

MORPHOSYNTACTIC VARIATION IN AN EMERGING DIALECT OF EASTERN NORTH
CAROLINA SPANISH

by

Anna A. Lawrence

A Senior Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

by

Anna A. Lawrence

Greenville, NC

May 2016

Approved by:

Dr. Stephen A. Fafulas

Department of Foreign Languages and Literatures, Thomas Harriot College of Arts and Sciences

Abstract

For the past several decades, the Spanish-speaking population in eastern North Carolina (eNC) has been increasing rapidly. While much is known of other Spanish varieties spoken in major cities in the US, little is known about the formation of Spanish communities in the rural South. This project aims to fill that void by adding to our knowledge of Spanish morphosyntax by noting specific properties that make this emerging Spanish community's expression of explicit subject in Spanish different than other bilingual Spanish varieties around the U.S. Methodology for the study included: (a) collection of speech samples through sociolinguistic interviews, and (b) participant profiles through a Background Language Proficiency (BLP) Questionnaire. Results from our study conducted on 1,309 tokens from 18 first- and second-generation speakers of Mexican Spanish indicated a statistically significant difference in subject expression based on time spent in the region of eNC, bilingual dominance in English, country of birth, and individual rates of expression. Additionally, preliminary results indicate an influence of verbal person as a predictor of subject expression. Overall, the findings from this study indicate an emerging effect of bilingual contact between English and Spanish in the region of eNC. This justifies the need to investigate whether the region plays a different role in constructing the linguistic identities of its Hispanic inhabitants separately from other historically bilingual U.S. cities.

Acknowledgements

The findings and insights I was granted from this thesis would have been impossible without the help, support, and collaboration of several individuals. First, it is necessary to thank Dr. Stephen Fafulas, my professor, mentor, and supervisor for this Senior Honors Project, who inspired me to pursue the study of Hispanic linguistics both during my time at East Carolina University and for my future graduate studies. Additionally, much of the data collection process for this investigation would have been impossible without the funding I was given as a recipient of an East Carolina University Undergraduate Research and Creative Activity (URCA) award. The entire process of data analysis would not have been possible without the East Carolina University Study of Communities, Involvement & Outreach and Linguistics (SoCIOLing) Lab. I would also like to thank Dr. Amanda Boomershine at the University of North Carolina at Wilmington, along with the UNCW Linguistics Lab, for their additional assistance in coordinating data collection. Finally, I would like to thank the East Carolina University Honors College for the overwhelming scholastic support I have received throughout my undergraduate career.

Table of Contents

Introduction	5
Purpose of Study	5
Background.....	6
Summary of Previous Literature.....	8
Methodology	12
Research Questions and Variables.....	12
Participants and Corpus	14
Instruments of Investigation.....	16
Data Coding and Analysis.....	17
Results and Discussion	18
Total Distribution of Subject Expression	18
Distribution of Social Variables	19
Distribution by BLP score.....	20
Expression by country of origin.....	21
Time in eNC and subject expression.....	22
Individual variation of subject expression.....	24
Distribution of Linguistic Variables	25
Subject expression and plurality of subject.....	26
Subject expression and verbal tense	27
Conclusions.....	29
Limitations of the Study	29
Future Directions	30
References	33

Morphosyntactic variation in an emerging dialect of eastern North Carolina Spanish

Introduction

Purpose of Study

It is an indisputable fact that throughout the past several centuries, the United States has transformed into a bilingual epicenter in which the use of English and Spanish has been progressively shared and has come into contact across multiple generations of Spanish-speaking immigrants and their children. Although currently there exists an extensive collection of existing literature exploring the varieties of Spanish spoken in the United States, especially the Spanish of those who acquire it as a heritage language, most studies focus have historically focused on urban regions with historically established Hispanic populations such as New York City or Los Angeles (Escobar & Potowski, 2015). However, though these studies have thoroughly examined the effects of bilingual contact between English and Spanish in communities with long-residing Spanish speakers, they are unable to examine how language changes in regions that have historically failed to contain a visible Hispanic presence. In order to fill this void in sociolinguistic literature, this study investigates the ways in which the speech of both native and heritage speakers of Spanish changes in a region outside of those that have been previously studied.

The area in which this study was completed is eastern North Carolina, which henceforth will be abbreviated as eNC. This zone stands out within the United States in the fact that its Spanish-speaking population has grown considerably in the last few decades, both in urban regions as well as in more rural areas due to the agricultural industries that attract migrant workers from other countries (Escobar & Potowski, 2015). Given that the region of eNC has only recently become prevalently bilingual, it can be concluded that there is a necessity to

investigate its speech community due to the recent influx of Hispanic immigrants that have newly located themselves in rural areas of the state. Additionally, the findings from this study and others conducted in eNC have the potential to determine whether the variety of U.S. Spanish found in eNC and other rural communities in the country differs from the Spanish of more heavily populated urban centers, forming a new dialect as a result of Spanish-English contact in an area of less dense population.

Background

This study focuses specifically on the linguistic variable of subject expression in eastern North Carolina Spanish as a dependent variable. Subject expression is a topic that has received considerable attention in the literature on U.S. Spanish (see Carvalho, Orozco & Shin, 2015 and Escobar & Potowski, 2015). However, while the primary linguistic factors that condition subject expression in Spanish have been mostly identified (e.g., person and number of the referent; tense, mood and aspect (TAM) of the verb; and previous mention of the referent or switch reference; see Otheguy, Zentella & Livert, 2007) the extent to which extra-linguistic factors such as levels of bilingualism impact subject expression are less well known. Furthermore, both the social and linguistic factors that affect subject expression have yet to be studied in rural regions such as eastern North Carolina to determine whether expression in these regions is consistent with the patterns found in previous literature on urban areas with a greater Hispanic presence.

