CHAPTER 9

L2 proficiency and L2 dialect processing during study abroad

Christine Shea
University of Iowa

In this chapter, we examine how proficiency and L2 dialect processing interact over the course of a three-month study abroad program in Buenos Aires, Argentina in L1 English/L2 Spanish learners. To measure dialect processing changes, participants completed an auditory form-priming task in which primes and targets were taken from Mexican and Buenos Aires Spanish. The pretest results show differences within and across proficiency levels for dialect processing, as do the posttests. This study adds to the literature on advancedness and language acquisition by highlighting how dialect-specific knowledge develops over the course of study abroad and how advanced learners (vs. less-advanced) develop dialect-specific representations.

Keywords: advanced L2 proficiency, dialect processing, dialect representation, study abroad

There has been much research examining the role of study abroad programs in the development of advanced language proficiency. Researchers have considered how study abroad experience leads to changes in fluency, grammatical knowledge, vocabulary, and cultural/pragmatic knowledge (Leonard & Shea, 2017; Marijuan & Sanz, 2018; Xu, 2019; Zaytseva, Pérez-Vidal & Miralpeix, 2018; McManus, Mitchell & Tracy-Ventura, 2020, among many others). Together, this work has painted a rich picture of how students develop linguistic knowledge over long-term (greater than four weeks, typically) and short-term foreign language study-abroad experiences.

Research comparing language development in study abroad contexts to at-home instructed learning shows that oral fluency and vocabulary skills tend to improve more in the former context than in the latter (e.g., Mitchell, Tracy-Ventura & McManus, 2017; Isabelli, 2007). As mentioned by McManus et al. (2020), this may be due to differences between classroom and study abroad learning contexts and interactional opportunities, which in turn may interact with levels of linguistic knowledge exhibited by learners. In a recent study by McManus et al., the authors argue that Skill Acquisition Theory (SAT, DeKeyser, 2017) successfully captures a

https://doi.org/10.1075/ihll.31.09she

crucial difference between study abroad and at-home language learning. SAT explicitly considers the interaction between knowledge and opportunities to practice and rehearse this knowledge. During study abroad, participants can immediately put into practice each aspect of the language they acquire and have abundant opportunities to do so, across a wide variety of speakers and contexts. Moreover, and particularly relevant to the present study, learners' proficiency is predicted to play a key role in development under SAT. Specifically, more advanced learners already have a greater store of declarative knowledge and can more fully benefit from the opportunities available during study abroad to practice and automatize this knowledge (McManus et al., 2020).

In the present study, we expand the focus of most previous study abroad research beyond complexity, accuracy and fluency (CAF, Housen, Kuiken, & Vedder, 2012) to include the development of dialect processing. For L2 learners, developing proficiency in the target language entails processing and storing input that builds knowledge of the L2 and can serve as the basis for more efficient processing. One potential impediment to this process is linked to the speech signal, which contains numerous sources of variability (indexical and linguistic, see Bent & Holt, 2017). This may have at least two consequences. First, the L2 learner may not be able to segment lexical items from the acoustic signal, or 'break into' the speech signal, preventing them from recognizing words they already know. Alternatively, L2 learners may not be able to segment new words that they do not yet know, leading to difficulties with vocabulary acquisition and grammatical processing. This is a particularly relevant question for L2 learners who are studying abroad and are exposed to new dialect variability for the first time. In the present study, we examine how dialect processing relates to overall changes in language knowledge and processing over the course of a three-month study abroad program. When students study abroad, it is often the first time they are fully immersed in one particular dialect of their target language, one to which they may not have been exposed previously. Study abroad provides one of the first opportunities for learners to be immersed in a relatively uniform L2 dialect community for a meaningful length of time.

Background

Processing

Language comprehension involves a sophisticated process of speech segmentation, lexical activation and meaning extraction with the goal of combining elements into larger structures. The precise details of this process lie beyond the scope of this chapter, but the initial stages – that of recognizing and activating lexical items in the L2 input – are relevant to L2 dialect processing. The ability to swiftly segment potential lexical candidates and activate their meaning is necessary to process incoming speech, and this process can be impeded when the incoming signal does not align with the representations created by the listener (Bradlow & Bent, 2002; Broersma, 2012; Weber & Cutler, 2004). This is precisely the case that arises when listeners are exposed to a new dialect – whether in the L1 or the L2. There is an extensive body of work showing that native listeners' processing is slowed down when they are processing a non-native dialect (Floccia, Goslin, Girard, & Konopczynski, 2006) or otherwise non-native accented speech (Bent, Bradlow, & Wright, 2006). These difficulties can be overcome, however, as listeners gain experience with the variability, suggesting that adaptation and learning are possible (Adank, Evans, Stuart-Smith, & Scott, 2009; Cooper & Bradlow, 2018). As the listener gains experience with the input, however, processing new dialect features becomes easier.

Evidence of this type shows that when perceiving speech, listeners simultaneously process linguistic and indexical information contained in the acoustic signal and encode speaker information regarding gender, social class, and dialects (Foulkes & Docherty, 2006; Pierrehumbert, 2002, 2006, 2016). Theoretically, these results have been accounted for by means of 'episodic' approaches to speech perception (Johnson, 1997, 2006) or, more recently, hybrid approaches to speech perception (McClelland, Luce, & Charles-Luce, 2003; Pierrehumbert, 2003). Exemplar-based and hybrid approaches do not regard variability in the speech stream as something that should be eliminated or abstracted away during speech processing but rather as part of the representations themselves. Phonetic details that occur in the speech of a particular dialect community can be stored along with the social information that identifies the speakers themselves. In terms of processing items from different dialects, processing is faster and more accurate for native and standard dialects than for unfamiliar, non-standard dialects (Clopper, Pierrehumbert, & Tamati, 2010; Clopper, Tamati, & Pierrehumbert, 2016; Floccia, Goslin, Girard, & Konopczynski, 2006). The native variety is robustly represented because it has been encountered many times, over a long period of time and across many different speakers. When listeners are processing a different, non-native dialect (for example, Southern US vs. Midwest English), the non-native dialect is less robustly represented and in the case of the non-standard dialect, less socially salient as well (Sumner, Kim, King, & McGowan, 2014).

