THE PHONOLOGY OF THE ABSOLUTE INITIAL STATE OF L3 ACQUISTION

By

KYLE PARRISH

A Dissertation submitted to the

School of Graduate Studies

Rutgers, the State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in The Department of Spanish and Portuguese, Second Language Acquisition and Bilingualism

Written under the direction of

Joseph Casillas

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2023

ABSTRACT OF THE DISSERTATION

The phonology of the absolute initial state of L3 acquistion

by

KYLE PARRISH

Dissertation Director:

Joseph Casillas

This is where the body of your abstract goes. The abstract should summarize your work. The abstract for a dissertation or document may be longer than one page; word count is more important than page length in this section.

ACKNOWLEDGEMENTS

Couldn't find sections/acknowledgements.tex to include as your acknowledgements.

DEDICATION

I dedicate this dissertation to my loving wife, Marinna, and to my 3 fur children, Olive, Kevin, and Phyllis. I could not have done it without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

```r
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 -

## v tibble  3.1.1     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() -
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Chapter 1: Literature Review

1.1 Introduction

The difficulty of learning a new language in adulthood has a well documented history. This is especially true in the case of phonological acquisition. For example, many studies over the course of the past five decades have demonstrated that adult second language (L2) learners often produce and perceive the sounds of the target language in a non-native manner (**Flege & Eefting, 1988, Flege (1991)**). This difficulty arises, in part, because bilinguals often navigate complex communicative situations in which they produce and perceive speech from both of their languages in real time. Much less is known about the acquisition of a third language (L3A), particularly with regard to L3 production, perception, and phonological learning. The present dissertation explores the production patterns and perceptual categorization routines of adult bilinguals during the initial stages of L3 acquisition.

Empirical studies in L3 phonological acquisition have found evidence of multi-directional influence in language production (CITE), with the impact of previously learned languages on a third varying. In some cases, L2 influence on L3 production has been found, where in others, L1 influence or simultaneous influence of both languages on L3 production has been reported. It is unclear what factors could be at the root of these varied findings. On one hand, it might be sampling error. At large, empirical studies in L3 acquisition have used low samples to build models. One issue with this approach is related to sampling error. On the other, factors such as L2 and L3 proficiency, language dominance and choices of methods and analysis may play a role in the conclusions drawn by these studies.

The present dissertation aims to address the issue of low sample sizes and varied statistical analyses by, firstly, recruiting bilinguals who do not yet know a third language, and, secondly, by using more fitting statistical analyses made more appropriate by a larger sample size. It is likely that the relatively homogeneous populations of bilinguals are in greater supply that trilingual populations, particularly when suggested

methodological practices are to be used, such as the use of mirror-image groups. With the increase in sample size, more precise observations and conclusions may be drawn about the very starting point of L3 phonological acquisition which are far less likely to be explained by statistical limitations such as sampling error. To deal with the potential impact of proficiency and dominance on L3 production and perception, L2 proficiency was measured using the LexTALE in English (citation) and Spanish (citation). The Bilingual Language Profile (BLP) was used to measure language dominance and background (citation).

1.2 Models of third language acquisition

Research in third language acquisition has attempted to model the interplay between L1 and L2 language systems in the acquisition of a third. Among questions asked by third language models is whether the L1 or L2, or a combination of both languages, serves as the basis in L3 acquisition. This question is complicated in the context of multilingualism due to the widespread diversity in bilingual populations that include wide variation in ultimate attainment in adult L2 learners, and, in the case of phonological acquisition, wide variation in the production patterns of L2 segments. The predictions and theoretical underpinnings of models of L3 acquisition are covered in the following section, as well as the empirical evidence for each model.

1.2.1 The Cumulative Enhancement Model

The Cumulative Enhancement Model (CEM) (**Flynn et al., 2004**) was one of the first formal models of L3 acquisition in adulthood. The authors argued that L3 acquisition can provide unique insights about language learning that is not possible in L1 or L2 research alone. The key question in this article was whether the properties of the L1 maintain a privileged status in L3 acquisition, or if L3 acquisition is cumulative process, in which all grammatical properties of previously known languages impact subsequent language acquisition.

Within this seminal article, the authors conducted a study to provide evidence for

