

Spanish Language Maintenance and Shift in a Newly-Forming Community in the Southeastern United States: Insights From a Large-Class Survey

Jim Michnowicz¹, Sonya Trawick², Rebecca Ronquest³

¹ Foreign Languages and Literatures, NC State University, ² Spanish, Italian and Portuguese, Pennsylvania State University, ³ Foreign Languages and Literatures, NC State University

Keywords: Sociolinguistics, Language Contact, Language Maintenance, Language Shift, Spanish in the U.S., Spanish in the southeast

Hispanic Studies Review

Vol. 7, Issue 2, 2023

The southeastern United States has seen unprecedented growth in the Latino/Hispanic population over the past two decades. This growth creates a new language contact situation, that can provide insight into not only newly forming, incipient bilingual communities, but also the early stages of language contact that would have also occurred in more established bilingual communities. One of the primary findings of previous work is the shift to English among US Latinos. These studies, however, have largely been carried out in established communities with longstanding Latino populations (i.e. in the Latino ‘Heartland’ (Villa & Rivera-Mills, 2009)). The present study examines patterns of Spanish language maintenance and shift (LMLS) in the emerging bilingual community in North Carolina. Based on more than 1000 surveys with NC Latinos, results show clear patterns of language shift, along with suggestions that this shift is happening more quickly than in more established regions. In particular, younger, second-generation speakers are leading the shift to English, although indications of shift are also found among first-generation immigrants, in ways that both connect and distinguish NC from other regions.

1. Introduction: Spanish in North Carolina and the Southeast

The southeastern U.S. has experienced some of the fastest growth in the Latino population in the U.S. over the past two decades. North Carolina (NC) provides one such example, from approximately 75,000 Hispanics in 1990 to over 800,000 in 2010, more than 900% growth (*North Carolina’s Hispanic Community: 2021 Snapshot*, 2021). This rapid growth has created a new language contact situation, where it has been argued that the initial steps of Spanish-English contact can be observed, providing insight into not only incipient communities, but also into the development of long-standing communities elsewhere in the U.S. (e.g., Carter, 2005; Michnowicz et al., 2018; Ronquest et al., 2020). Southeastern states such as NC and other “New Destination” communities (Zúñiga & Hernández-León, 2005) serve as a space where recent first-generation (G1) immigrants from various countries are in constant contact not only with L1 English speakers, but also with second-generation (G2) Spanish-speaking populations, which make up a small majority of Spanish-speakers in NC (61% as of 2021; *North Carolina’s Hispanic Community: 2021 Snapshot*, 2021). As of 2020, there were close to one million residents identifying as Hispanic/Latino in NC, representing almost 10% of the total population of the state. Of these, people of Mexican descent make up a little more than half (55%), with another quarter composed of Central American (16%) or Puerto Rican (11%) origin, followed by other smaller groups representing Spain and Latin America (U.S. Census Bureau, 2020).

This newly forming, diverse community of Spanish-speakers in NC provides an ideal context in which to study language contact phenomena, including questions of language maintenance and language shift (hereafter LMLS) among G1 and G2 populations, which is the focus of the present study. Based on survey data with more than 1000 participants, this study examines patterns and indices of LMLS across generations and in a variety of contexts/domains in order to provide insight into the future direction of Spanish-English bilingualism in NC.

2. Language Maintenance/Language Shift

LMLS has been widely studied among immigrant groups throughout the U.S. and beyond, and at their core these studies seek to answer one primary question: has there been a change in the language(s) used by a speech community over a certain period of time (e.g., Fishman, 1964, 1965, 1966; Porcel, 2011)? Differing rates of language use over time can indicate LS, frequently to the dominant language, whereas continued use of the immigrant language (across or within particular domains) would constitute evidence for LM among a given community. Importantly, LS can be bidirectional, with generations moving back and forth between different degrees of shift and maintenance, although the most common result, particularly for multilingual communities in the U.S., is loss of the immigrant language (Porcel, 2011). LM, on the other hand, has been defined as stability in language use “that has persisted without dramatic change for more than three or four generations and that shows no sign of incipient change” (Thomason, 2001, p. 23). In the context of the U.S., research suggests Spanish LM on a societal level despite LS on an individual level (Escobar & Potowski, 2015; Silva-Corvalán, 1994). In this situation, Spanish is maintained as an active language at the community level due to the continued immigration of G1, Spanish-dominant speakers, but G2 speakers show a marked shift to English. This suggests that in the future, language attrition among Spanish-speaking communities may continue to accelerate, perhaps leading to language loss in some communities, as has also been found for other immigrant languages in the U.S. (see, e.g., Grosjean, 1982).

Studies on LMLS have indicated a three-generation model of LS, where G1 speakers are dominant in the home language, their G2 children show a more balanced bilingualism between the home language and the community language, and their third-generation (G3) grandchildren are English dominant - a situation which leads to the likely loss of the home language (Fishman, 1964, 1965, 1966; Grosjean, 1982; Romaine, 1995). Other evidence, however, implies that LS happens more quickly, as both G1 and G2 speakers can show evidence of LS (Grosjean, 1982; Haugen, 1969; Veltman, 1983). This accelerated shift is supported by national survey data by the Pew Research Center which finds a decrease in reported Spanish and a concomitant increase in reported English across three generations, a pattern beginning even among G1 speakers (Pew Research Center, 2017). Likewise, G. Bills et al. (2000) argue that in regard to Spanish in the U.S., the three-generation model “is

an oversimplified account of the actual state of affairs and in many respects underestimates the rapidity with which linguistic absorption into the dominant society is actually taking place” (15).

Alternatively, some studies have found that the three-generation model underestimates LM. Inspecting various bilingual groups from different geographic regions in the U.S., Mora, Villa, & Dávila (2005, 2006) report high rates of Spanish transmission from G1 immigrant parents to their G2 children. Research has also documented the maintenance of Spanish far beyond the three-generation window. For example, Anderson-Mejías (2005) finds that Spanish was maintained into the 5th generation in Texas, and Villa & Rivera-Mills (2009) find Spanish maintenance on some level into the 7th generation in New Mexico. Of note here is the fact that these studies provide examples of communities where the presence of Spanish and Hispanic/Latino heritage is well-established. The multi-generational maintenance in these regions can often involve a type of cyclical bilingualism, where later generations, in some sense, “reclaim” their heritage language through education or other means, thereby complicating theories of what constitutes LMLS (e.g., Silva-Corvalán, 2001).

Population density and other demographic and social factors such as attitudes toward the language(s) in contact can all determine the outcomes of LMLS (Porcel, 2011). For example, it has been shown that geographic distance from the Mexican border affects LM in some parts of the country, as Spanish in border regions is constantly renewed by patterns of cyclical immigration and bilingualism (G. D. Bills et al., 1995; Mora et al., 2005). Villa & Rivera Mills (2009) refer to the importance of the “heartland factor”, where “[t]he ‘heartland’ can be defined as a region in which those of Spanish-speaking origin have a historic presence, form a demographic majority in many areas and move back and forth across national and international political borders, thus creating a bilingual dynamic in which Spanish is lost or maintained in relation to its affective and instrumental values” (29; see above examples of communities in Texas and New Mexico). Rivera-Mills (2007, as cited in Villa & Rivera-Mills, 2009) describes an “identity link” that can lead Heritage speakers to reinforce or learn their heritage language, further complicating the model of generational shift. At the same time, in spite of continued maintenance on some level across generations due to these societal factors, Villa & Rivera Mills (2009) find no monolingual Spanish speakers in the U.S. after G1, demonstrating that speakers have already begun the shift to English by G2, at least in some domains.

Even more important than physical proximity to Spanish-speaking countries may be the presence of a large, stable Spanish-speaking population, even if it is distant from a geo-political border (for example, New York City or Chicago; see McCullough & Jenkins, 2005; Mills, 2005; Rivera-Mills, 2000b, 2000a). Important to our understanding of LMLS in newly forming regions like NC is the fact that this part of the southeastern U.S. is neither geographically close to the border, which would allow for cyclical immigration and bilingualism,

nor possessing of a long-established, stable Latino population like that found in other regions of the U.S. Finally, conflicting results and patterns uncovered by previous studies on LMLS serve to reinforce that sociolinguistic generations are not homogeneous, and a wide level of variation is to be expected not only between generations but also within speakers of the same generation (Anderson-Mejías, 2005).

Based on patterns discovered in previous research on Spanish in the US, the present study seeks to provide an initial answer to the following two research questions:

1. What are the reported patterns of language use among Spanish-speakers in NC?
2. What do these patterns suggest regarding LMLS in NC?

