Sound communities

A quantitative proposal for studying bilingual speech

Charlie Nagle, Melissa M. Baese-Berk, Mark Amengual, and Joseph V. Casillas
University of Texas at Austin | University of Chicago | University of California, Santa Cruz | Rutgers University

Bilingualism researchers have intensively studied how learning and using multiple languages affects all levels of linguistic structure. In this strand, examining diversity in the bilingual experience and the extent to which variables like language dominance regulate crosslinguistic interaction has been of special interest. However, most studies sample small groups of bilinguals from a single research site, creating a twofold generalizability problem. First, with small samples it is unlikely that researchers will be able to fully capture and quantify the range of variables known to affect findings. Second, when bilinguals are recruited from a single site, it is impossible to determine if findings are site-specific or apply to bilinguals more broadly. To address these issues, we propose a large(r)-scale, multisite approach to bilingualism research. We believe that such an approach, when informed by open science practices, has the potential to significantly advance the state of the art.

Keywords: bilingualism, multisite research, big team science, open science

Introduction

In recent years, interest in bilingualism and second language acquisition has rapidly expanded. Researchers have recognized that referencing a monolingual norm ignores most individuals' linguistic experience because bi/multilingualism is more common than its monolingual counterpart (Grosjean, 1997; Kroll & Bialystok, 2013). One major challenge in this line of research is terminological because the terms "bilingual" and "multilingual" often refer to individuals from radically different linguistic and sociocultural backgrounds who show non-trivial

https://doi.org/10.1075/jslp.25016.nag | Published online: 5 February 2026
Journal of Second Language Pronunciation ISSN 2215-1931 | E-ISSN 2215-194X

differences in the experiential and attitudinal variables that are considered fundamental to the bi/multilingual experience. Indeed, this is precisely why Bialystok (2021) characterized bilingualism as a "package of Swiss cheese, with different manifestations of bilingual experience placing the holes in different places that together define the experience" (p. 2). In this sense, one can see how characterizing someone as "bilingual" without a rich understanding of their background can lead to a false impression that one "bilingual" should somehow resemble another when those individuals might actually be drawn from distinct bilingual populations. Historically, bilingualism research has focused on the individual but asked questions about bilingualism as a whole. This approach assumes that bilingualism is a single, coherent characteristic, when, in reality, it is a highly multidimensional, variable phenomenon whose manifestation depends on myriad factors. We contend that, to date, individual studies of bilingualism reveal more about the experiences and behaviors of specific bilingual samples than about bilingualism itself. Studying bilingualism in all of its complexity therefore requires thinking about both bilingual individuals and the (bilingual) communities, geographies, and social strata they inhabit. In this paper, we address bilingual speech research as a whole while focusing especially on Spanish-English bilingualism because, as individuals who have learned and actively use both languages, we are especially interested in that community. Below, we discuss the state of the art in bilingual speech research, explore a specific relevant case of bilingualism research, provide general and specific recommendations, and lay out an example of how the research we advocate for might proceed.

State of the Art in Bilingual (Speech) Research

Bilingualism across space and time

A bilingual is typically defined as an individual who has command of more than one language (Grosjean, 1997). One challenge is that the division between monolingual and bilingual is not clear cut because the two categories represent a continuum rather than a binary classification. Most researchers would agree that an individual who knows one language well (i.e., their native language) and only knows how to count in a second language is likely not bilingual. However, researchers also understand that a "fully-balanced" bilingual, who demonstrates equal proficiency in both their languages in all domains and for all contexts and situations, is theoretically possible but not common (see Grosjean, 1989). Indeed, the continuum from monolingual to bilingual encompasses a wide range of speakers, including simultaneous and sequential childhood bilinguals, heritage language learners, and second language learners, among many others. The terminology we have adopted as a field reflects our collective effort to derive discrete categories that capture important experiential and acquisitional characteristics about bilinguals, even if those terms ultimately fall short of capturing the totality of the bilingual experience. In speech research,[1] even more complex terminology has been adopted in recognition of the fact that early experience with the sounds and rhythm of a language may leave a subtle, yet indelible mark on the perception and production systems of the future user (see, e.g., childhood overhearers in Au et al., 2002).

Between-subjects variability is also matched by substantial within-subjects variability. Throughout their lives, bilinguals often move between physical spaces with distinct language properties and through life stages with distinct language (learning) affordances. Language dominance is one way of capturing and quantifying such differences. In the context of bilingualism, dominance "refers to observed asymmetries of skill in, or use of, one language over the other" (Birdsong, 2014, p. 374). In concrete terms, one second-generation English-Spanish bilingual speaker in California may be more dominant in English than another English-Spanish bilingual in the same community. Likewise, that same individual may show dominance shifts (relative dominance changes as a matter of degree, but not reversal) or even dominance switches (one language replaces the other as dominant). These shifts and switches can be brought about by diverse circumstances, including temporary relocation or permanent immigration to a different linguistic community, change of employment, change of educational setting, caregiving arrangements, or even engagement with a new social network. In terms of pronunciation, shifting language contexts can bring about temporary phonetic drift, where the production of one language drifts toward the phonetic norms of the other (Leeuw & Chang, 2023; e.g., Sancier & Fowler, 1997; Tobin et al., 2017). There is also a compelling body of longitudinal case studies documenting accent change across the lifespan in celebrities like Steffi Graff (Leeuw, 2019), Arnold Schwarzenegger (Kornder & Mennen, 2021), and Ole Gunnar Solskjaer (Kelly, 2022).

Though as researchers we are accustomed to thinking about speakers, the same logic applies to communities, insofar as synchronic variation between communities is also accompanied by diachronic variation within them. In an era of globalization and mass migration, the bilingual speech communities that have historically been the object of study may look and therefore behave very differently

1. Here, and throughout the paper, we use speech broadly to include phonetic aspects of pronunciation at the segmental and suprasegmental level, as well as global measures of speech including fluency, accentedness, comprehensibility, and intelligibility.

over time. On the one hand, this variation presents a research opportunity: If we can systematically tag studies by decade and recover time-locked sociodemographic data associated with particular communities at particular points in time, then we can gain a powerful lens into diachronic patterns of stability and change in demographics, attitudes, and behaviors (e.g., Mejías et al., 2003). On the other hand, this variation presents challenges because it cannot be assumed that speech data collected from a bilingual community 30 or 40 years ago is representative of that community today (e.g., the Canadian French-English bilinguals included in Caramazza et al., 1973 versus Canadian French-English bilinguals now). In short, the very nature of what it means to be bilingual has changed and will continue to do so. Thus, there is an important temporal dimension to bilingualism research that, in our view, we do not adequately account for with current research practices. That is not to say that findings from older studies are somehow invalid, but rather that those findings are at least partially a property of when (and where) the data were collected. One need only consider the rapid evolution of technology to realize that bilinguals today have opportunities for language learning and interaction that were inconceivable a few decades ago.

Indeed, many previous studies have noted these issues. For example, more than a decade ago, Baum and Titone (2014) advocated for understanding bilingualism as a complex, multifaceted process, and encouraged researchers to explore individual variability among bilinguals. Similarly, Pliatsikas et al. (2020) have advocated for a multifaceted approach to neurolinguistic investigations of bilinguals. Furthermore, many studies have taken multisite approaches to research and have demonstrated the crucial impact of sociocultural context on findings (e.g., Freeman et al., 2022). In the current manuscript, we hope to build on these recommendations and approaches by advocating for a specific approach to bilingual speech research that extends beyond what is typical practice.