L2 learners and dialect processing

For L2 learners, various questions arise as to what exactly a "native" L2 dialect may be. In post-secondary classroom contexts, learners are not typically immersed in the L2 outside the classroom and are often taught by instructors from many different dialect regions (Schmidt, 2018). Consequently, most L2 learners are not exposed to one specific dialect over the course of their classroom experience with the L2 to the L2 and, perhaps more importantly, many L2 learners may lack the proficiency to even have a specific 'dialect identity' in their second language. In a recent study, Schmidt (2018) examined classroom learners of Spanish and showed how proficiency interacted with the perceptual identification of /s/-aspiration in that language. Dialects spoken in low-land areas of the Spanish-speaking world exhibit aspiration of /s/ when it occurs in syllable coda position (e.g., *viste* 'saw' /biste/→ [bi**h**.te].) Schmidt's participants carried out a classification experiment in which they heard nonwords that either exhibited /s/-aspiration in word-medial, pre-consonantal position or did not. They then had to select the word they believed they heard from among six written options on the screen, each of which had a different consonant in the internal coda position, with the final option being 'none'. Schmidt found that contact factors such as study abroad location, native speaker contact and awareness of the different dialect features predicted L2ers' perception of the target nonwords and their subsequent word-identification accuracy. However, this only held for learners who had progressed beyond intermediate-level courses. Based upon Schmidt's results, it appears that learners may need to reach a certain proficiency/experience level before they can start to identify dialect-specific features such as Spanish /s/-aspiration.

In this study, we use auditory primary task to examine how L2 dialect familiarity affects processing changes over the course of a three-month study abroad in Buenos Aires, Argentina by L1 English-L2 Spanish learners. The specific research questions that guided the study are as follows:

1. Do L1 English/L2 Spanish speakers process lexical stimuli from Mexican and Buenos Aires dialects in the same way?
2. Does processing of the two dialects change after a three-month immersion in Buenos Aires? And if so, how?
3. What is the role of proficiency in processing the two dialects?
4. How do changes in proficiency interact with dialect processing before and after study abroad?

Research questions 1 and 2 relate to changes that occur in the processing of the Buenos Aires dialect while questions 3 and 4 relate to how these changes may manifest across different proficiency levels, with question four considering the interaction between proficiency and dialect processing. The dialect feature we examine is the well-studied process of *rehilamiento* that is a salient marker of speakers from the Río de la Plata region of Argentina and Uruguay. *Rehilamiento* involves the pronunciation of the palatal fricative [ʝ] as the alveopatal fricative [ʃ] or [ʒ] in words such as *playa* 'beach' /plaʝa/→ [plaʃ/ʒa] and *calle* 'street' /kaʝe/→ [kaʃ/ʒe] (Rohena-Madrazo, 2015). In Mexico City Spanish, the target sound is realized as either a palatal fricative or palatal glide (or often along a continuum between the two, see Hualde, 2004). The process of *rehilamiento* does not occur outside this region of the Spanish-speaking world.

Because the advanced group presumably has more experience with Spanish, we predict that they will already have established more robust lexical representations of the Mexican variety, which may mean slower processing of the Buenos Aires accent on the pretest but faster processing on the posttest, due to experience with the new dialect. We do not claim that Mexican Spanish is the default variety to which learners are exposed in classroom contexts. Instead, we use Mexican Spanish as the control variety because it does not exhibit rehilamiento and, importantly, pronunciation of the target forms are similar to their phonetic realizations in other varieties of Latin American Spanish. Thus, we predict that the advanced learners may show an advantage for the Mexican dialect at the pretest stage but this advantage will disappear by the posttest stage. The less advanced learners, on the other hand, may lack dialect-specific representations at the pre-study abroad stage, which means that the extended experience with one dialect may result in an advantage for the Buenos Aires dialect overall but this advantage is predicted to be greater for the less advanced group than for the more advanced group. For the advanced group, the advantage of the Buenos Aires dialect over the Mexican dialect is predicted to be smaller because the advanced learners will have both dialects stored as part of their lexical representations, distinct from the lower-level learners who will have the Buenos Aires dialect as dominant at posttest. The key difference, in quantitative terms, is the *difference* observed between the pre and posttests on the different priming trials, across the two groups.

In the next section we present the methodology, beginning with the auditory form-priming task. We then present the tasks used to collect proficiency data. Following this, we present the results from the auditory form-priming task (hypothesis 1) and then the combined analysis of data from the auditory form-priming task and proficiency tasks (hypothesis 2).

Method

Participants completed a pretest (week 1), which consisted of an auditory priming task with lexical decision along with five proficiency tasks (see below), which served to divide the participants into advanced and intermediate proficiency levels. The posttest was identical to the pretest and carried out in weeks 14 or 15 of their stay in Buenos Aires. All tasks were the same in the pre- and posttests. Experiments were counterbalanced across participants for the pretest (half completed the proficiency test first) and those who did the proficiency tests first on the pretest completed it second for the posttest.[1]

1. While we recognize that repeating the same tasks may have had an effect on performance, we decided to proceed with this method so as to isolate as much as possible the effects for study abroad.

Participants

Participants were 39 English native speaker undergraduates from the United States (33), Canada (4), and Australia (2) studying abroad at three private universities in Buenos Aires, Argentina. None of the participants had any previous experience with this dialect nor had they ever studied a class in Spanish dialectology prior to their study abroad session or lived in a Spanish-speaking country. Of the 39 participants, 13 were living with an Argentinian host family, 14 lived with Argentinian students and the rest lived with students from the study abroad program. Table 1 provides detailed information about the participants.

Table 1. Participant information

	Advanced group		**Lower level group**	
	Mean	***SD***	***Mean***	***SD***
Age at pretest	21.1	1.1	20.9	0.3
Age of onset (learning Spanish)	11.1	2.1	14.1	3.2
Years of formal study of Spanish	8.0	1.4	5.2	2.5
Previous time abroad in Spanish-speaking countries (months)	0.0		0.0	
Number of other languages studied	0.5	0.2	0.2	0.7

As Table 1 shows, the two groups were similar on the age variables but differed on the years of formal study. Three members of the advanced group had studied languages other than Spanish (German and Mandarin) and one member of the less advanced group had done so (French). None of these multilingual participants had lived abroad for longer than two weeks nor had they reached greater than second-year college classes in their other language.