the newly proposed CEM. The study examined the production of restrictive relative clauses in adults and children who spoke L3 English (L1 Kazakh-L2 Russian). The results revealed similar performance on the experimental task by adults and children, and that the production of relative clauses was influenced by the participants' L2 Russian. This influence was facilitative, since this syntactic structure is common in Russian and English, but not Kazakh. The authors note, however, that the results of this study could also be explained by a special status for the *L2*, rather than the L1.

Berkes & Flynn (2010) conducted a further study to empirically test the predictions of the CEM. German L1 and Hungarian L1-German L2 speakers were tested in English (their L2 and L3). The authors found evidence that the L3 group performed better than the L2 group, when the syntactic structure in question was L1-like, oppositely from **Flynn et al. (2004)**, which found facilitation when L2 and L3 elements were similar. The authors argue that the results of these two studies taken together suggest that both the L1 and the L2 may influence the L3 grammar, and that these languages' influence help to produce target-like L3 productions, when the L3 has a common feature with either the L1 or the L2.

An issue with this model is non-facilitation, as argued by Rothman.

1.2.2 The L2 Status Factor Model

In contrast to the CEM, the L2 Status Factor Model (L2SF) predicts that the L2 will influence the L3 by default, due to the proposed cognitive similarity between the L2 and the L3. These proposed cognitive similarities stem from the Declarative-Procedural model, which posits that late learned languages are largely subserved by the declarative memory system, whereas early learned languages are subserved by procedural memory.

Evidence for the L2 Status factor has been reported...

Counterevidence to this model has also been found in empirical studies that would serve as the basis for subsequent models.

1.2.3 The Typological Primacy Model

The Typological Primacy Model, like the L2SF, predicts that a single language will influence the L3, but differs in that this language may be either the L1 or the L2. The choice of language is thought to be driven by individually perceived psycho-typology. In this view, the choice of language is also not thought to be conscious, but rather driven by input in L3 learning.

There has been some support in the literature for the TPM, but this evidence often simultaneously supports other models. For instance, in a recent systematic review of 92 studies in L3 acquisition, `@puig-mayenco_systematic_2020` concluded that, in support of the TPM, that either the L1 or L2 influenced L3 performance in 59 out of 92 studies. On the other hand, 29 of the studies out of 92 found that the L2 influenced the L3. However, these findings were not mutually exclusive; the authors coded 25 total studies as being explained both by L2 status and typology, meaning that the results of these studies reported that the L2 transferred to the L3, but could not rule out the possibility that psycho-typological transfer could also explain the results, since the studies did not use mirror image groups (i.e. L3 groups with the same languages, but the opposite order of acquisition).

Counter-evidence and issues with this model include lack of cross-linguistic support, and lack of specificity when it comes to the term "initial stages." For example, evidence for the TPM has been found primarily in studies which contain Romance languages and a non-Romance language. To the author's knowledge, no study has yet shown support for the TPM's predictions outside of language combinations which contain two romance languages and a non-romance language which also employ the criteria suggested by Rothman (CITE) of mirror-image groups. The studies coded in the systematic review by Puig-Mayenco et al. as providing evidence for L2 Status and Typological proximity cannot be said to conclusively provide evidence for the TPM for the same reasons that they do not provide conclusive evidence for the L2SF, since L2 status and typological primacy could explain the results. In short, the TPM does

not currently have support from a study which uses mirror-image groups and does not involve 2 romance languages.

A second potential shortcoming of the TPM is the vagueness associated with the term "initial stages" of L3 acquisition. Rothman (2015) describes this time as some 20-25 hours of instruction, before which time access to either langauge is possible. This vague criterion makes it quite difficult to derive the predictions of the TPM in the case of learners who first encounter an L3.

1.2.4 The Linguistic Proximity Model and the Scalpel Model

These models propose that both language systems may influence the L3, but also that they may not.

Unlike the TPM and the L2SF, the Linguistic Proximity Model

*(**westergaard__crosslinguistic__2017?**) suggests that transfer occurs on a property-by-property basis, rather than holistically. The LPM suggests that there is a "full transfer possibility," meaning that any individual structure may transfer at any time, but that it also may not. Scholars have argued that this vague prediction creates a problem in modeling L3 transfer acquisition (**bardel__l1__2020?**; **wrembel__multilingual__2020?**), since it is unclear when transfer of a particular structure occurs and when it does not. Likewise researchers have argued that, unlike the TPM and the L2SF, the LPM is not easily falsifiable (**bardel__l1__2020?**). Given the lack of predictive power of the LPM, it may only receive post-hoc support.*

Evidence for these models.. Westergaard et al. (2017), Mitronova & Westergaard (2018)

Issues come from lack of predictive power of the models.

1.3 Models of L2 phonological acquisition

Many accounts for L2 phonological acquisition exist. In general, these models all have in common that language-specific L1 categories drive L2 speech development. In the section that follows, a brief overview of each model, along with its evidence, will be covered.