Current experimental design and analysis in bilingual speech research

Most studies on bilingual speech use some variation on a common design. Typically, a modest convenience sample of bilingual speakers (e.g., 20–40) is recruited and tested in both of their languages. There may be a language mode manipulation, where participants are exposed to and given instructions in one of the languages to encourage language-specific processing (Lozano-Argüelles et al., 2021), or the target stimuli themselves may be designed with the goal of influencing processing strategies (Casillas & Simonet, 2018). The participants may differ from one another across a range of key variables such as age of onset of L2 learning, quantity and quality of L2 input and exposure, or composite variables such as language dominance (Birdsong et al., 2012). Then, their performance on perception and production tasks is related to their individual differences profiles, processing mode, or another independent variable that is the focus of the study.

This research paradigm has generated a large and important body of work on the (socio)linguistic and contextual variables that affect the extent to which bilinguals develop, maintain, and deploy language-specific representations in speech perception and production (for examples of early work on this topic, see Flege & Eefting, 1987, 1988). Researchers have now begun to use the tools of metaanalysis to sift through this research, demonstrating, for instance, that the "compromise" phonetic categories that bilinguals appear to show in their languages may be more an artifact of study design than a generalizable characteristic of bilingual behavior (Casillas, 2021). Crucially, although we focus intensively on the bilingual speakers that provide data, it bears mentioning that the same argumentation applies to studies involving listener-based evaluations of (bilingual) speech. After all, listeners are at least as complex as speakers, insofar as they bring with them expectations and biases that shape their perception of the speech and the speaker (Lindemann & Subtirelu, 2013), biases that may become active in certain situations (Brown et al., 2023; Reid et al., 2019).

In our view, current practice has created two critical gaps. First, the modest sample size of most studies makes it difficult, if not impossible, to accurately and comprehensively characterize the individuals included in the research. Thus, whether those individuals represent a normally-distributed continuum of language dominance, use, and other key quantitative variables remains an open question. It also bears mentioning that in the US research context, many researchers have relied on college-aged undergraduate student participants, and for work on heritage speakers or speakers of minority languages, these speakers commonly find themselves in a predominantly English-speaking environment for most of their lives. Another issue is that studies have been conducted at a single site, which makes it impossible to separate the characteristics of the individuals from the characteristics of the communities they live in.

We recognize that there are good reasons for both trends. Finding bilingual speakers who have particular linguistic or social characteristics may be difficult, and recruiting participants and bringing them into a lab for experiments is a time consuming process for both the participant and experimenter. Coordinating these efforts across sites can be an added burden in terms of time and expense. At the same time, technological advances in data storage, processing, and analysis have made larger-scale, multisite projects more feasible than ever. If individual bilingual behaviors coalesce into community standards and practices, then large-sample, longitudinal, multicommunity studies are an important frontier for future research.

Opportunities for open speech science

Open science is not new but has gained increasing traction over the last decade. In 2016, Heidi Byrnes, then editor of *The Modern Language Journal*, described an uptick in the collective methodological consciousness of applied linguistics researchers, stating that "methodological issues… demand a kind of professional scrutiny that goes directly to the core of what we do and what we know and what we can tell our publics that we know" (p. 825). One important dimension of methodological scrutiny is the quality, transparency, and reproducibility of research, and applied linguistics initiatives such as the Instruments and Data for Research in Language Studies (IRIS, 2011–, www.iris-database.org) and SLA Speech Tools (Mora-Plaza et al., 2022) are signs that researchers are deeply concerned with these issues (e.g., Plonsky, 2013). While open science is an umbrella term encompassing a range of research principles (for an overview, see Marsden & Morgan-Short, 2023), one especially compelling feature of open science is an emphasis on multisite research (Morgan-Short et al., 2018).

Multisite research brings with it the potential for recruiting a more diverse and representative sample. Scholars have noted several forms of sampling bias in applied linguistics, including an overreliance on college-aged convenience samples (Andringa & Godfroid, 2020). Where bilingualism is concerned, sampling tends to capture bilinguals at a particular moment in their language trajectories, a moment which is unlikely to be representative of the dynamics of their language experience. These considerations can also be inscribed into current broader conversations on understanding language learning, processing, and use in diverse groups of individuals (Kutlu & Hayes-Harb, 2023).

Recruiting a diverse and representative sample is challenging because researchers are bound by the geographic and institutional contexts in which they work and their bureaucratic norms. Yet, where a single researcher or a research team at a single site might struggle, multisite research can succeed. Multisite studies are all but guaranteed to include a larger, more diverse, and therefore more representative sample Moranski and Ziegler (2021), which means that they have a higher chance of achieving generalizability and robust external validity compared to single-site studies. Another advantage to multisite research is that at the analytical stage, researchers can examine the extent to which findings vary across clustering variables such as research sites, classes, or other relevant units. In this way, both the aggregate effect, pooling across units, and unit-specific variation can be modeled. Such an approach has clear interpretational advantages given that it affords the research group the opportunity to compare findings at multiple levels of data structure (Morgan-Short et al., 2018). Applied to bilingual speech research, such an approach could provide insight into the extent to which bilinguals located in distinct communities maintain robust phonetic differences between the languages they know and use. It is easy to imagine how in some communities, these differences might be larger than in others, and those differences could in turn be driven by community-level practices that transcend the level of individual speakers.

Another clear advantage to a multisite approach is the fact that it demands clarity with respect to research tasks and procedures, insofar as they must be aligned across sites. Transparent, reproducible practices ensure that observed differences are due to true differences rather than to non-trivial deviations in study implementation. Such deviations are not the product of researcher apathy or negligence. Indeed, most, if not all, researchers want to carry out high quality studies that have the greatest chance of lending real insight into the research topic with real implications for the target domain(s). Yet, even under the best circumstances, "the lifecycle of any research study is beset by a series of decisions, many of which are essentially arbitrary, whose consequences are usually unknown" (Bolibaugh et al., 2021, p. 804). Making the research process transparent and reproducible pays dividends during the project and afterwards: during the project, taking the time to create reproducible workflows ensures fidelity of implementation and afterwards, future researchers can replicate the project while critically reassessing the methodological decisions that were made at the time. In an ideal scenario, if research tasks and procedures are well documented and publicly available, then researchers outside of the core research network may be able to participate and contribute data to the study (Morgan-Short et al., 2018). This has the added benefit of creating an equitable research landscape where scholars from around the globe can collaborate and interact, which in turn should lead to even more representative samples and research findings.

As data sets grow larger and more complex, so too do analytical choices. Even if tasks and procedures are rigorously implemented and followed, researchers face a garden of forking paths at the analysis stage, where researcher degrees of freedom can lead experts to engage in radically different analytical strategies. These strategies, like their methodological and procedural correlates, can be completely justified (Roettger et al., 2019) yet still have a serious effect on the conclusions that researchers reach (Coretta et al., 2023). The issue is further complicated in bilingualism research, as it is an inherently complex and heterogeneous phenomenon. It is no surprise, then, that analyzing bilingual data involves additional steps that often present the researcher with even more decision points, or researcher degrees of freedom, all of which can have downstream consequences. The case is magnified for speech research given additional decisions that need to be made related to measuring and quantifying a variety of speech phenomena.

A Case in point: Spanish-English Bilinguals in the United States

As Bialystok's "Swiss cheese" metaphor alludes to, grouping units are important because individuals are embedded into layers of structure that shape their behaviors. Bilingualism can certainly be conceived of as an individual characteristic because there is no doubt that each bilingual holds attitudes towards and makes choices about their bilingual identity, and those attitudes and choices affect their behaviors, linguistic and otherwise. At the same time, those attitudes, choices, and behaviors are also constrained, or at least influenced, by community-level language ideologies, patterns of language dominance and use, and so on. We have no doubt that bilingual speech research must continue to focus on individuals and their characteristics, but it will be difficult to understand the full scope of bilingual behavior if the geographies where bilinguals live and work are not considered. To date, most studies have sampled bilinguals from a single community, which means that it is currently not possible to understand the extent to which individual behavior is shaped by community-level patterns. This might explain why, for instance, studies sometimes report contradictory findings. In fact, those findings may not be contradictory at all. Instead, they may simply reflect variation in the bilingual communities from which speakers are drawn, underscoring the fact that bilingualism in one community or region may not look anything like bilingualism in another.