Tasks

Spanish proficiency tasks

The following section describes the tasks that were used at the beginning of their study abroad session to divide the participants into high and low proficiency categories.

Task one: Monologue

Participants carried out three monologue tasks in Spanish (narrating a past event based upon a series of pictures, describing personal activities, explaining advantages and disadvantages of a particular activity). Instructions were presented on a computer screen and after a 30 second preparation time, participants had a maximum of two minutes to respond. Thirty second excerpts were taken from each monologue (from second 5 to second 35) and, after eliminating fillers (fillers were items such as ‘uhh…’, ‘that is…’), the productions were analyzed for pause length and location. Based upon results from Leonard and Shea (2017), we counted only long pauses (>.49 seconds) and mid-clause pauses for the current study, since there were found to be greater in number for less-proficient learners. The total number of long pauses and mid-clause pauses was calculated for each participant.

Tasks two and three: Grammatical and lexical knowledge

Grammatical knowledge was measured by an untimed 30-item grammar test in which each item contained an error (e.g., verb tense, aspect or mood, agreement, incorrect pronoun) and participants had to detect the error and write in the correct form. Vocabulary knowledge was tested by an untimed 30-item vocabulary test adapted from the *Diploma de Español como Lengua Extranjera* (DELE), used previously in second language acquisition research (e.g., Slabakova, Rothman, & Kempchinksy, 2011; Leonard & Shea, 2017). A vocabulary and grammar score was calculated for each participant, out of a maximum of thirty points each.

Tasks four and five: Timed processing speed

To measure participants’ ability to process grammatical and lexical items under time pressure, they completed a picture-naming task and a sentence-matching task. In the picture-naming task, they saw a picture (from the Snodgrass & Vanderwart [1980] picture set; all pictures had 90% naming agreement among native Spanish speakers) and had to name it as quickly as they could. In the sentence-matching task, participants heard a sentence and had to indicate whether it accurately described the picture they were looking at. The goal of this task was to measure morphosyntactic processing abilities under time pressure. For example, participants heard the sentence *Mira la televisión* (‘He/she watches television’). A mismatched picture showed two people watching television. A matching picture would show one person watching television. The sentences were presented in auditory form since our intent was to represent online speaking processing as realistically as possible. None of the sentences or picture names had the target *rehilamiento* sound in them. They were produced by a native speaker of Mexican Spanish.[2] Participants heard a sentence, immediately followed by the presentation of a picture. Reaction time

2. Since none of the pictures or sentences in the processing task were produced by native speakers of Argentinian Spanish, we did not measure their processing of dialect-specific features. The processing of specific dialect features was analyzed in the auditory priming task only.

was measured from the point at which the picture appeared. The task included 34 sentence picture pairs in the main set and six pairs in the warm-up set. In order to guarantee that we were in fact measuring reaction times to items familiar to participants, items in both the picture-naming and sentence-accuracy tasks that had less than 75% accuracy (7 of 50 items for the picture-naming task, 6 of 34 items for the sentence-accuracy task) were excluded. The remaining items had 94.8% accuracy and 87% accuracy, respectively. Reaction time scores that were more than two standard deviations above or below the participant's mean score (4.3% of the data on the pretest and 4.4% of the data on the posttest) for that task were eliminated. Additionally, 0.99% of the pretest data and 1.49% of the posttest data had to be eliminated due to false activation of the voice key or failure of the voice key to detect a response.

Auditory form priming with lexical decision

To measure changes in dialect processing, we used an auditory priming task with lexical decision. In this task, listeners heard a prime, followed by a target and had to decide if the target was a real word in the language. The prime serves to activate all items in the listeners lexicon that phonetically align with it, and the prime-target combinations that exhibit the greatest overlap will experience the greatest activation, or facilitation. When the prime and target are the *same*, reaction times will be fastest. When the prime and target are the same word but differ in a particular phonetic feature, such as a dialect difference, reaction times will vary; listeners who process each dialect equally well will show faster reaction times to the target than listeners for whom one dialect inhibits activation of the other. This method has been used to investigate cross-dialect lexical activation because it allows researchers to determine what listeners consider to be similar lexical items (Llompart & Simonet, 2018; Shea, under review; Sumner & Samuel, 2009). That is, if listeners process words from different dialects equally well, they will prime each other equally well. This would mean that each dialect activates lexical candidates that could, in theory, include the other dialect.[3] If, however, one dialect is not activated by the other (as shown by longer lexical decision times to the target), we can assume that the listener does not perceive them as equal and did not activate the target dialect lexical item as a possible candidate after hearing the prime.

3. It is also possible that listeners have less well-defined representations and accept variable pronunciations of the same sound because they do not have a clear or robust representation of the sound itself against which to compare the input.

For the auditory priming experiment, the prime and target could be from the same dialect ('matched') or different dialects ('unmatched'). Half were produced by Mexican speakers (Mexico City) and the other half were produced by speakers from Buenos Aires. We used speakers of Mexican Spanish as the contrast for two reasons. First, Mexico City Spanish is a conservative dialect in which sounds correspond very closely to their orthographic equivalents, a fact that makes the phonology-orthography encoding process more straightforward and facilitates lexical activation for non-native speakers (Veivo & Järvikivi, 2013; Shea, 2017). Second, for the participants from the United States and Canada (39/42), this was the dialect with which they had had the most direct experience and was most used in their classes.[4]

Stimuli

The stimuli were recorded by four female native Spanish speakers (two from Mexico City, two from Buenos Aires). At the time of recording, all were graduate students in a Spanish Department at a university in the United States. The stimuli were recorded in a sound-proof booth, using a Marantz PMD671 solid state recorder and a Sure SM58 cardioid microphone. Speakers were asked to read the word list twice, using the carrier phrase '*La palabra es* _____' ('The word is ____'). The clearest version of the word was included in the experiment, as judged by the author.

Lexical items were matched for frequency and length. Frequency data was taken from NIM, a web-based software (Guasch et al., 2013) that provides online access to the *Léxico informatizado del Español* ('Spanish Computerized Lexicon') (LEXESP; Sebastián-Gallés et al., 2000). This corpus includes 5,629,279 Spanish tokens and 166,494 word-types. The average length of the stimuli was 5.7 letters (range: 4–8) and the average log frequency was 1.83 (range: 1.67–2.9). The filler items were taken from the NIM web corpus as well and reflected the same frequency and length range as the target items and ranged between 5–9 letters in length. None of the filler items had the target sounds in them. The nonword items were created by switching the penultimate or ultimate vowel in a real word. For example, instead of the word *mesa* ('table'), the nonword became *mesu*, a word that does not exist in Spanish. Items included verbs, nouns and adjectives.