1.3.1 The Speech Learning Model

The original SLM **Flege (1995)** focuses on the acquisition of segments in the L2 and emphasizes the importance of cross-linguistic similarity in L2 speech learning. In this view, new segments are predicted to be easier to learn, where segments which have close matches in the L2 will be much harder to acquire at a native-like level. Essentially, the SLM proposes that L1 and L2 sounds are linked through a process referred to as "interlingual identification," in that the L2 sound encountered as first exposure are seen as either good or bad phonetic variants of a native category **(Flege, 1995)**. In this view, acquisition of L2 segments occurs after increased exposure to these segments, where segmental complexity modulates the rate of learning **(Flege, 2021)**. Additionally, the SLM posits that allophonic variation and segment position matter, as opposed to the learning of a phoneme generalizing across positions.
For instance, **Iverson, Hazan, and Bannister (2005)** conducted a study in which they succesfully trained Japanese speakers to better identify /r/ and /l/ in word initial position, but that this training was not effective for word-medial or word-initial clusters containing liquids.

The original SLM also suggested that the age of first exposure (the younger the better), and L2 experience were important predictors to L2 learning success. **add studies of this**

The SLM makes a distinction between identification as opposed to categorization. The importance of this distinction involves the learning of phonetic variants during the learning process, which lead to new L2 phonetic category formations. For example,

using a two-alternative forced choice task, **Bohn and Flege (1993)** asked Spanish monolinguals to identify English stops and provided them with pre-voiced, short lag and long lag stops as tokens. Importantly, Spanish monolinguals identified english stops identified long lag /t/ as /t/, despite the lack of a long-lag category in their native inventory.

1.3.1.1 The revised SLM

The SLM was recently revised, and has updated some assumptions and predictions. The revised model maintained many of the same assumptions as the original, and importantly focuses on sequential bilinugals, where speech learning begins when the phonetic categories of the L1 have been established. According to **Flege (2021)**, the updated tenets of the Revised Speech Learning Model (SLM-r) are that, first, phonetic categories are formed based on their statistical regularities in the input, such that greater exposure to a particular phonetic cue should be correlated with the re-tuning of that cue towards the input. Second, all learners make use of the same learning mechanisms in L2 learning that they do in native language learning. That is, the SLM-r refutes the idea of a critical period (**CITE**) for speech learning, and suggests that non-native production in perception is due to differences the quality and quantity of input between native speakers and L2 learners and L1 effects. Another important update to the SLM-r is the idea that perception and production co-evolve. In other words, the SLM-r does not predict that perception will precede production.

The revised model has not yet received empirical support, due to its recent revision. It is outside the scope of the present dissertation to evaluate the predictions of the SLM-r when it comes to statistically driven re-tuning of phonetic categories. However, the phonetic categories that are in place in a bilingual at the first exposure to L3 learning could be treated as an integrated inventory when it comes to L3 learning. Based on the idea that the same mechanisms are in place in L1 and L2 phonetic category formation, the same mechanisms should be in place for L3 phonetic category formation. In this view, the acoustic similarity of an L3 segment to an L1 or L2

segment should predict how difficult learning an L3 segment will be. In this view, there is not reason to predict a blocking of one language system (L1 or L2) provided that these systems have well-established phonetic categories.

1.3.2 The Perceptual Assimilation Model

Relative to the SLM, the Perceptual Assimilation Model (SLM) focuses on the perception of sound contrasts by L2 learners (CITATIONS). The PAM involves various scenarios in which the cross-linguistic inventories of specific languages predict how easily L2 sounds will be to acquire. For instance, if a L2 learner's native language contains a contrast that also exist in the L2, then this contrast will be easily perceived in the L2 (two-category assimilation). In the event that two native language categories correspond the a single L2 category (single category assimilation), discrimination is predicted to be intermediate. In the event that a native category must be split into two categories (X assimilation) discrimination is predicted to be poorer.

Evidence for the predictions of the PAM has been found in studies which involve the presentation of sounds to naive learners of an L2 (first exposure), or to more experiened L2 learners. Experiments typically present the subjects with native-language vowel categories in written form and auditory stimuli of L2 vowel sounds and participants are tasked with choosing the closet mathcing native-language vowel category given the options, and to rate the goodness of fit of this decision.

This methodological paradigm has found evidence that **two category assimilation** is difficult for L2 learners. For example,

Evidence for PAM...

Escudero and Chládková (2010) found that Spanish L1 speakers assimilated SSBE /ae/ and /a/ to Spanish /a/ (that discrimination of this contrast was difficult).