There are myriad forms of bilingualism that could be investigated, but Spanish-English bilingualism presents a unique and significant opportunity because Spanish is the most frequently spoken additional language in the United States (US Census Bureau, 2022). Its diverse speakers include individuals raised in a Spanish-speaking country who have immigrated to the US, individuals raised in the US who have grown up speaking Spanish, and English speakers who have acquired Spanish through a combination of immersion and/or formal instruction. States like California, Arizona, Texas, and Florida have large and well-documented bilingual populations, but there are also populations of bilingual speakers around prominent US urban centers throughout the country and large populations of bilingual speakers in agricultural regions. These communities have their own characteristics in terms of language contact and use, as do their speakers. As such, they offer an unprecedented opportunity for large-scale, multisite research into bilingual speech.

Consider a major urban center such as Chicago. Opportunities to use Spanish and English might vary neighborhood-by-neighborhood, and that variation might condition attitudes toward the use of one language or the other, as well as attitudes toward language mixing and code-switching. Individuals also come with their own personal and intergenerational histories. Perhaps some are recent immigrants to the US, whereas others are first or second generation immigrants. Some later generation speakers might be proficient in Spanish, whereas others might be childhood overhearers. Compare this sample to Spanish-English bilinguals in major urban centers in another state, such as Texas, where community attitudes toward bilingualism are likely to be quite different, given proximity to Mexico and prevailing attitudes toward Spanish in the region. Additionally, in Chicago there are more opportunities for dialectal leveling and accommodation, due to extensive contact between Puerto Rican and Mexican Spanish speakers (Potowski & Torres, 2023), whereas in cities like Houston, El Paso, and San Antonio, despite substantial sociodemographic diversity, there may be fewer features that arise from dialect contact due to the presence of a larger Mexican(-American) Spanish-speaking community (but see Bayley et al., 2012). To put a point on it, what is possible, acceptable (and even typical) language use for a bilingual in Chicago may not be possible and acceptable for a bilingual in El Paso and vice versa. Likewise, what is possible and acceptable for one bilingual in El Paso may not be possible and acceptable for another, depending on the social groups they interact with, the beliefs they hold, and the identities they construct for themselves (Achugar & Pessoa, 2009; Velázquez, 2013). This means that we should perhaps not expect to observe consistent effects across geographic regions and communities, but rather highly nuanced, complex, and even contradictory (i.e., context-specific) ones. In practical terms, by investigating the diverse contexts of Spanish-English bilingualism in the US, we can gain insight into how effects vary across many organizational units, including regions, cities, and neighborhoods, examining how the characteristics of those units affect the range of observed behaviors, outcomes, and effects. In the next section, we give examples of two areas where a large-scale, open science-driven approach to speech research in Spanish-English bilinguals could be particularly informative.

Examples of opportunities for speech research

Bilingual speech research has commonly tested bilinguals in both languages to examine crosslinguistic influence in speech perception and speech production. One basic test of the extent to which sound categories interact is whether bilinguals maintain a robust difference in each language, which in the case of categorical perception might manifest as different crossover boundaries in each language and in the case of production might manifest as differences in production patterns. Another way to view this issue is through the lens of phonetic drift, or the extent to which perception and production are pulled toward the norms of each language (Tobin et al., 2017). Typically, crosslinguistic influence has been studied in terms of the effect of the first language on the second, or of the dominant language on the non-dominant language, but researchers have also grown interested in when and how the first language might be affected by the learning and use of an additional language (Kartushina et al., 2016; Olga et al., 2024). Scholars have shown that crosslinguistic interaction is evident even from the earliest stages of language learning (Chang, 2012) and might therefore contribute to sound change (Chang, 2019). In other words, the learning and use of multiple languages may lead to a progressive blending of the sound systems, such that the sounds come to resemble one another more and more over time. This may be especially true for crosslinguistically similar sounds, which can be defined as sounds that occur in both languages but show subtle yet important phonetic differences in their realization and/or distribution (Flege, 1995; Flege & Bohn, 2021).

Until now, scholars have examined crosslinguistic interaction and phonetic drift in individual speakers (Sancier & Fowler, 1997) or in bilingual samples of modest size (Tobin et al., 2017). Yet, to understand how bilingualism potentially drives sound change and the time course over which that happens, large-scale research is needed. Studying Spanish-English bilingualism presents a golden opportunity because of the size and diversity of the population of Spanish-English bilinguals living in the US and the historical and ongoing importance of this community. More concretely, the size of this population will allow us to understand crosslinguistic influence and its moderators at scale and in precise terms, and the fact that this population is only gaining importance suggests that any large-scale benchmark studies conducted now can serve as a time-stamped baseline against which similar studies carried out in the future can be compared. As an example, stop consonants are among the most frequently researched target structures because they occur in all languages and their acoustic properties are well understood. For this same reason, they are often used as target structures in bilingual speech studies (Chang, 2012; Sancier & Fowler, 1997), including studies involving Spanish and English (Amengual, 2012; Casillas & Simonet, 2018; Gorba & Cebrian, 2023). Imagine a study in which, instead of 20–40 bilingual speakers who are recruited from one site, research was expanded to dozens of sites, each with 20–40 speakers, yielding a total sample of hundreds of individuals. If information about the characteristics of the sites (and the individuals within them) were collected, then much higher-powered analyses could be carried out, and such analyses could include site as a random effect to understand by-site variation. This model could be expanded to other structures, which might show different crosslinguistic effects (because the extent of crosslinguistic interaction may depend on the type of structure being studied), such as the production of laterals (Amengual, 2018). A logical extension of such a model would be examining the extent to which passive exposure to bilingual speech might bring about changes in perception and production in individuals with limited or no functional capacity in the additional language, in the same way that childhood overhearers seem to display a propensity for pronunciation when learning the language in adulthood (Au et al., 2002). In this sense, community overhearers might also be an important conduit of long-term sound change. In fact, bilingualism has often been credited as the source of sound change (Sankoff, 2002; Weinreich, 1968), and recent research has focused on integrating community- and individual-level perspectives into theories of bilingualism and contact-induced sound change (Filipović & Hawkins, 2019; Hawkins & Filipović, 2024; Yao & Chang, 2016).

Another area of special opportunity is related to accent. A speaker's accent is used by listeners to determine facts about the speaker, including, perhaps most obviously, their likely place of origin (e.g., Carmichael, 2016). However, a speaker's accent is also used by listeners for a variety of other aspects of perception including properties of the speaker that are not explicit in the speech, such as intelligence, trustworthiness, and competence (Hanzlíková & Skarnitzl, 2017; Jackson & Denis, 2024; Lev-Ari & Keysar, 2010; Lorenzoni et al., 2024). These features then impact how listeners perceive the speech produced by the speaker (Jiang et al., 2020). That is, a speaker's accent is very likely to significantly impact both how people perceive the speaker and how they perceive their speech.

