

A comparative analysis of first-person singular subject pronoun expression in Facebook and sociolinguistic interviews: Studying variation via social media

Un análisis comparativo de la expresión de sujeto en la primera persona singular entre Facebook y entrevistas sociolingüísticas: Analizando la variación en las redes sociales

Ryan M. Bessett

University of Texas Rio Grande Valley, United States of America

ryan.bessett@utrgv.edu

Original recibido: 13/09/2021

Dictamen enviado: 06/12/2021

Aceptado: 31/01/2022

Abstract

This study explores Facebook posts and compares them to oral data from sociolinguistic interviews to investigate variation in social media. The participants in the Facebook group were verified through acquaintances to be born and raised in Hermosillo, making for an accurate comparison, leaving differences to be due to data source and not dialectal variation. The variable used for the comparison is first-person subject pronoun expression. The results show differences in terms of overall frequency (16.7 % expressed pronouns in the interview data and 10.2 % in the Facebook data), as well as the variables that condition subject pronoun expression in Hermosillo Spanish. These analysis findings further our understanding of variation in social media and variation in written Spanish.

Keywords: language variation, Spanish, subject pronoun expression, written variation.

Resumen

Este estudio explora publicaciones de Facebook y las compara a datos orales de entrevistas sociolingüísticas para investigar la variación en las redes sociales. Los participantes del grupo de Facebook fueron verificados por conocidos para asegurar de que fueron nacidos y crecidos en Hermosillo, creando una comparación rigurosa, permitiendo que las diferencias entre una fuente de datos y otra se puedan atribuir al tipo datos y no a la variación dialectal. La variable empleada para la comparación es la expresión de sujeto de la primera persona singular. Los resultados demuestran diferencias en términos de frecuencia (16.7 % de pronombre expreso en los datos de

las entrevistas y 10.2 % en los datos de Facebook), tanto como en las variables que condicionan la expresión de sujeto en el español hermosillense. Los resultados de este análisis promueven nuestro conocimiento de la variación en las redes sociales y de la variación en el español escrito.

Palabras clave: español, expresión de sujeto, variación lingüística, variación escrita.

Introduction

Recent variationist work has become increasingly interested in finding new ways of measuring natural, everyday speech that goes beyond the traditional Labovian sociolinguistic interview (see papers from *New Ways of Analyzing Variation*, 2016, 2017). While many of the new techniques involve recording participants during multiple interaction types (at work, with friends, with family, etc.), another avenue has been found in the use of language in social media platforms. Social media provides an interesting venue for analyzing variation in that it is not only informal, casual speech that is produced at the moment and usually without much thought to how it is worded, but also that it is written versus the traditional spoken language analyzed through sociolinguistic interviewers (Yus, 2002, 2003).

One major concern when working with large corpora or from data-mining social media pages is that it is difficult, if not impossible, to gather pertinent extralinguistic data of the people producing the utterances that are being examined (Díaz-Campos, 2014). The present study seeks to eliminate this, and other issues related to working with social media by manually collecting the data from specific open profiles pages. We ensured an accurate place of birth and residence of the participants by using friends and friends of friends to verify.¹ In this study, we will explore the differences in the production of subject pronoun expression in Spanish via data from Facebook posts from residents of Hermosillo, Sonora, Mexico, and from sociolinguistic interviews (also from Hermosillo residents) to contribute to the dialogue of how social media can be used to analyze variation.

Previous research on subject pronoun expression in Spanish

Subject pronoun expression is an ideal test case for the comparison in this study because it is a variable that has been extensively studied in Spanish variationist research in both monolingual (Cameron, 1993; Solomon, 1999; Soares da Silva, 2006; Manjón-Cabeza Cruz, Pose Furest & Sánchez García, 2016; Repede, 2019; Martínez-Lara, Guerrero González & González Vergara, 2021; among

¹ This is sometimes referred to as the “snowball” method.

others) and bilingual (Elizaincín, 1995; Bayley, 1997; Flores-Ferrán, 2004; Otheguy, Zentella & Livert, 2007; Torres Cacoullós and Travis, 2010; Carvalho & Child, 2011; Otheguy & Zentella, 2012; Michnowicz, 2015; Roselló Veredeguer, 2021; among others) communities. In Spanish, verbs can be expressed with a pronoun like in (1a) or unexpressed as seen in (1b).