It is not by coincidence that such a wide array of previous sociolinguistic studies conducted in the U.S. have focused on subject expression in Spanish; rather, this variable is in fact ideal for observing the effects of bilingual contact with English over time because of the differences in its manifestation between languages. Generally, native speakers of English, which in this case is the language in contact with Spanish, almost exclusively must express the subject

of a phrase explicitly, while the rules of Spanish permit that the subject be omitted, and that it be communicated implicitly through the morphological flexion of the verb in conjugation. For example, in the English sentence *They buy flowers*, the subject *they* is obligatory in order to produce a statement that communicates all of the intended information. In contrast, Spanish allows two ways of expressing the same statement: *Ellos compran flores* (“They buy flowers”) or *Compran flores* (“Ø Buy flowers”**). Although the second option does not explicitly state the subject *ellos*, it is communicated implicitly by the verb’s conjugation that reflects the plural third person of the subject. This study examines and compares the dependent variable of subject expression for two participant groups, first-generation (G1) native speakers and second-generation (G2) heritage speakers of Spanish. As outlined in Escobar & Potowski, 2015, G1 speakers include those who were born and educated in a Spanish-speaking country and moved to the U.S. after adulthood, while G2 speakers were born in the U.S. to at least one parent from the G1 category, and grew up speaking Spanish in the household. Subject expression in these groups is analyzed in as affected by various independent variables. Within the category of linguistic variables, there is an interest in examining the effects of verbal tense and plurality of the subject person on subject expression. Additionally, the study measures the dependent variable as it is affected by social factors such as country of birth or generation in the U.S., level of bilingualism, and time spent in eastern North Carolina. The primary research question, as a result, is whether aforementioned factors impact subject expression significantly in order to demonstrate new patterns of Spanish language use in the region of eNC due to increased bilingual contact. Given that English typically requires explicit subject expression, this study hypothesizes that in general, participants who have spent more time in eNC; who were born in the United States; and who are more dominant in English will have higher rates of explicit subject expression. Additionally, it is

hypothesized that verbal tense and subject person will have visible effects on subject expression among all speakers, given that certain linguistic cases require that one clarify the subject more specifically.

In sum, this study includes data from 28 participants that are, to some extent, considered native speakers of Spanish, whether by birth and education in a Spanish-speaking country or by heritage. In all cases, the principal dialect to which the participants have been exposed is Mexican Spanish; however, 13 of the participants were born in the U.S. with Mexican parents, while five were born in Mexico. The data from this study was obtained through sociolinguistic interviews in order to elicit individual speech samples for each participant, in conjunction with written data from a questionnaire designed to measure the background of bilingual proficiency of the speakers (the *Bilingual Language Profile*). The corpus as a whole was analyzed through a system of individual codes that recognize the cases in which the opportunity arises to choose to express the subject in order to determine if there is a difference in the expression of this phenomenon according to the parameters discussed in our hypothesis.

Summary of Previous Literature

As stated above, it should be noted that there already exists a body of previous empirical studies that have been completed to describe the phenomenon of subject expression in U.S. Spanish. Although these investigations provide data that can aid in determining the possible results of this specific study, they only give information appropriate for making inferences on our hypothesis based on the patterns that tend to be seen in other regions in the country. However, the majority of this literature is concentrated in the field of major U.S. cities such as New York City, Chicago, and Los Angeles; they do not demonstrate the effect of bilingual contact between English and Spanish in more rural regions. For this reason, although existing

studies may help investigators to create informed questions for research, they also justify the necessity for new research in the targeted region of this study, of which currently there are few antecedents.

Bayley, Cardenas, Schouten, and Velez completed an empirical study in San Antonio, Texas with native speakers of Spanish. In this case, the investigators interviewed speakers who had been born in the Caribbean, Puerto Rico, New York, and Mexico in order to analyze the frequency and prevalence of their use of explicit and implicit subject pronouns in their speech. It was proposed in the study that since the population of Puerto Rican background in San Antonio is relatively much smaller than the Mexican-American population, there would be a convergence of dialects due to the contact between them and the predominance of the group with Mexican and U.S. influence. Methodology included the use of sociolinguistic interviews and Goldvarb analysis. Data from the study revealed that there were various linguistic factors influencing the speech of those interviewed, including person and tense in a phrase. In general, Caribbean Spanish speakers and Puerto Ricans use explicit subject pronouns with more frequency than Mexicans. In the case of social factors, the investigators determined that in order for there to be convergence between dialects, it was necessary that there be frequent and consistent contact between groups of speakers (Bayley, Cardenas, Schouten, and Vélez, 2012). Therefore, this study illustrates the potential effect that the contact between groups of speakers can have over the expression of subjects. Additionally, it demonstrates subject person and verbal tense as linguistic factors with the capacity to influence the dependent variable of subject expression.

Another study completed by Flores-Ferrán examines the use of explicit and implicit subject pronouns used by Spanish speakers in bilingual communities of Spanish and English in a specific case of speech – conflict. The investigators proposed that since discourse based in

conflict demands that speakers use verbs in order to defend themselves, to reject certain ideas, or to threaten, this case would affect the expression of subject for the narrator. The study was completed in New York City with 41 native Spanish speakers of Puerto Rican origin. In their sociolinguistic interviews, the participants had to narrate a story containing a primary conflict. The data were analyzed in order to determine the prevalence of explicit subject depending on the case of conflict and on the subject person. The results demonstrated that the narratives contained explicit subject in 55% of the possible cases, and 57% of the cases with conflict, in all of the personal cases. The investigators determined that this specific factor in discourse had a significant effect on the linguistic factor of subject expression in speech; thus, they suggest that conflict is an additional social extralinguistic factor that can affect subject expression in U.S. Spanish (Flores-Ferrán, 2010). The results of this study are especially relevant in order to demonstrate that our dependent variable can be influenced by social factors. Although this particular factor relates to extralinguistic factors associated with discourse and differs from the themes investigated in the current study, it still draws awareness toward the possibility of its effects.