Escudero et al. (2014) Salento Italian L1 speakers assimilated SSBE /ae/, /a/ and /ˆ/ to Salento Italian /a/.

Escudero & Vasiliev (2011) Spanish speakers assimilated Canadian English /ae/ and /e/ to Spanish /a/

Escudero & Williams (2011) Spanish listeners categorized Dutch /a/ and /a:/ in terms of their native /a/

The PAM in L3...

The present dissertation tests whether assimilation acquired during second language acquisition apply to third language sound perception. In other words, it tests whether L3 perception mirrors L2 perception in similar ways that has been observed in L3 production. Specifically, it tests whether L3 speakers categorize L3 sounds similary to L2 sounds, and whether phonetic discrimination of sounds that would be phonemic in the L1 is accessible in the perception of L3 words at first exposure, or whether there is an initial blocking or the L1 or L2 bias effect.

1.3.3 The Second Language Linguistic Perception Model

The Second Language Linguistic Perception Model (L2LP_ is a computational model of L2 speech learning and is similar to the PAM in that it focuses on sound contrasts as the basis for L2 speech learning, rather than single segments (**van Leussen & Escudero, 2015**, Escudero, 2005, 2009). In this view, it is difficult for L2 learners to make contrasts which are not present in their L1. In their revision of the L2LP, van Leussen & Escudero support this claim with several empircal studies. Namely, that these studies have provided evidence that L2 learners experience difficult with the contrasts of /r/ and /l/ in Japanese, (**Aoyama et al., 2004**) "beat" and "bit" in Spanish and Portuguese (**Flege et al., 1997; Rauber et al., 2005**) "bet" and "bat" in Dutch (**Broersma, 2005**). The authors argue that these results suggest that linguistic experience is at the heart of L2 learning. Specifically, cross-linguistic comparisons of L1 categories and L2 categories are thought to predict the ease of L2 category learning. The L2LP, unlike the PAM and the SLM, also aims to model the entire learning process, rather than the beginning stages of L2 speech learning.

This entire learning process is predicted computationally and is baed on Stochastic Optimality Theory (Boersma, 1998).

An important tenet of the postulate of the L2LP is the optimal perception hypothesis. Essentially, this proposal suggests that the initial perception of L2 sounds is the result of L1 acquisition. The developmental of L2 learners is predicted similarly to PAM, where a single category assimulation from PAM, in which a native category must be split, is called a new category scenario in the L2LP. This scenario is predicted by both models to be difficult for the L2 learner. On the other hand, the PAM and the L2LP predict that the case when two L1 sounds correspond well to two L2 sounds that discrimination of these sounds will be relavitely easier. This is referred to as a single category assimilation in PAM and a similar scenario in the L2LP. The final scenario of the L2LP is the subset scenario. In this case, a single L2 phoneme is perceived as two L1 categories. The same case is referred to as uncategorized or categorized-uncategorized in the PAM, and both models predict that discrimination of these sounds will be better than the case of new scenarios, but not as good as discrimination of subset scenarios.

Empirical support has been found for these predictions. In two-category assimilation/similar scenario, **Escudero and Boersma (2002)**, native Dutch speakers assimilated the Spanish /i/ to their native Dutch /i/, and the Spanish /e/ to their Dutch /I/.

However, the L2LP predicts that the similar (two-category assimilation) scenario may lead to inappropriate lexical contrasts, and argues that pre-lexical and lexical contrasts should be taken into account when it comes to sound discrimination. Evidence for this claim stems from studies which found that L2 learners could not perceive a contrast in lexical items that they could disriminate outside of lexical items (**Curtin et al., 1998**). Other studies show that some lexical items can also be relaibly distinguished by L2 learners that could not be told apart pre-lexically **(Weber and Cutler, 2004; Cutler et al., 2006; Escudero et al., 2008)**

As a result, an important tenet of the revised L2LP (**cite**) is that meaning-driven learning predicts the developmental path of L2 phoneme perception. The L2LP simulates the entire trajectory of L2 learning based on the various learning scenarios. In the revised L2LP, **van Leussen & Escudero, 2015** suggest that, despite past null results (**Weiand, 2007**), category reduction is possible when it is driven by meaning based learning.

The predictions of the L2LP which are important for the present dissertation are the Full Copying hypothesis (**Escudero, 2005**). If L3 phonological learning is another instance of L2 learning, then the L2LP and the TPM should share the prediction that the initial state of L3 learning is the end state of either L1 or L2 learning, but not both. In this view, bilingual participants who are first exposed to an L3 should produce and categorize L3 sounds as similarly to a single language, rather than producing or perceiving some sounds as L1-like and others as L2-like. Additionally, these participants' behavior on the experimental tasks should a) resemble either their own L2 behavior or L1 behavior or b) in the case of L1 influence, resemble a monolingual comparison group who is first exposed to an L2 (e.g. Spanish L1, English L2, exposed to German, and Spanish L1 exposed to German should behave similarly if Spanish is language which is "fully copied" at first exposure.)

