, J. E., & Loftin, J. D. (2000). The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics, 28,* 27-42.

Gupton, T., & Leal Méndez, T. (2013). Experimental methodologies: Two case studies investigating the syntax-discourse interface. *Studies in Hispanic and Lusophone Linguistics, 6*(1), 139-164.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In B. Bel & I. Marlien (Eds.), *Proceedings of the 1st International Conference on Speech Prosody* (pp. 47-57), Aix-en-Provence, France: Laboratoire Parole et Langage.

Hakimzadeh, S., & Cohn, D. (2007). *English usage among Hispanics in the United States*. Washington, DC: Pew Hispanic Center. Retrieved from http://www.pewhispanic.org/2007/11/29/english-usage-among-hispanics-in-the-united-states/

Hakuta, K., & D'Andrea, D. (1992). Some properties of bilingual maintenance and loss in Mexican background high-school students. *Applied Linguistics*, *13*(1), 72-99.

Harada, T. (2003). L2 influence on L1 speech in the production of VOT. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15Th International Congress of Phonetic Sciences* (pp. 1085-1088), Barcelona, Spain: Universitat Autònoma de Barcelona.

Harris, M. J., Gries, S. T., & Miglio, V. G. (2015). Mexican and Chicano Spanish prosody: Differences related to information structure. In J. Levis, R. Mohammed, M. Qian & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference* (pp. 40-50), Santa Barbara, CA: University of California.

He, A. W. (2006). Toward an identity theory of the development of Chinese as a heritage language. *Heritage Language Journal*, *4*(1), 1-28.

Heller, M. (2007). *Bilingualism: A social approach*. Hampshire/New York: Palgrave Macmillan.

Henriksen, N. (2015). Acoustic analysis of the rhotic contrast in Chicagoland Spanish: An intergenerational study. *Linguistic Approaches to Bilingualism, 5*(3), 285-321.

Henriksen, N., & Willis, E. W. (2010). Acoustic characterization of phonemic trill production in Jerezano Andalusian Spanish. In M. Ortega-Llebaria (Ed.), *Selected proceedings of the 4th conference on laboratory approaches to Spanish phonology* (pp. 115-127). Somerville, MA: Cascadilla Proceedings Project.

Hertel, T. (2003). Lexical and discourse factors in the second language acquisition of Spanish word order. *Second Language Research, 14*, 273-304.

Hing, B. O., & Lee, R. (1996). *Reframing the immigration debate. The state of Asian Pacific America series. A public policy report.* Los Angeles: LEAP Asian Pacific American Public Policy Institute: UCLA Asian American Studies Center.

Hinton, L., Bremner, S., Corcoran, H., Learner, J., Luthin, H., Moonwomon, B., et al. (1987). It's not just valley girls: A study of California English. In J. Aske, N. Beery, L. Michaelis & H. Filip (Eds.), *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society* (pp. 117-127), Berkeley, CA: Berkeley Linguistics Society.

Hoot, B. (to appear). Narrow presentational focus in Mexican Spanish: Experimental evidence. *Probus.* doi: 10.1515/probus-2014-0004

Hoot, B. (2016). Narrow presentational focus in heritage Spanish and the syntax–discourse interface. *Linguistic Approaches to Bilingualism.* doi: 10.1075/lab.14021.hoo

Hornberger, N. H., & Wang, S. C. (2008). Who are our heritage language learners? Identity and biliteracy in heritage language education in the united states. In D. Brinton, O. Kagan & S.

Bauckus (Eds.), *Heritage Language Education: A New Field Emerging* (pp. 3-35). New York: Routledge.

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication, 10*(2), 163-169.

Hualde, J. I. (1999). *Basic intonational contours in Spanish.* Paper Presented at the First Sp-ToBI Workshop, Columbus, OH.

Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge: Cambridge University Press.

Hualde, J. I. (2002). Intonation in Spanish and the other Ibero-Romance languages: Overview and status quaestionis. In C. Wiltshire, & J. Camps (Eds.), *Romance phonology and variation* (pp. 101-115). Amsterdam: John Benjamins.

Hulsen, M. E. H. (2000). *Language loss and language processing: Three generations of Dutch migrants in New Zealand* (Doctoral dissertation). University of Nijmegen, the Netherlands.

Hurtado, A., & Vega, L. A. (2004). Shift happens: Spanish and English transmission between parents and their children. *Journal of Social Issues, 60*(1), 137-155.