4. By this we do not mean that their teachers were necessarily all from Mexico but rather that teachers of elementary level Spanish tend to modify their own dialect features to facilitate comprehension by beginner learners and this modified Spanish resembles that of Mexico (highland dialects) because of the close phonology-orthography correspondence. We do not mean to imply any implicit value in this dialect over others, simply that anecdotally, Spanish instructors have often commented that they modify their own dialects to approximate that of a (non-existent) norm, which most closely approximates the Standard Mexican variety.

There were seven trial types in total (see Table 2), of which four were of experimental interest (matched/unmatched and controls) and the rest were fillers, and fillers + nonwords (nonwords were never used as primes, only as targets). For the matched trials, the prime and target were the *same word* and also shared the *same dialect* (with different speakers). For the unmatched trials, the prime and target were the *same word* but from *different dialects*. For the control trials, the words differed across the prime and target. Participants heard three trials for each of the matched, unmatched and control conditions (= 24 total). There were 72 trials for the filler-nonword condition, 36 for the matched fillers, and 36 for the unmatched fillers. This gave a total number of 168 trials.

Table 2. Sample stimuli for the auditory form priming experiment

Trial type	**Prime**		**Target**	
matched for dialect	millón	[miʃon]	millón	[miʃon]
	cabello	[kaβeʝo]	cabello	[kaβeʝo]
unmatched for dialect	caballo	[kaβaʝo]	caballo	[kaβaʃo]
	cebolla	[seβoʃa]	cebolla	[seβoʝa]
control – matched for dialect	llamar	[dʒamaɾ]	llave	[dʒaβe]
	bella	[beʝa]	pasillo	[pasiʝo]
control – unmatched for dialect	lluvia	[dʒuβi̯a]	lloro	[ɟʝoɾo]
	botella	[bot̪eʝa]	anillo	[aniʃo]
filler – nonwords	casa	[kasa]	fresu	[fɾesu]
fillers – same words	gato	[gat̪o]	gato	[gat̪o]
fillers – different words	brazo	[bɾaso]	árbol	[aɾβol]

Procedure

The study took place in a quiet room with only the experimenter and participant present, on the campus of a private university in Buenos Aires. All interaction took place in Spanish. Participants were paid $ 20.00 for the pretest and $ 40.00 for the posttest, prorated to guarantee participation in the posttest (39/40 participants returned for the posttest).

Experiment 1 was carried out using Superlab experimental software (Cedrus Co., V.5.0). Participants were told they were going to listen to pairs of items in Spanish and they had to decide if the second member of the pair was a real word. They were given five trials, with feedback, before starting the experimental trials. All instructions on the screen were in Spanish.

For each trial, participants saw a '+' on the computer screen and then heard the auditory prime, followed by a 500 ms interstimulus interval, and then heard the auditory target. Stimuli were randomized across participants. Stickers were placed on the keyboard to indicate their decision of 'word' or 'nonword'. Reaction times were recorded as soon as the button was pressed following the presentation of the target. Participants had 2 seconds to respond before the trial timed out. The start of the next trial was indicated by the '+' in the center of the screen.

Results

Because the results from the auditory priming tasks are discussed in terms of proficiency groups, we first present the results from the proficiency tasks, followed by those of the auditory priming task.

Proficiency tasks

Because the scores for the different measures were on widely varying scales, we first converted raw scores to z-scores and then scaled them to a value from 1–10. These were the scores that served to group the participants into high and low proficiency and were used for the regression analysis discussed in the next section. In order to avoid arbitrary cutoffs, we eliminated the participant closest to the cutoff for the high proficiency group and two from the low proficiency group, which gave us 20 participants in each group. Table 3 presents the proficiency task results.

Table 3. Proficiency task results (raw data)

Proficiency	Pretest score Mean (SD)		Posttest score Mean (SD)		Significance pre- vs. posttest (2-tailed)	
	high	**low**	**high**	**low**	**high**	**low**
Monologue task: Number of long pauses	6.2 (2.7)	11.1 (4.2)	4.2 (1.1)	10.1 (3.2)	<.001	n.s
Number of mid-clause pauses	4.3 (1.4)	9.1 (3.2)	2.8 (1.5)	8.2 (3.3)	<.001	.03
Grammar (max:30)	22.1 (2.6)	16.8 (3.4)	23.8 (.99)	20.1 (1.3)	n.s.	.002
Vocabulary (max:30)	23.2 (4.1)	17.3 (4.2)	25.7 (2.9)	21.2 (.97)	.001	<.001
Sentence-matching (log)	2.29 (.88)	3.56 (.45)	1.88 (.19)	2.9 (1.0)	.039	<.001
Picture-naming (ms)	887 (142)	1566 (264)	752 (192)	1242 (201)	<.001	.004

Auditory priming task

In total there were 6,662 possible responses each for the pre and posttest, or a total of 13,104 overall. The error rate overall across each group and each condition are presented in Table 4, for the pretest and the posttest.

Table 4. Pretest and posttest reaction times across groups and conditions

PROFICIENCY CONDITION	PRETEST				POSTTEST			
	Reaction time (ms)		Accuracy (% error)		Reaction time (ms)		Accuracy (% error)	
	high	low	high	low	high	low	high	low
matched for dialect Arg-Arg	966 (275)	1293 (306)	1.2%	2.3%	626 (215)	1139 (272)	.6%	.9%
Mex-Mex	771 (133)	899 (143)	.8%	1.1%	665 (244)	851 (151)	.9%	1.3%
unmatched for dialect Mex-Arg	999 (327)	997 (285)	2.2%	3.1%	727 (298)	997 (285)	1.3%	1.9%
Arg-Mex	902 (308)	1007 (292)	1.8%	3.2%	635 (277)	887 (246)	1.1%	2.3%

For the analysis that follows, we only used the matched and unmatched trials with correct responses on the lexical decision task. We carried out two sets of data analyses. The first set included data from the auditory priming experiment only. To determine initial and post-study-abroad differences in dialect processing across and within each group, we employed a linear mixed model analysis of the pretest priming data and then a second model of the posttest priming data to test the hypothesis that in fact, there were differences at the pretest between both groups and at posttest as well. We then present a second set of models that examine the differences for each priming condition across proficiency groups and within each group to test for changes in priming of the Buenos Aires dialect.