3. Methods

In order to answer the research questions, data were collected as part of a larger classroom-based project on Spanish in NC over a five-year period (2014-2016; 2019-2020). Multipart surveys were distributed by undergraduate students in two of the authors' senior seminar courses on Spanish in the U.S. The survey asked which language(s) participants use most often - Spanish, English, or Both Spanish and English Equally (henceforth *Both Equally*) - across four contexts or domains of use: with their families, with their friends, at work, or when watching television. These answers were analyzed in light of the results of a short demographic questionnaire included in the survey¹.

Each student in the course administered a minimum of 10 surveys. In order to avoid a bias towards participants more comfortable with technology, approximately 50% of the surveys were distributed on paper, while the remaining 50% were collected online via Google Forms. A total of 1081 surveys were collected. After removing surveys that lacked responses to the language use questions, 1054 surveys were included in the final analysis.

Data were analyzed using a multinomial logistic regression with the *nnet* package (Venables & Ripley, 2002) in R (R Core Team, 2021). Multinomial logistic regression is used when a discrete dependent variable has more than two levels, allowing for all data to be analyzed at once, rather than subdivided for binary comparisons. A minimal main effects model was fit via model comparison with *ANOVA*, and interactions were modeled and visualized with conditional inference trees run via the *partykit* package (Hothorn & Zeileis, 2015). Plots were created with the packages *ggplot2* (Wickham, 2016) and *sjPlot* (Lüdtke, 2021).

¹ The survey also included questions related to loanwords and forms of address, as discussed in Michnowicz et al. (2018). As the survey was originally designed to be distributed and analyzed by numerous undergraduate students within one course and across multiple semesters, the survey is purposefully brief and focused. A focused survey such as this can provide a clear initial picture of language use patterns, serving as motivation for future, more detailed research. A copy of the survey used for these studies and an English translation can be accessed at: <https://drive.google.com/drive/folders/1rzZaRg7TuqKE0AB1xB-JRmQV1UPL6855?usp=sharing>

As with any self-reported survey data of this sort, participants may over- or under-estimate their patterns of use in a particular context, or may respond with what they think they should say (see discussions in, e.g., Delgado et al., 1999). Responses will be interpreted in this light, and taken to represent attitudes toward particular language(s) or ways in which speakers view themselves vis-à-vis different linguistic communities in NC.

4. Results and Analysis

[Table 1](#) provides the variables included in the quantitative analysis, as well as the token counts for the 1054 survey participants. Bolded levels represent the reference levels in the regression models.

The results of the multinomial logistic regression analyses are found in Tables A1 and A2 in the Appendix. In order to more effectively determine which variables correlate to a shift to English in NC, the reference level for each variable was set to the group that most favored Spanish, discussed below and bolded in [Table 1](#). Sex was not significant in any model, and based on the model comparisons with an *ANOVA*, was removed from the statistical models. The minimal model found significant effects of Age Group, Generation, Education Level, and Context/Domain. Time in the US was also significant for G1 speakers, as determined by a separate model run with G1 speakers only. As with a binary logistic regression, a positive coefficient signifies more use of a particular language compared to the reference level. The difference between the middle age group and older speakers was not significant in the overall data set. All other comparisons produced significant results.

We now continue with a discussion of the results by variable of interest, followed by a discussion of the results of a regression analysis.

4.1 Survey Responses

As mentioned, data were collected over a 5-year period. Before continuing, we present the results across the years of data collection in [Figure 1](#).

We found a similar pattern in all 5 years: more Spanish for G1 participants, more English for G2 participants, and similar rates of *Both Equally* for both groups. Closer inspection shows a small increase in reported English and *Both Equally* among G1 participants across time. Still, the overall similarity across years justifies including all five years of data together, but with continued longitudinal data collection these real-time variables merit continued attention.

Results from the present study suggest both language maintenance and language shift in different ways and contexts. [Figure 2](#) shows the overall results of language use across the contexts and domains of language use (family, friends, work, TV).

English makes up 43% of respondents' answers, with Spanish representing a further 30%, and *Both Equally* making up the remaining 26%. The overall finding is suggestive of language shift, as English was the most common response in the data by a fairly wide margin. Further analysis demonstrates that language use is highly dependent on a variety of factors, detailed below.

Table 1. Variables in the quantitative analyses and the results of questions in the survey, in which bolded levels represent the reference levels in the regression models (total surveys = 1054)^a

Variables	Levels	Survey Responses	
		N	% per variable
Dependent Variable			
Reported Language Use			
	Spanish English Both Equally	<i>Responses Vary by Context/Domain of Language Use</i>	
Independent Variables			
Participant Sex			
	Male Female Other ^b	426 620 1	41% 59% <1%
Age Group			
	Older (50+) Middle (30-49) Younger (18-29)	152 359 507	15% 35% 50%
Education Level			
	Less than High School ^c High School University	99 207 725	10% 20% 70%
Generation			
	G1 (born outside of the U.S.) G2 (born in the U.S.)	682 359	66% 44%
Time in the U.S. (for G1 only)			
	A (0-10 years) B (11-19 years) C (20+ years)	180 280 224	26% 41% 33%
Geographic Region ^d			
	Mexico Northern S America Caribbean Central America Southern Cone US Unspecified Spain	276 237 156 100 52 40 8	32% 27% 18% 12% 6% 5% 1%
Context of Language Use			
	Family Friends Work Television	<i>Responses Vary by Language</i>	

^aThe N for each variable does not sum to the total number of completed surveys due to participants having the option to leave certain questions blank in the survey.

^bThe survey administered during the last two years of data collection, 2019 and 2020, allowed participants to choose “Other” for their sex/gender. In the present data, only one person chose to identify as “Other”. In order to not skew the results for speaker sex, this participant was excluded from the statistical analysis.

^cBecause the vast majority of respondents had at least a high school education, levels below High School were collapsed into Less than High School.

^dDue to imbalances and NAs, Geographic Region was not included in the analysis, but should be explored further in future research.

[Figure 3](#) provides the reported language use by generation, a significant result in the regression analysis ($p < 0.001$).

G2 speakers report far more English and far less Spanish than G1 speakers, again suggestive of language shift in NC.²

[Figure 4](#) provides responses by participant sex.

As indicated previously, participant sex was not found to be a significant predictor of language use, as men and women show very similar reported patterns of use. This is a somewhat surprising finding, given that men were found to report significantly more English loanword use in NC (Michnowicz et al., 2018), an indicator of language contact affecting language use, and women were shown to have more positive attitudes toward Spanish in Georgia, another high-growth state in the Southeast (2009). Further research should continue to examine sex-based patterns across time.

[Figure 5](#) shows the language use results by age group.

Older (50+) and middle (30-49) age groups show similar patterns, and the differences between these two age groups were not found to be significant (Appendix Table A1). Highly significant differences were found for the younger age group (18-29), however, with younger speakers favoring both English and *Both Equally* compared to Spanish (the reference level) ($p < 0.001$). The most striking difference is the 25-point jump in reported English among the younger age group. This finding suggests a rapid shift to English among younger Spanish speakers in NC. The present finding is consistent with a Pew Research Center study (Krogstad et al., 2015), that reported higher rates of speaking English “well” and “English only” among younger U.S. Latinos relative to older speakers.

The results by education level reveal significant differences among all levels ($p < 0.001$), shown in [Figure 6](#).

First, Spanish is the primary language by far among lesser-educated respondents, who also report the lowest rates of English. This speaks to the role of education in promoting the dominant language, seen among the other two education groups (see Krogstad et al., 2015, for national trends). Next, high school-educated respondents show remarkable balance across languages, as these speakers represent a transitional group between Spanish dominance on one side, and English dominance on the other. Finally, college-educated respondents show the inverse pattern of the lower educated group, although to a more moderate degree. Interestingly, these speakers show the same rate of *Both Equally* as the least educated participants. The fact that all three groups report at least a combined 50% of reported Spanish or equal language use suggests language maintenance. Based on these results however, as more G2 (and in the future G3) speakers attend college, we would expect to see an increased shift to English.

[Figure 7](#) shows language use by context/domain, a highly significant predictor in the regression analyses ($p < 0.001$).

The only domain that shows majority Spanish use is Family, which is not unexpected, as younger, English-dominant speakers employ Spanish to communicate with older, Spanish-dominant family members. Some previous

² Age is a factor in this result. A cross-tabulation of Age Group and Generation shows that G2 participants skew young in the present data (79% Younger, vs. 17% Middle and only 4% Older). The age distribution for G1 participants, however, is more balanced (21% Older, 45% Middle, 35% Younger). This difference should be kept in mind when interpreting the results for these two variables.