1.4 Previous literature in L3 phonology

The models discussed to this point, with the exception of Wrembel's model, have focused largely on morpho-syntax. The findings in L3 phonology have largely yielded mixed results and do not quite have a comprehensive model. Despite the lack of a model, some patterns emerge from the body of research. The following sections cover the empircal studies that have been done in L3 production studies accross L3 proficiency levels, and the fewer studies done in L3 perception.

1.4.1 L3 Production Studies

The findings in empirical studies of L3 phonological cross-linguistic influence have varied. One of the first studies to examine progressive influence of the L1 or L2 on L3 production was the seminal case study of **Williams and Hammarberg (1993)**. This study elicited the production of an adult L1 British English, L2 German, and L3 Swedish speaker in the L2 and L3 upon her arrival in Sweden. The speech samples were rated for native-likeness by native speakers of German and Swedish respectively, with low ratings (i.e. non-nativeness) being elaborated upon. In the event of non-native speech, the raters guessed where the speaker in the recording might be from. The informant had near native productions in her L2 German, while her L3 Swedish was rated as being non-native like and to be German-accented. The experiment was repeated after 6 months in Sweden, however, and the Swedish raters then judged the informant's Swedish to be British English accented.

This study constituted evidence of an L2 status effect in the initial stages of L3 phonological acquisition, in which the second learned language influence L3 production. This notion has been called the 'foreign language effect' (**meisel_transfer_1983?**), which refers to the idea that speakers who learn a second non-native language are biased to sound unlike a native speaker of their native language.

The default L2 status effect has received some empirical support in the literature. In a study of global accent production, heavier L2 influence in L3 productions was found by L1 Polish, L2 German and L3 English speakers based on ratings of EFL instructors. Specifically, 53% of L3 productions were identified as having come from a German native speaker (**wrembel_l2-accented_2010?**). S vowel production (**kamiyama_acquisition_2007?**) vowel reduction and speech rhythm (**gut_cross-linguistic_2010?**).

Similar patterns of L2 influence have been found in VOT productions. (**llama_influence_2010?**) examined L3 Spanish VOT production by French-English mirror-image bilingual groups and found that both groups had L2-like productions of

the L3.

Williams & Hammarberg 1998

Hammarberg and Hammarberg 1993, 2005, Cenoz 2001 Hammarberg 2001 Bannert 2005, Jessner 2006, Fernandes-Boëchat and Siebeneicher Brito 2008,

Other findings in L3 production, however, have yielded mixed results. Several studies have found that acoustic properties of the participants' productions fall between L1 and L3 values, suggesting that both the L1 and the L2 have some influence on L3 productions, rather than solely one language. For instance, (**wrembel_vot_2014?**) measured VOT and aspiration in all languages of participants with two different language combinations: L1 Polish, L2 English, and L3 French; (2) L1 Polish, L2 English, and L3 German. The results showed that each language had a specific stop-value, and that the L3 VOT productions were intermediate, falling between the L1 and L2 values. Similarly, (**wrembel_cross-linguistic_2011?**) examined thirty-two learners of L3 French with L1 Polish and L2 English who were recorded reading lists of words in carrier phrases. As in previous studies (**wrembel_vot_2014?**), combined transfer from the L1 and the L2 in VOT productions was found.

Findings of combined L1 and L2 influence in VOT productions were also reported by (**wunder_phonological_2010?**) in L3 Spanish speakers, and by (**blank_transferencia_2009?**) in L3 English speakers who spoke L1 Brazilian Portuguese and L2 French. Other studies have found an L1 influence on production despite L3 proficiency (**wrembel_foreign_2012?**), or in advanced L3 learners (**llama_revisiting_2018?**).*

Parrish (2021) examined Mexican Spanish-English bilinguals who produced French words in isolation at first exposure to the language. The results found that the relative VOT of the L3 fell between their own L1 and L2 values, in line with previous research (citations). However, a subsequent analysis of the data revealed that wide individual variation existed in the data, in which some participants produced L3 French as L1

Spanish like, and other produced intermediate, L2-like values.

1.4.2 L3 Perception studies

Balas (2019), Liu

*Perception studies in L3 acquisition have been much more scarce than those on production, and primarily in young participants. One of few L3 perception studies in adults is (**liu_effects_2019?**), which examined the perceptual boundary of a VOT continuum in trilinguals. The participants were L3 Spanish speakers who spoke L2 English and L1 Chinese. Though the authors focused their analysis on regressive transfer and comparisons to all speakers, the reported boundaries in the study (n = 10, 28ms for Chinese, 24.6ms for English and 23ms for Spanish) suggest that the participants were using their English boundaries for L3 Spanish categorization.