In the context of bilingualism, this is especially interesting because bilingual speech is often classified as "other" in both languages. For example, Hispanic English is a specific dialect and has a specific accent associated with it, even for speakers who do not speak Spanish. This variety is distinct from what one might consider "Spanish-accented English" (e.g., Mendoza-Denton, 2014). Similarly, Spanish spoken by Heritage Spanish Speakers is sometimes recognized as being distinct from "native" varieties of Spanish (e.g., Potowski, 2012). In bilingual contexts, one can imagine that the accents in each language variety may differ as a function of both issues of dominance, as described above, and a variety of social and identity factors. However, crucially, the varieties of each language spoken will also likely differ as a function of the specific context. For instance, Wolfram et al. (2004) described an emergent variety of Hispanic English in the American South that shares some phonological features with other varieties of Hispanic English but is also distinct from those varieties in important ways. Studying production and perception of accents in bilingual communities can provide insight into both the linguistic and social features of those communities.

The multisite approach championed here would allow researchers to investigate what properties of perception and production of accents are specific to a given context and which are broader features of a bilingual community more generally. For example, some acoustic-phonetic features of Spanish-English bilingual speech might be the result of language contact and thus would likely be observed across several geographic regions. Other features may be specific to the regional varieties of either language spoken in the area or might be indicative of specific linguistic communities, in which case those features would not be consistently observed across communities in a multisite investigation. This could help us understand how accents are formed and what aspects of community identity they are likely to reflect. Furthermore, multisite studies can provide important insights into accent adaptation across different communities and in different participant populations (e.g., Gu et al., 2025).

We have used the term "bilingual" intentionally, with the goal of cultivating an inclusive understanding of the multidimensionality of bilingual speakers, including individuals who have grown up speaking two (or more) languages as well as individuals who have learned an additional language in adulthood via immersion, classroom instruction, or both. Indeed, it is our belief that recruiting bilinguals who vary in terms of the quantity, quality, and consistency of their language contact and use (and the ways in which they predominantly developed their proficiency in the languages they know) holds the best potential for illuminating the variables that affect how sound systems interact in the bilingual mind, a goal that has long been central to speech research and the many models that have been developed to describe L2 speech learning (e.g., Flege & Bohn, 2021). In the next section, we offer general considerations and specific recommendations for carrying out work on bilingual speech, before describing an example of a multisite project under development.

Recommendations

General considerations

It is critical for us to understand precisely who the participants in a given study are and what bilingual populations they best represent. It therefore seems necessary to collect and report on more extensive language background information. In addition to questions around dominance and proficiency, it would be helpful for the research community to collect information about social networks and their language use, geographic factors including the participant's neighborhood, as well as language attitude information. Basic sociodemographic data would also be useful because, as Sugden and Moulson (2015) observed, "even if not interested in the impact of race, culture, and ethnicity on the phenomenon under study, including these variables can add depth to researchers' understanding of a phenomenon" (p. 2). We believe this to be true of bilingualism because it likely interacts with several other social constructs and categories, which may covertly shape how bilingualism manifests in a particular community or individual.

It bears mentioning here that L2 speakers, particularly instructed language learners (who represent one "type" of bilingual), tend to be viewed as somewhat homogenous, possibly because they are enrolled in a common curriculum and therefore tend to have shared sources of input (e.g., their instructors) and patterns of use (e.g., with one another). Perhaps true in the broadest sense, the reality is likely to be far more complex because adult learners pursue language study for diverse reasons and may rely on diverse sources of input, particularly in a globalized and highly digital world where individuals might supplement traditional forms of input with (social) media in the target language and language learning apps. Thus, detailed language information should be collected from classroom learners as well, although the precise information that gets collected may be qualitatively different than, for instance, the information that would be collected from individuals who have grown up using and/or currently actively use multiple languages outside of an academic context.

Pressures of academic publishing often encourage researchers to make broad generalizations about their work in order to demonstrate "impact." However, this can result in sweeping generalizations that do not actually represent the bilingual community that is being studied. We recommend that researchers closely consider how their results may be specific to the population they are examining by including, for instance, a Constraints on Generality statement (Simons et al., 2017). These statements also drive researchers toward a meta-analytic mindset, within which one research goal is to understand how differences in sampling, methodology, and analytical choices might drive differences in outcomes and conclusions. When dealing with a phenomenon as complex as bilingualism, it seems fair to assert that most, if not all studies, reveal some aspect of the true underlying variable of interest.

Much work in our field is produced with an eye toward eventual pronunciation teaching. Therefore, it is critical to explore how this approach to multisite research may impact both how instructors want to address pronunciation and the actual targets they try to teach to their students. This issue is not a new problem [for discussions of inner circle, expanding circle, and outer circle Englishes, see Jenkins (2005)]; however, it is a crucial consideration in current pronunciation instruction (e.g., Hansen Edwards et al., 2021). Understanding appropriate targets for a local context can only be done if those targets have been assessed. Similarly, understanding the range of variability within a given target for a particular context will be easier to do with a more complete understanding of the range both within and across individuals and communities. This can reinforce the shift away from normative models to an intelligibility-based approach (Levis, 2005) anchored in the diverse ways bilinguals perceive and produce language (Sakai, 2018).

Instruments to assess language dominance

Dominance is perhaps the most salient and widely investigated bilingual characteristic. Several standardized instruments have been developed in the past twenty years to systematically assess this construct (Birdsong et al., 2012; Dunn & Fox Tree, 2009; Lim et al., 2008). Dominance-related indices are also embedded in broader measures of bilingual experience, such as the Language Experience and Proficiency Questionnaire (Marian et al., 2007), the Q-BEx Questionnaire (De Cat et al., 2025), the Language and Social Background Questionnaire (Luk & Bialystok, 2013), the Bilingual Input-Output Survey of the Bilingual English-Spanish Assessment (Peña et al., 2021), and the L2 Language History Questionnaire (Li et al., 2006). Alongside these protocols, researchers frequently employ ad hoc tools, ranging from self-ratings of language use and proficiency (Kartushina & Martin, 2019), to detailed background and exposure questionnaires (Tomé Lourido & Evans, 2019), to modality-specific self-assessments of production, comprehension, literacy, and fluency (Antoniou et al., 2012; Flege et al., 2002). This diversity of methods reflects the complexity of the field but also complicates cross-study comparisons (for an example, see supplementary materials). Looking forward, we suggest two priorities for research on language dominance.

First, researchers should align dominance measures with the specific outcomes under investigation. Standardized instruments and bespoke measures of linguistic experience capture different facets of bilingual competence and should not be treated as interchangeable (Flege et al., 2002). We suggest that researchers avoid using language dominance instruments uncritically simply because they are convenient or align with a multifactorial view of language dominance. Instead, they should evaluate whether their study's specific goals would be better achieved by classifying participants using performance-based measures such as mean length of utterance (MLU) or lexical diversity (Cantone et al., 2008), speaker accent ratings (Flege et al., 2002), or other experimental criteria (Hazan & Boulakia, 1993). Additionally, the predictive power of dominance indices can vary: some may show strong correlations with outcomes like phonemic vowel contrasts or voice onset time, while others, such as unstressed vowel reduction, may exhibit weaker or no associations (Amengual & Simonet, 2020). Second, studies should treat dominance as a gradient rather than a categorical variable. Group comparisons between, for instance, Language X- and Language Y-dominant bilinguals should also examine within-group variation, reflecting the scalar, continuous nature of dominance and the heterogeneity of bilingual trajectories (e.g., Beatty-Martinez & Titone, 2021; Luk & Bialystok, 2013). Gradient approaches allow for more precise modeling of individual differences in speech learning, phonetics, and phonology.

Statistical modeling and sampling

Mixed-effects modeling is now the standard for quantitative studies in the language sciences due to the hierarchical, or clustered, nature of the data we collect. In pronunciation research, common designs include at least two hierarchical levels: speakers produce items. However, many additional levels are possible. In listener-based designs, listeners are another grouping variable, and in multisite studies, speakers and listeners might be nested within sites. In all designs, the goal is not simply to characterize the sample of speakers, listeners, and items we have sampled. Instead, we seek to generalize beyond them, to the population of speakers, listeners, and items (and sites, and so on) from which these samples have been drawn. This makes hierarchical modeling a necessary tool. Yet, beyond its necessity to account for variability within and correlations between these clusters, it is also advantageous insofar as the random effects structure of models can be purposefully manipulated to test hypotheses related to the amount of variance associated with grouping variables.