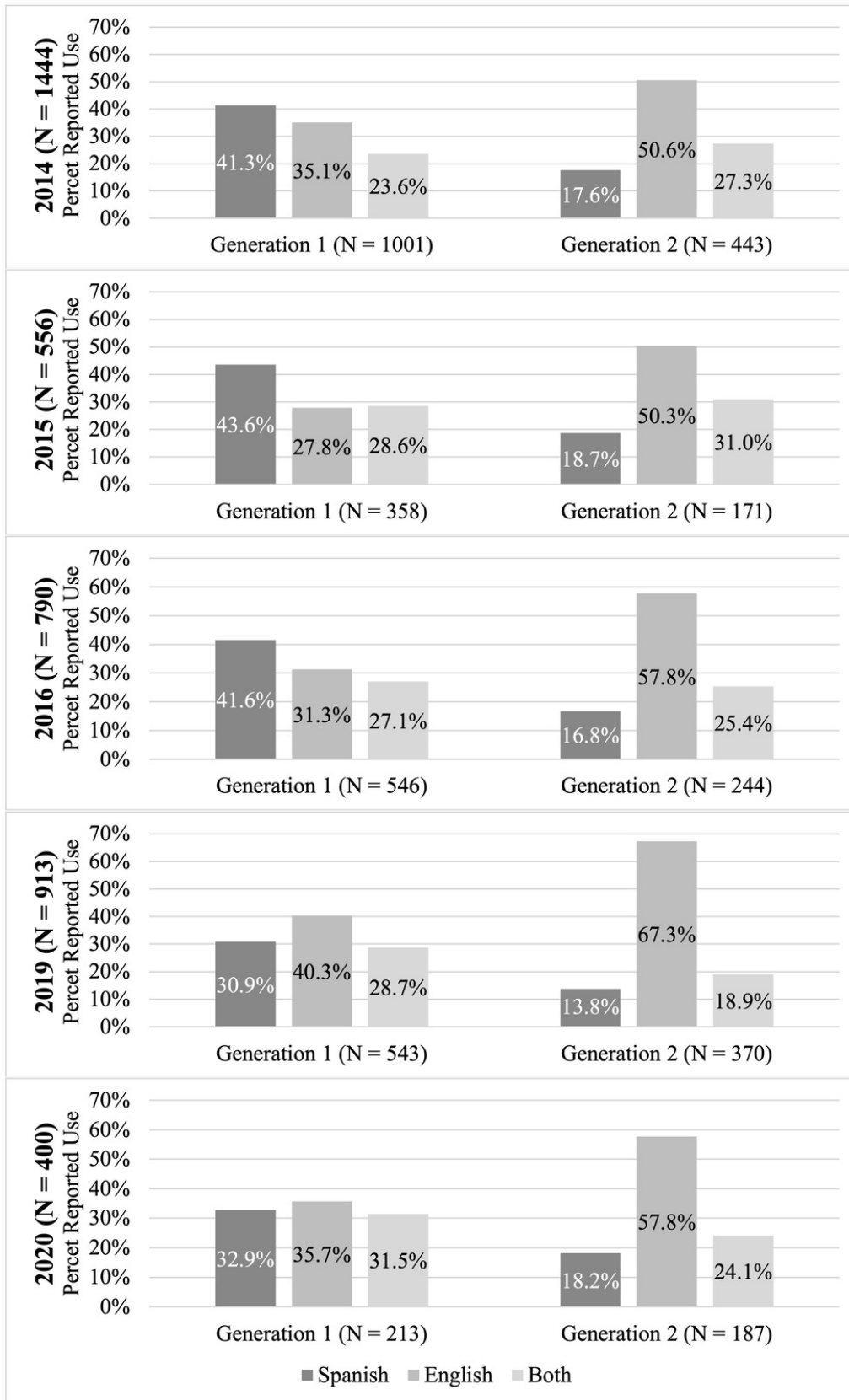


Figure 1. Reported language use by Generation and Year Survey Collected (N = 4103 survey responses).

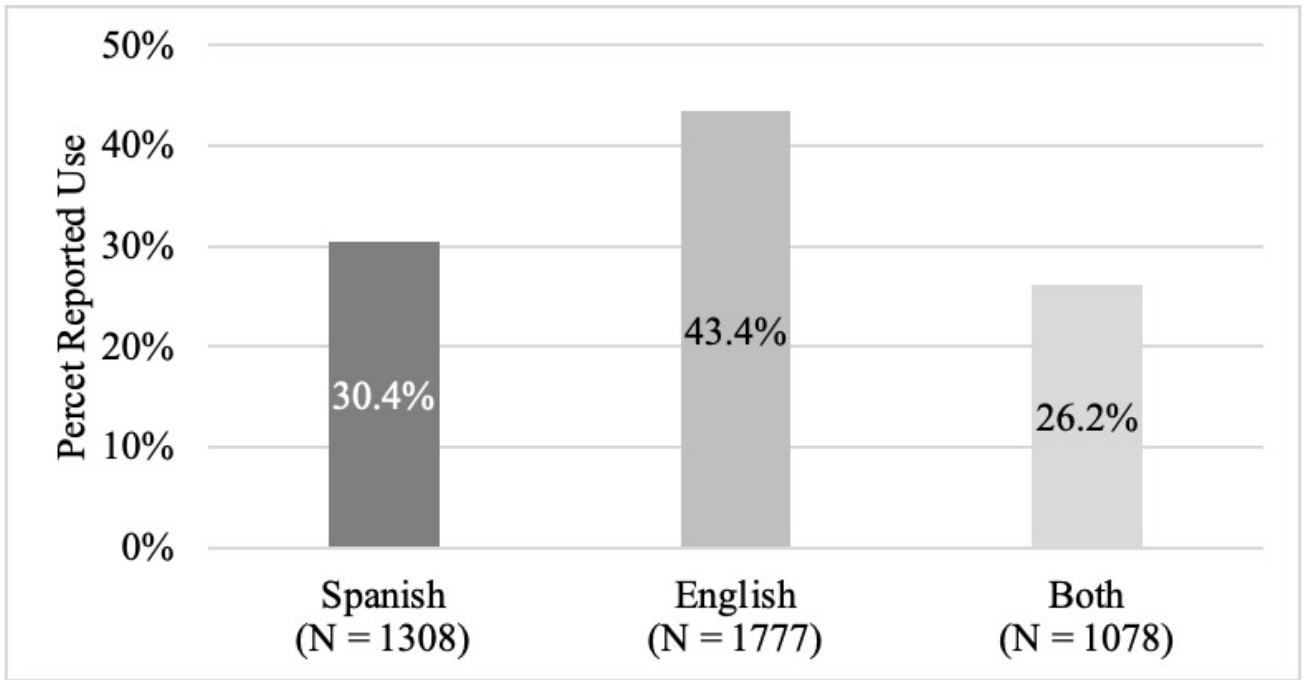


Figure 2. Reported language use by all participants across all contexts of language use (family, friends, work, TV) (N = 4163 survey responses).

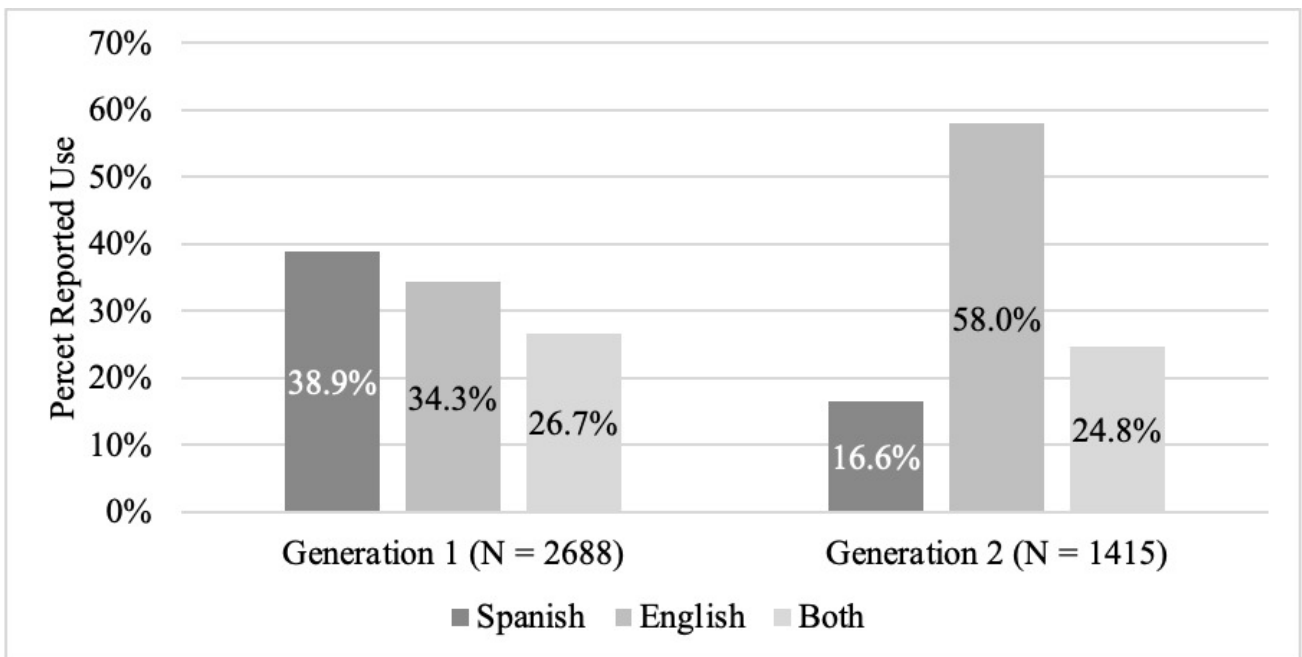


Figure 3. Distribution of reported language use by Generation (N = 4103 Survey Responses).

studies suggest that as long as the heritage language is used in the family, language maintenance is assured (Fishman, 1985, among others). Other studies, however, suggest that diglossia, where Spanish is *only* used with (presumably older, G1) family members, indicates future language shift to English (Eckert, 1980).