The first set of models were created using generalized linear mixed effects models (implemented in [R Core Team] R, V. 3.5, using the lmerTest package [Kuznetsova et al., 2014]). We normalized the reaction time measures to z-scores, to control for the non-normal distribution in reaction time values. Pretest and posttest reaction times (log transformed) were the dependent measures. Trial and participant were included as random factors. Predictors included condition (matched/ unmatched for dialect) and proficiency level. The best-fitting converging model justified by the data included the interaction between the two fixed effects and by-trial slopes. Maximal random effects structure was tested via likelihood ratio comparison. All predictors were tested individually as fixed main effects and since both predictors were categorical, they were dummy coded, with high proficiency as the reference level for proficiency. Table 5 presents the results for these two models. Since the hypothesis focuses on interactions across groups and within groups, Table 5 provides only the results from proficiency * condition interactions.

Table 5. Auditory form priming results across groups and within groups

Across groups								
	PRETEST				POSTTEST			
Random effects trial	**Var. .048.**		**St.D. .07**		**Var. .040**		**St.D. .06**	
Fixed effects	**Est.**	**Std. error**	***t* value**	***p* value**	**Est.**	**Std. error**	***t* value**	***p* value**
(Intercept)	−.5011	.034	14.7	<.001***	−.581	.031	−18.48	<.001***
Proficiency	.996	.036	27.5	<.001***	1.14	.033	34.12	<.001***
Prof. (ref=high) by Arg-Arg:Arg-Mex	−.390	.058	−7.61	<.001***	−.274	.047	−5.772	<.001***
Arg-Arg:Mex-Arg	−.372	.0512	−7.27	<.001**	−.303	.048	−6.37	<.001***
Arg-Arg:Mex-Mex	−.148	.043	−2.89	.004**	−.300	.048	−6.13	<.001***
Arg-Mex:Mex-Arg	−.175	.033	−3.36	.037**	−.188	.048	−4.6	.012**
Arg-Mex:Mex-Mex	−.537	.059	−10.49	<.001***	−.174	.051	−2.9	<.001***
Mex-Arg:Mex-Mex	−.520	.061	−10.16	<.001***	−.603	.48	−12.7	<.001***

Within groups								
High proficiency	PRETEST				POSTTEST			
Random effects trial	**Var. .071**		**St.D. .081**		**Var. .040**		**St.D. .06**	
Fixed effects	**Est.**	**Std. error**	***t* value**	***p* value**	**Est.**	**Std. error**	***t* value**	***p* value**
(Intercept)	−.508	.037	−13.6	<.001***	−.587	.036	−16.5	<.001***
Arg-Arg:Arg-Mex	.199	.367	5.41	<.001***	.143	.034	4.21	<.001***
Arg-Arg:Mex-Arg	.190	.036	5.25	<.001***	.156	.896	4.67	<.001**
Arg-Arg:Mex-Mex	.076	.036	2.1	.03 **	−.154	.033	4.6	<.001 ***
Arg-Mex:Mex-Arg	−.008	.036	−.241	.805	.013	.033	.394	.693
Arg-Mex:Mex-Mex	.087	.031	4.6	.021**	−.297	.033	−8.88	<.001***
Mex-Arg:Mex-Mex	.27	.036	7.39	<.001***	−.310	.033	−9.32	<.001***

(*continued*)

Table 5. (*continued*)

Within groups								
High proficiency	PRETEST				POSTTEST			
Random effects trial	**Var. .071**		**St.D. .081**		**Var. .040**		**St.D. .06**	
Fixed effects	**Est.**	**Std. error**	***t* value**	***p* value**	**Est.**	**Std. error**	***t* value**	***p* value**
Low proficiency	PRETEST				POSTTEST			
Random effects trial	**Var. .071**		**St.D. .088**		**Var. .054**		**St.D. .08**	
Fixed effects	**Est.**	**Std. error**	***t* value**	***p* value**	**Est.**	**Std. error**	***t* value**	***p* value**
(Intercept)	.499	.038	13.13	<.001***	.567	.036	15.8	<.001***
Arg-Arg:Arg-Mex	−.014	.043	−.580	.421	−.152	.034	−4.42	<.001***
Arg-Arg:Mex-Arg	−0.194	.037	−5.32	<.001***	−.157	.034	−4.62	<.001***
Arg-Arg:Mex-Mex	.160	.036	1.65	.100	.14	.033	4.01	<.001***
Arg-Mex:Mex-Arg	.02	.03	.552	.581	−.045	.034	−.14	.886
Arg-Mex:Mex-Mex	.274	.364	7.52	<.001***	.288	.033	8.49	<.001***
Mex-Arg:Mex-Mex	.254	.036	7.00	<.001***	.300	.034	8.62	<.001***

Discussion of auditory priming results

The across-group results from the auditory form priming experiment show that on both the pretest and the posttest, the high proficiency group was significantly faster overall on all conditions. The within-group comparisons revealed different patterning across each group. For the pretest results, the advanced group had showed slower reaction times for the Arg-Arg trials compared to all other trials. Only one trial type did not reach significance for the advanced group, specifically, the comparison between the two unmatched trials. For the posttest, however, the results for the advanced group shifted in favor of the Arg-Arg trials. For the posttest, the matched Arg-Arg trials were faster than the matched Mex-Mex trials, indicating a change in dialect processing speed across the study abroad session (pretest: $\beta = -.076$, $p < .001$; posttest: $\beta = -.154$, $p < .001$). This shift shows that the high proficiency group became faster at processing the Argentinian dialect after their study abroad session.

The within-group results for the low proficiency learners revealed more complex patterns. For the pretest, the matched Arg-Arg trials were faster compared to the unmatched trials but slower when compared to the Mex-Mex trials. The unmatched trials were also significantly slower than the Mex-Mex trials. On the posttest this tendency remained consistent. These results suggest that when the target is not consistent with the prime, low proficiency listeners cannot recover from the representations they have activated and exhibit less priming for the different variant.

