Methodological Issues in L3 Phonology

Jennifer Cabrelli Amaro
University of Florida

Abstract

The design and implementation of a meticulously developed methodology is the foundation of empirical research, as it allows the researcher to collect data and answer a research question in a logical and systematic way, while controlling for myriad possible intervening variables. In the field of second language phonology, there are several macro issues that surface in discussions of procedure and best practices, including the selection of participant populations, testing paradigms and stimuli development, and data analysis, and the ways in which they have been addressed thus far in the literature. This article focuses first on the application of these issues to the nascent field of L3 phonology, and then on the methodologies I have implemented to test acquisition of L3 phonology and its influence on the stability of native and non-native phonological systems, addressing the aforementioned issues.

1. Methodological issues in L3 phonology

1.1 Introduction

Just as the study of L2 acquisition has used the established field of L1 acquisition as a methodological point of departure, the field of L3 acquisition has used existing L2 phonology research in the same way. In spite of the small number of studies that have been published in the area (see Cabrelli Amaro 2012 for a review), the methodology employed has varied as much as it has in the study of L2 phonology. Reviewing the limited literature, there is a clear need to experiment with different methodologies in an effort to build a solid body of empirical research, always taking into account the macro issues that will be discussed here. In doing so, the possibilities of making direct comparisons across studies that have implemented vetted methodologies and of building a holistic picture of the processes behind the acquisition of a third language become more realistic. In the following discussion, I acknowledge that the issues of participant selection, test design and data analysis that apply to L2 phonology methodology in many ways are very similar for L3 methodology. However, there are some important differences that should be reviewed here before turning to the methodologies that I have implemented in my
investigation of L3 phonology. I will first address the selection of experimental and control groups, after which I will turn to the assessment of language experience and proficiency. I will then focus on issues related to testing paradigms, and will end the section with a brief discussion of methods of data analysis in L3 phonology research.

1.2 Selection of experimental and control groups

Two of the issues subject to the most debate and criticism are the selection and categorization of participants and the use of control groups, both of which will be highlighted in this section. The number of studies that do not provide background information or administer objective proficiency measurements is concerning, given that both measures are paramount in constructing homogeneous participant groups. I recognize that a truly homogenous group is an impossibility given the variation among L2 and especially L3 learners, but there are a number of ways to control for confounding variables that fall out of such variation.

1.2.1 Experimental groups

First and foremost is the need to consider a variable not found in L2 phonology research: the existence of an established non-native phonological system in addition to the native phonological system. While addressing all possible learner variables in L2 participant selection is already daunting, imagine the task at hand once the number of variables has virtually doubled. The presence of more than one system comprising the L3 initial state has several implications for the selection of participants, including proficiency in the L2, whether the L2 is still in the process of being acquired or has reached a so-called steady state, potential modification of the L1 and L2, and whether the context of acquisition of the L2 and L3 has been formal or naturalistic.

Another factor to consider is the number of experimental groups necessary to test a particular hypothesis or hypotheses. Given that the source of transfer is one of the main questions in L3 phonology, there is a need in cases of L3 initial stages research to use more than one experimental group so as to tease apart a possible privileged status for the L2 from typological similarity as the driving force behind transfer. A practice primarily used in L3 morphosyntax studies, Llama, Cardoso & Collins’ (2010) study of L3 Spanish VOT was the first to implement an L1/L2 mirror-image methodology in which one group was made up of L1 English/L2 French speakers and the other, L1 French/L2 English speakers. This practice must also be implemented when investigating regressive transfer, that is, transfer that originates in the L3 and affects an existing linguistic system, if the goal of the research is to investigate what determines transfer from an L3 to an L2 and/or L1.

Potential deterministic factors include the status of the existing system (L1, L2) and typological proximity, among others.

A final issue to consider is that arises given the criteria to be fulfilled is the difficulty in finding subject pools large enough to design a study with sufficient statistical power. This is especially challenging when the intent is to look at the initial stages of the L3, given the small window of time in which data can be collected. Few researchers have put together samples that go beyond single digits (but cf. Wrembel 2010, who created a corpus; and Llama, Cardoso & Collins 2010, who had 22 participants), and thus collaboration will be necessary to make larger subject pools available and to eventually create corpora that can be used by researchers worldwide.