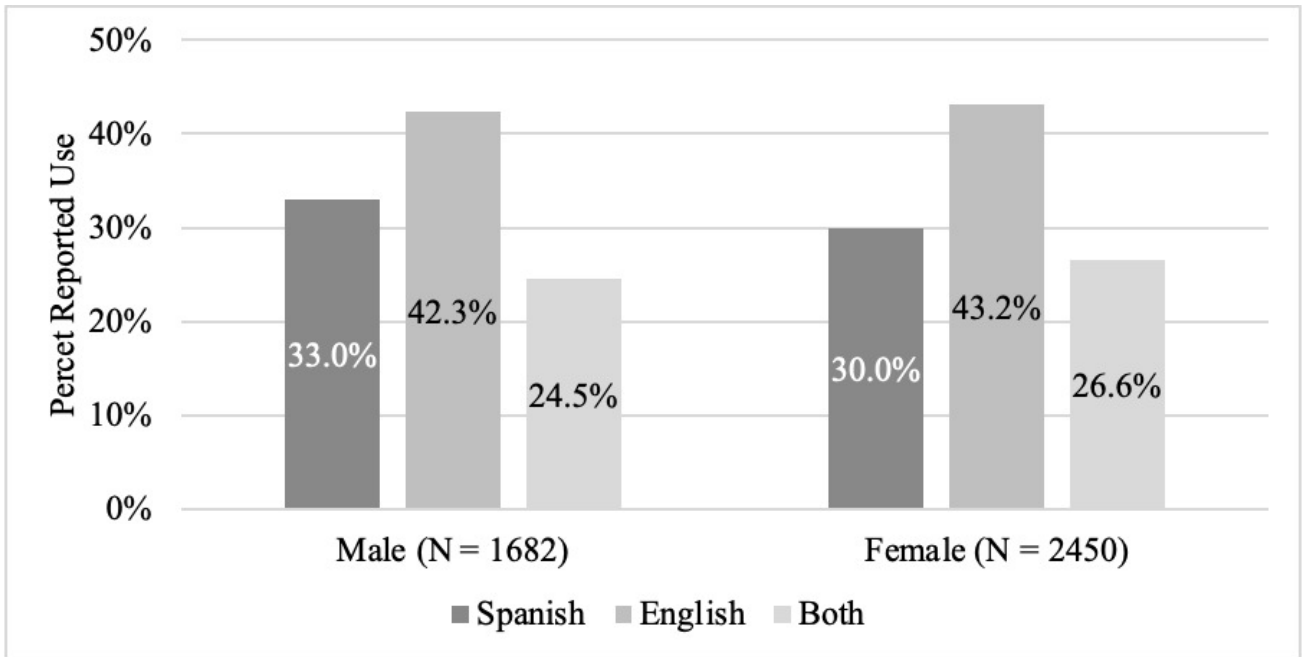


Figure 4. Reported language use by participant sex (N = 4132 survey responses).

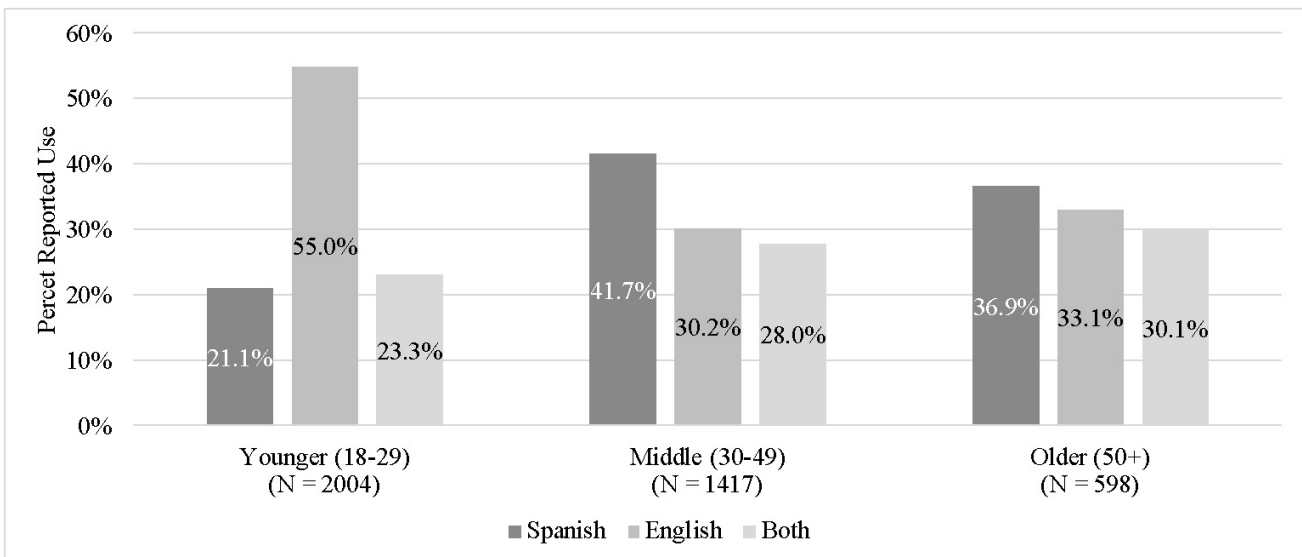


Figure 5. Reported language use by age group (N = 4020 survey responses).

The possibility of this diglossia can be seen most clearly in the Work domain. English accounts for two thirds of the responses, showing that English is required in the workplace in NC and indicating future language shift. As will be further explored in the discussion below, this result differs across several other social factors.

Respondents show similar trends with friends and television - around 50% English, with lesser rates of Spanish or *Both Equally*, although one third of participants report watching TV in both languages, which could be indicative of language maintenance.

The final variable which had a significant main effect in the regression analysis is time in the US for G1 participants, shown in [Figure 8](#) ($p < 0.001$).

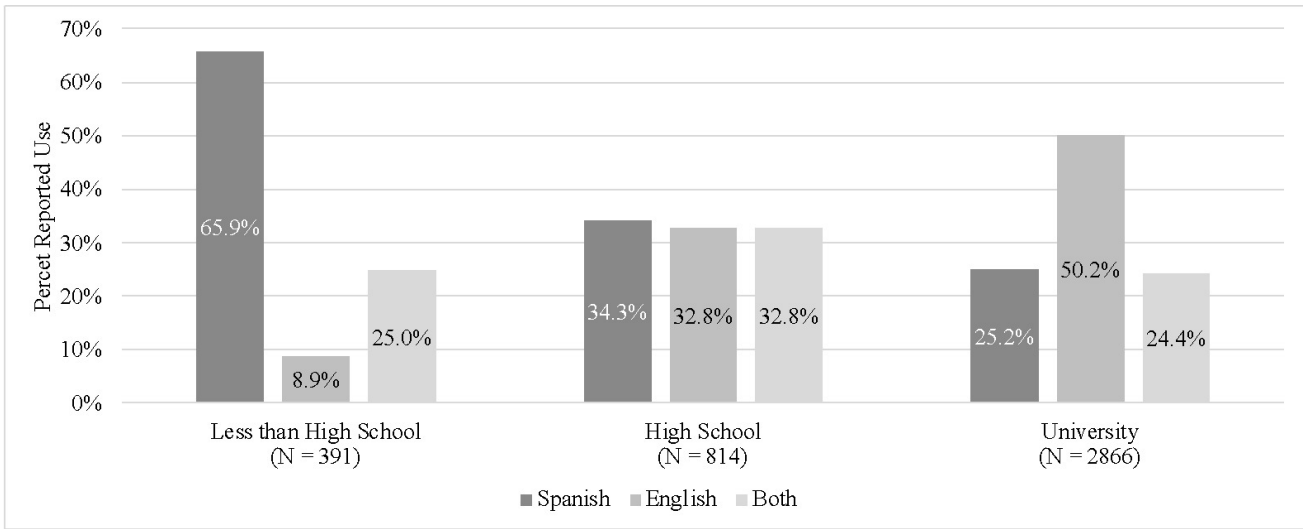


Figure 6. Reported language use by level of education attained (N = 4071 survey responses).