Difference score and proficiency measures

We created a generalized linear mixed effects model (implemented in R, V. 3.5, using the lmerTest package [Kutznesova et al., 2014]) to analyze how changes in reaction time data across participants' study abroad session aligned with changes in the proficiency data. The dependent variable in this model different from pure reaction time (as above) and instead considered the degree of difference between the pre- and posttests. To do this, we subtracted the posttest reaction time from the pretest reaction time and then divided that amount by the pretest reaction time itself. For example, for one of the matched Arg-Arg trials, Participant 3 had a reaction time of 688 ms on the pretest. On the posttest, this participant had a reaction time of 621 ms for the same trial, a difference of 67 ms. For the same trial, Participant 27 had a reaction time of 488 ms on the pretest and 389 ms on the posttest, a difference of 66 ms, an absolute difference that was close to that of Participant 3. However, 67 ms faster on an initial reaction time of 488 ms represents a greater proportional change (.15) than the same absolute difference with an initial reaction time of 621 (.07), a difference that would be obscured if only absolute reaction time difference is taken into account. Thus, for the second analysis, we used as our dependent variable a *difference score proportion*, that allowed us to control for the initial differences in reaction times across participants and get a truer measure of the changes that occurred across the three-month study abroad session.

Trial was included as a random factor. Predictors included proficiency (high vs. low), and pre-posttest difference scores for the picture naming, vocabulary, sentence matching, mid-clause pauses, long pauses and grammar scores. Because the scores were on different scales, all continuous predictors were normalized and then converted to a scale of 1–10. The dependent variable was the difference score proportion. Higher difference score proportions represent faster reaction times on the posttest trials compared to the pretest trials. The best-fitting converging model justified by the data included vocabulary and interactions among proficiency (high vs. low) and sentence matching, picture naming, long pauses and mid-clause pauses. Maximal random effects structure was tested via likelihood ratio comparison and the best model included only trial as a random effect. The proficiency predictor was dummy coded, with high proficiency as the reference level. Table 6 presents the results from this model.

Table 6. Difference scores by proficiency level and tasks

Random effects: Trial	**Variance**	**Standard deviation**		
	.055	.007		
Fixed effects	**Estimate**	**Standard error**	***t*-value**	***p*-value**
Intercept	.27	.06	16.2	<.001***
proficiency (ref=high)	.187	.071	26.4	<.001***
vocabulary	−.023	.031	−2.05	.041*
Interactions between proficiency				
and picture naming	.043	.032	2.17	.030*
sentence matching	.081	.011	2.70	.012*
mid-clause pauses	−.112	.023	−2.6	.0012**
long pauses	−.07	.011	−3.1	.017*

Discussion of difference score ~ proficiency results

The positive coefficient for proficiency indicates that the high proficiency group exhibited a significantly greater change in the difference score proportion overall from pretest to posttest ($\beta = -.187, p < .001$) than did the lower proficiency group. In other words, their overall reaction time differences for all conditions was greater than the differences for the low proficiency group. The high proficiency group improved (proportionately) more on dialect processing speed than the lower proficiency group. There was a significant effect for vocabulary ($\beta = -.023, p = 041$) whereby vocabulary difference scores fell as reaction time difference score proportions went up, for both groups. Sentence matching did not reach significance.

Of the four interactions included in the model, all reached significance. Sentence matching ($\beta = .081, p = .012$), picture naming ($\beta = .043, p = .03$), mid-clause pause ($\beta = -.112, p = .0014$), and long pause difference scores ($\beta = -.07, p = .017$) all interacted significantly with proficiency. For picture naming and sentence matching, the results indicate that as the difference score proportion went up, so did the difference score for these two variables. In other words, greater overall priming (faster on the posttest than on the pretest) means faster picture naming and sentence matching for the advanced group. For the pause variables, the results also show that higher difference score proportions means fewer mid-clause pauses and fewer long pauses.

Together, these results paint an interesting picture for how changes in proficiency measures across a three-month study-abroad program interact with dialect processing for learners of different proficiency levels, and particularly for advanced learners. For both groups, vocabulary scores went down as reaction time proportions went up, suggesting that more efficient processing of the new dialect was not explained by gains in vocabulary for either group – or at least the vocabulary measured by our task.

For the processing variables of picture naming and sentence matching, higher difference score proportions meant higher posttest reaction times on these two variables. Relatedly, the pausing measures – most closely connected to the fluency component of CAF – showed a negative relationship with difference scores on the priming task, suggesting that as fluency increases, dialect perception also improves. Picture naming and the number of mid-clause pauses both reflect the speed at which speakers can retrieve activated lexical items from their second language lexicon. The finding that these two proficiency variables predicted dialect priming results may reflect overall improvements in the automatization of lexical access, which may, in turn, be due to increased experience with interacting in the target language over the study abroad session (not necessarily due to dialect-specific factors). As argued below, we suggest that improved dialect processing is a key part of this.

General discussion

Changes in L2 dialect processing

The research questions guiding this study focused on how advanced versus low proficiency learners represent and process new dialects to which they are exposed during study abroad and how changes in different proficiency measures interact with changes in dialect perception. The first hypothesis related to how high versus low proficiency participants process lexical items produced by native speakers of Mexican Spanish (familiar at pretest) and native speakers of Buenos Aires dialect (unfamiliar at pretest). The across-group results showed that overall, more advanced learners were faster than the lower-level learners at processing the Buenos Aires dialect on the pretest and posttest. Within-group results revealed important differences, however, between the pretest and posttest reaction times. On the pretest, the advanced proficiency group was faster at processing the Mexican matched trials than either the unmatched or Argentinian matched trials. This suggests that the advanced group was more familiar with the Mexican dialect than the Argentinian dialect and activated it more quickly on the lexical decision task. On the posttest, however, the matched Mexican dialect trials lost this advantage when compared to both the matched Argentinian and unmatched trials.