Ilieva, G. N. (2012). Hindi heritage language learners' performance during OPIs: Characteristics and pedagogical implications. *Heritage Language Journal*, *9*(2), 18-36.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., & Kuhl, P. K. (2006). Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *Neuroreport*, *17*(10), 957-962.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*(2), 541-573.

Jensen, J. T. (1993). *English phonology*. Amsterdam: John Benjamins Publishing.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*, 60-99.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of english fricatives. *The Journal of the Acoustical Society of America*, *108*(3 Pt 1), 1252-1263.

Jun, S. A. (2005). Prosodic typology. In S. A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430-458). Oxford: Oxford University Press.

Jusczyk, P., Friederici, A., & Wessels, J. (1993). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory & Language, 32*, 402-420.

Kang, K. H., & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *The Journal of the Acoustical Society of America, 119*(3), 1672-1683.

Kang, K. H., & Nagy, N. (2012). VOT merger in heritage Korean in Toronto. In P. Cajax (Ed.), *Proceedings of the 2012 Annual Conference of the Canadian Linguistic Association* (pp. 1-15). Toronto: Canadian Linguistic Association.

Khattab, G. (2000). VOT production in English and Arabic bilingual and monolingual children. *Leeds working papers in linguistics*, 8, 95-122.

Kim, J. Y. (2010). The phonological effect of single-word insertion by Korean heritage speakers: Vowel epenthesis in word-final position. *Korean Journal of Linguistics, 35*(1), 49-70.

Kim, J. Y. (2011). L1-L2 phonetic interference in the production of Spanish heritage speakers in the US. *The Korean Journal of Hispanic Studies, 4*, 1-28.

Kim, J. Y. (2012). *Discrepancy between the perception and production of stop consonants by Spanish heritage speakers in the United States* (Master's thesis). Korea University, South Korea.

Kim, J. Y. (2014). Use of suprasegmental information in the perception of Spanish lexical stress by Spanish heritage speakers of different generations. In N. Campbell, D. Gibbon & D. J. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 453-456). Dublin: Trinity College.

Kim, J. Y. (2015). Perception and production of Spanish lexical stress by Spanish heritage speakers and English L2 learners of Spanish. In E. Willis, P. Martín Butragueño & E. Herrera Zendejas (Eds.), *Proceedings of the 6th conference on laboratory approaches to romance phonology* (pp. 106-128). Somerville, MA: Cascadilla Proceedings Project.

Kim, M. R. (2011). *The phonetics of stress manifestation: Segmental variation, syllable consistency and rhythm.* (Doctoral dissertation). Stony Brook University, USA.

Klee, C. A. (1998). Communication as an organizing principle in the national standards: Sociolinguistic aspects of Spanish language teaching. *Hispania,* 81(2) , 339-351.

Knightly, L. M., Jun, S., Oh, J. S., & Au, T. K. (2003). Production benefits of childhood overhearing. *The Journal of the Acoustical Society of America, 114*(1), 465-474.

Kondo-Brown, K. (2003). Heritage language instruction for post-secondary students from immigrant backgrounds. *Heritage Language Journal, 1*(1), 1-25.

Krogstad, J. M., & Lopez, M. H. (2014). *Hispanic nativity shift*. Washington, DC: Pew Hispanic Center. Retrieved from http://www.pewhispanic.org/2014/04/29/hispanic-nativity-shift/

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences*, 97, 11850–11857.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews.Neuroscience*, *5*(11), 831-843.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 979-1000.

Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period". *Language Learning and Development, 1*, 237-264.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “Perceptual magnetic effect”. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timodium, MD: York Press.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America, 100*, 2425-2438.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*(4577), 1138-1141.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13-F21.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America, 100*(15), 9096-9101.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Kupisch, T., Barton, D., Klaschik, E., Stangen, I., Lein, T., & van de Weijer, J. (2014). Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal, 11*(2), 123-150.

Kupisch, T., Lein, T., Barton, D., Schröder, D. J., Stangen, I., & Stoehr, A. (2014). Acquisition outcomes across domains in adult simultaneous bilinguals with French as weaker and stronger language. *Journal of French Language Studies, 24*, 347-376.

Labov, W. (2008). Triggering events. In S. M. Fitzmaurice, & D. Minkova (Eds.), *Empirical and analytical advances in the study of English language change* (pp. 11-54). Berlin, New York: Mouton de Gruyter.