As a concrete example, imagine a study with ten sites. At each site, 30 bilinguals are recruited and asked to complete the BLP to quantify their dominance, as well as a battery of speech perception and production measures. A key advantage of using a hierarchical model with sites as a random effect is that we can account for and estimate the extent to which the relationship between dominance and speech perception and production varies across sites. This would allow us to more faithfully generalize our results beyond a particular site, as well as represent critical first steps in learning whether the dominance-perception or dominance-production relationship is relatively stable or, alternatively, site-specific. If, for instance, the variance estimate were large, perhaps suggesting the effect was site-specific, this might imply that the effect of dominance on speech perception and production depends not just on the characteristics of the individual, but also on the patterns observed in the communities themselves, motivating important downstream analyses and/or future investigations.

There are two primary challenges with estimating such effects. Practically, one needs to sample enough units at each level to estimate the associated variance components. If the goal is to ensure that the sample is representative of the population, then, up to a certain point, the more sampled units, the better, but researchers need to balance conceptual (and analytical) adequacy with practicality. Another challenge with estimating effects lies in the modeling process itself. Random effects, particularly complex random effect structures involving multiple random slopes and/or random slopes for interaction terms, are notoriously difficult to fit because they often fail to converge. As Barr et al. (2013) pointed out, the issue is not whether such variance exists – it certainly does – but rather "whether or not that variation is actually detectable in the sample" (p. 274). Gomes (2022) offers guidance on when to treat a factor as fixed or random based on the number of units, and Brauer and Curtin (2018) include a list of 20 remedies to attempt when models do not converge.

Alternatively, researchers might consider using Bayesian Data Analysis. A full treatment of such an approach is beyond the scope of this paper, but, in general, it can be advantageous for estimating grouping variables in a hierarchical model when priors are appropriately specified. In addition, because using a Bayesian approach provides the researcher with a distribution of samples from the full joint posterior of the model, any parameter of interest, including grouping variables, can be scrutinized and described like any other distribution of data. As a consequence, meaningful questions that go above and beyond simple point estimates can be asked about any random effect, including estimated correlations, regarding the likelihood of observing that effect in the population given the data, the model, and the researchers' prior assumptions (e.g., What percentage of speakers, listeners, etc. would be expected to show that effect?). For a list of helpful references in designing studies and carrying out statistical analyses in frequentist and Bayesian frameworks, see supplementary materials.

Multisite collaborations

While collaboration is a hallmark to the type of research we advocate for here, it is not without challenges on a variety of dimensions ranging from logistical to fiscal to methodological. Addressing these challenges is not trivial, but examples of very large-scale data analysis and collection do exist in our field. Researchers tend to assume that such projects are high in cost and time, but that is not always the case. For example, in a recent paper, Coretta et al. (2023) present coordinated data analysis with 211 authors across 46 teams of researchers. The initiating authors provided a model and infrastructure for the data to be disseminated and analysis conducted and documented. This is one of several successful examples one could model a larger project after.

Of course, this project bypassed one of the more complicated issues in collaboration and large scale research projects because all analysts used the same data set in their analysis. Ensuring that all parties collect data in a comparable way and that the quality of that data is high is also a challenge. However, given the benefits of open science, there are several possible methods that could result in more uniform data collection. Perhaps a team of initiating authors could create a set of stimuli to be used by all researchers. Perhaps they could also provide researchers with apps that could run on any computer and allow for data collection to happen across labs (e.g., via Gorilla research software or another system). Open source software tools which are commonly used in our field could be leveraged for this. Screening for data quality would be a more complicated endeavor, requiring an initiating team of researchers (or perhaps a set of researchers across teams dedicated to quality checks on a given project) to examine the data in more detail, ensuring data from multiple sites are comparable. As these examples illustrate, there is no single best approach, but rather a multitude of approaches with respect to roles, responsibilities, and timelines that can be applied.

Crucially, one must also address ethical considerations in conducting multisite research. It may be difficult to coordinate ethical review across multiple sites, and the potentially conflicting ethical considerations that various institutions may have might necessitate changes to the methods themselves. However, there are many good guidelines for conducting multisite research from, for example, the United States National Science Foundation. These guidelines and individual institutions can share templates for conducting multisite ethical review of projects. In some cases, one institution may be able to provide ethical oversight for all project sites. Financial considerations may also be complicated for multisite research. Given the shifting funding landscape, it is critical to seek out guidance from program officers and other research administrators to determine potential funding sources.

Acknowledging these challenges is important, but it is equally important for us to realize that we, as speech researchers, do not need to reinvent the wheel. Many fields are able to conduct multi-site large-scale projects through collaboration and can serve as a template for our work (see, e.g., the Psychological Science Accelerator). One potential solution to scaling up work in this area would be to pilot projects with mid-sized, multisite teams and use this experience to pilot and develop infrastructure for larger teams. Alternatively, one could begin with multiple sites doing more basic tasks (e.g., analyzing single data sets, as in Coretta et al., 2023), and then expand to additional tasks within these pre-established teams and groups. Further, tools and instruments that are used in a single lab can be modified to be appropriate for multisite research, and there is also potential for materials, tasks, and protocols to be used in their current form. That is, our field has a long and rich history of data collection and many of the same tools we currently use can be repurposed for multisite research.

An Example of a multisite project in development

As a concrete example of how a project could logistically work, we submit an example of a planned project from our own team. Significant work has assessed how individuals use phonological categories within and across their two languages. One well-studied example is that of stop-voicing contrasts in Spanish and English, which differ with regard to the phonetic realization as measured by voice-onset time (VOT). Spanish is typically described as a true-voicing language in which the phonologically voiced stops, /b, d, g/, manifest with pre-voicing or lead VOT, and the voiceless counterparts, /p, t, k,/, are phonetically voiceless with short-lag VOT. English, on the other hand, is often described as an aspirating language in which the category distinction is the same as Spanish (i.e., voiced versus voiceless stops at bilabial, coronal, and velar place), but the phonetic realizations include short- versus long-lag VOT for voiced and voiceless stops, respectively. However, it is possible that these categories within each of these languages might shift as a function of how the two languages are used within communities or how an individual uses their two languages. That is, one could imagine that an individual who routinely switches between English and Spanish may show less differentiation between their English and Spanish stop categories. Similarly, a community where the two languages are used together more frequently, where codeswitching is more prevalent, and so on might show similar patterns at a group level. It may also be the case that the specific varieties of each language being used may impact how phonological categories are shaped and used within and across languages.

Our pilot project under a broader "Sound Communities" umbrella, will collect data at four sites (in Austin, Texas, Chicago, Illinois, New Brunswick, New Jersey, and Santa Cruz, California, where the four authors are located). We will use the same methodology to collect perception data across a number of tasks (e.g., ABX discrimination data, identification data, and visual analog scale data) and production data also across multiple tasks (e.g., repetition, delayed repetition, reading, and picture naming). Additionally, participants will complete extensive language use questionnaires, as well as questionnaires about social networks. We will couple this with available census data about language use in the participants' zip codes. Together, this will allow us to more clearly understand variation in individual and community level language patterns. Because our data collection tools will be developed using open source materials (e.g., Gorilla research software) and our analysis scripts will be shared publicly, they also represent a series of tools that other researchers can use in areas where the Spanish-English bilingual situation may differ from those we are able to capture in our current locations (e.g., bilinguals in Miami where the influence of Cuban Spanish is greater than in any of the aforementioned communities). This would also allow researchers outside of the United States to use the same tools to investigate how language dominance and language variety may impact categories in both languages (e.g., for Spanish-English bilinguals in Spanish-speaking countries around the world, for other language combinations and bilingual scenarios where stop consonant comparisons are relevant and interesting). This approach has rich benefits as described above, and represents a specific question that could not be addressed by a single researcher at a single site.