1.2.2 Control groups

The adequate selection of a baseline measure is also fundamental to the study of L3 phonology. In several studies (e.g. Gut 2010, Llama, Cardoso & Collins 2010), L3 learners have been treated as cases of multiple monolinguals and monolingual production values are used as a baseline measure, although it has been shown (e.g. Cruz-Ferreira 2010), as for late L2 learners (e.g. Fowler, Sramko, Ostry, Rowland & Hall 2008), that monolingual productions are not comparable to those of speakers with multiple systems that yield bidirectional influence. As Hopp & Schmid (2011) note, “the choice serves to move the yardstick of nativelikeness to a point which may, by definition, be out of reach for most bilinguals” (p.4).

Ideally, properties should be analyzed in the L1, L2 and L3 of each learner and compared against one another over time, such that the learner acts as his or her own control. Doing so informs us of what the learner has available for transfer from the L1 and crucially the L2. When investigating regressive transfer, this is especially important, since it is necessary to have evidence of what the learner’s systems looked like at the L3 initial state. Otherwise, how would you know what has changed? In the case that logistics do not allow for a longitudinal investigation, the ideal control group will consist of native speakers of the L3 that are advanced speakers of the experimental groups’ L1 and L2. Given the ample evidence of bidirectional cross-linguistic interaction in the phonological systems of both early bi/multilinguals (e.g. Fabiano & Goldstein 2005) and late bi/multilinguals (e.g. Bullock, Toribio, Gonzalez & Dalola 2006), employing speakers with these profiles as a control might provide a more accurate glimpse of the best possible scenario of what an end state might look like for a non-native speaker of the L3 under investigation.
1.3 Assessing language experience and proficiency

Researchers must take a broad history of participants' language background, starting with a questionnaire such as that used by Flege & Mackay (2011). A detailed language history provides vital information regarding variables related to foreign accent (see Jesney (2004) for a review of these variables). Information should minimally include age of acquisition, length of residence in the country where the L2 is spoken (in the case that the learner is in an immersion environment), education in the L1 and L2, self-reported use (context and frequency) of the L1 and L2, as well as scalar self-ratings of reading, writing, speaking and aural comprehension in the L1 and L2 as well as an L3/Ln where applicable. Concerning the manner in which the L2 was acquired, metalinguistic awareness often yielded by formal L2 acquisition can complicate direct comparisons between L3 learners with different contexts of L2 acquisition, and this information is vital for control of such a variable. Finally, one of the most important elements of a background questionnaire that is often left out is that of other languages spoken. These languages and their phonological inventories are then not taken into consideration and can quite easily be an intervening variable of which the researcher is unaware.

In conjunction with background information and more subjective self-ratings, an objective proficiency measurement is necessary to determine the group to which a participant belongs and is crucial to the validity of any study in which proficiency is a variable. In addition, phonological proficiency (perception and production) testing of the properties under investigation in the L1 and L2 must be employed as a critical complement to a global proficiency measurement, and most L3 phonology research does not take this into account (but cf. Tremblay 2007). There is already research published that acts as a warning against failure to do so: in a study of L3 VOT's, Llama, Cardoso & Collins (2010) used vocabulary tests to test overall proficiency, following the assumption that the number of words known is directly related to a learner's proficiency. However, they did not specifically test phonological proficiency, and discovered post-hoc that one of the learner groups might not have acquired the L2 phonological system completely.

1.4 Testing paradigms

The issues that arise in L3 phonology testing methodology are comparable to those in L2 research, but again, the addition of another linguistic system to the mix complicates the experimental design. Regardless of the paradigm(s) selected, when deciding upon a property to use to investigate the source of transfer to an L3, the researcher will need to develop tests in all three languages. Given the individual variation among L3 learners, and particularly successive (as opposed to simultaneous) bilingual learners, it is necessary to test the L1 and L2 as well as the L3 to be sure one is observing what s/he intends, and crucially to be certain that the L2 properties under investigation have been acquired at the time of testing. While some studies may aim to compare L2 learners of varying proficiency levels acquiring an L3, others may intend to test near-native L2 learners at the L3 initial stages. L1 attrition research has established that modification of a previously acquired linguistic system can occur over time (e.g. Bullock, Toribio, González & Dalola 2006, Lord 2008, Major 1992 *inter alia*), and thus it is necessary to test the state of the L1 as well. While this has become common practice in the more established area of L3 morphosyntax, Tremblay (2007) is the only study to date to report data from all three languages. To make direct comparisons across languages, a master set of tokens should be created, composed of nonce words, and each language should be tested on separate occasions. Another challenge arises when testing source(s) of progressive transfer. To execute the mirror image methodology that has been mentioned, the phenomenon needs to be present (or absent) in the L1 and L3, but crucially not in the L2. In the case of regressive transfer, the property investigated in the L3 must present differently in the L1 and L2 so that influence can be measured in terms of modification to the representations in the existing phonological systems.