Lacorte, M., & Canabal, E. (2003.) Interaction with heritage language learners in foreign language classrooms. In C. Blyth (Ed.), *The Sociolinguistics of Foreign-Language Classrooms: Contributions of the Native, the Near-Native, and the Non-Native Speaker* (pp. 107-129). Boston: Heinle.

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

Ladefoged, P. (2001). *A course in phonetics* (4th ed.). Fort Worth, TX: Harcourt College Publishers.

Lenneberg, E. H. (1967). *The biological foundations of language*. New York: Wiley.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242-1255.

Lleó, C., Martin R., & Margaret K. (2004). Acquisition of language-specific pitch accent by Spanish and German monolingual and bilingual children. In T. Face (Ed.), *Laboratory approaches to Spanish phonology* (pp. 3-27). Berlin: Mouton de Gruyter.

Llisterri, J., Machuca, M., de la Mota, C., Riera, M., & Ríos, A. (2003). The perception of lexical stress in Spanish. In M. Solé, D. Recasens & J. Romero (Eds), *Proceedings of the 15th International Conference of Phonetic Sciences* (pp. 2023-2026). Barcelona: Casual Productions.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Association for Research in Otolaryngology, 49*, 606-608.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251-285.

Lopez, M. H., & González-Barrera, A. (2013). *What is the future of Spanish in the United States?* Washington, DC: Pew Hispanic Center. Retrieved from http://www.pewresearch.org/fact-tank/2013/09/05/what-is-the-future-of-spanish-in-the-united-states/

Lord, G. (2007). The role of the lexicon in learning second language stress patterns. *Applied Language Learning, 17*, 1-14.

Low, E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43*, 377-401.

Lozano, C. (2006). Focus and split intransitivity: The acquisition of word order alternations in non-native Spanish. *Second Language Research, 22*, 145-187.

Lynch, A. (2003). The relationship between second and heritage language acquisition: Notes on research and theory building. *Heritage Language Journal*, *1*(1), 26-43.

Lynch, A. (2014). The first decade of the heritage language journal: A retrospective view of research on heritage languages. *Heritage Language Journal, 11*(3), 224-242.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics, 46*(2), 187-200.

Mack, M. (2000). Phonetic transfer in a French-English bilingual child. In P. H. Nelde (Ed.), *Language attitudes and language conflict* (pp. 107-124). Bonn: Dümmler.

Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica, 56*, 158-185.

Major, R. C. (1992). Losing english as a first language. *The Modern Language Journal, 76*, 190-208.

Marín Gálvez, R. (1995). La duración vocálica en español. *Estudios De Lingüística Universidad De Alicante, 10*, 213-226.

Martin, C., Swender, E., & Rivera-Martinez, M. (2013). Assessing the oral proficiency of heritage speakers according to the ACTFL proficiency guidelines 2012–Speaking. *Heritage Language Journal,* 10(2), 211-225.

Martínez Celdrán, E. (1984). Cantidad e intensidad en los sonidos obstruyentes del castellano: Hacia una caracterización de los sonidos aproximantes. *Estudios De Fonética Experimental, 1*, 71-129.

Martínez Celdrán, E. (1991). Sobre la naturaleza de los alófonos de /b, d, g/ en español y sus distintas denominaciones. *Verba, 18*, 235-253.

Martinez, G. (2003). Classroom based dialect awareness in heritage language instruction: A critical applied linguistic approach. *Heritage Language Journal, 1*(1), 1-14.

Martinez, G., & Schwartz, A. (2012). Elevating" low" language for high stakes: A case for critical, community-based learning in a medical Spanish for heritage learners program. *Heritage Language Journal, 9*(2), 37-49.

Mayberry, R. I., & Fischer, S. D. (1989). Looking through phonological shape to lexical meaning: The bottleneck of non-native sign language processing. *Memory & Cognition, 17*(6), 740-754.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Mehler, J., Dupoux, E., Nazzi, T., & Dehaene-Lambertz, G. (1996). Coping with linguistic diversity: The infant's viewpoint. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 101-116). Mahwah, NJ: Erlbaum.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, *29*, 143-178.

Menke, M., & Face, T. (2010). Second language Spanish vowel production: An acoustic analysis. *Studies in Hispanic and Lusophone Linguistics, 3*(1), 181-214.

Menn, L., & Stoel-Gammon, C. (1995). Phonological development. In P. Fletcher, & B. MacWhinney (Eds.), *The handbook of child language* (pp. 335-360). Cambridge, MA: Blackwell.