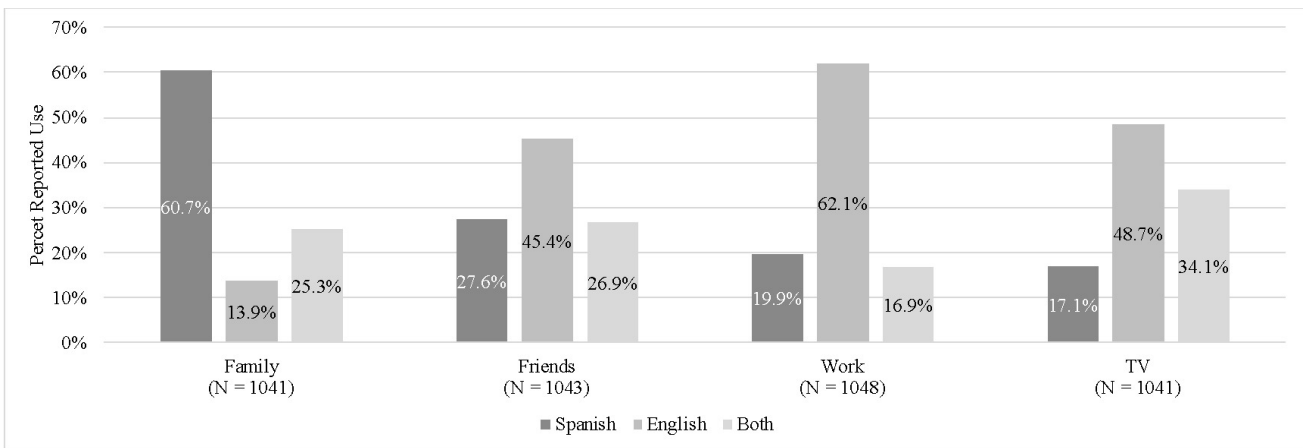


Figure 7. Reported language use by context of language use (N = 4163 survey responses).

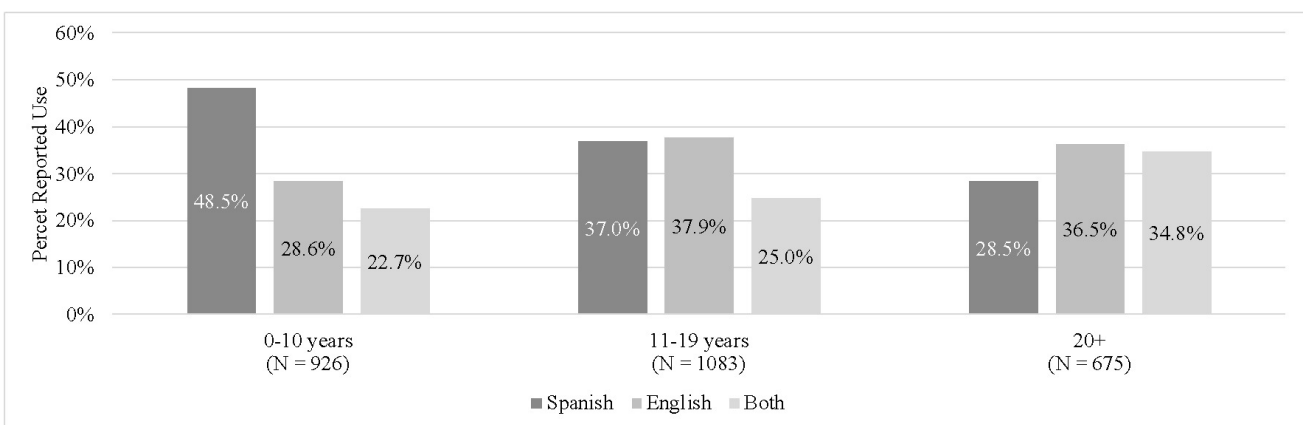


Figure 8. Reported language use by Time in the US for G1 participants (N = 2684 survey responses).

There is a decrease in reported Spanish use with more time in the US (10% per 10-year period), as well as a rise in English and *Both Equally*. Those in the U.S. for 11-19 years (Group B) and those for 20+ years (Group C) are fairly stable, suggesting that the patterns of language use are largely set during an

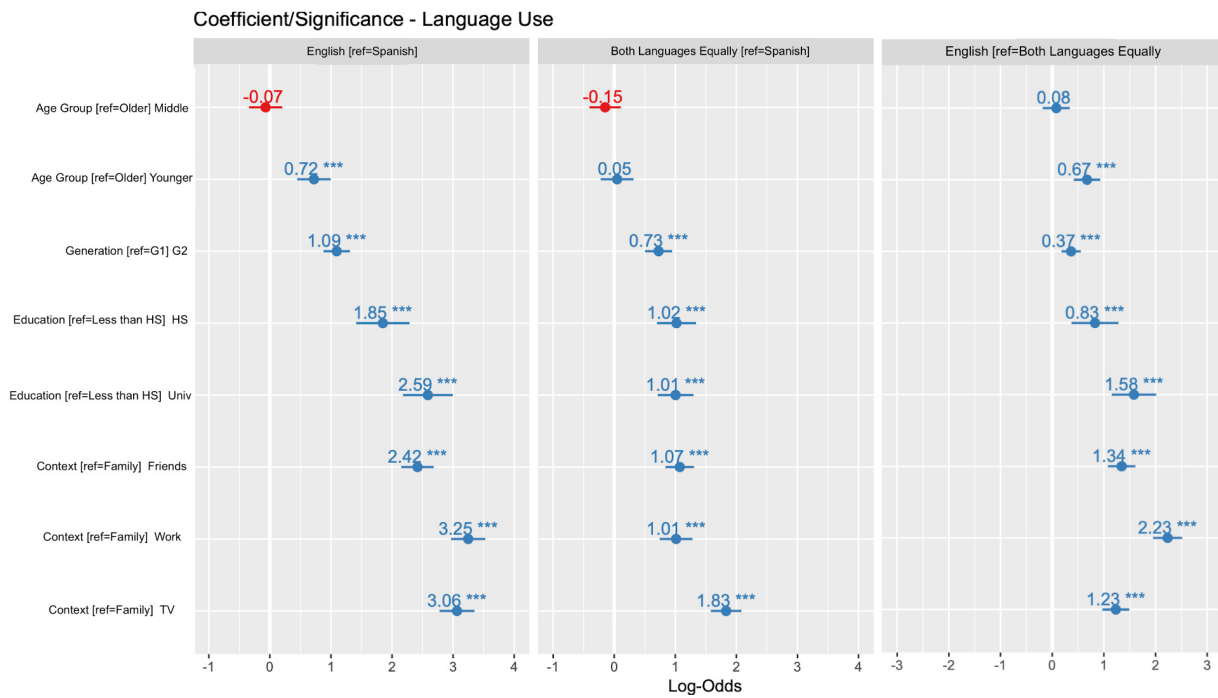


Figure 9. Coefficients of language use. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05.

immigrant's first 10 years in the United States/North Carolina. These results show that Spanish is maintained across time by G1 speakers, although as we will see, it is not consistent across contexts.

4.2 Further Statistical Analyses

Additional insight is gained by plotting the log-odds coefficients from the multinomial logistic regressions. A positive coefficient favors that language (as opposed to the reference level) for that group. In [Figure 9](#), the first panel shows coefficients for English compared to the reference level, Spanish; the second panel compares *Both Equally* to the reference level Spanish; and the third panel compares English to the reference level *Both Equally*.

As shown in the first and second panels of [Figure 9](#), middle aged speakers (30-49) statistically favor Spanish as compared to older speakers (50+), although the differences between these two age groups are not significant. The coefficients in the leftmost panel show that every other group and level significantly favors English more than Spanish. Likewise, the results in the middle panel show a similar pattern, with every variable (again, except age - a nonsignificant result) favoring *Both Equally* over Spanish. Combined, these results suggest LS, as Spanish-only uses are disfavored almost across the board in favor of English or *Both Equally*. Finally, the rightmost panel shows English versus *Both Equally*. All of the groups statistically favor English over *Both Equally*, again a significant result for all but the middle age group. In other words, given the choice between English and Spanish or both languages, English is the language of choice for younger, more highly-educated speakers in non-familial contexts. To summarize, we see a significant preference for English over Spanish and *Both Equally*, and a preference for *Both Equally* over Spanish, a result suggestive of language shift.