- (1) a. **Yo soy** aficionado al buen comer. –FB²01
 ‘I am a fan of eating good food.’
 b. **Soy** el Dr. House de los posts, carnal. –FB01
 ‘(IØ) am the Dr. House of these posts, bro.’

Previous studies have shown rates of first-person singular pronouns from 21 %-33 % in Mexican Spanish. The lowest rate comes from Yucatan Spanish with 21 % (Michnowicz, 2015, p. 112), followed by Mexico City at 24.7 % (Lastra & Martín Butragueño, 2015, p. 43), and lastly Veracruz Spanish with the highest level at 33 % (Orozco, 2016, p. 5). The rates are much lower compared to Caribbean dialects; for example, first-person singular pronouns are expressed 51.6 percent of the time in the Dominican Republic (Alfaraz, 2015, p. 10).

In addition to overall frequency, several linguistic factors have been known to condition pronoun expression in Spanish. First, co-reference is consistently a high-ranking factor in Spanish dialects where a switch in the reference from the previous verb favors an expressed pronoun while a subject with the same reference as the previous verb disfavors an expressed pronoun (Cameron, 1993; Travis, 2007; Lastra & Martín Butragueño, 2015; Alfaraz, 2015; among many others). This pattern can be seen in the examples in (2), where (2a) shows a context where the previous subject is the Øsame and (2b) is a context where there was a change in referent of the prior verb.

- (2) a. No entendí, soy lento. –FB02
 ‘(IØ) didn’t understand, (IØ) am slow.’
 b. Nada que ver. El Gallo Negro es pro. **Yo soy** aficionado al buen comer.–FB01
 ‘Not even close. El Gallo Negro is a pro. **I am** a fan of good eating.’

² FB and H represent participants in the two data sets analyzed in this study. FB is the label for participants in the Facebook group and H is for the participants in the sociolinguistic interview group.

While switch reference places the focus on when a referent is changed or maintained in the discourse, other researchers have explored the priming effect that results from a previously expressed pronoun triggering the following verb also to have an expressed pronoun, regardless of a change in referent (Travis, 2007; Torres Cacoullós & Travis, 2010; Carvalho & Child, 2011; Carvalho & Bessett, 2015; among others). The priming effect can be seen in (3), wherein (3a) a sequence of two previously expressed pronouns triggers an expressed pronoun of the same referent in a third verb, but in (3b), a similar sequence of unexpressed pronouns results in an unexpressed pronoun in the third verb in the sequence.

- (3) a. Si no voy a llegar a dormir, yo hablo. Pero si yo no hablo... bueno, yo hablo. –H20
 ‘If (IØ)’m not going to go home to sleep, I call. But if I don’t call... well, I call.’
- b. Uuuuuufff te lo juro que te iba a etiquetar y vi que tú lo publicaste. –FB11
 ‘Uuuuuufff (IØ) swear to you that (IØ) was going to tag you and (IØ) saw that you posted it.’

Another factor that conditions pronoun expression in Spanish is the tense, mood, and aspect of the verb. While the ranking differs between communities, the general trend is that ambiguous verb forms favor expressed pronouns (Travis, 2007; Lastra & Martín Butragueño, 2015; Alfaraz, 2015; Orozco, 2016; among others). This can be seen in (4) wherein (4a) the verb in the imperfect tense has an expressed pronoun but the verb in (4b), which is conjugated in the morphological future, has an unexpressed pronoun.