There is a general consensus among L2 phonologists that investigation of the processes of both perception and production is necessary to fully understand the nature of the acquisition of a non-native phonological system, although there is disagreement about whether results from perception and production research can be compared. To my knowledge, L3 phonology researchers have only tested production, and have relied on paradigms frequently used in L2 phonology research. Such paradigms have ranged from simple elicitation tasks such as word lists (e.g. Llama, Cardoso & Collins 2010, Tremblay 2011) and a delayed repetition task (Tremblay 2011) to more open-ended tasks such as spontaneous speech samples (e.g. Gut 2010, Hammarberg & Hammarberg 2005) and story retelling (e.g. Gut 2010). While there are plusses and minuses for each task type, the use of various task designs is potentially very positive for the field, since we can assume that once data from varying tasks converge, findings will more accurately reflect the representations of non-native phonological systems.

1.5 Data analysis

New challenges present themselves once data has been collected. The method of data analysis might differ depending on the objectives, but I will focus on that which is most applicable to the study I will present here, the implementation of acoustic measurement and/or use of trained raters. Once criteria for measurement of any relevant acoustic elements have been decided upon, acoustic measurement can be a rather objective way to examine quantifiable elements of L2 production, and Zampini (2008) proposes that it can provide parameters for natural speech
characteristics and inform areas of differences between L1 and L2 speakers (and in this case, L3 speakers), such that results can be used to formulate new or similar research questions. In addition to analysis of segmental phenomena, it is a particularly effective way to analyze suprasegmental data such as intonation and speech rate (e.g. Gut 2010).

The other option for analysis is the more subjective use of expert raters to track and/or rate phenomena in a speech sample. The key is to ensure good interrater reliability, indicative of the fact that listeners share the same response to what they hear. Unfortunately, ‘good’ interrater reliability is not always quantified in published research, but should generally be above 70%. The use of acoustic analysis and perceptual judgments by trained judges in tandem is potentially very valuable, as employment of both methods allows for a balance between the inherent subjectivity of perceptual judgments and the bias that can come from sole use of acoustic analysis.

As seen, when designing methodology for the study of L3 phonology, the variables increase as the number of languages in play increase and there are then myriad factors to consider when attempting to paint an accurate picture of L3 phonology. I have proposed that testing of the L1, L2, and L3, perception measurements and data analysis combining acoustic analysis and perceptual judgments by trained raters are all considerations that should be made going forward. While this selection is by no means exhaustive, it presents a solid foundation from which progress in this field can continue, and I will now outline the way in which I have addressed these macro issues of participant selection, testing paradigms and data analysis in my investigation of L3 influence on native and non-native phonological systems.

2. A methodology to test the Phonological Permeability Hypothesis

2.1 Introduction

In what follows, I will describe how the L3 phonology methodology described above can be implemented. The methodology for the study that I outline here was designed to test the acquisition of Brazilian Portuguese (BP) as an L3 by bilingual speakers of English and Spanish, and specifically to investigate potential regressive phonological influence from BP to Spanish. The hypothesis I test is called the Phonological Permeability Hypothesis (PPH) (Cabrelli Amaro & Rothman 2010), and maintains that adult L2 phonological systems, even when they evidence nativelike target constraint rankings (assuming an Optimality Theoretic framework) are insurmountably different from native systems with regards to stability, a position that can be supported or falsified by comparing their susceptibility to cross-linguistic influence when a novel language of typological proximity is introduced as an L3.

To test the PPH, I investigated two phenomena: the stop/continuant alternation, present in Spanish but not in BP, and word-final vowel reduction, present in BP, but not in Spanish. In Spanish, underlyingly voiced stops surfaces as approximants in postvocalic position, within and across word boundaries. However, in BP the underlying voiced stops /b d g/ categorically surface as [-continuant] in postvocalic position. This is exemplified in (1).