Mikulski, A. M. (2010). Age of onset of bilingualism, language use, and the volitional subjunctive in heritage learners of Spanish. *Heritage Language Journal, 7*(1), 28-46.

Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research, 21*, 199-249.

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor* John Benjamins Publishing.

Montrul, S. (2010). Dominant language transfer in adult second language learners and heritage speakers. *Second Language Research, 26*(3), 293-327.

Montrul, S. (2011). Morphological errors in Spanish second language learners and heritage speakers. *Studies in Second Language Acquisition, 33*, 163-192.

Montrul, S. (2012). Is the heritage language like a second language? *Eurosla Yearbook*, *12*(1), 1-29.

Montrul, S. (2013). How "Native" are heritage speakers? *Heritage Language Journal, 10*(2), 153-177.

Montrul, S. (2014). Structural changes in Spanish in the united states: Differential object marking in spanish heritage speakers across generations. *Lingua, 151*, 177-196.

Montrul, S., & Foote, R. (2014). Age of acquisition interactions in bilingual lexical access: A study of the weaker language of L2 learners and heritage speakers. *International Journal of Bilingualism, 18*(3), 274-303.

Moore, J. K., & Guan, Y. L. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *Journal of the Association for Research in Otolaryngology, 2*(4), 297-311.

Muntendam, A., & Torreira, F. (2016). Focus and prosody in Spanish and Quechua: Insights from an interactive task. In M. E. Armstrong, N. Henriksen & M. del Mar Vanrell (Eds.), *Intonational grammar in ibero-romance: Approaches across linguistic subfields* (pp. 69-90). Amsterdam: John Benjamins Publishing.

Nagy, N., & Kochetov, A. (2013). Voice onset time across the generations: A cross-linguistic study of contact-induced change. In P. Siemund, I. Gogolin, M. E. Schulz & J. Davydova (Eds.), *Multilingualism and language diversity in urban areas: Acquisition, identities, space, education* (pp. 19-38). Amsterdam: John Benjamins Publishing.

Nava, E., & Zubizarreta, M. L. (2010). Deconstructing the nuclear stress algorithm: Evidence from second language speech. In N. Erteschik-Shir, & L. Rochman (Eds.), *The sound patterns of syntax* (pp. 291-316). Oxford: Oxford University Press.

Navarro-Tomás, T. (1914). *Manual de pronunciación española* (3rd ed.). Madrid, Spain: Editorial Hernando S.A.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1-19.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht, The Netherlands: Foris Publications.

Nibert, H. J. (2000). *Phonetic and phonological evidence for intermediate phrasing in Spanish intonation* (Doctoral dissertation). University of Illinois at Urbana-Champaign, USA.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189-234.

O'Grady, W., Lee, O., & Lee, J. (2011). Practical and theoretical issues in the study of heritage language acquisition. *Heritage Language Journal, 8*(3), 23-40.

Oh, J. S., & Nash, B. A. (2014). Attitudes and motivations of adult Spanish language learners: A comparison of heritage language learners and second language learners. *Heritage Language Journal, 11*(1), 29-44.

Oh, J. S., Jun, S., Knightly, L. M., & Au, T. K. (2003). Holding on to childhood language memory. *Cognition, 86*(3), B53-B64.

Ortega-Llebaria, M. (2006). Phonetic cues to stress and accent in Spanish. In M. Díaz-Campos (Ed.), *Selected proceedings of the 2nd conference of laboratory approaches to Spanish phonology* (pp. 104-118). Somerville, MA: Cascadilla Press.

Ortega-Llebaria, M., Gu, H., & Fan, J. (2013). English speakers' perception of Spanish lexical stress: Context-driven L2 stress perception. *Journal of Phonetics, 41*, 186-197.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in de-accented syllables. In P. Prieto, J. Mascaró & M. Solé (Eds.), *Segmental and prosodic issues in romance phonology* (pp. 155-176). Amsterdam/Philadelphia: John Benjamins.

Ortega-Llebaria, M., & Prieto, P. (2009). Perception of word stress in Castilian Spanish: The effects of sentence intonation and vowel type. In M. Vigário, S. Frota & M. J. Freitas (Eds.), *Phonetics and phonology: Interactions and interrelations* [*Phonetics and phonology: Interactions and interrelations*] (pp. 35-50). Amsterdam: John Benjamins.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic correlates of stress in central Catalan and Castilian Spanish. *Language and Speech, 54*(Pt 1), 73-97.