Conclusion

In order to understand bilingual speech, we must understand that a single site and a single sample are unlikely to be representative of the bilingual experience broadly speaking, just as a single individual is not necessarily representative of an entire community of language users. At first glance, this may seem a truism, but the reality is that we may tacitly adopt this perspective when designing, implementing, and interpreting the results of our studies. One site is incapable of capturing all of the complexity associated with bilingual language use, and each study is likely to inadvertently capture a heterogeneous sample. As siloed researchers, we are likely to continue to discover findings that are, on their face, divergent, when divergence was in fact an inevitable outcome due to non-trivial differences in sampling and study design.

In addition to the methodological concerns raised above it is important to recognize the opportunities that multisite research would create in terms of theoretical advancement in our field. For example, examining what is ostensibly the same population in multiple locations would allow us to investigate synchronic variation that may be driven by un(der)investigated factors that fly under the radar in many linguistic studies (e.g., geography, cf: work in sociolinguistics, Reed, 2020). This work will also pave the way for relatively novel areas of exploration. Language contact via bilingualism is creating new language varieties, with new lexical items, phonetic and phonological targets, and even syntactic properties. The ability to investigate language change in progress and to ask complex questions about diachronic variation is a strength of bilingualism research as a whole, a strength that in our view can best be leveraged through larger-scale, collaborative, open research projects. This type of work can also inform instructional practices, such as what appropriate pronunciation targets might be, by shedding light on the diverse ways bilingual communities perceive and produce speech. Without such a firm understanding, instructional resources and instructors may inadvertently teach to a nonrepresentative model or otherwise elevate one particular way of speaking as normative.

It is also important to remember that bilingualism research can have a real impact on language and education policy and therefore a real impact on people. Findings that are based on narrowly constructed samples, if taken to be representative of all bilinguals, could have negative consequences on the lives of many individuals, including in the availability of translators in critical settings such as healthcare and legal environments or in services for and assessment of bilingual children. Thus, the challenge facing the bilingual research community is bigger than a single phonetic variable or syntactic structure. Instead, it is a challenge that strikes at the very heart of science and research ethics itself. As a community, if we fail to ensure that our methods are sound and that our findings are widely and appropriately applicable to the communities we research, we are not appropriately serving those communities. Rethinking our methodology means reappraising and reinvesting in the ethical side of our work.

Funding

Open Access publication of this article was funded through a Transformative Agreement with Rutgers, The State University of New Jersey.

References

doi Achugar, M., & Pessoa, S. (2009). Power and place: Language attitudes towards Spanish in a bilingual academic community in Southwest Texas. *Spanish in Context*, *6*(2), 199–223.

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

doi Amengual, M. (2018). Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics*, *69*, 12–28.

doi Amengual, M., & Simonet, M. (2020). Language dominance does not always predict cross-linguistic interactions in bilingual speech production. *Linguistic Approaches to Bilingualism*, *10*(6), 847–872.

doi Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics*, *40*, 134–142.

doi Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, *40*(4), 582–594.

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

doi Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

doi Baum, S., & Titone, D. (2014). Moving toward a neuroplasticity view of bilingualism, executive control, and aging. *Applied Psycholinguistics*, *35*(5), 857–894.

Bayley, R., Cárdenas, N. L., Schouten, B. T., & Salas, C. M. V. (2012). Spanish dialect contact in San Antonio, Texas: An exploratory study. *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, 48–60.

doi Beatty-Martinez, A. L., & Titone, D. A. (2021). The quest for signals in noise: Leveraging experiential variation to identify bilingual phenotypes. *Languages*, *6*(4), 168.

Bialystok, E. (2021). Bilingualism as a slice of Swiss cheese. *Frontiers in Psychology*, *12*, 1–6.

Birdsong, D. (2014). Dominance and age in bilingualism. *Applied Linguistics*, *35*(4), 374–392.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

Bolibaugh, C., Vanek, N., & Marsden, E. J. (2021). Towards a credibility revolution in bilingualism research: Open data and materials as stepping stones to more reproducible and replicable research. *Bilingualism: Language and Cognition*, *24*(5), 801–806.

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411.

Brown, B., Tusmagambet, B., Rahming, V., Tu, C.-Y., DeSalvo, M. B., & Wiener, S. (2023). Searching for the "native" speaker: A preregistered conceptual replication and extension of Reid, Trofimovich, and O'Brien (2019). *Applied Psycholinguistics*, *44*(4), 475–494.

Bureau, U. S. C. (2022). *The american community survey: 50 years of personal stories*. U.S. Department of Commerce. https://www.census.gov/content/dam/Census/library/publications/2022/acs/acs-50.pdf

Byrnes, H. (2016). Notes from the editor: Celebrating 100 years of" the modern language journal". *The Modern Language Journal*, 3–18.

Cantone, K., Kupisch, T., Müller, N., & Schmitz, K. (2008). Rethinking language dominance in bilingual children. *Linguistische Berichte*, *215*, 307–343.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421–428.

Carmichael, K. (2016). Place-linked expectations and listener awareness of regional accents. In A. M. Babel (Ed.), *Awareness and control in sociolinguistic research* (pp. 123–151). Cambridge University Press.

Casillas, J. V. (2021). Interlingual interactions elicit performance mismatches not "compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops. *Languages*, *6*(9), 1–20.

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64.

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

Chang, C. B. (2019). Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research. *Journal of Phonetics*, *74*, 96–113.

Coretta, S., Casillas, J. V., Roessig, S., Franke, M., Ahn, B., Al-Hoorie, A. H., Al-Tamimi, J., Alotaibi, N. E., AlShakhori, M. K., Altmiller, R. M., Arantes, P., Athanasopoulou, A., Baese-Berk, M. M., Bailey, G., Sangma, C. B. A., Beier, E. J., Benavides, G. M., Benker, N., BensonMeyer, E. P., … Roettger, T. B. (2023). Multidimensional signals and analytic flexibility: Estimating degrees of freedom in human-speech analyses. *Advances in Methods and Practices in Psychological Science*, *6*(3), 1–29.

doi De Cat, C., Gusnanto, A., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2025). How detailed do measures of bilingual language experience need to be? A cost-benefit analysis using the Q-BEx questionnaire. *Bilingualism: Language and Cognition*.

doi Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition*, *12*(3), 273–289.

doi Filipović, L., & Hawkins, J. A. (2019). The Complex Adaptive System Principles model for bilingualism: Language interactions within and across bilingual minds. *International Journal of Bilingualism*, *23*(6), 1223–1248.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

doi Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge, MA: Cambridge University Press.

doi Flege, J. E., & Eefting, W. (1987). Production and perception of English stops by native spanish speakers. *Journal of Phonetics*, *15*, 67–83.

doi Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of english and spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America*, *83*(2), 729–740.

doi Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*(4), 567–598.

doi Freeman, M. R., Robinson Anthony, J. J., Marian, V., & Blumenfeld, H. K. (2022). Individual and sociolinguistic differences in language background predict stroop performance. *Frontiers in Communication*, *7*, 1–18.

doi Gomes, D. G. (2022). Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model? *PeerJ*, *10*, 1–16.

doi Gorba, C., & Cebrian, J. (2023). The acquisition of L2 voiced stops by English learners of Spanish and Spanish learners of English. *Speech Communication*, *146*, 93–108.

doi Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15.

doi Grosjean, F. (1997). The bilingual individual. *Interpreting*, *2*(1–2), 163–187.