These results do not support the hypotheses laid out at the beginning of this study predicting no significant differences between the dialects would emerge for the advanced group at the posttest stage and that instead, the advanced learners would 'add' an additional dialect to their already established L2 dialect knowledge (an 'L2D2'). We further predicted that the lower-level group would show a clear advantage for the Argentinian dialect at the posttest stage, given that they were building Spanish lexical representations based upon their experience in Buenos

Aires and had no previous L2 dialect to guide their perception. The results from the pretest tend to support this hypothesis. There was no significant difference between the matched Arg-Arg and Mex-Mex trials on the pretest but a significant difference did emerge for the posttest, whereby the Arg-Arg trials were faster than the Mex-Mex trials. Therefore, it appears that the lower-level group did begin to create representations consistent with the Argentinian dialect post-study abroad, but did not exhibit as strong a shift as the more advanced learners did. In general, it is difficult to confidently state that the lower-level group had any specific dialect representation at the beginning stages of their study abroad experience – the only comparisons that showed significant differences between conditions were those that had matched trials as one of the comparison pairs (Arg-Arg and Mex-Mex trials were always faster than any of the unmatched trials for the pretest). This suggests that the lower-level learners did not have any dialect specificity in their representations at the beginning of their time in Buenos Aires. Otherwise, we would have observed differences across the sets of unmatched trials, as we did for the advanced group.

We now ask how to account for these results. It is important to keep in mind the processes by which variability is encoded in the lexicon. In the exemplar approach, the robustness of lexical representations varies as a function of the number and strength of the exemplars they encode (Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2002, 2006). More frequent words have more robust representations because the listener has encountered them more often and therefore have a greater number of exemplars included in the exemplar space. Clopper et al. (2016) distinguish between recognition and encoding of lexical tokens. Recognition involves identifying the category of the current token (i.e., determining which category the token is most similar to) while encoding involves adding the token to the listener's lexical representations (Clopper et al., p. 2). For example, it is possible that a listener may recognize a specific exemplar of a word but not necessarily encode it into their representations. This might occur when a listener hears a mispronunciation or speech produced with a foreign accent. Varieties that are less familiar are encoded less robustly and will take longer for listeners to process.

The present results suggest that the low-level learners may be recognizing words produced using different dialects but they may not be encoding these differences in their lexicon as robustly as the advanced learners. The advanced listeners show a significant difference from the pretest to the posttest for the Arg-Arg trials and actually exhibit a disadvantage for the Mexican dialect trials on the posttest. Thus, it is possible that the advanced learners have encoded the Argentinian variety at the end of their study abroad experience and the Mexican dialect, perhaps because they are no longer exposed to it, is not activated as fully.

The advanced L2 listeners began their study abroad experience with more robust representations of the Mexican variety than the Argentinian variety. The evidence for this was clearly shown by the greater priming effect for the Mexican dialect. The Mexican dialect was selected as the $L2D2_1$, or 'first dialect' because it is the variety most L2 Spanish learners are exposed to (arguably outside of the US Northeast and Florida) and is a variety in which the pronunciation aligns very closely to orthographic representations. It is important to note, however, that we are not claiming that listeners necessarily had Mexican Spanish consciously encoded as their specific dialect representation prior to study abroad – indeed, in the case of low-proficiency learners, it may be that they had no robust L2 dialect specific-representation at all, and it was only after being exposed to the Buenos Aires dialect for three months that they began to develop an identifiable dialect-specific basis for lexical activation in Spanish.

On a more speculative note, it is also possible that the advanced group had greater opportunities for interaction with native speakers of Buenos Aires Spanish. Because the advanced group began their study-abroad session with a higher level of proficiency, they may have experienced a swifter integration into the culture than the lower-level learners. The lower-level learners might not have experienced such a high degree of integration and not felt as identified with the particular *porteño* dialect features. As an anonymous reviewer pointed out, the lower level learners might have had higher levels of anxiety because of difficulties (perceived or real) of communication and this heightened anxiety may have had effects on abilities to process input. Consequently, it is possible that the more advanced group had greater opportunities to interact with Argentinians while in Buenos Aries, potentially leading to greater identification with this dialect as one that is socially indexed to a speech community they feel part of.

The advanced group also had another important advantage over the low-level group in terms of acquiring and processing a new dialect, an advantage that is related to research question 2. By definition, these participants were more proficient in Spanish than the low-level group, which meant that processing a new dialect of Spanish may have represented less of a challenge, since the lower-level learners faced the dual task of establishing Spanish language knowledge and developing dialect-specific processing as well. The advanced learners added dialect-specific details onto an already well-developed system – more similar to what happens with L1 new-dialect acquisition. We hypothesize that better overall knowledge of the language granted more advanced listeners the ability to pay attention to dialect-differences and minimize the processing costs associated with it. Combined with the actual shift in priming effects across the pretest and posttest for the advanced learners, we can further hypothesize that this group did in fact develop new, dialect-specific listening and that this skill developed in tandem with other processing and fluency-related proficiency measures.

In addition to dialect processing, L2 learners also experienced changes in terms of their overall linguistic knowledge. We predicted that each group would exhibit different gains across the dialect processing task (confirmed by the auditory priming task analysis) and, consequently, the gains (difference scores) in reaction time efficiency for auditory priming would be explained by an interaction between each proficiency task and proficiency level – in other words, that pre-posttest reaction time differences would be explained by different proficiency measures for each group. The results from the proficiency-proficiency tasks interaction showed that the advanced group improved significantly on all the tasks across the pre-posttest measures.

As mentioned in the introduction, Skill Acquisition Theory (SAT, DeKeyser, 2017; McManus et al., 2020) can provide a theoretical understanding of the results from this study. Specifically, SAT proposes that L2 development is driven by language use and based in three types of L2 knowledge: declarative knowledge ('about' language), procedural knowledge (rule-governed, effortful, emerges through practicing declarative knowledge) and automatized knowledge that arises through opportunities for extensive and meaningful language use. We argue that the more advanced learners began their study abroad experience with greater procedural knowledge (and possibly more automatized knowledge as well), which facilitated the processing of a new dialect. Picture naming tasks require an ability to quickly access lexical items in the second language and produce the correct phonological sequence that corresponds to the lexical representation and, importantly, inhibit the first language equivalent when the task requires the L2. The pause variables can also be understood as reflecting lexical retrieval, particularly since we took the measurements at clause-medial position, which is not a natural place for a pause boundary to occur (Felker, Klockmann, & De Jong, 2019).