Otheguy, Zentella, and Livert also described the effects of contact between Spanish and English in the case of the declaration of explicit subject with 142 native speakers of Spanish of various Spanish-speaking origins in six communities in New York City. Through the completion of sociolinguistic interviews, data on pronouns were collected for more than 63,500 verbs to reveal the influence of English in their use. These data were examined with respect to geographic origin of the speaker and the number of years that each speaker had spent in the U.S. The geographic groups were divided by Caribbean Spanish speakers and Continental Central and South American Spanish speakers. Additionally, the data were analyzed through various

linguistic factors, which were organized in a hierarchy determined by the effect that they had toward the use of explicit subject pronouns. It was found in the study that in general, the linguistic factor with the greatest effect was the subject person; that those who had spent more time in the U.S. used explicit subject pronouns with more frequency than those recently immigrated to the U.S.; and that in all cases, Caribbean Spanish speakers explicitly expressed the subject with the greatest frequency. Therefore, the investigators concluded that external factors such as dialectal origin of Spanish and the time spent in the country, and consequently English contact, had significant effects on subject expression (Otheguy, Zentella, and Livert, 2007). Also, linguistic factors such as person may affect the use of subject in any case. Additionally, Livert and Otheguy also investigated in 2010 whether time spent in the U.S. affected subject expression for native first generation Spanish speakers from a group of 113 sociolinguistic interviews taken from the Otheguy-Zentella corpus of New York Spanish. A multivariable analysis of 51,478 tokens determined the three greatest social predictors of increased subject expression to be the age of arrival in the U.S., the time the participant had spent in New York, and the region of origin (Livert & Otheguy, 2010). The results of these studies performed in New York City can serve as a tool for this current investigation to demonstrate how contact between languages can produce a universal effect toward a speech community, regardless of whether speakers are first or second generation.

Finally, it should be noted that the current investigation is motivated most strongly by the existing results of an ongoing project on eastern North Carolina, to which this study also contributes. Lawrence, Meehan, and Fafulas (2016) demonstrated a significant difference on subject expression between first- and second-generation speakers of Mexican Spanish from a corpus of 10 participants. Additionally, an observable difference was observed for the variables

of tense and person as predicted. However, the results of this study demanded the need for a more robust corpus representing a greater geographic expanse of eNC, as well as further analysis of social variables including time spent in the region, and the level of bilingualism on speech. As a whole, the body of existing literature cited above to support this study contributes to a collection of antecedents that collectively justify the variables that have been chosen for the current investigation. As it can be seen that the linguistic factors of subject person and verbal tense can affect subject expression in a phrase and that there is validity in predicting that social variables such as bilingualism and residency in a specific region such as eNC can have an effect on morphosyntax, we can predict the outcomes of our research with confidence. Additionally, though previously completed research in eNC demonstrates significant results that related to our current research questions, there is currently a lack of investigation principally on the variable of bilingualism and the effect of time in eNC specifically on our dependent variable, further justifying the need for expansion of research in the region.

Methodology

Research Questions and Variables

The principal purpose of this study is to examine the linguistic and social variables that have the possibility to affect whether a person will explicitly express the subject in a phrase. The speech community of eNC is a center of bilingual contact between Spanish and English. Therefore, it is proposed with the support of evidence from previous studies surrounding the same linguistic phenomenon of subject expression the following research questions:

- 1) How can a participant's country of birth influence subject expression for native speakers of Spanish, and what relationship does this the time that a person spends in the region of

eNC and the resulting bilingualism between English and Spanish that ensues have toward this expression?

2) What is the effect of verbal tense and the subject person toward subject expression?

For this study, it is necessary to define and operationalize the variables that are investigated. What is searched for in the investigative process is the manifestation of the dependent variable, which is identified as the explicit (overt) or implicit (null) expression of the subject in a declarative phrase. To review, this concept refers to cases in which the speaker has the option to either overtly or implicitly state the subject based on free variation. For example, in the statements *Fui al supermercado* (“ \emptyset [I - null] went to the supermarket”**) and *Yo fui al supermercado* (“I went to the supermarket”), the speaker has the option to explicitly say who the subject *yo* is because the morphology of the verb communicates the person. This investigation attempts to identify the prevalence of the cases that do contain the explicit expression of the *yo* subject amongst others. It should also be noted that the current study does not only examine the expression of subject pronouns, and also includes noun phrases that act as explicit subjects without the use of the pronoun, e.g. *El doctor entró la sala* (“The doctor entered the room”) vs. *Entró la sala* (“ \emptyset [he – null] entered the room”).

On the other hand, the results of this manifestation of the dependent variable are examined through two categories of independent variables: linguistic and social. The linguistic variables attempt to document the inherent and internal traits of the participants’ language that impact subject expression. In this study, the linguistic variables that are examined are the subject person and verbal tense. It is proposed that there will be a greater tendency amongst the data for the explicit subject to be used more often with the subject is either the first or third person; in the

same vein, it is also proposed that there will be a difference in the explicit expression of subject depending on whether the verbal tense is in the past, present, or future.

The social variables of this study examine cultural or external factors affecting the language that can influence the manifestation of subject expression regardless. In this study, it was elected to examine participant country of birth, level of bilingualism, and the time each individual had spent in the region of eNC (regardless of country of birth). It was hypothesized that participants that had been born in the U.S. and were more dominant in English over Spanish would display higher rates of explicit subject expression than those who were born in Mexico or who were Spanish-dominant. Additionally, it was suggested that there could be a positive correlation between the number of years an individual had spent in eNC and the individual's rate of explicit subject expression; better stated, the more time a participant had spent in the region, the higher explicit subject declaration for that individual would be.