Other studies in L3 perception, which have examined young trilinguals, have found a wide range of results. For instance, (**wrembel_cross-linguistic_2020?**) found that cross-linguistic influence is structure dependent and varies among individuals, while (**balas_perception_2019?**) found CLI to be modulated by a complex combination of factors such as markedness of the segment under examination, proficiency in the L2 and the L3, and L1 typology. Finally, another study on L3 perception in young learners (Polish L1, English L2, German, L3) revealed that the L2 predominantly influences L3 perception in both rhotic sounds and devoicing of word final stops (**wrembel_cross-linguistic_2020?**). These variable findings of the source of influence in L3 perception and production, including single language influence, combined languages influence, or structure-dependent influence, point to a lack of homogeneity in multilinguals.*

1.5 Methods and anylsis in previous work

The widely varied findings in previous work do not have a clear cause, but may be related to issues related to sampling issues combined with methodological choices. In this subsection, an overview of sample sizes in previous work will be given, followed

by potential issues related to small samples. Subsequently, common methods in the body of research used to analyze L3 data will be discussed, followed by their potential shortcomings.

Overall, the sample sizes in the body of research across domains have been quite small. An analysis of the sample used in the empircal studies in the systematic review of Puig-Mayenco et al. (2019) revealed that the average sample size per group was **X**, (sd = **Y**). In the empirical studies in L3 phonology cited in the body of this dissertation, the average sample size per group was **X** (sd = **Y**). For a full list of studies included in L3 group calculation, see **list**, and for a full list of studies in the Puig-Mayenco et al. systematic review see Puig-Mayenco et al. (2019). **figure** shows the distribution of sample sizes in L3 research by domain. As the figure suggests, the majority of L3 research has been constrained by relatively small sample sizes.

Unfortunately, the tradition in L3 research has involved model building with small samples. The use of small samples are associated with higher sampling error, and, as a result, a higher risk of type 1 error. In other words, a single study with a small sample cannot rule out the possibility that their results are due to sampling error, or a non-representative pool of participants from an assumed population distribution. As **brysbeart** argues, low samples lead to low statistical power, and in turn provide a metaphorical blurred picture of our desired outcome.

In addition to issues associated with low sample size, L3 research to date has used statistical methods which provide dubious evidence for their claims. Among these issues is the use of an Analysis of Variance (ANOVA) as a basis for determining whether groups or individuals perform experiment tasks in a practically equivalent manner rather than a statistical Test of Equivalence (**Lakens**). At the heart of this issue is a criticism which may apply to frequentist methods of statistical analysis in general; testing against the null hypothesis. If the null hypothesis is rejected, evidence is provided that the difference between or within groups is non-zero. On the other hand, if a non-zero difference is not found, there is not evidence for practical

equivalence. Such as assumption has been made in the L3 literature, and in particular in L3 model building.

For example, in his seminal article introducing the Typological Primacy Model, Rothman (2011) concluded that mirror-image groups of Spanish-English bilinguals did not perform differently on the L3 Brazilian Portuguese acceptability judgment task and took this as evidence for similar performance between the groups, and evidence of typological similarity effects in L3 judgments. There are two possible issues with this conclusion. First, sampling error, it can't be said that the low samples did not draw primarily from participants who were influenced by their Spanish, second a wide confidence interval from the low sample size would make providing evidence for statistical equivalence in a test of equivalence likely impossible. A power analysis of the descriptive data in Rothman (2011) reveals that the statistical power in a test of equivalence with the report sample sizes (12 and 15) is zero. This suggests that, given the means and standard deviations reported in this study and equivalence bounds of -.4 and .4 standard deviations from the mean, 100 simulations (random sampling from distributions with those means and standard deviations) of the repeated experiment yield a statistically equivalent result **0** percent of the time. On the other hand, non-zero differences are found **X** percent of the time. These post-hoc analyses of the data from Rothman (2011) suggest that sufficient evidence was not provided for practical equivalence between groups. **brysbeart** argues that, for between group comparisons, samples of at least 80 per group are necessary.

It seems that large sample sizes are quite difficult to find in L3 research. Plonsky (2015) suggests that combating issues associated with low sample size is possible in ways other than simply increasing sample size. For instance, he suggests that the use of descriptive statistics, including effect sizes and confidence intervals, would be an improvement in L2 research in general. This advice is in line with the idea that frequentist analysis, and linguistic research, has relied on p-values to determine the presence of a so-called statistical difference. Plonsky, along with others, have argued that the use of p-value alone to make real-world inferences in problematic due to issues

associated with sampling error and the presence of the magnitude of an effect.