- a. Este es el motivo de mi gordura... ¿ves XY³? Ya **decía yo** que algo pasaba. –FB10
 ‘That is the motive for my chubbiness... See XY? **I was saying** that something was going on.’
- b. Me ofende que pienses que me **iré a roquear** sin ti. –FB11
 ‘(ItØ) offends me that (youØ) think that (IØ) will go rock without you.’

³ XY is used here to replace identifiable information, in this case the name of a friend of the participant.

While morphological ambiguity is not always a significant predictor variable in Spanish, it was found to condition pronoun expression in Mexico City, where morphologically ambiguous verb forms favored an expressed pronoun (Lastra & Martín Butragueño, 2015). In some studies, the type of clause has also been shown to affect pronoun expression. Expressed pronouns are least likely to be realized in coordinate clauses and most likely to be realized in main clauses (Abreu, 2009; Flores-Ferrán, 2009; Otheguy & Zentella, 2012; among others). The factor groups outlined here will be used in the present study to compare subject pronoun expression in Hermosillo to other dialects to aid in the comparison of Facebook versus sociolinguistic interview data.

Of particular interest in the case of this study is that genre effects are important in the use of expressed versus unexpressed pronouns in Spanish-speaking communities (Solomon, 1999; Flores-Ferrán, 2002, 2010; and Travis, 2007). This observation has two main variants: the theme of the narrative being discussed and the type of conversation. First, narratives about conflicts have shown higher rates of expressed pronouns than other narrative types (Solomon, 1999; Flores-Ferrán, 2002, 2010). Additionally, conversational interviews (with more dialogue between interviewer/interviewee) showed to have higher expressed pronouns than a narrative (a more monologue storytelling of the interviewee) style (Travis, 2007). These findings hold implications for the present study since Facebook posts are of a more monologue nature than the conversational style of the sociolinguistic interviews, so we expect to find a significant difference in the overall use of expressed pronouns between the two data sets.

Lastly, very little is known about variation in written contexts in Spanish. Martínez Mira (2009) examined the use of subjunctive vs. the indicative in contexts where the subjunctive is expected and found that heritage speakers of Spanish favored the subjunctive or used it more in oral data as opposed to in their writing (p. 114). In terms of subject pronoun expression in writing, heritage speakers have shown to use more expressed pronouns in formal writing samples than in informal (free writing) samples (Martínez, 2007, p. 36-37). In a monolingual context, Fernández Flórez (2019) found a rate of 12 % expressed pronouns for Sonoran speakers in a written narrative activity, a rate much lower than the one found in sociolinguistic interview data from Mexico (21 %-33 %, as outlined above). These results indicate that the genre effect found in speech is also present in written data. Additionally, there is reason to believe that written data results in a lower overall frequency of expressed pronouns.

The present study

This study will examine subject pronoun expression in Hermosillo, Sonora, Mexican Spanish, from both spoken and written speech. The spoken data comes from sociolinguistic interviews and the written data from Facebook posts. Based on the previous research outlined above, we can form the following hypotheses. Hypothesis 1: The Facebook data set will have a lower overall percentage of expressed first-person singular pronouns than the sociolinguistic interview group. Hypothesis 2: There will be a significant difference between the data sources where the Facebook data set will disfavor expressed pronouns, and the sociolinguistic interview data set will favor expressed pronouns. Hypothesis 3: The factors that condition pronoun expression will be different for the Facebook and sociolinguistic interview data sets.

Methodology

Participants

Data was gathered from 32 monolingual participants from and living in Hermosillo, the capital of the state of Sonora in northern Mexico. These participants are equally divided between men ($n = 16$) and women ($n = 16$) and by data source, sociolinguistic interview ($n = 16$) and Facebook posts ($n = 16$). Table 1 shows the distribution of participants by sex and data source.

TABLE 1. DISTRIBUTION OF PARTICIPANTS BY GENDER AND DATA SOURCE

GENDER	DATA SOURCE	
	INTERVIEW	FACEBOOK
Male	8	8
Female	8	8
Total	16	16

The participants in the sociolinguistic interview group were friends of the investigator or friends of friends. Each interview was semi-structured, consisting of narratives and conversational styles, and lasted approximately one hour. The Facebook participants were friends of friends of the investigator whose profiles stated they were born or lived in Hermosillo. Each one was confirmed by a friend of the investigator to have been born, raised, and currently living in Hermosillo.