Participants and Corpus

The sociolinguistic corpus that was analyzed for this study was composed of oral and written data from 18 participants. All of these participants were native speakers of Spanish in that they had acquired the language from the onset of infant language development; at the very minimum, each had at least two parents who spoke Spanish as their first language. It was elected to examine only the data from individuals who had Mexican parents in order to control for dialect and to avoid skewing the data with results from regional dialects whose speakers typically display higher rates of subject expression. The participant group was divided in two based on social factors: the first-generation (G1) speakers who were born in Mexico, and the second-generation (G2) heritage speakers, who were born in the U.S. to Mexican parents. The G1 group includes 5 speakers, whereas the G2 group includes 13. Of the G2 group, 8 participants

completed the data collection process in Greenville, North Carolina, while 5 completed the process in Wilmington, North Carolina. Table 1 illustrates the social background of each participant. It should be noted that each participant had spent at least 10 years in the U.S., but the range of time spent in eNC is ampler – in this category, the collection of participants had lived in the region between 2 and 24 years.

Table 1. Sociolinguistic background of participants.

Participant	Age	Country of Birth	Years in ENC	Highest Level of Education	Total Language Dominance
1	23	USA	3	University	-101.708
5	20	USA	2	Community College	-73.94
10	24	USA	24	Masters	-64.35
15	20	USA	20	Some University	-64.35
18	29	Mexico	2	Some University	5.076
14	21	USA	3	High School	6.54
4	48	Mexico	9	University	67.468
3	35	Mexico	16	University	101.25
6	34	Mexico	7	University	123.666
16	37	Mexico	8	Some University	183.432
33	22	USA	4	Some University	-71.564
36	20	USA	20	Some University	-44.136
37	19	USA	2	Some University	-74.378
38	19	USA	2	Some University	-70.112
40	22	USA	12	Some University	-29.972

Instruments of Investigation

The data from this study were obtained primarily through the collection of oral samples that were collected through the completion of sociolinguistic interviews. These interviews consisted of a semi-structured format based off a question list designed to collect social information about each participant (e.g. “What is the highest level of education that you have completed?”) as well as stimulating the participating to recall narrative and cultural information (e.g. “What did you like to do as a child?”; “[Why] do you think it is hard to be latino/a in eNC?”). These interviews were recorded with digital audio recorders for the reference of the investigators. Later, the data were collected in written form through the process of transcription that documented all speech interactions between the interviewers and participants including features such as discourse markers.

Apart from the interviews, the variable of bilingualism was measured with the aid of a written questionnaire called the Bilingual Language Profile (BLP). Participants completed sections of the survey assessing the language history, use, attitudes, and proficiency in English and Spanish to determine a score of Total Language Dominance that indicated whether the individual was more dominant in one language over another (Birdsong, Gertken, & Amengual, 2012). Results from the BLP for each participant are indicated in the “Total Language Dominance” column of Table 1. In this study, negative values indicate an English-dominant score, while positive values are Spanish-dominant. The closer a participant’s score is to zero, the more truly bilingual he or she is based on the profile score. Although social variables such as gender and age are also recorded in the BLP, it should be noted that these factors were not used in the analysis of this specific study.

Data Coding and Analysis

To measure the prevalence of the dependent variable quantitatively in the interview transcriptions, there was the necessity to develop a process of coding to organize the data for statistical analysis. The investigators filtered the entire corpus of transcriptions in order to identify all of the cases in which there was the option of free variation between the use of explicit subject and implicit/null subject expression. These tokens were recorded and categorized electronically using Microsoft Excel. Figure 1 illustrates the method by which verbs were selected from the corpus of transcriptions.

Figure 1. Example of variable subject expression in transcription.

P37 – 20 F USA:

“Cuando es algo relacionado a la cultura o comida, lo que sea... es cuando más uso el español pero, I mean, por lo tanto, yo uso inglés.”

“When it’s something related to culture, or food, whatever, is when [\emptyset - I - null] use Spanish most, but, I mean, for the most part, I use English.”

In order to eliminate sources of linguistic cases that could affect the preference for a form of expressing the subject, several verbs were not coded. In sum, verbs were only coded that occurred in main clauses and that did not fall within the following exceptions: 1) subordinate clauses with the subjunctive mood; 2) verbs like *gustar* (“to please”), e.g. *interesar* (“to interest”), *aburrir* (“to bore”); 3) passive and impersonal expressions, which require use of the reflexive pronoun *se*; and 4) subjects represented by demonstrative pronouns (e.g. *eso*, “that”). Table 2 provides an example of the coding system that was used in order to classify the different variables from the study. For each token, it was determined whether the subject of the phrase was explicit or implicit (e.g. *Yo me llamo* = Explicit/*exp*), whether the subject was first, second, or

third person singular or plural (e.g. *Tengo veinte tres* = First person singular/*I Y*), and if the verbal tense was in the past, present, or future (e.g. *Yo nací en Winston-Salem* = Past/*pas*). It should be acknowledged that the classification of “Past” tense in this study included both Preterite and Imperfect forms, and that the category of “Future” incorporates the morphological future (in which *hablar* is conjugated *hablaré, hablarás*, etc.) and the Conditional to represent cases that refer to abstract events that have yet to take place.

Table 2. Sample of coding system used in this study.

Token	Participant	Age	Country of Birth	Years in eNC	Subject explicit/implicit (exp/imp)	Person (1/2/3)	Person plural? (Y/N)	Verbal tense (pas/pre/fut)	BLP
Yo me llamo	1	23	USA	3	exp	1	n	pre	-101.708
Tengo veinte tres	1	23	USA	3	imp	1	n	pre	-101.708
Yo nací en Winston Salem	1	23	USA	3	exp	1	n	pas	-101.708
Voy a agarrar mi bachelor's en mayo	1	23	USA	3	imp	1	n	pre	-101.708
Tengo un apartamento	1	23	USA	3	imp	1	n	pre	-101.708
Vivo en ***	1	23	USA	3	imp	1	n	pre	-101.708

Results and Discussion

Total Distribution of Subject Expression

Overall, the process of coding generated 1,309 tokens that could be classified as either containing explicit or implicit subject expression. 333 of these tokens contained explicit statement of the subject, while 976 were implicit, demonstrating a greater prevalence proportionally for the null declaration of subject through morphological flexion of the verb in main clauses. This rate of subject expression is more or less consistent with previously studied

native speech rates, which typically range between 15 and 40 percent, depending on the speaker's dialect of origin (Escobar & Potowski, 2015). Table 3 shows the prevalence for this sample to compare the manifestation of subject for the entire corpus of data.