(1) Spanish and BP postvocalic stops

\[
\begin{array}{lll}
| \text{Spanish} | \text{BP} | \\
|----------------|----------------|
| /b/ cabra       | [ka.ˈbra]      | [ka.bra]       & ‘goat’ \\
| /d/ padre       | [pa.ˈdre]      | [pa.drə]       & ‘father/priest’ \\
| /g/ agua        | [ˈa.gwa]        | [a.gwa]        & ‘water’ \\
\end{array}
\]

The second phenomenon, vowel reduction (also known as vowel neutralization), is present in BP but not in Spanish and is conditioned by stress, a suprasegmental property. BP vowel reduction is a merging of members of the vocalic inventory in unstressed syllables, and reduction of the stressed vowel inventory /e e a o o u/ occurs in accordance with the grade of weakening of the syllable. Pretonic syllables evidence a neutralization of tense and lax mid vowels, resulting in the merging of /e e / to [e] and the merging of /o o/ to [o]. In posttonic final position, considered the weakest, the mid vowels are neutralized to [o o] and /e raises to [e], resulting in a surface form inventory of [e i o]. The present study focuses on reduction in posttonic final position, which does not occur in Spanish, whose 5-vowel inventory /a e i o u/ is fully realized independent of position. This is exemplified in (2).

(2) BP and Spanish word-final vowels

\[
\begin{array}{lll}
| \text{BP} | \text{Spanish} | \\
|----------------|----------------|
| /u/ casa       | [ka.ˈza]       | [ˈka.sa]       & ‘house’ \\
| /e/ frase      | [ˈfra.se]       | [fra.se]       & ‘phrase’ \\
| /o/porto       | [ˈpar.to]       | [ˈpar.to]      & ‘birth’ \\
\end{array}
\]

I now turn to the methodological issues that have been the focus of this discussion as they relate to testing of the PPH.

2.2 Experimental and control groups

Given the predictions of the PPH, it was necessary to test and compare two groups of L3 learners. Group 1 was composed of native speakers of English that acquired Spanish as an L2 after the offset of a critical/sensitive period and have advanced proficiency (see Section 2.2 for a detailed account of their proficiency assessment).
Group 2 is composed of simultaneous bilingual speakers of English and Spanish. There are two subgroups within both group 1 and group 2: learners with intermediate BP proficiency and learners with advanced BP proficiency. These experimental groups composed the optimal testing scenario under which this hypothesis could be tested, because by using two groups across which all three languages remain constant, the only variables that separated one group from the other are timing and context of acquisition of Spanish. Table 1 shows the experimental groups.

<table>
<thead>
<tr>
<th>L3 BP proficiency</th>
<th>Group</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate</td>
<td>1 Successive English/Spanish bilinguals</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2 Simultaneous Spanish/English bilinguals</td>
<td>15</td>
</tr>
<tr>
<td>Advanced</td>
<td>1 Successive English/Spanish bilinguals</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>2 Simultaneous Spanish/English bilinguals</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 1. Experimental groups

The Spanish control group (n=15) consisted of L1 Spanish speakers who had acquired English after the offset of the critical/sensitive period and had advanced English proficiency. The BP control group (n=15) consisted of L1 BP speakers who had acquired both English and Spanish after the offset of the critical/sensitive period and had advanced proficiency in both languages.