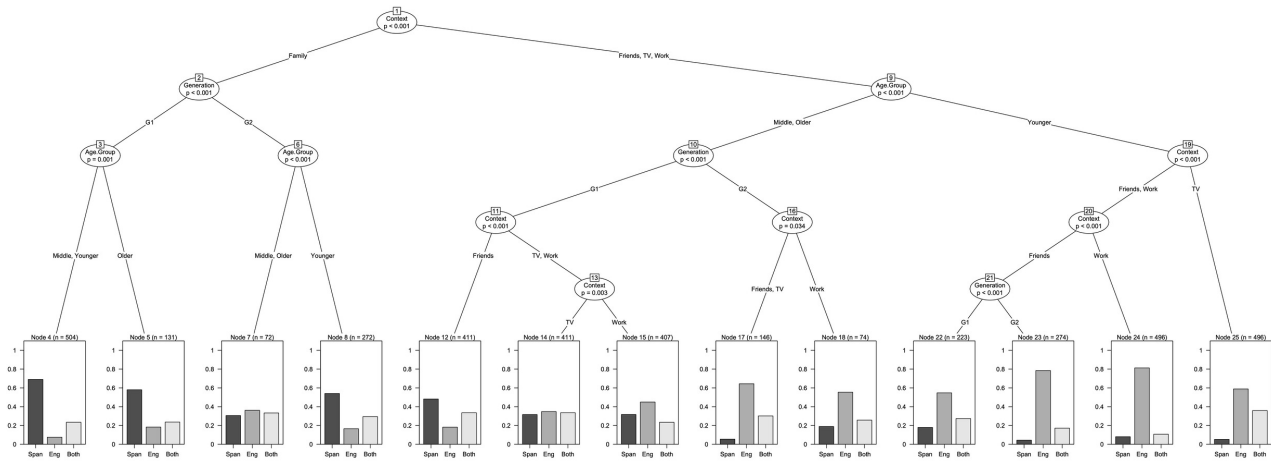


Figure 10. Conditional Inference Tree of Context, Generation and Age Group

Finally, conditional inference trees of the interactions among independent variables were created to shed further light on patterns of language maintenance and language shift in NC. Conditional inference trees provide a visual representation of binary, significant breaks in the data, with individual nodes representing significant differences between levels of a variable. The most important variables are at the top of the tree, with embedded or less impactful variables ranking lower on the tree.³ Figure 10 shows the interplay of context/domain, age group and generation.

As indicated in the tree, the highest rates of Spanish were reported for the family domain - in particular among G1 speakers (Node 3). The highest rates of English were reported for younger speakers at work (both generations - Node 20), as well as younger G2 speakers with friends (Node 21). G2 speakers continue to use Spanish with their families, but have essentially shifted to English with their friends and at work, i.e., the important “community” domain mentioned as so essential to language maintenance by Pease-Alvarez (2002), further suggesting language shift in the future. At no point in Figure 10 does *Both Equally* exceed 50%, but it is at its highest among younger speakers for the TV context (node 25).

One point of interest is that younger G2 speakers report more Spanish with their families than do older or middle-aged speakers (Node 8), likely because the younger speakers are using Spanish with older members of their families, while older speakers are using more English or bilingual speech with their children. This is relevant for studies such as Shin (2013), who found that bilingual children play an important role in introducing English or contact-forms into their families, as caretakers are exposed to and use more English with their bilingual children. This trend may predict increased contact language forms in the future.

3 See Tagliamonte (2012) for more information on the benefits of conditional inference trees and how to interpret them.

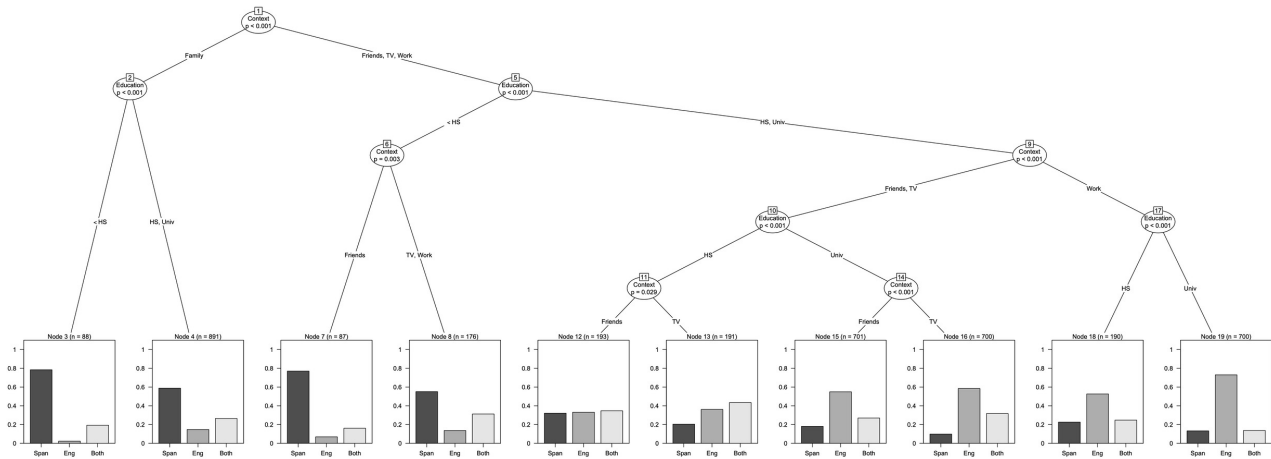


Figure 11. Conditional Inference Tree of Context and Education Level

Likewise, the only respondents that show a majority Spanish use with friends are older and middle-aged G1 speakers (Node 11), whereas older and middle-aged G2 speakers (Node 16), as well as younger speakers of both generations (Node 21), show a predominance of English, suggesting expanding social circles outside of the Latino community that may lead to increased language shift.

Overall, these results show that G2 speakers in NC do not show a balance between Spanish and English, at least for the broad categories shown here. This may suggest a faster shift to English than the 3-generation model would predict, as also indicated by Bills et al. (2000)

Further exploration of the relationship between context and educational level is depicted in the conditional inference tree in Figure 11. Spanish again dominates in the family domain, but less so for more educated participants (Node 2). We can clearly see the role that education plays in more exposure to English - and opportunities to use English - if we compare Nodes 11 and 14 for social interactions with friends and entertainment (TV), and Node 17 for the role of English at work. In both cases, university-educated participants report more English and less Spanish than participants with a high school education. This potentially sets up an interesting dichotomy, whereby higher educated individuals may have the cultural and economic prestige needed to strengthen the use of the home language in the larger community, while at the same time being the least likely to report using Spanish.

Figure 12 shows a conditional inference tree with context, age group and time in the US, with data from G1 speakers only. The amount of reported Spanish in the family domain shows a sharp drop with increased time in the US (Nodes 4, 5 and 6). Community-based domains - friends and work - show similar patterns, with the amount of reported English increasing across Time in the US, indicating language shift even among G1 speakers. At the same time, we also observe an increase in reported balanced use (*Both Equally*) in both

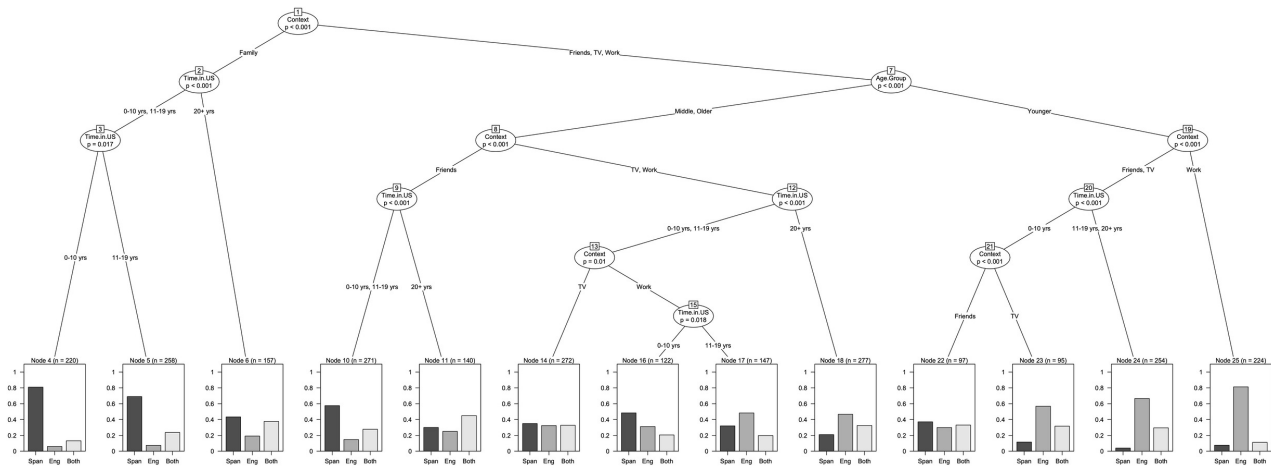


Figure 12. Conditional Inference Tree of Context, Age Group and Time in the US (G1 participants only).