Table 3. Total distribution of subject expression.

Subject Expression	Frequency/(%)
Explicit	333 (25.4%)
Implicit	976 (74.6%)
Total	1309 (100%)

What can be seen from the results of distribution from the entire corpus is that there is a strong preference for speakers to implicitly communicate the subject in speech. These results can be explained primarily by the morphosyntactic system of Spanish, which allows speakers the option to not overtly declare the subject in many cases in discourse. Since the use of the implicit subject demands less effort and abbreviates the phrase, it is possible that this fact may cause the speakers as a community to continue omitting subject expression in the majority of cases despite increased contact with English. Below, it will be shown how additional variables affect this pronoun expression rate.

Distribution of Social Variables

As stated above, a selection of various social and extralinguistic variables and their effects on subject expression among participants were chosen for analysis in this study. Those that were examined included the language dominance of participants as indicated by BLP score, the participant country of origin, and the time that participants had spent in the region of eastern North Carolina (eNC). After the coding process, there were also significant individual results of

subject expression that prompted additional analysis aside from the factors outlined in the research questions

Distribution by BLP score. One of the primary purposes of this investigation was to determine the effects of bilingualism on subject expression. The variable of BLP score, took into account the values of Total Language Dominance in either English or Spanish, yielded significant results that indicate the effects of increased proficiency in one language over another as a source of influence toward subject expression. From the entire sample, 996 tokens were generated from English-dominant individuals, 275 of which were verbs with explicit subjects, while 313 came from Spanish-dominant individuals, 58 of which were explicitly declared verbs. These frequencies resulted in a 27.6% rate of explicit expression for the English-dominant participants and an 18.5% expression rate for the Spanish-dominant participants. This distribution is illustrated in Table 4.

*Table 4. Subject expression by Total Language Dominance. ****

BLP	Subject Expression (N₀/%)	
	Explicit	Implicit
English-dominant	275 (27.6%)	721 (72.4%)
Spanish-dominant	58 (18.5%)	255 (81.5%)

p = .001 ***

Inferential statistical analysis was completed on these values through a Pearson Chi-square test, which yielded a p-value of .001. This result indicates that the difference in subject expression between English-dominant and Spanish-dominant individuals was highly significant. Furthermore, this value is consistent with our hypothesis, which stated that individuals more dominant in English would display higher subject expression rates than those who were Spanish dominant. Because dominance in English indicates a higher influence of English overall on the participant's bilingual background, it is likely that this difference can be attributed to language

contact. This variable is independent from the individual's country of origin, which as indicated by the sample did not entirely go hand-in-hand with language dominance.

Expression by country of origin. The frequencies and rates of subject expression for the social variable of country of origin can be seen in Table 5, which collects both the number of tokens for the explicit declaration of subject for all of the participants of both the first generation, Mexican-born group, and the second-generation heritage speaker group. The majority of all the coded tokens in terms of frequency occurred with the heritage speakers, given the fact that the majority of participants (13 of 18) were in this group. Heritage speakers generated 1,076 of the total 1,309 tokens. The explicit rate of subject expression for heritage speakers was 26.6% compared to the 20.2% of first-generation speakers from Mexico, meaning that the second-generation group did display an observably higher rate of expression.

*Table 5. Subject expression based on country of origin.**

Country of Origin	Subject Expression (N ₀ /%)		
	Explicit	Implicit	Total
Mexico	47 (20.2%)	186 (79.8%)	233 (100%)
USA	286 (26.6%)	790 (73.4%)	1076 (100%)

*p=.042**

Furthermore, with this variable, an inferential statistical test was performed in order to determine whether there was a significant difference in explicit subject use for participants born in the U.S. and those who were born in Mexico. The results of a Pearson Chi-square test provided a p-value of .042, which demonstrates that there is a slightly significant difference between the explicit subject expression of heritage speakers and native speakers from Mexico. This finding is also crucial in answering the research questions for this study, which proposes that speakers born in the U.S. would declare the explicit subject with more frequency. It can also be compared to previous literature, which identifies social variables as influential in bilingual

communities (Flores-Ferran, 2010). This being said, the results of this test were not as significant as expected for the current study when compared to the results generated by Lawrence, Meehan, and Fafulas (2016), which can possibly be attributed the more imbalanced corpus regarding country of birth among other social factors such as individual variation.

Time in eNC and subject expression. The rates of explicit declaration were also analyzed based on the number of years each participant had spent in the specific region where this study was conducted, eastern North Carolina (eNC). In Table 7, the individual expression rates can be seen organized by the ascending number of years that each participant had lived in the region. From this sample, it can be seen that individuals had spent a range between 2 and 24 years in eNC. The average number of years spent in eNC across participants was 10.2 years.

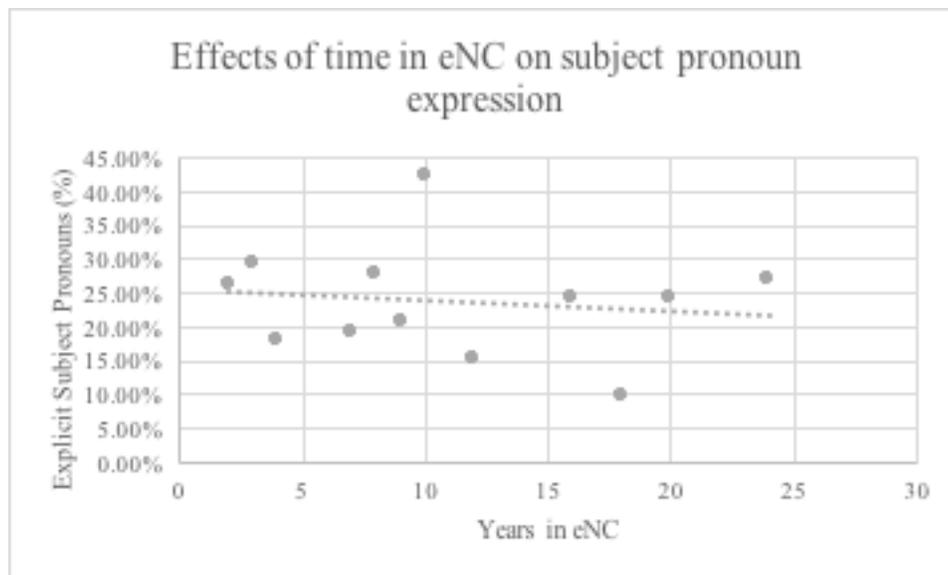
Table 6. Years in eNC and subject expression rate by participant. **