2.3 Proficiency assessment

Learners minimally had to have advanced proficiency in Spanish, which was based on three criteria: a 50-item written Spanish proficiency test used extensively in L2 research over the last decade (e.g. Montrul & Slabakova 2003, Rothman & Cabrelli Amaro 2010), ratings by three trained native-speaker judges of overall foreign accent on a 1-9 scale based on a 20-second excerpt of a spontaneous speech sample, and production of the two phenomena under investigation in the speech sample. The control participants each provided a speech sample that was also submitted to a foreign accent rating, and used as a point of comparison with the experimental participants for production of the speech phenomena under investigation. While it is optimal to measure perception, a perceptual measurement was not used to avoid priming. To be included in the study, all participants had to score a minimum of 40/50 on the Spanish proficiency test, a minimum average of 7/9 on the accent rating scale, and production of the two phenomena in the speech sample in comparison to the controls’ speech samples could not be statistically significant. To assess BP proficiency, participants with the equivalent of at least one semester of BP exposure took a 100-point proficiency test used in Rothman & Iverson (2011), and a speech sample was again taken to measure overall phonological proficiency and production of the two phenomena. Intermediate participants had a written test score of 60-79, a score of 4-6 on the 9-point scale, and production of the phenomena compared with that of the controls was statistically significant. Advanced L3 BP learners scored a minimum of 80 on the proficiency test, a minimum of 7 on the accent scale, and production of phenomena was not statistically significant compared with control production. In the case of a mismatch between the written test score and speech sample evaluation, the speech sample evaluation was used to determine proficiency. For the purpose of statistical power, the ideal number of participants for each subgroup was 15, or 60 experimental participants total. At the time in which this went to press, testing was ongoing. Tables 2 and 3 show the criteria for advance proficiency in Spanish and BP respectively.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Criterion scale/value</th>
<th>Minimum requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proficiency test</td>
<td>0-50</td>
<td>40</td>
</tr>
<tr>
<td>Accent rating</td>
<td>1-9</td>
<td>7</td>
</tr>
<tr>
<td>Production: stop/continuant alternation</td>
<td>statistical significance when compared with control</td>
<td>difference as compared with control must be statistically significant</td>
</tr>
<tr>
<td>Production: vowels</td>
<td>statistical significance when compared with control</td>
<td>difference as compared with control must be statistically significant</td>
</tr>
</tbody>
</table>

Table 2. Advanced Spanish proficiency criteria

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Criterion scale/value</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proficiency test</td>
<td>0-100</td>
<td>60-79</td>
<td>80-100</td>
</tr>
<tr>
<td>Accent rating</td>
<td>1-9</td>
<td>4-6</td>
<td>7-9</td>
</tr>
<tr>
<td>Production: stop/continuant alternation</td>
<td>statistical significance when compared with control</td>
<td>difference is statistically significant</td>
<td>difference is not statistically significant</td>
</tr>
<tr>
<td>Production: vowels</td>
<td>statistical significance when compared with control</td>
<td>difference is statistically significant</td>
<td>difference is not statistically significant</td>
</tr>
</tbody>
</table>

Table 3. BP proficiency criteria
2.4 Methodology

The following battery of tests, which includes a delayed repetition task, a syllable concatenation task, and a naturalness decision task, was designed to examine participants' perception and production of Spanish phonological phenomena and novel BP phonological phenomena at different points throughout the L3 acquisition process, the comparison of which will contribute towards the verification or falsification of the PPH. Tests in Spanish and BP were administered on different days at least one day apart to control for language mode (see Grosjean 1998 for a discussion), and the order of languages tested was counterbalanced across participants. Before testing, a language background questionnaire was completed and participants were recorded for 10 minutes conversing with a native speaker. The tasks were then administered and task order was counterbalanced across participants, although production always preceded perception testing to limit priming effects. A master set of stimuli was created for use across the production and perception tests in both Spanish and BP, consisting of disyllabic nonce words presented auditorily to avoid lexical and orthographical interference.

2.4.1 Production testing

2.4.1.1 Production test 1: Guided spontaneous speech

As mentioned, testing began with a 10-minute recorded conversation between the participant and a native speaker of the language being tested. In addition to providing a speech sample for rating of overall phonological proficiency and subsequent isolation and analysis of individual occurrences of production of the phenomena under investigation, this task was fundamental in moving the participants into the relevant language mode. While conversation was spontaneous, topics covered were consistent across participants. To minimize potential rater bias resulting from grammatical errors, topics were chosen requiring grammatical structures of minimal difficulty.