In order to address these potential issues, the present dissertation recruited bilinguals, rather than trilinguals, at first exposure to a third language in order to pull from a likely higher and more homogeneous population of participants. This higher sample, coupled with the use of frequentist tests of equivalence and Bayesian inference allow for both a categorical and gradient interpreation of the data. In doing so, less reliance is put on a narrative interpretation of the results, and the results lend themselves to a more objective outcome.

Examples of non-main effects as equivalence:

with ancova "The non-significant value for the 2-way language group*level interaction (p = .410) suggests that the proficiency effect was statistically equivalent for the two language groups, Moreover, the 3-way language group*level*type of relative interaction (p = .088) gave a non-significant value, which suggests that the effect of proficiency is not significantly different across levels of the two studies and sentence type. " Berkes and Flynn (2010, p. 153)

L3 Study: HungarianL1/GermanL2/EnglishL3: Berkes and Flynn (2010, p. 156) As evidence for similar performace on all 3 types of relative clause production: lack of sig. p-value, no main effect or sig. pariwise comparison

1.6 Cross-lingusitic language features

In order to gain insights into the relative influence of a first and second language the present dissertation examines (relative) voice-onset time (VOT) in L3 production, the formant values of vowels in perception and production. In the following sections, the vocalic systems of Spanish, English and French, as well as their respective uses of voice-onset time, are overviewed.

1.6.1 Vocalic systems

The Spanish vowel space is the smallest of the present study and consists of 5 distinct vowel monopthings (**LIST THEM**) categorizations. English and French have a larger relative vowel space.

1.6.2 Use of Voice-Onset Time

The description of cross-lingusitic uses of VOT.

*Voicing in these segments can be measured acoustically using voice-onset time (VOT). VOT refers to the time interval between the release of a stop consonant and the onset of vocal fold vibration (**lisker__cross-language__1964?**). For example, in Spanish, the difference between [p] and [b] is manifested as a difference in VOT. Where [b] is voiced (a negative VOT), [p] is voiceless (a positive VOT). Languages which contain this distinction, in which the realizations of /p/ and /b/ are phonetically voiceless and phonetically voiced, respectively, are referred to as "true-voicing languages" (**lisker__cross-language__1964?**). Relevant to the present study, Spanish, French, and Hungarian fall under this category. English, on the other hand, contrasts stop consonants with only positive VOT via a long versus short lag distinction. That is to say, both /p/ and /b/ are phonetically voiceless, but /p/ is realized as a long lag stop and is aspirated, and /b/ is a short-lag stop that is not aspirated (**lisker__cross-language__1964?**).*

1.7 Bringing L2 speech models and L3 models together and evaluating predictions

By using the methods used to test the prediction of models of L2 phonological acquisition (The SLM, the PAM, and the L2LP), more nuanced evidence may be obtained to evaluate the predictive power of L3 models. With the revision of the SLM, at least three L2 speech models advocate for the study of naive or beginning learners in L2 speech learning research. The present dissertation adopts this point

of view and applies it in a third-language context in perception and production. By measuring the perception and production patterns of the first (or, at least, very early) exposure to a third language, coupled with measurements in each language, an the variation of cross-linguistic influence can be observed/studied in L3 perception and production. If, as the SLM-r predicts, category formation is driven by input and retuning of L1 categories in L2 acquisition, and the same mechanisms that are used in L1 phonetic category formation in L2 phonetic category formation, then it is likely that the SLM would predict that those same mechanisms are at play in L3 phonetic category formation, and that this process is guided by L3 input.

1.8 Perception

Following this logic, the question becomes whether L1 or L2 categories are retuned to L3 categories, and what conditions determine which language category is initially chosen and its rate of retuning. The present dissertation focuses on which category is chosen as the initial L3 category, while the rate of change in these categories is left for future research. One possibility which may influence whether an L1 or L2 category influences an L3 is the acoustic similarity of the L3 segment relative to an individuals' L1 and L2 phonetic categories. The TPM predicts that phonetic cues do play a role in L3 input parsing, in that they are parsed and used in order to make a decision of which language system to holistically transfer. However, the TPM would (likely) not predict that two segments would be assimilated to two distinct languages, and it would not predict that the same segment would be categorized as different sounds. Additionally, if one language holistically impacts the acquisition of a third at first exposure, then the behavior of L3 learners should resemble L2 learners, provided the L1 is the source of influence in L3 acquisition, or should be practically equivalent in a within-subject comparison. In the case of the predicted behavior of L2 influence on L3 productions, L2 production and perception should be practically equivalent in a within-subject comparison.

The predictions of the Linguistic Proximity Model lack predictive power, but would

be able to account for the same L3 sound being categorized differently by the same subject, and by different subjects, and for two different L3 sounds being categorized as an L1 and L2 category respectively.