Participant	Years in eNC	Subject Expression Rate
5	2	35.14%
18	2	10.71%
37	2	25%
38	2	24.40%
1	3	35.24%
14	3	16.33%
33	4	18.00%
6	7	19.05%
16	8	22.50%
30	8	30.70%
4	9	20.93%
23	10	42.50%
40	12	15.30%
3	16	24.32%
25	18	9.70%
15	20	23.33%
36	20	24.40%
10	24	27.14%

$p=.005^{**}$

A Pearson Chi-square test analyzing the number of years spent in eNC versus the average subject expression rate for individuals who had spent a selected number of years in the region yielded a p-value of .005. Given this value, the amount of time an individual had resided in eNC did have a statistically significant effect on subject expression rate. However, results differed from the expected when individual values were displayed visually. Figure 3 illustrates subject expression rates for the sample based on the number of years spent in eNC. The line of Best Fit indicates that there is a slight negative correlation between time in eNC and subject expression, which contradicts our original hypothesis stating that more time in the region would increase explicit subject expression. It should be acknowledged that this social variable does not take age or country of origin into account when measuring statistical significance, meaning that the number of years spent in eNC does not measure what fraction of an individual's life time that number of years may represent. Future results may differ when the proportions of participants' lifetimes spent in eNC are considered.

Figure 2. Time in eNC and subject expression. **



$p=.005^{**}$

Individual variation of subject expression. Aside from the rates of subject expression seen in the groups of participants defined by other social variables, the individual rates of subject expression for each participant were also calculated based on the number of explicitly declared tokens within each entire set of tokens per participant. For example, the first participant (1) had 105 total coded tokens, 37 of which were explicit. This proportion resulted in a 35% individual rate of subject expression from the set of tokens. The average rate of individual expression was 25.4%. The entire list of individual rates can be seen in detail in Table 4.

Table 7. Rates of subject expression by participant. ***

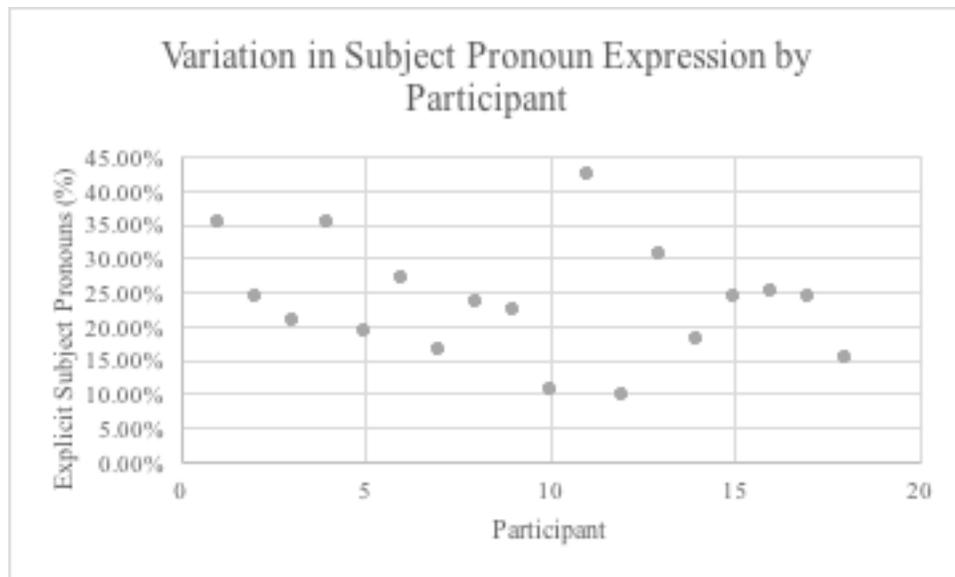
Participant	Explicit (N ₀)	Implicit (N ₀)	Total (N ₀)	% Explicit
1	37	68	105	35.24%
3	9	28	37	24.32%
4	18	68	86	20.93%
5	26	48	74	35.14%
6	8	34	42	19.05%
10	54	145	199	27.14%
14	8	41	49	16.33%
15	7	23	30	23.33%
16	9	31	40	22.50%
18	3	25	28	10.71%
23	34	46	80	42.50%
25	3	28	31	9.70%
30	23	52	75	30.70%
33	16	73	89	18.00%
36	22	68	90	24.40%
37	25	75	100	25%
38	20	62	82	24.40%
40	11	61	72	15.30%
Average rate of subject expression: 25.4%				

p = .001 ***

Additionally, a Pearson Chi Square performed in SPSS determined that the differences between the individual expression rates of participants was very significant (*p* = .001). When

individual rates of expression for the corpus are taken into account, it can be seen that there was a rather large range within these expression rates, between 9.7% and 42.5%. It is possible that the significant differences between individual rates, especially when taking into account the vast majority of participants were second-generation speakers, can be explained by the fact that it is likely each of the participants possess different linguistic backgrounds behind their bilingualism between English and Spanish, including different attitudes, histories, and proficiencies in English and Spanish. As can be seen by the similarly significant variable of BLP scores toward subject expression, it is possible that degrees of bilingualism may impact these individuals differently in their morphosyntax. In Figure 2, a visual representation of the variation between individual scores is also shown to further illustrate the fact that rates are not necessarily consistent between participant groups, which is likely due to additional external social factors.