Coding practices

The first 100 tokens of verbs conjugated for first person singular ‘yo’ were coded for each sociolinguistic interview. The precedent for limiting the scope of the study to one grammatical person stems from the realization that each grammatical person behaves differently (e.g., Lapidus & Otheguy, 2005; Torres Cacoullós & Travis, 2010a, 2010b; Gudmestad, House, & Geeslin, 2013; de Prada Pérez, 2015; Bessett, 2018). The first 100 tokens starting with the most recent and working back, were included for the Facebook group. The number of tokens was limited to keep a more balanced representation of tokens among participants and is in line with previous studies (Carvalho & Bessett, 2015; Bessett, 2018). Potential Facebook participants who did not have open profiles were discarded as well as profiles containing less than 35 tokens, again to maintain an equal representation.

Each token was coded for the dependent variable of expressed or unexpressed pronoun. Additionally, the independent variables that have been proven to condition subject pronoun expression in monolingual communities, as outlined in section 1.2 of the introduction, were included in the analysis. The first factor group was tense, mood, aspect (TMA) of the verb and was comprised of the following within-factor groupings: present indicative, preterit, imperfect indicative, periphrastic future, morphological future, conditional, present subjunctive, past subjunctive, and the perfect tenses were combined into one category. Second, co-reference measured the relation of the referent of the verb to that of the previous verb and was made up of three categories, co-reference with the subject of the previous verb (no switch), switch with the subject of the previous verb but co-reference with the object, and switch with the subject and the object(s) of the previous verb (a complete switch). The next factor, parallelism, accounted for the pronoun expression of the previous two verbs in the same utterance and was coded for the following categories: first token, previous subject was unexpressed (N_), previous subject was expressed (E_), penultimate subject was expressed and previous subject was unexpressed (EN_), last two subjects were expressed (EE_), penultimate subject was unexpressed and previous subject was expressed (NE_), and the last two subjects were unexpressed (NN_). Several clause types were also coded: main, subordinate relative, subordinate other, and coordinate. The morphological ambiguity of the verb was coded as being ambiguous or unambiguous. The data was also coded for data sources, sociolinguistic interview, or Facebook. The data was analyzed through a multivariate model in the statistical program GoldVarb.

Results

This section will explore the differences in subject pronoun expression in the sociolinguistic interviews and the Facebook posts. First, we discuss the overall frequencies, compare them to the previously studied Mexican communities, and respond to Hypothesis 1. Then, we test Hypothesis 2 by reporting on whether the data source is a significant predictor variable for subject pronoun expression in Hermosillo. Next, we explore the linguistic conditioning of the dependent variable by data source to explore Hypothesis 3 and determine if both the sociolinguistic interview and the Facebook data follow the same patterns. Lastly, we briefly comment on patterns found about the extralinguistic factor of participant sex.

Overall frequency by data source

In terms of the overall frequency of first-person singular ‘yo’ in Hermosillo, there is a difference between the interview and Facebook data. The interview data reveals 16.7 % (202 of a total of 1 211 tokens) expressed pronouns while the Facebook data only contains 10.2 % (138 of a total of 1,355 tokens), as shown in Table 2.

TABLE 2. OVERALL FREQUENCY OF 1ST PERSON SINGULAR SPE IN HERMOSILLO BY DATA SOURCE.

INTERVIEW		FACEBOOK	
%	n	%	n
16.7	202/1211	10.2	138/1355

Hypothesis 1 predicted that there would be a higher overall rate of expressed ‘yo’ for the sociolinguistic interview data set than the Facebook data set. Given the results in Table 2, we find support for Hypothesis 1.