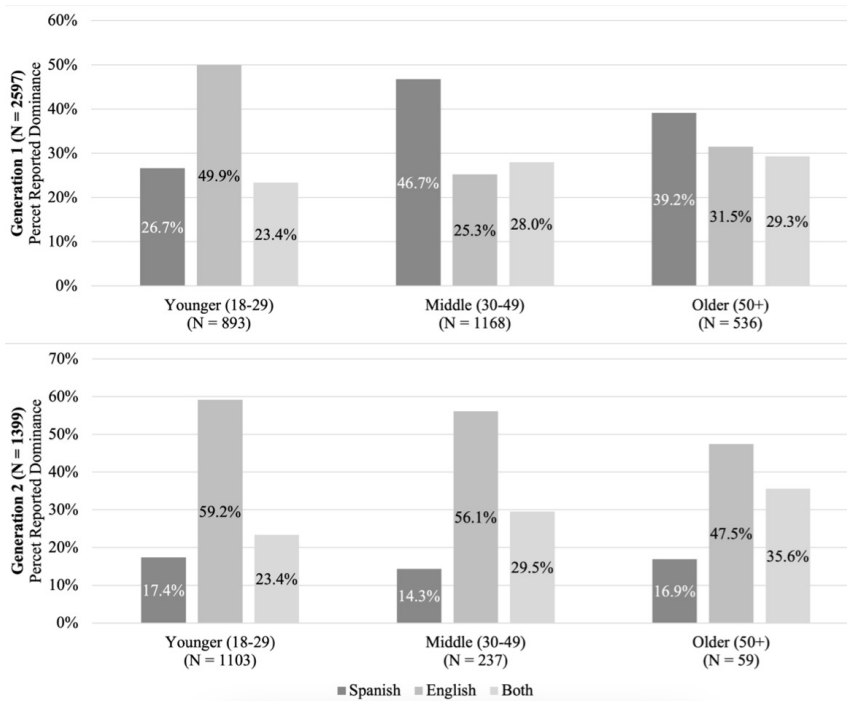


Figure 13. Self-reported language dominance by Generation (G1 vs. G2) and Age Group (Older vs. Middle vs. Younger) (see Table 1) (N = 3996 survey responses).

family and friend contexts, suggesting that if Spanish is preserved, it will be as one option in a bilingual environment, rather than as the only available code among G1 speakers.

4.3 Participants by Self-Reported Language

In addition to the segments described above, the survey included the question of which language(s) participants believed that they speak best. Figure 13 provides the results, separated by Generation (G1 vs. G2) and Age Group (Older, Middle, Younger).

The results in Figure 13 are not surprising, with G1 speakers overall reporting higher rates of dominance in Spanish, but there are two important trends that merit further attention. First, younger G1 speakers report rates of

Spanish in-line with G2 speakers; age of arrival is likely an important factor that should be included in future research. This group also reports the highest rates of *Both Equally*, again indicating that the shift from English to Spanish does not happen immediately, but instead passes through a period of more or less balanced bilingualism (at least as perceived by speakers themselves). Second, while around 40% of middle and younger G2 speakers report balanced bilingualism, no G2 groups report Spanish dominance, meaning that the future of LMLS in NC will likely depend on community factors, like endogamy or exogamy in families, friend groups and the perceived utility of maintaining Spanish, etc. In particular, future research is needed to see how these trends develop among G3, a group which is only beginning to form in NC (Michnowicz et al., 2018). Qualitative self-reports of this sort likely both over- and under-estimate language proficiency, and as with the rest of the survey data presented here, respondent answers should be taken as attitudes toward their own self-perceived language use.

5. Discussion and Conclusions

Results from the present study indicate that the shift to English in NC is largely complete by G2 for contexts studied here, although importantly no group has shifted away from Spanish completely. Still, several factors point toward an ultimate result of language shift in NC: The domains in which Spanish is reported are narrowed for younger, more educated and G2 speakers, and Spanish is largely being relegated to primarily the familial domain, a trend that while stronger among G2 participants, was also found for G1 participants as well. As has been argued by Eckert (1980), the use of a language in the family is necessary but may not be sufficient for LM, as diglossia points strongly toward LS in the future.

Previous research discussed above has indicated that the future of LM/LS in a region can be determined by looking at four key groups/domains: younger speakers, G2 speakers, language use outside of the family, and language use among higher educated speakers (see Porcel, 2011 and sources therein). As indicated by the coefficients for these groups in the multinomial logistic regression (see [Figure 9](#)), a statistically significant hierarchy of *English* > *Both Equally* > *Spanish* emerges in the data. Young G2 speakers with high levels of education prefer English over the other options, and bilingual forms over monolingual Spanish, meaning that all of these groups statistically favor indices of LS by a significant margin.

Additional factors point toward rapid LS to English. NC and the southeastern U.S. lack the conditions favorable to cyclical bilingualism, as the important “Heartland Factor” (Villa & Rivera-Mills, 2009) is not applicable to the sociolinguistic context in the region. The newly developing Latino communities in the Southeast may lack the critical mass of Spanish-speakers necessary to preserve the language long term, and the absence of a well-established, focused bilingual community, as in NYC, Chicago, or the Southwest, points toward fairly rapid language shift in NC, as also indicated by the present data.

Nevertheless, the present data also show that complete shift to English is not inevitable in NC, as all groups report maintaining Spanish in at least some contexts (see also Hurtado & Vega, 2004). Future longitudinal studies are needed to chart the direction of LMLS in these newly forming communities, with the goal of not only documenting how these processes respond to differing pressures across communities, but also to better understand how to target educational and community resources to help preserve Spanish among future generations. Anecdotally, our linguistic outreach efforts with the Spanish-speaking community in NC often find negative attitudes toward bilingual forms, which forces speakers into an ‘all or nothing’ position regarding Spanish language use, and the future development of Spanish in NC will depend on speaker attitudes and the role of Spanish in expressing Latino or Hispanic identity, among other factors (see also Howe & Limerick [2020] on attitudes toward Spanish in Georgia).⁴ While this study provides an initial picture of LMLS in NC, future research should focus on both attitudinal factors, as well as on patterns of actual language use in the community.

Finally, the present study presents a concrete example of the type of substantive, community-based research that is accessible to, and manageable for, undergraduate students. Through the use of surveys, undergraduate students are able to inspect both linguistic variables and the attitudes that surround them. Such methods allow even students inexperienced with the empirical method to engage with the various parts of the research process, including collecting data and understanding relevant variables of interest, on real research projects with meaningful findings. In this way, involving relatively large numbers of undergraduate students (more than 100 over a five-year period) in hands-on research can have innumerable benefits for students, such as increased engagement and retention, as well as personal feelings of satisfaction and confidence in their abilities to undertake research, while at the same time providing valuable data and insights into important research questions in the field (see also Lopatto, 2010; Van Herk, 2008 for more on the benefits of (class-embedded) undergraduate research).⁵

Submitted: May 24, 2022 EDT, Accepted: October 24, 2022 EDT

⁴ For examples of our linguistic outreach efforts, see: <https://linguistica.fl.chass.ncsu.edu/think-and-do/>

⁵ Interested readers should feel free to contact the author(s) for more information on integrating large-scale research projects into the classroom.