Figure 3. Individual variation and subject expression.



Distribution of Linguistic Variables

Aside from the selection of social variables that were examined and tested for significance in this study, we also observed the distribution of linguistic variables including

subject and verbal tense. Previous literature identifies these variables as influential toward bilingual morphosyntax, especially in subject expression (see Carvalho, Orozco, & Shin, 2015), but it was of interest to observe the behavior of these variables in the more recently studied community of eastern North Carolina.

Subject expression and plurality of subject. During the coding process, tokens were organized for the linguistic variable of person by dividing between the first, second, and third person singular and plural. The majority of all tokens from the sociolinguistic interviews were identified as the first person singular, with 781 total verbs, 137 of which were explicit. Table 8 shows the frequency and rate for each form of grammatical person for all coded tokens.

Table 8. Subject expression and plurality of subject.

Person	Singular (N ₀ /%)		Plural (N ₀ /%)	
	Explicit	Implicit	Explicit	Implicit
1 st	137 (17.5%)	644 (82.5%)	5 (5.4%)	88 (94.6%)
2 nd	0	45 (100%)	0	0
3 rd	111 (50.5%)	109 (49.5%)	79 (46.5%)	91 (53.5%)

It was expected that more tokens would occur in the first person singular due to the sociolinguistic interview instrument used for data collection, which contained questions focusing more often on the participant than from other perspectives. However, an observable difference between subject expression can be seen with the grammatical person when one takes into account the proportion of explicit expression within each form. Here, the rates for each person demonstrate that there was a much greater prevalence of explicit subject use in both singular and plural third person cases, with a 50.5% explicit rate for third person singular and a 46.5% rate for third person plural. These rates are noticeably higher than the average rate of explicit subject expression for the entire corpus, which as seen above was 25.4%. In contrast, rates of declaration

for the first person were much lower than the average for the whole corpus, with a rate of 17.5% explicit subject expression for first person singular, and 5.4% for first person plural. There were only 45 total cases of the second person, all of which were singular and implicit. Although this rate is visibly the lowest for explicit subject expression rate at 0%, there were also too few cases total to analyze this distribution.

The findings for this linguistic variable have significant importance to this study. In terms of the proportions observed for person in Table 8, a notable difference can be seen between the first and third person, with a greater preference for subject expression in the third person. This finding is consistent with the suggestions of previous literature, which identifies person as the most influential variable for explicit declaration of subject (Otheguy, Zentella, & Livert, 2007). Additionally, the results for this variable address a hypothesis for this study, which proposed that there would be a greater prevalence for explicit declaration in the third person. It is possible that this finding can be explained by the inherent nature of the grammatical third person, which can refer to any subject that is not in proximity to the speaker and often requires more specification through detailed noun phrases to provide the appropriate context within a statement.

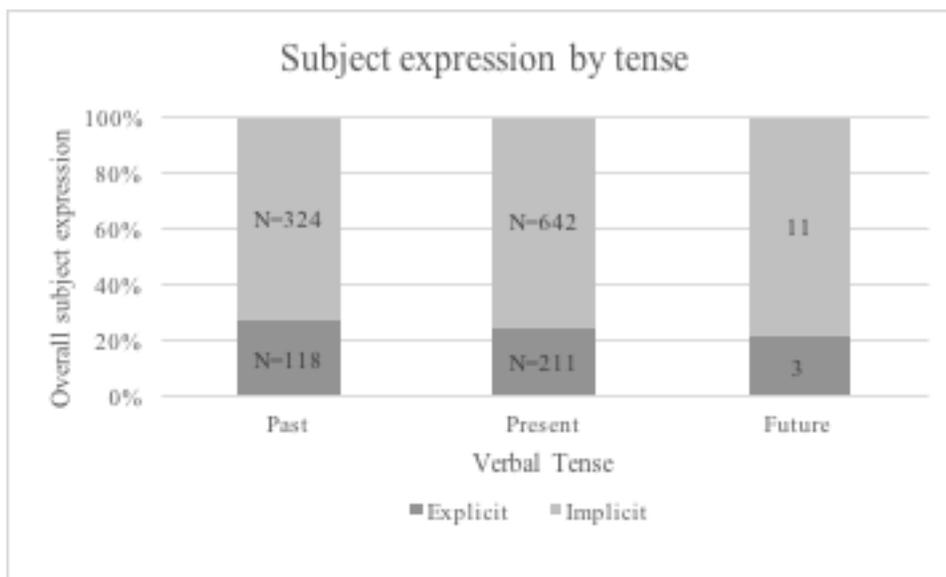
Subject expression and verbal tense. Table 9 shows the distribution by number of tokens for verbal tense as well as the rates for explicit declaration for each tense. 442 verbs from the corpus were conjugated in the past, 853 in the present, and 14 in the future. The past tense represented 33.8% of all cases, the present 65.1%, and the future 1.1%. Therefore, it can be determined that a greater number of the total cases occurred in the present. Additionally, more explicit tokens also were conjugated in the present.

Table 9. Subject expression and verbal tense.

Tense	Subject Expression (N ₀ /%)		
	Explicit	Implicit	Total
Past	118 (26.7%)	324 (73.3%)	442 (100%)
Present	211 (24.7%)	642 (75.30%)	853 (100%)
Future	3 (21.4%)	11 (78.6%)	14 (100%)

However, when the proportion of explicit tokens is examined within each verbal tense category, there is not as visible a difference between the rates of explicit subject expression for each tense. When referring to Figure 4, it can be seen that the rate of subject expression is similar within each category, in the range of 20% to 26%. With the future tense, the lowest rate of subject expression was found, at 21.4%. Despite this rate, it should also be acknowledged that there were very few cases that were coded in the future category within the corpus compared with other tenses, so this rate may not be representative of typical subject expression rates within forms of the future tense.

Figure 4. Subject expression by verbal tense.