Otheguy, R. (2016). The linguistic competence of second-generation bilinguals: A critique of "incomplete acquisition". In C. Tortora, M. den Dikken, I. L. Montoya & T. O'Neill (Eds.), *Romance linguistics 2013: Selected papers from the 43rd linguistic symposium on romance languages (LSRL), New York, 17-19 April, 2013* (pp. 301-319). John Benjamins Publishing.

Otheguy, R., & Zentella, A. C. (2011). *Spanish in New York.* Oxford: Oxford University Press.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64*(3), B9-B17.

Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*, 399-410.

Peirce, J. W. (2007). PsychoPy--psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8-13.

Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping, 19*(3), 170-182.

Pérez, D. C. (2010). Heritage language instruction in the 21st century. *New Waves-Educational Research & Development*, *13*(1), 70-82.

Peyton, J. K., Ranard, D. A., & McGinnis, S. (Eds.) (2001). *Heritage languages in America: Preserving a national resource. Language in education: Theory and practice.* Washington, DC: Delta Systems; McHenry, IL: Center for Applied Linguistics.

Phinney, J. S., Romero, I., Nava, M., & Huang, D. (2001). The role of language, parents, and peers in ethnic identity among adolescents in immigrant families. *Journal of Youth and Adolescence, 30*(2), 135-153.

Polinsky, M. (1997). American Russian: Language loss meets language acquisition. In W. Brown, E. Dornisch, N. Kondrashova & D. Zec (Eds.), *Proceedings of the Annual Workshop on Formal Approaches to Slavic Linguistics* (pp. 370-406). Ann Arbor, MI: Michigan Slavic Publications.

Polinsky, M. (2000). A composite linguistic profile of a speaker of Russian in the US. In O. Kagan, & B. Rifkin (Eds.), *The learning and teaching of Slavic languages and cultures* (pp. 437-465). Bloomington, IN: Slavica.

Polinsky, M. (2006). Incomplete acquisition: American Russian. *Journal of Slavic Linguistics*, *14*, 161-219.

Polinsky, M. (2008). Gender under incomplete acquisition: Heritage speakers' knowledge of noun categorization. *Heritage Language Journal, 6*(1), 40-71.

Polinsky, M., & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass, 1*(5), 368-395.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics, 52*(1), 37-52.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421-435.

Potowski, K. (2004). Spanish language shift in Chicago. *Southwest Journal of Linguistics, 23*(1), 87-116.

Potowski, K. (2013). Language maintenance and Shift. In R. Bayley, R. Cameron & C. Lucas (Eds.), *The Oxford Handbook of sociolinguistics* (pp. 321-339). Oxford: Oxford University Press.

Prieto, P., & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics, 35*, 473-500.

Putnam, M. T., & Sánchez, L. (2013). What's so incomplete about incomplete acquisition?: A prolegomenon to modeling heritage language grammars. *Linguistic Approaches to Bilingualism, 3*(4), 478-508.

Quilis, A. (1971). Caracterización fonética del acento español. *Travaux De Linguistique Et De Littérature, 9*, 53-72.

Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Biblioteca Románica Hispánica, Gredos.

Quilis, A., & Esgueva, M. (1983). Realización de los fonemas vocálicos españoles en posición fonética normal. In M. Esgueva, & M. Cantarero (Eds.), *Estudios de fonética I* (pp. 137-252). Madrid: Consejo Superior de Investigaciones Científicas.

Quintanar-Sarellana, R., Huebner, T., & Jensen, A. (1993). Tapping a natural resource: Language minority students as foreign language tutors. In B. J. Merino, H. T. Trueba & F. A. Samaniego (Eds.), *Language and Culture in Learning: Teaching Spanish to Native Speakers of Spanish* (pp. 208-221). London: Falmer.

Ramírez, A. G. (1988). Spanish in the United States. In E. Acosta-Belen, & B. R. Sjostrom (Eds.), *The Hispanic experience in the united states: Contemporary issues and perspectives* (pp. 187-206). New York: Praeger.

Rao, R. (2014). On the status of the phoneme /b/ in heritage speakers ofSspanish. *Sintagma, 26*, 37-54.

Rao, R. (2015). Manifestations of /bdg/ in heritage speakers of Spanish. *Heritage Language Journal*, *12*(1), 48-74.