References

- Anderson-Mejías, P. L. (2005). Generation and Spanish language use in the Lower Río Grande Valley of Texas". *Southwest Journal of Linguistics*, 24, 1–12.
- Bills, G. D., Hernández-Chávez, E., & Hudson, A. (1995). The geography of language shift: Distance from the Mexican border and Spanish language claiming in the Southwestern U.S. *International Journal of the Sociology of Language*, 114(1), 9–27. <https://doi.org/10.1515/ijsl.1995.114.9>
- Bills, G., Hudson, A., & Hernández-Chávez, E. (2000). Spanish home language use and English proficiency as differential measures of language maintenance and shift". *Southwest Journal of Linguistics*, 19, 11–27.
- Carter, P. M. (2005). Quantifying rhythmic differences between Spanish, English, and Hispanic English. *Theoretical and Experimental Approaches to Romance Linguistics*, 4(272), 63–75. <https://doi.org/10.1075/cilt.272.05car>
- Delgado, P., Guerrero, G., Goggin, J. P., & Ellis, B. B. (1999). Self-assessment of linguistic skills by bilingual Hispanics. *Hispanic Journal of Behavioral Sciences*, 21(1), 31–46. <https://doi.org/10.1177/0739986399211003>
- Eckert, P. (1980). Diglossia: separate and unequal. *Linguistics*, 18(11–12), 1053–1064. <https://doi.org/10.1515/ling.1980.18.11-12.1053>
- Escobar, A. M., & Potowski, K. (2015). *El español de los Estados Unidos*. Cambridge University Press. <https://doi.org/10.1017/cbo9781316091326>
- Fishman, J. A. (1964). Language maintenance and language shift as fields of inquiry: A definition of the field and suggestions for further development. *Linguistics*, 2(9), 32–70. <https://doi.org/10.1515/ling.1964.2.9.32>
- Fishman, J. A. (1965). Who speaks what language to whom and when? *La Linguistique*, 1(2), 67–88.
- Fishman, J. A. (1966). *Language loyalty in the United States*. Mouton.
- Fishman, J. A. (1985). Bilingualism and biculturalism as individual and as societal phenomena". In *The rise and the fall of the ethnic revival* (pp. 39–56). Mouton de Gruyter.
- Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Harvard University Press.
- Haugen, E. I. (1969). *The Norwegian language in America: A study in bilingual behavior*. Indiana University Press.
- Hothorn, T., & Zeileis, A. (2015). partykit: A Modular Toolkit for Recursive in R. *Journal of Machine Learning Research*, 16, 3905–3909.
- Howe, C., & Limerick, P. P. (2020). Understanding Language Attitudes among Members of a New Latino Community in the Southeastern United States: From Speech to Tweets. In *Spanish across Domains in the United States* (pp. 364–387). Brill. https://doi.org/10.1163/9789004433236_017
- Hurtado, A., & Vega, L. A. (2004). Shift happens: Spanish and English transmission between parents and their children. *Journal of Social Issues*, 60(1), 137–155. <https://doi.org/10.1111/j.0022-4537.2004.00103.x>
- Krogstad, J. M., Stepler, R., & Lopez, H. L. (2015, May 12). *English Proficiency on the Rise Among Latinos*. <https://www.pewresearch.org/hispanic/2015/05/12/english-proficiency-on-the-rise-among-latinos/>
- Lopatto, D. (2010). *Undergraduate research as a high-impact student experience*.
- Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.10.
- McCullough, R. E., & Jenkins, D. L. (2005). Out with the old, in with the new?: Recent trends in Spanish language use in Colorado". *Southwest Journal of Linguistics*, 24, 91–110.

- Michnowicz, J., Hyler, A., Shepherd, J., & Trawick, S. (2018). Spanish in North Carolina: English-origin loanwords in a newly forming Hispanic community. In J. Reaser, E. Wilbanks, W. Wolfram, & K. Wojcik (Eds.), *Language Diversity in the New South* (pp. 289–305). University of North Carolina Press.
- Mills, S. V. (2005). Acculturation and communicative need in the process of language shift: The case of an Arizona community”. *Southwest Journal of Linguistics*, 24, 111–125.
- Mora, M. T., Villa, D. J., & Dávila, A. (2005). Language maintenance among the children of immigrants: A comparison of border states with other regions of the U.S. *Southwest Journal of Linguistics*, 24, 127–144.
- Mora, M. T., Villa, D. J., & Dávila, A. (2006). Language shift and maintenance among the children of immigrants in the U.S.: Evidence in the Census for Spanish speakers and other language minorities”. *Spanish in Context*, 3(2), 239–254. <https://doi.org/10.1075/sic.3.2.04mor>
- North Carolina’s Hispanic Community: 2021 Snapshot. (2021). Carolina Demography. <https://www.ncdemography.org>
- Pease-Alvarez, L. (2002). Moving beyond linear trajectories of language shift and bilingual language socialization. *Hispanic Journal of Behavioral Sciences*, 24(2), 114–137. <https://doi.org/10.1177/0739986302024002002>
- Porcel, J. (2011). Language maintenance and language shift among US Latinos. In *The Handbook of Hispanic Sociolinguistics* (pp. 623–645). John Wiley & Sons. <https://doi.org/10.1002/9781444393446.ch29>
- R Core Team . (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rivera-Mills, S. V. (2000a). Intraethnic attitudes among Hispanics in a Northern California community”. In A. Roca (Ed.), *Research on Spanish in the United States: Linguistic Issues and Challenges* (pp. 377–389). Cascadilla Press.
- Rivera-Mills, S. V. (2000b). *New perspectives on current sociolinguistic knowledge with regard to language use, proficiency, and attitudes among Hispanics in the U.S.: The case of a rural Northern California community*. E. Mellen Press.
- Rivera-Mills, S. V. (2007). The fourth generation: Turning point for language shift or mere coincidence? *XXI Conference on Spanish in the US and The VI Spanish in Contact with Other Languages*.
- Romaine, S. (1995). *Bilingualism* (2nd ed.). Blackwell.
- Ronquest, R. E., Michnowicz, J., Wilbanks, E., & Cortes, C. (2020). Examining the (mini-) variable swarm in the Spanish of the Southeast. In A. Morales-Front, M. Ferreira, R. Leow, & C. Sanz (Eds.), *Hispanic linguistics: Current issues and new directions* (pp. 303–325). John Benjamins Publishing Company. <https://doi.org/10.1075/ihll.26.15ron>
- Shin, N. L. (2013). Women as leaders of language change: A qualification from the bilingual perspective. In A. M. Carvalho & S. Beaudrie (Eds.), *Selected proceedings of the 6th Workshop on Spanish Sociolinguistics* (pp. 135–147). Cascadilla Proceedings Project.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford University Press.
- Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Georgetown University Press.
- Tagliamonte, S. (2012). *Variationist sociolinguistics: Change, observation, and interpretation*. Wiley-Blackwell.
- Thomason, S. G. (2001). *Language contact*. Edinburgh University Press.

- U.S. Census Bureau. (2020). *2016-2020 American Community Survey 5-Year Estimates*.
<https://data.census.gov/cedsci/table?q=Latino%20Origin%20NC&tid=ACSDT5Y2020.B03001>
- Van Herk, G. (2008). The very big class project: Collaborative language research in large undergraduate classes. *American Speech*, 83, 222–230.
- Veltman, C. (1983). Anglicization in the United States: Language environment and language practice of American adolescents. *International Journal of the Sociology of Language*, 1983(44), 99–114.
<https://doi.org/10.1515/ijsl.1983.44.99>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer.
- Villa, D. J., & Rivera-Mills, S. V. (2009). An integrated multi-generational model for language maintenance and shift: The case of Spanish in the Southwest. *Spanish in Context*, 6(1), 26–42.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Zúñiga, V., & Hernández-León, R. (Eds.). (2005). *New destinations: Mexican immigration in the United States: Community Formation, Local Responses and Inter-Group Relations*. Russell Sage Foundation.

Appendix

Table A1. Minimal multinomial logistic regression (main effects only); N = 3917 survey responses

Coefficients		Ref = Older		Ref = G1	Ref = <HS		Ref = Familia		
Language (ref = Spanish)	Int	Age.Group = Middle	Age.Group = Younger	Generation = G2	Edu = HS	Edu = Univ	Context = Friends	Context = Work	Context = TV
Inglés	-4.739262	-0.06955902	0.72224300***	1.0941303***	1.848956***	2.587560***	2.417322***	3.247285***	3.064135***
Las dos	-2.006811	-0.14858643	0.04804197***	0.7272255***	1.020246***	1.006007***	1.073437***	1.013745***	1.832848***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

Table A2. Minimal multinomial logistic regression for Time in the US with G1 respondents (main effects only); N = 2534 survey responses

Coefficients		Ref = Older		Ref = <HS		Ref = Family			Ref = Time in US 0-10 yrs	
Language (ref=Spanish)	Int	Age.Group = Middle	Age.Group = Younger	Edu = HS	Edu=Univ	Context = Friends	Context = Work	Context = TV	Time in US = 11-19 yrs	Time in US = 20+ yrs
Inglés	-5.990811	0.4565610 **	1.7683583***	1.734704***	2.627624***	2.028554***	3.413100***	3.027744***	0.7665127***	1.78906***
Las dos	-3.184636	0.4617861**	0.9365641***	1.092513***	1.160366***	1.028901***	1.150188***	1.694154***	0.4229308***	1.523154***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05