In any case, it can be seen that these data do not demonstrate a preference for explicit declaration in any verbal tense more than in others. Although with the results it can be observed that the speech analyzed in this corpus preferred to use the present tense, this manifestation cannot be tied to the dependent variable. This finding is notable for this investigation because it demonstrates possible evidence against previous literature, which suggests that verbal tense can affect explicit declaration of subject (Bayley, Cardenas, Schouten, & Vélez, 2012). Additionally, when these data are compared with our hypothesis, it should be admitted that in this study a difference in subject expression between verbal tenses was not found.

Conclusions

Limitations of the Study

Ultimately, it should be acknowledged that though this study yielded a set of significant studies that generate a meaningful direction for linguistic research on eNC Spanish, there still remain gaps in a study like this one, which can only examine a selection of many possible variables in a corpus. In regards to the results for the linguistic variable of person, there was a much greater preference among the sample for the use of the first person because the questions from the sociolinguistic interviews, our primary instrument, were designed to elicit personal responses. Additionally, the system of coding did not take into account the difference between several verbal forms, within the categories that we defined, specifically the preterite and imperfect that were both classified as past tenses, and the morphological future and conditional that collectively represented the “future” category. It is possible that if additional analysis was completed comparing results between these individual tenses, that the results could be distributed differently. More specifically, it is likely that dividing the “past” category between preterite and

imperfect would yield different rates of explicit subject expression as a result of the different morphological features found in the flexion of these tenses.

There is also much left to explore within the categories of social variables for bilingual Spanish speakers in eNC. The corpus was limited by the imbalance between background of participants – while there were 13 heritage speakers in our sample, there were only 5 first-generation speakers from Mexico, which could have skewed the results for the country of origin against the dependent variable. Additionally, the results that were found on the individual variation between speakers demonstrate that the use of overall BLP scores may not properly illustrate the complex linguistic backgrounds possessed by each of the participants – rather, it is necessary to conduct further research examining how more defined aspects of participants’ bilingualism, such as language history, use, attitudes, and proficiency in both Spanish and English may affect subject expression rates. Finally, the variable of time spent in the region of eastern North Carolina is needed to be investigated in further detail, as our current analysis does not take into account country of birth or BLP along with the time that individuals had spent in the regions of investigation. A multivariable analysis will likely be needed to determine whether the impact of time spent in this region is skewed by other social variables.

Future Directions

By recognizing the limitations of this study, this investigation can also clear a path toward future research questions involving the topic of subject expression in eNC Spanish. For example, though inferential statistics were conducted only for social variables, there is still a need to see whether results are significant for the distribution of the linguistic variables of verbal tense and grammatical person in this sample. Additionally, as has been noted above, future studies should create a division between different verbal forms and tenses, especially by

distinguishing between the preterite and imperfect past tenses in Spanish. It is also possible to incorporate additional analysis on other social variables. This study only took into account the effects language contact on Mexican Spanish; however, though this is one of the most prevalent Spanish origin dialects in eNC, it is not the only one represented in the region. It is necessary to also investigate the theme of Spanish dialect contact in the region and subject expression, not only the impact of English on this variable. Finally, though the Total Language Dominance score used in the study to measure individuals' preference for English or Spanish gave insight into the effects of bilingualism on our dependent variable, this score does not take into account the specifics of participants' linguistic backgrounds. A future study is necessary to determine whether certain aspects of these backgrounds found in the BLP instrument, such as Language History or Language Attitudes, can serve as additional predictors of subject expression when separated from the overall scores from the BLP.

Overall, the findings from this study as detailed in this thesis, though they have not all fulfilled the predictions outlined in our hypotheses, still demonstrate a visible effect between the investigated variables toward the declaration of subject in the U.S. Spanish of this rural region. The study is one of few that has tested these variables in a rural region as compared with what has been outlined in previous literature on the effects of language contact in U.S. Spanish. Furthermore, our results indicate that bilingual contact between English and Spanish has a notable impact on on subject expression across generations of Spanish speakers in eNC, which only further justifies this region in the field of Hispanic linguistics as one worthy of future study. Given that many first- and second-generation Spanish speakers develop their own linguistic identities based on the areas where they live in the U.S., eNC's rural background and the patterns

of change that we have documented in the region suggest its own Hispanic residents may be in process of constructing their own emerging identities characteristic of the region.

References

- Bayley, R., Cardenas, N.L., Treviño Schouten, B., & Vélez Salas, C.M. (2012). Spanish dialect contact in San Antonio, Texas: An exploratory study. En Geeslin, K. & Díaz-Campos, M. (Eds.) *Selected Proceedings of the 14th Hispanic Linguistics Symposium* (pp. 48-60). Somerville, MA: Cascadilla Proceedings Project.
- Birdsong, D., Gertken, L.M., & Amengual, M. (2012, January). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. Retrieved from: <https://sites.la.utexas.edu/bilingual/>.
- Carvalho, A.M, R. Orozco & N. L. Shin (eds.). 2015. *Subject Pronoun Expression in Spanish: A Cross-dialectal perspective*. Georgetown University Press.
- Escobar, A. M. & Potowski, K. (2015). *El español de los Estados Unidos*. Cambridge: Cambridge University Press.
- Flores-Ferrán, N. (2010). ¡Tú no me hables!: Pronoun expression in conflict narratives. *International Journal of the Sociology of Language* 2010 (203), 61-82.
doi:10.1515/IJSL.2010.022
- Lawrence, A., Meehan, A., & Fafulas, S. (2016, February). Morphosyntactic variation of subject pronoun expression in an emerging dialect of Spanish in eastern North Carolina. *Spanish Linguistics in North Carolina (SLINKI) Conference*. Charleston, SC: College of Charleston.
- Livert, D. & Otheguy, R. (2010). A multilevel statistical analysis of changes in language use among first-generation immigrants in a bilingual setting. *International Journal of the Sociology of Language* 203(1), 83-99.

Otheguy, R., Zentella, A.C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 83(4), 770-802.

Retrieved from <http://www.jstor.org/stable/40070965>.