Gu, Y., Xie, X., & Kurumada, C. (2025). *The accent atlas: A geolocation-based assessment of non-native accent familiarity and linguistic diversity*. OSF Preprint. https://osf.io/gf9jy

doi Hansen Edwards, J., Chan, K. L. R., Lam, T., & Wang, Q. (2021). Social factors and the teaching of pronunciation: What the research tells us. *RELC Journal*, *52*(1), 35–47.

doi Hanzlíková, D., & Skarnitzl, R. (2017). Credibility of native and non-native speakers of English revisited: Do non-native listeners feel the same? *Research in Language*, *15*(3), 285–298.

doi Hawkins, J., & Filipović, L. (2024). Bilingualism-induced language change: What can change, when, and why? *Linguistics Vanguard*, *10*(s2), 115–124.

doi Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, *36*(1), 17–38.

doi Jackson, S., & Denis, D. (2024). What I say, or how I say it? Ethnic accents and hiring evaluations in the Greater Toronto Area. *Language*, *100*(2), e27–e62.

Jenkins, J. (2005). Teaching pronunciation for English as a lingua franca: A sociopolitical perspective. In C. Gnutzmann & F. Intemann (Eds.), *The globalisation of english and the english language classroom* (pp. 145–158). Gunter Narr Verlag Tübingen.

Jiang, X., Gossack-Keenan, K., & Pell, M. D. (2020). To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quarterly Journal of Experimental Psychology*, *73*(1), 55–79.

Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning*, *66*(S2), 155–186.

Kartushina, N., & Martin, C. D. (2019). Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production. *Journal of Phonetics*, *77*, 1–21.

Kelly, N. (2022). Change across time in L2 intonation vs. Segments: A longitudinal study of the English of Ole Gunnar Solskjaer. *Languages*, *7*(3), 1–17.

Kornder, L., & Mennen, I. (2021). Longitudinal developments in bilingual second language acquisition and first language attrition of speech: The case of Arnold Schwarzenegger. *Languages*, *6*(2), 1–25.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, *25*(5), 497–514.

Kutlu, E., & Hayes-Harb, R. (2023). Towards a just and equitable applied psycholinguistics. *Applied Psycholinguistics*, *44*(3), 293–300.

Leeuw, E. de. (2019). Native speech plasticity in the German-English late bilingual Stefanie Graf: A longitudinal study over four decades. *Journal of Phonetics*, *73*, 24–39.

Leeuw, E. de, & Chang, C. B. (2023). Phonetic and phonological L1 attrition and drift in bilingual speech. In M. Amengual (Ed.), *The cambridge handbook of bilingual phonetics and phonology*. Cambridge University Press Cambridge.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, *38*, 202–210.

Lim, V. P. C., Liow, S. J. R., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Determining language dominance in English-Mandarin bilinguals: Development of a self-report classification tool for clinical use. *Applied Psycholinguistics*, *29*(3), 389–412.

Lindemann, S., & Subtirelu, N. C. (2013). Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning*, *63*(3), 567–594.

Lorenzoni, A., Faccio, R., & Navarrete, E. (2024). Does foreign-accented speech affect credibility? Evidence from the illusory-truth paradigm. *Journal of Cognition*, *7*(1), 26.

Lozano-Argüelles, C., Arroyo, L. F., Rodríguez, N., López, E. M. D., Pozú, J. J. G., Markovits, J., Varela, J. P., Rocafiguera, N. de, & Casillas, J. V. (2021). Conceptually cued perceptual categorization in adult L2 learners. *Studies in Second Language Acquisition*, *43*(1), 204–219.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, *25*(5), 605–621.

doi Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967.

doi Marsden, E., & Morgan-Short, K. (2023). (Why) are open research practices the future for the study of language learning? *Language Learning*, *73*, 344–387.

doi Mejías, H. A., Anderson-Mejías, P. L., & Carlson, R. (2003). Attitude update: Spanish on the South Texas border. *Hispania*, 138–150.

doi Mendoza-Denton, N. (2014). *Homegirls: Language and cultural practice among latina youth gangs*. John Wiley & Sons.

doi Moranski, K., & Ziegler, N. (2021). A case for multisite second language acquisition research: Challenges, risks, and rewards. *Language Learning*, *71*(1), 204–242.

doi Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). *Tools for second language speech research and teaching*.

doi Morgan-Short, K., Marsden, E., Heil, J., Issa Ii, B. I., Leow, R. P., Mikhaylova, A., Mikołajczak, S., Moreno, N., Slabakova, R., & Szudarski, P. (2018). Multisite replication in second language acquisition research: Attention to form during listening and reading comprehension. *Language Learning*, *68*(2), 392–437.

Olga, D., Celata, C., Natalia, K., et al. (2024). Native speech perception in the context of multilingualism and language learning. *Languages*, *7*.

Peña, E. D., Bedore, L. M., & Torres, J. (2021). Assessment of language proficiency and dominance in monolinguals and bilinguals. In W. S. Francis (Ed.), *Bilingualism across the lifespan: Opportunities and challenges for cognitive research in a global society* (pp. 88–105). Routledge.

doi Pliatsikas, C., DeLuca, V., & Voits, T. (2020). The many shades of bilingualism: Language experiences modulate adaptations in brain structure. *Language Learning*, *70*(S2), 133–149.

doi Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*(4), 655–687.

Potowski, K. (2012). Identity and heritage learners: Moving beyond essentializations. In S. M. Beaudrie & M. Fairclough (Eds.), *Spanish as a heritage language in the United States: The state of the field* (pp. 179–202). Georgetown University Press.

Potowski, K., & Torres, L. (2023). *Spanish in chicago*. Oxford University Press.

doi Reed, P. E. (2020). Place and language: Links between speech, region, and connection to place. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(3), 1–11.

doi Reid, K. T., Trofimovich, P., & O'Brien, M. G. (2019). Social attitudes and speech ratings: Effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition*, *41*(2), 419–442.

doi Roettger, T. B., Winter, B., & Baayen, H. (2019). Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. *Journal of Phonetics*, *73*, 1–7.

doi Sakai, M. (2018). Moving towards a bilingual baseline in second language phonetic research. *Journal of Second Language Pronunciation*, *4*(1), 11–45.

doi Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436.

Sankoff, G. (2002). Linguistic outcomes of language contact. In J.K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (Vol. 1, pp. 638–668). Blackwell.

Simons, D.J., Shoda, Y., & Lindsay, D.S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128.

Sugden, N.A., & Moulson, M.C. (2015). Recruitment strategies should not be randomly selected: Empirically improving recruitment success and diversity in developmental psychology research. *Frontiers in Psychology*, *6*, 523.

Tobin, S.J., Nam, H., & Fowler, C.A. (2017). Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics*, *65*, 45–59.

Tomé Lourido, G., & Evans, B. (2019). The effects of language dominance switch in bilinguals: Galician new speakers' speech production and perception. *Bilingualism: Language and Cognition*, *22*(3), 637–654.

Velázquez, I. (2013). Individual discourse, language ideology and Spanish transmission in El Paso, Texas. *Critical Discourse Studies*, *10*(3), 245–262.

Weinreich, U. (1968). *Languages in contact: Findings and problems*. Mouton de Gruyter.

Wolfram, W., Carter, P., & Moriello, B. (2004). Emerging Hispanic English: New dialect formation in the American South. *Journal of Sociolinguistics*, *8*(3), 339–358.

Yao, Y., & Chang, C.B. (2016). On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese. *Language*, *92*(2), 433–467.