2.4.2.2 Production test 2: Delayed repetition

A delayed repetition task was used to obtain a baseline measurement of participants' production and importantly to avoid visual presentation of the stimuli and possible orthographic interference. A series of trials was presented auditorily, in which a target word was presented at the end of the carrier phrase, followed by the distracter question ¿En referencia a qué? ¿En referencia a qué?, minimizing imitation effects and requiring the answer to be the target word in the carrier phrase. A 3 (place of articulation) x 2 (word position) design was used to create the stimuli to test the stop/continuant alternation word medially and across word boundaries, and 5 tokens were created for each type (n = 30). All tokens had a CaCa structure and the final segment of the carrier phrase was also /ai/. To determine whether the participants had acquired the full range of the alternation and to observe possible modifications to this alternation, 15 additional tokens were included with a CanCa/CamCa structure, given acoustic and EPG evidence that a nasal consonant conditions a high degree of constriction in a proceeding underlying voiced stop (e.g. Hualle, Shosted & Scarpone 2011). To test vowel reduction, 10 disyllabic nonce words with a word-final front vowel ([e] in Spanish, [I] in BP) and 10 with a back vowel ([o] in Spanish, [o] in BP) were selected. The experimental stimuli (n = 65) and fillers (n = 40) were presented in random order by E-prime, following a 5-item practice session and a 5-item warm-up that did not form part of the data set for analysis.

2.4.2.3 Production test 3: Concatenation

A concatenation task originally used to test L2 stress was adapted from Guion, Clark, Harada & Wayland (2003). In this task, the participant heard two isolated nonce monosyllables with an interstimulus interval (ISI) of 500ms. The participant then combined the syllables to produce the disyllabic nonce words used across tasks within the same carrier phrase used in the delayed repetition test. After a 5-item practice, the task was completed in two blocks with the same experimental tokens (n = 65) and fillers (n = 40), once with penultimate stress and once with final stress, and the block order was counterbalanced across participants to avoid a task effect. Each block began with a 5-item warm-up that was not part of the data set. Given that vowel reduction in BP is most salient in unstressed word-final position, the two blocks informed whether learners had acquired the alternation conditioned by stress and allowed for observation of any regressive influence on the Spanish system. Similarly, there is evidence (e.g. Cole, Hualle & Iskarous 1999) that the stop/continuant alternation is conditioned by stress, and this task afforded the observation of stop/continuant production in stressed and unstressed environments.

2.4.3 Perception testing

Learners then completed a naturalness decision task, and reaction time was recorded to observe potential processing differences (or lack thereof) among the groups. Such a measurement will be especially helpful in the case that behavioral measures do not provide evidence of a statistically significant difference between the groups. While accuracy scores might not reflect a difference, reaction time might, which would then indicate processing differences. The same set of nonce words was used to control for lexical frequency and semantic competition. After careful consideration of the evidence surrounding the use of synthetic versus natural speech and one versus multiple speakers for perception stimuli, all recorded stimuli consisted of...
natural speech produced by a female simultaneous bilingual speaker of Spanish and BP. By doing so, the study is more ecologically valid, while limiting cognitive load.

Pairs of experimental (n = 65) and filler (n = 40) disyllabic words in carrier phrases were presented auditorily with an ISI of 500ms. Each pair was identical, except that one token contained an appropriate allophone in the given environment and the other an inappropriate allophone. The order of presentation of target and non-target stimuli was counterbalanced, and the participant had 3000ms to select the sentence that sounded more natural within the carrier phrase by pressing 1 or 2 on a keyboard.

2.5 Data analysis

As noted in Section 1.4, a fundamental question with respect to analysis of production data is whether to use raters, acoustic analysis, or both. It is clear that there are advantages and disadvantages for each method of evaluation, and it is most beneficial to combine them.

The two phenomena I am investigating have been studied extensively, and there is sufficient literature that outlines acoustic parameters and measurement criteria for vowels and stops/continuants. In the case of vowel reduction, I used F1 and F2 measurements as described in Herrick (2003) to determine whether a vowel was reduced, measuring the average space between the F1 and F2. Duration was also measured. To observe production of postvocalic and postnasal underlying stops, I calculated the difference in intensity between the CV segments obtained via manual segmentation of the corresponding sound wave in Praat (Boersma & Weenink 2012) following Hualde, Simonet, Shosted & Nadeau (2010). Given the gradient nature of the phenomenon, it was important to measure the degree of occlusion, rather than treating the phenomenon as binary.

While there is no ideal way to compare the human ear to a machine, the use of trained raters brings us closer to discovering if there is a correlation between the measurements we use and the cues that the listener uses. To corroborate acoustic analysis, three trained native speakers of each language evaluated data from five participants chosen at random. Raters used a 1-5 scale to evaluate each production, given that both phenomena are gradient. To compare the acoustic analysis and rater evaluations, the acoustic analysis data was converted to a 1-5 scale.