Chapter 2: Production at first exposure to an L3

2.1 Overview

The present study investigates the production of words in a language by bilinguals that they do not yet know. In so doing, the present dissertation aims to investigate the very initial stage of acquisition by providing an overview of the initial relative influence of both L1 and L2. In this study, sequential, late, Spanish-English bilinguals of both order of acquisition will be given an aspirating L3 (German) or a true-voicing L3 (French) to shadow.

2.1.1 Research Questions

RQ1: Will Spanish-English bilinguals produce L3 words with Spanish or English like VOT?

RQ2: Will Spanish-English bilinguals produce vowels with F1 and F2 values more like English or Spanish?

2.2 Methods

2.2.1 Participants

8 total groups of participants will take part in the experiment. There will be 4 bilingual groups with opposite orders of acquisition, Spanish L1-English L2, and English L1-Spanish L2, with each of these groups hearing either L3 German or L3 French, and 4 monolingual English and monolingual Spanish groups, assigned to either French or German. The bilingual groups will contain 75 participants per group, where the monolingual groups will contain 50 participants per group, for a total of 500 participants.

2.2.2 Tasks

Oragnization paragraph: An elicted production task & A shadowing task in the L3.

2.2.2.1 Bilingual Language Profile

To measure language use and language dominance, the Bilingual Language Profile (BLP) will be used.

2.2.2.2 LexTALE

The LexTALE will be used to measure English and Spanish proficiency. The LexTALE is a lexical decision task, in which participants see either words or pseudowords on a screen one at a time. Participants are supposed to then decide whether the word presented is a real word in the language or a pseudoword. The task is intended to be a measure of vocabulary size and thus a proxy of proficiency. ADD SENTENCE JUSTIFYING THE LINK BETWEEN VOCAB AND PROF. Either the English and Spanish versions of the task will be used in the study, depending upon the L2 of participant.

2.2.2.3 Elicited Production Tasks

In the elicited production task, the participants read words which appeared on the screen one at a time and submitted their recording of the read words by clicking on a record button to begin the recording, a stop button to end the recording, and an upload button upon completion of the recording. Participants were able to listen to their recording after completing a trial and were able to re-record in the event of an error.

2.2.2.4 Shadowing Task

For the shadowing task, participants were given a French word on screen and were able to play a recording of that word up to three times. The participant would then record themselves repeating the word that was uttered one at a time in a procedure identical to the elicited production task. The experiment was programmed and given in the online platform Labvanced. In order to control for language mode, the language specific word lists were presented in separate sessions, and the order of the languages (Spanish

first or English first) was counter-balanced across participants. All participants ended the experiment with the L3 shadowing task.

2.2.3 Analysis plan

The results will be analyzed using a Bayesian Multilevel Regression where the outcome will by VOT and the fixed effect, categorical predictor will be language, and the fixed effect continous predictor variables LexTALE scores, and a dominance metric.

2.3 Simulated Results

Chapter 3: Vowel Production at first exposure to an L3

3.1 Research Questions

3.2 Methods

3.2.1 Tasks

3.2.2 Participants

3.3 Simulated Results

Chapter 4: Perception of L3 vowels

4.1 Research Questions

RQ1: Will Spanish-English bilinguals assimilate L3 sounds to L1 or L2 categories?

RQ2: Will Spanish-English bilinguals perceive differences on an acoustic continuum using L1 or L2 boundaries?

4.2 Methods

4.2.1 Participants

4 total groups of participants will take part in the experiment. There will be 2 bilingual groups with opposite orders of acquisition, Spanish L1-English L2, and English L1-Spanish L2, and monolingual English and monolingual Spanish participants. The bilingual groups will contain 75 participants per group, where the monolingual groups will contain 50 participants per group, for a total of 400 participants. These sample sizes were justified by a power analysis, in which a power level of .8 was desired (CITATION).

4.2.2 Tasks

4.2.2.1 AX Discrimination Task

The /i/ - /u/ continuum will be used, in which French /y/ is an intermediate. At this time, a second continuum is also being considered.

4.2.2.2 Category Identification Task

The purpose of this task is to identify whether French segments are assimilated to Spanish or English categories. Four conditions are proposed for vowels (1) The same vowel sound is found in both languages, such as /i/, which can be found in all 3 languages. (2) A sound which is found in English, but not in Spanish, such as /e/ in "bed," or the short /I/. (3) Only Spanish, such as the rounded back-vowel /o/, and

finally, (4) a vowel which is not present in English or Spanish, but is in French, such as /y/.

The stimuli will be created by splicing natural utterances of the French vowels together, in closed syllables, with a consonant (e.g. gVg). The stimuli will be played 5 repetitions each, which will create 20 tokens per participant.

4.2.2.3 Analysis plan

The results will be analyzed using a Bayesian Logistic Regression where the outcome of language choice (English or Spanish) will be modeled as a function of presented vowel sound, LexTALE score, and a continuous language dominance metric.

4.3 Simulated Results

REFERENCES