Supplementary materials

Comparison of Participant Data in Ten Studies on Spanish-English Bilinguals

Table 1 highlights ten studies of Spanish-English bilinguals in two U.S. regions: Arizona (Southwest) and Florida (Southeast). It shows publication dates and whether each study included family history, language use, dominance, and objective proficiency, four variables central to bilingual experience. The goal is not to critique what is missing, but to underscore the diversity of approaches to defining and measuring bilingualism, with particular attention here to one key construct: language dominance.

Table 1. Example studies on bilingual communities in the US Southwest and Southeast

Paper	Region	Language	Period	n	Family history	Language use	Language dominance	Objective proficiency
Lynn (1940)	Southwest (Arizona)	Spanish	1940	150	×	✓	✓	×
DuBord (2004)	Southwest (Arizona)	Spanish	2000	18	×	×	✓	×
Christoffersen (2014)	Southwest (Arizona)	Spanish	2010	30	✓	✓	✓	×
Gonzales & Lotto (2013)	Southwest (Arizona)	Spanish	2010	32	×	✓	✓	×

Paper	Region	Language	Period	n	Family history	Language use	Language dominance	Objective proficiency
Rodríguez-Guerra et al. (2023)	Southwest (Arizona)	Spanish	2020	61	×	×	×	✓
Byers & Yavas (2017)	Southeast (Florida)	English	2010	60	✓	✓	✓	✓
Brice et al. (2009)	Southeast (Florida)	Spanish/ English	2000	16	✓	✓	✓	✓
Alvord & Rogers (2014)	Southeast (Florida)	Spanish	2010	10	✓	✓	×	×
Carter et al. (2020)	Southeast (Florida)	English	2020	25	×	×	×	×
Flege & Hammond (1982)	Southeast (Florida)	English	1980	50	✓	×	×	×

Of the ten studies in Table 1, four of them reported participants' language dominance. However, methods varied widely: some used standardized instruments (e.g., the Bilingual Dominance Scale, Dunn & Fox Tree, 2009), others relied on self-reported fluency or language use, and some assigned dominance classifications without methodological detail. This variability makes it difficult to compare findings across studies and limits the generalizability of conclusions.

Although ad hoc measures remain common, the past two decades have produced several standardized tools that are reliable, valid, and accessible — notably the Bilingual Language Profile (BLP, Birdsong et al., 2012), the Bilingual Dominance Scale (BDS, Dunn & Fox Tree, 2009), the Self-Report Classification Tool (SCRT, Lim et al., 2008), and the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian et al., 2007). Comparative evaluations of these instruments are provided by Gertken et al. (2014) and Peña et al. (2021).

Statistical modeling resources

For researchers interested in following some of our suggestions, there are many resources available to help determine best practices. Winter (2019) is a user-friendly introduction to regression modeling for linguistics in R. Winter's book builds from simple models to the multilevel models that have become the standard in the field. Nagle (2024) addresses multilevel modeling for pronunciation data. Gelman & Hill (2007) is a general treatment of multilevel modeling. Gomes (2022) deals with treating a factor as a fixed or random effect depending on the number of units, and Austin & Leckie (2018) is also useful when considering cluster size in conjunction with power and Type 1 error.

With respect to Bayesian modeling, Levshina (2022) is a comparison of frequentist and Bayesian approaches, which can help researchers understand differences between the two, and Vasishth et al. (2018) and Nalborczyk et al. (2019) are Bayesian tutorials dealing with phonetic data.

References

doi Alvord, S. M., & Rogers, B. M. A. (2014). Miami-Cuban Spanish vowels in contact. *Sociolinguistic Studies*, *8*(1), 139–170.

doi Austin, P. C., & Leckie, G. (2018). The effect of number of clusters and cluster size on statistical power and Type I error rates when testing random effects variance components in multilevel linear and logistic regression models. *Journal of Statistical Computation and Simulation*, *88*(16), 3151–3163.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin. https://sites.la.utexas.edu/bilingual/

doi Brice, A. E., Carson, C. K., & Dennis O'Brien, J. (2009). Spanish-English articulation and phonology of 4-and 5-year-old preschool children: An initial investigation. *Communication Disorders Quarterly*, *31*(1), 3–14.

doi Byers, E., & Yavas, M. (2017). Vowel reduction in word-final position by early and late Spanish-English bilinguals. *PloS ONE*, *12*(4), 1–24.

doi Carter, P. M., López Valdez, L., & Sims, N. (2020). New dialect formation through language contact: Vocalic and prosodic developments in Miami English. *American Speech*, *95*(2), 119–148.

Christoffersen, K. O. (2014). Does child code-switching demonstrate communicative competence?: A comparison of simultaneous and sequential bilinguals. *Journal of Second Language Acquisition and Teaching: Arizona Working Papers in SLAT*, *21*, 20–40.

DuBord, E. M. (2004). Gender assignment to english words in the spanish of southern arizona [PhD thesis]. University of Arizona.

doi Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition*, *12*(3), 273–289.

doi Flege, J. E., & Hammond, R. M. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*, *5*(1), 1–17.

doi Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

doi Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual Language Profile. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 208–225). Multilingual Matters.

doi Gomes, D. G. (2022). Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model? *PeerJ*, *10*, 1–16.

doi Gonzales, K., & Lotto, A. J. (2013). A bafri, un pafri: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychological Science*, *24*(11), 2135–2142.

Levshina, N. (2022). Comparing Bayesian and frequentist models of language variation: The case of *help + (to-)infinitive*. In O. Schützler & J. Schlüter (Eds.), *Data and methods in corpus linguistics: Comparative approaches* (pp. 224–258). Cambridge University Press.

doi Lim, V. P. C., Liow, S. J. R., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Determining language dominance in English-Mandarin bilinguals: Development of a self-report classification tool for clinical use. *Applied Psycholinguistics*, *29*(3), 389–412.

Lynn, K. (1940). A phonetic analysis of the English spoken by Mexican children in the elementary schools of Arizona [PhD thesis]. Louisiana State University Historical Dissertations; Theses.

doi Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967.

Nagle, C. (2024). *A guide to quantitative research methods in second language pronunciation*. Routledge.

doi Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P.-C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242.

Peña, E. D., Bedore, L. M., & Torres, J. (2021). Assessment of language proficiency and dominance in monolinguals and bilinguals. In W. S. Francis (Ed.), *Bilingualism across the lifespan: Opportunities and challenges for cognitive research in a global society* (pp. 88–105). Routledge.

doi Rodríguez-Guerra, M., Colina, S., & Fabiano-Smith, L. (2023). Interaction in bilingual early speech acquisition: Acceleration in the bilingual acquisition of English liquids for English-Spanish bilinguals. *Lingua*, *281*, 1–21.

doi Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161.

Winter, B. (2019). *Statistics for linguists: An introduction using R*. Routledge.

Address for correspondence

Joseph V. Casillas
College of Arts and Sciences, Department of Spanish & Portuguese
Rutgers University
Academic Building-CAC
15 Seminary Place
New Brunswick, NJ 08901
United States

joseph.casillas@rutgers.edu
https://orcid.org/0000-0001-8735-9910

Co-author information

Charlie Nagle
College of Liberal Arts, Department of Spanish & Portuguese
University of Texas at Austin
cnagle@austin.utexas.edu
https://orcid.org/0000-0003-2712-2705

Melissa M. Baese-Berk
Division of the Humanities, Department of Linguistics
University of Chicago
mmbb@uchicago.edu
https://orcid.org/0000-0002-4855-9319

Mark Amengual
Humanities Division, Department of Languages and Applied Linguistics
University of California
amengual@ucsc.edu
https://orcid.org/0000-0003-2007-9687

Publication history

Date received: 3 March 2025
Date accepted: 2 December 2025
Published online: 5 February 2026