Conclusions

This study examined how learners of advanced and lower-level proficiency processed an L2 dialect feature at the beginning and again at the end of a three-month study abroad session in Buenos Aires, Argentina. Changes across different priming trials for high and low proficiency learners were then analyzed in terms of five tasks. Results show that indeed, significant differences exist between advanced and low proficiency learners on the auditory priming task and also on the way proficiency interacts with the proficiency tasks. Advanced learners have more robust lexical representations than the low proficiency learners and may have arrived at the beginning of their study abroad session with a particular dialect already representationally established. This changed, however, over the course of their time in Buenos Aires: the advanced learners processed the Argentinian dialect more quickly at the end of the three-month session than the Mexican dialect and the matched Mexican prime-target pairs exhibited greater processing cost at the end of participants' time in Buenos Aires than at the beginning when compared to the unmatched trials as well.

The present study adds to the literature on second language dialect perception and also to the literature on study abroad by specifically testing how learners of different proficiency levels shift their processing of L2 dialect features and, importantly, how this relates to overall changes in proficiency during the study abroad session itself. The results highlight a relatively understudied area of L2 acquisition, specifically, what exactly serves as the baseline for dialect-specific processing over the course of L2 development. This study shows that not only do more advanced learners have greater vocabulary and grammatical knowledge, they also have more robust lexical representations, that allow them to process a new dialect differently from less proficient learners. While intuitively this may not be surprising, the evidence presented here helps us understand, qualitatively and quantitatively, what the basis for L2 dialect acquisition is. Advanced learners do indeed have a baseline dialect they use to process a new L2 dialect and this baseline shifts as experience accumulates.

In terms of how we might understand advancedness, this study shows that more advanced learners do process dialect differences distinctly from less advanced learners and that the notion of 'advancedness' itself should be expanded to include lexical processing and encoding. Many studies have examined how learners of different proficiency levels perceive individual sounds in their target language (see work by Flege) but very few have considered how advanced learners differ from less-advanced learners in terms of how these sounds are encoded in the lexicon and how the variability encoded by dialect differences is acquired. As we stated, this is still a relatively understudied area of SLA. First language cross-dialect perception has received a great deal of recent attention (see work by Clopper) but second language learners are also exposed to dialect variability over the course of L2 acquisition. It is important that moving forward we expand our notion of advancedness to include L2 dialect perception.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology. Human Perception and Performance*, 35, 520–529. https://doi.org/10.1037/a0013552

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 97–121.

Bent, T., & Holt, R. F. (2017). Representation of speech variability. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8, e1434.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, 112, 272–284. https://doi.org/10.1121/1.1487837

Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology*, 1, 65–92. https://doi.org/10.1515/labphon.2010.005

Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. (2016). Variation in the strength of lexical encoding across dialects. *Journal of phonetics*, 58, 87–103. https://doi.org/10.1016/j.wocn.2016.06.002

DeKeyser, R. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *Routledge handbook of instructed second language acquisition* (pp. 15–32). London, UK: Routledge. https://doi.org/10.4324/9781315676968-2

Felker, E. R., Klockmann, H. E., & De Jong, N. H. (2019). How conceptualizing influences fluency in first and second language speech production. *Applied Psycholinguistics*, 40, 111–136. https://doi.org/10.1017/S0142716418000474

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34, 409–438. https://doi.org/10.1016/j.wocn.2005.08.002

Guasch, M., Boada, R., Ferre, P., & Sanchez-Casa, R. (2013). NIM: A Web-based Swiss-army knife to select stimuli or psycholinguistic studies. *Behavior Research Methods*, 45, 765–771. https://doi.org/10.3758/s13428-012-0296-8

Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 22, 1166–1183.

Hualde, J. I. (2004). Quasi-phonemic contrasts in Spanish. In *WCCFL 23: Proceedings of the 23rd West Coast Conference on Formal Linguistics* (pp. 374–398). Somerville, MA: Cascadilla Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, 34, 485–499. https://doi.org/10.1016/j.wocn.2005.08.004

Leonard, K. R., & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, 101, 179–193. https://doi.org/10.1111/modl.12382

Marijuan, S., & Sanz, C. (2018). Expanding boundaries: Current and new directions in study abroad research and practice. *Foreign Language Annals*, 51, 185–204. https://doi.org/10.1111/flan.12323

McManus, K., Mitchell, R., & Tracy-Ventura, N. (2020). A longitudinal study of advanced learners' linguistic development before, during and after study abroad. *Applied Linguistics* (preprint). https://doi.org/10.1093/applin/amaa003

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2, 33–52. https://doi.org/10.1146/annurev-linguistics-030514-125050
Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, 34, 516–530. https://doi.org/10.1016/j.wocn.2006.06.003
Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology, Vol. 7* (pp. 101–139). Berlin, Germany: Mouton de Gruyter. https://doi.org/10.1515/9783110197105.101
R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/> (8 July, 2020).
Rohena-Madrazo, M. (2015). Diagnosing the completion of a sound change: Phonetic and phonological evidence for/ʃ/in Buenos Aires Spanish. *Language Variation and Change*, 27, 287–317. https://doi.org/10.1017/S0954394515000113
Schmidt, L. B. (2018). L2 development of perceptual categorization of dialectal sounds: A study in Spanish. *Studies in Second Language Acquisition*, 40, 857–882. https://doi.org/10.1017/S0272263118000116
Shea, C. (2017). L1 English/L2 Spanish: Orthography–phonology activation without contrasts. *Second Language Research*, 33, 207–232. https://doi.org/10.1177/0267658316684905
Shea, C. (under review). L1 dialect primes L2 words.
Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60, 487–501. https://doi.org/10.1016/j.jml.2009.01.001
Veivo, O., & Järvikivi, J. (2013). Proficiency modulates early orthographic and phonological processing in L2 spoken word recognition. *Bilingualism: Language and Cognition*, 16, 864–883. https://doi.org/10.1017/S1366728912000600
Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3, 387–401.
Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199–211. https://doi.org/10.1016/j.system.2018.11.008
Zaytseva, V., Pérez-Vidal, C., & Miralpeix, I. (2018). Vocabulary acquisition during study abroad: A comprehensive review of the research and some methodological considerations. In C. Sanz & A. Morales-Front (Eds.), *The Routledge handbook of study abroad research and practice* (pp. 210–224). New York, NY: Routledge.