3. Conclusion

One of the primary goals in scientific research is to obtain results that are both generalizable and replicable. I have reviewed some of the major methodological stumbling blocks in L3 phonology research, using the established body of work to carefully select the way in which I chose and categorized participants, how I tested the participants and designed the stimuli, and the most effective way to analyze the data. I have worked toward determining the most adequate methodology to test the predictions of the PPH, while minimizing the possibility of skewed results where possible. The PPH predicts that if target-like L2 phonological systems are represented mentally in the same way as native systems, both systems will be equally impervious (or not) to influences of an additional language acquired in adulthood. In line with the evidence of L1 phonological attrition, I do not assume a priori that phonological L1 attrition does not happen; however, the PPH predicts that phonological L2 attrition happens more quickly and pervasively than L1 attrition when the L2 is acquired after the critical period (as opposed to a child L2 learner). By measuring the perception and production of underlyingly post-vocalic voiced stops and word-final vowels in Spanish and BP via implementation of the methodology I have presented, I have been able to investigate not only the acquisition of L3 BP by different profiles of Spanish/English bilinguals, but also to observe at what point and to what extent, if any, there are modifications of the native and non-native Spanish speakers' phonological system as it relates to the stop/continuant alternation and vowel realizations. As with any empirical research, there are limitations to this study, but the goal has been to control for as many variables as possible while recognizing that logistics and practicality must also be regarded when testing this hypothesis of L3 regressive transfer.

Acknowledgments

This research has been supported by National Science Foundation grant BCS#1132289. I would like to thank the editor of this Viewpoints section, Tim Gupton, and Christine Shea for their extensive comments and suggestions. I am also grateful to Jason Rothman, Ratree Wayland, Gillian Lord and Caroline Wiltshire for their feedback and support with the development of this project. All remaining errors are my own.

Notes

1. A prospective power analysis is necessary to determine the minimum sample size to minimize the probability of a type I error. For tests used to detect differences between or among groups such as t-tests and ANOVA, the goal is a sample that will yield power of approximately 80% (Kraemer & Thiemann 1987), and minimum sample size will depend on the size of the effect under consideration.

2. There are no standard proficiency measurements used in the field, and L2 researchers have used a wide array of measures ranging from impressionistic judgments to standardized tests (see Thomas 1994, 2006 for a review). Few studies have employed a battery of tests of global proficiency in terms of phonological aptitude, syntax and lexical knowledge so as to group participants
in a way that is neither arbitrary nor impressionistic, such as those used by Slevc & Miyake (2006).

3 While research such as that of Flege, Mackay & Meador (1999) have found a direct relationship between perception and production of L2 sounds, others such as Archibald (1996) and Hancin-Bhatt (2008) caution that due to the nature of perception and production and the way in which these processes are mediated, results testing these aspects of phonology cannot be directly compared. Hancin-Bhatt (2008) insists that generalizations from perception cannot be used to test predictions for production unless testing the differences and similarities between the two processes, rather than the underlying grammar. Mack (1989) also points out differences in testing techniques that make such comparisons inadequate.

4 Although I mention in Section 1.3 that it is important to test all of the learners’ languages when investigating progressive transfer by successive bilinguals, it is not necessary to do so in this investigation of regressive transfer from L3 BP, because I am concerned with modification of the typologically similar Spanish system. Following ample evidence of typological primacy (e.g., Montrul, Dias & Santos 2011, Rothman 2010, 2011, Rothman & Cabrelli Amaro 2010), it is predicted that Spanish will transfer to L3 BP rather than English. Similarly, any modification to existing systems is expected to be seen primarily in the Spanish system.

5 Cole, Hualde & Iskarous (1999) and Hualde, Simonet, Shosted & Nadeau (2010) have found that the effect preceding vowel quality has on the degree of constriction of /b d g/ is a result of tongue body movement. Without tongue body movement, a closure is not realized, and therefore a vowel and underlying voiced stop with the same place of articulation will have less constriction than a vowel and stop with different places of articulation. However, to limit the number of variables in the current study, vowel quality was controlled for such that the underlying voiced stop was always flanked by the low vowel /a/. It is therefore predicted that there will be a POA effect, and this will be taken into account in my analysis.

References