Recasens, D. (1999). Acoustic analysis. In W. J. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, Data and techniques* (pp. 322-336). Cambridge: Cambridge University Press.

Reinhart, T. (2006). *Interface strategies: Optimal and costly computation*. Cambridge, MA: MIT Press.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of crosslinguistic speech perception in native English speakers. *Behavioural Brain Research*, *111*, 13-23.

Robles-Puente, S. (2014). *Prosody in contact: Spanish in Los Angeles*. (Doctoral dissertation). University of Southern California, USA.

Romanelli, S., & Menegotto, A. C. (2015). English speakers learning Spanish: Perception issues regarding vowels and stress. *Journal of Language Teaching and Research, 6*(1), 30-42.

Ronquest, R. (2012). An acoustic analysis of heritage Spanish vowels (Doctoral dissertation). Indiana University, USA.

Ronquest, R. (2013). An acoustic examination of unstressed vowel reduction in heritage Spanish. In C. Howe, S. E. Blackwell & M. L. Quesada (Eds.), *Selected proceedings of the 15th hispanic linguistics symposium* (pp. 157-171). Somerville, M: Cascadilla Proceedings Project.

Rothman, J. (2007). Heritage speaker competence differences, language change, and input type: Inflected infinitives in heritage Brazilian Portuguese. *International Journal of Bilingualism, 11*(4), 359-389.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism, 13*(2), 145-155.

Rumbaut, R. G., Massey, D. S., & Bean, F. D. (2006). Linguistic life expectancies: Immigrant language retention in Southern California. *Population and Development Review, 32*(3), 447-460.

Ryan, C. (2013). Language use in the united states: 2011. Washington, DC: U.S. Census Bureau. Retrieved from http://www.census.gov/content/dam/Census/library/publications/2013/acs/acs-22.pdf

Saadah, E. (2011). *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic* (Doctoral dissertation). University of Illinois at Urbana-Champaign, USA.

Saalfeld, A. K. (2009). *Stress in the beginning Spanish classroom: An instructional study* (Doctoral dissertation). University of Illinois at Urbana-Champaign, USA.

Saalfeld, A. K. (2012). Teaching L2 Spanish stress. *Foreign Language Annals, 45*(2), 283-303.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Portuguese and English. *Journal of Phonetics, 25*, 421-436.

Sasse, H. J. (1987). The thetic-categorical distinction revisited. *Linguistics, 25*, 511-580.

Sebastian-Galles, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition, 72*(2), 111-123.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching, 10*(1-4), 209-232.

Selkirk, E. (1984). *Prosody and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, E. (1995). Sentence prosody: Intonation, stress and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550-569). Cambridge: Blackwell.

Simon, E. (2010). Child L2 development: A longitudinal case study on voice onset times in word-initial stops. *Journal of Child Language, 37*(1), 159-173.

Sluijter, A. M., & van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in dutch and american english. In H. T. Bunnell, & W. Idsardi (Eds.), *Proceedings of the 4th international conference on spoken language processing* (pp. 630-633). New Castle, DE: Citation Delaware.

Sluijter, A. M., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America, 100*, 2471-2485.

Sluijter, A. M., van Heuven, V. J., & Pacilly, J. J. (1997). Spectral balance as a cue in the perception of linguistic stress. *The Journal of the Acoustical Society of America, 101*(1), 503-513.

Sorace, A. (2000). Syntactic optionality in non-native grammars. *Second Language Research*, 16, 193-202.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*, 412-432.

Stangen, I., Kupisch, T., Proietti Ergün, A. L., & Zielke, M. (2015). Foreign accent in heritage speakers of turkish in germany. In H. Peukert (Ed.), *Transfer effects in multilingual language development* (pp. 87-108). Amsterdam: John Benjamins Publishing.

Steedman, M. (2014). The surface-compositional semantics of english intonation. *Language, 90*, 2-57.

Stepler, R., & Brown, A. (2015). *Statistical portrait of Hispanics in the United States*, 1980–2013. Washington, DC: Pew Hispanic Center. Retrieved from http://www.pewhispanic.org/2015/05/12/statisticalportrait-of-Hispanics-in-the-United-States-1980-2013-Trends/

Stevens, G. (1992). The social and demographic context of language use in the United States. *American Sociological Review*, 57(2), 171-185.

Stevens, J. (2011). Vowel duration in second language Spanish vowels: Study abroad versus at-home learners. *Arizona Working Papers in SLA & Teaching, 18*, 77-104.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 3-45). Timodium, MD: York Press.

Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, *9*(1), 97-114.

Svantesson, J. (1986). Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics, 14*(1), 53-70.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., et al. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*, 247-250.

Tadmor, C. T., Galinsky, A. D., & Maddux, W. W. (2012). Getting the most out of living abroad: Biculturalism and integrative complexity as key drivers of creative and professional success. *Journal of Personality and Social Psychology, 103*(3), 520-542.

Takagi, N. (2002). The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. *Journal of the Acoustical Society of America, 111*, 2887-2896.

Tallon, M. (2011). Heritage speakers of Spanish and foreign language anxiety: A pilot study. *Texas Papers in Foreign Language Education, 15*(1), 70-87.

Taylor, P., Lopez, M. H., Martínez, J. H., & Velasco, G. (2012). *When labels don't fit: Hispanics and their views of identity*. Washington, DC: Pew Hispanic Center. Retrieved from http://www.pewhispanic.org/2012/04/04/when-labels-dont-fit-hispanics-and-their-views-of-identity/

Torreira, F. (2007). Tonal realization of syllabic affiliation in Spanish. In J. Trouvain & W. Barry (Eds.), *Proceedings of the 16th International Conference of Phonetic Sciences* (pp. 1073-1076). Saarbrücken, Germany.

Torreira, F., Simonet, M., & Hualde, J. I. (2014). Quasi-neutralization of stress contrasts in Spanish. In N. Campbell, D. Gibbon & D. J. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 197-201). Dublin: Trinity College.

Tse, L. (1998a). Ethnic identity formation and its implications for heritage language development. In S. Krashen, L. Tse & J. McQuillan (Eds.), *Heritage Language Development* (pp. 15-29). Culver City, CA: Language Education Associates.

Tse, L. (1998b). Seeing themselves through borrowed eyes: Asian Americans in ethnic Ambivalence/Evasion. *Multicultural Review, 7*(2), 28-34.

Ulbricha, C., & Ordin, M. (2014). Can L2-english influence L1-german? The case of post-vocalic/r/. *Journal of Phonetics, 45*, 26-42.

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In J. K. Preyton, D. A. Ranard & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 37-80). McHenry, IL: The Center for Applied Linguistics and Delta Systems.

Vallduví, E. (1990). The role of plasticity in the association of focus and prominence. In Y. No & M. Libucha (Eds.), *Proceedings of the Eastern States Conference on Linguistics 7* (pp. 295-306). Columbus, OH: Ohio State University.

van Maastricht, L., Krahmer, E., & Swerts, M. (2015). Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. *Language Learning, 66*(1), 124-158.

Veltman, C. J. (1983). *Language shift in the United States*. Berlin: Mouton Publishers.

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: The role of prosodic constituents. *Journal of Child Language, 29*(2), 225-250.

Werker, J. F., & Polka, L. (1993). Developmental changes in speech perception. New challenges and new directions. *Journal of Phonetics, 21*, 83-101.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*(3), 233-251.

Wieman, L. A. (1976). Stress patterns in early child language. *Journal of Child Language, 3*, 283-286.

Willis, E. W. (2005). An initial examination of southwest spanish vowels. *Southwest Journal of Linguistics*, *24*, 185-198.

Yang, J., Fox, R. A., & Jacewicz, E. (2015). Vowel development in an emergent Mandarin-English bilingual child: A longitudinal study. *Journal of Child Language, 42*, 1124-1145.

Yeni-Komshian, G. H., Flege , J. E., & Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition, 3*(2), 131-149.

Yip, V., & Matthews, S. (2006). Assessing language dominance in bilingual acquisition: A case for mean length of utterance differentials. *Language Assessment Quarterly*, *3*(2), 97-116.

Zampini, M. L., & Green, K. P. The voicing contrast in English and Spanish: The relationship between perception and production. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 23-48). Oxford: Blackwell Publishers.

Zubizarreta, M. L. (to appear). Information structure and nuclear stress. In C. Féry, & S. Ishihara (Eds.), *Encyclopedia on information structure*. Oxford University Press.

Zubizarreta, M. L. (1998). *Focus, prosody, and word order*. Cambridge, MA: MIT Press.

Zubizarreta, M. L., & Nava, E. (2011). Encoding discourse-based meaning: Prosody vs. syntax. implications for second language acquisition. *Lingua*, *121*, 652-669.