When comparing the overall frequencies of the Hermosillo interview data to previously studied communities, we find that the rate is comparable to other areas of Mexico in that Mexican Spanish consistently demonstrates low rates of expressed ‘yo’, as shown in Table 3. However, Hermosillo does appear to be at the lower than any other previously studied communities in Mexico, with 16.7 % expressed ‘yo’ versus the 21 % in Yucatan Spanish (Michnowicz, 2015), the 24.7 % in Mexico City (Lastra & Martín Butragueño, 2015), and the 33 % in Veracruz (Orozco, 2016).

While the Facebook data set shows an even lower frequency than other communities as measured by sociolinguistic interviews, it is comparable to the other known study on formal written narratives in Sonoran Spanish, 10.2 % for

Facebook and 12 % for written narratives (Fernández Flórez, 2019). This finding corroborates previous reports of genre effects in the overall frequency of expressed pronouns in Spanish oral data (Solomon, 1999; Flores-Ferrán, 2002, 2010; and Travis, 2007), and extends it to include written data. This finding is also consistent with the results of Martínez (2007), who found that Heritage learners of Spanish used more expressed pronouns in a formal written assignment than in an informal one (p. 36-37). The written narratives from the Fernández Flórez (2019) study were collected in an academic setting, while the Facebook data is from social media, a much more informal context. Having explored in detail the overall frequencies of expressed pronouns in the interview and Facebook data sets, we now turn to test Hypothesis 2, whether the difference in frequency between the two groups results in a significant predictor variable for pronoun expression in Hermosillo.

TABLE 3. FIRST PERSON SINGULAR SPE FREQUENCIES IN MONOLINGUAL SPANISH COMMUNITIES.

COMMUNITY	FREQUENCY
Hermosillo, Facebook (current study)	10.2 %
Sonora, written narratives (Fernández Flórez 2019)	12.0 %
Hermosillo, interview (current study)	16.7 %
Yucatan (Michnowicz 2015)	21.0 %
Mexico City (Lastra & Martín Butragueño 2015)	24.7 %
Veracruz (Orozco 2016)	33.0 %

Data source as a predictor variable

To test Hypothesis 2, a multivariate analysis was run on the entire data set using pronoun expression as the dependent variable and the factors outlined in the methodology section as predictor variables, with the addition of data source (interview or Facebook). The results shown in Table 4 demonstrate that data source is not only a significant factor group, but it is also the second-highest ranking variable with a range of 19, after co-reference (range = 21). This result agrees with the first part of hypothesis 2, which stated the difference in pronoun expression between the two data sources would be significant.

To better understand the direction of the effect of data source on subject pronoun expression, Table 5 shows the factor weights for the two groups (interview and Facebook).

The sociolinguistic interview group favors expressed pronouns (FW = 0.60), while the Facebook group disfavors expressed 'yo' (FW = 0.41). The second part

of Hypothesis 2 predicted that the Facebook group would disfavor expressed pronouns, and so we must fail to reject Hypothesis 2. The fact that data source is a high ranked predictor gives further evidence to the notion of genre effects (Solomon, 1999; Flores-Ferrán, 2002, 2010; and Travis, 2007). Not only do genre effects exist in terms of frequency, but when matched up to other linguistic variables that condition pronoun expression, these effects are highly important.

Since the interview and Facebook data sets are significant predictors in determining pronoun expression, the next step is to explore how each data set is conditioned by the other linguistic variables and in doing so test Hypothesis 3 that the two groups will show differences in their conditioning.

TABLE 4. FACTORS CONDITIONING 1ST PERSON SINGULAR SPE IN HERMOSILLO.

FACTOR GROUP	RANGE
Co-reference	21
Data source	19
Clause Type	14
Morphological ambiguity	9
Log like = -966.688	P < 0.05

TABLE 5. DATA SOURCE FACTOR GROUP FOR 1ST PERSON SINGULAR SPE IN HERMOSILLO.

	FW	%	N
Interview	0.60	16.7 %	202/1211
Facebook	0.41	10.2 %	138/1355
Log likelihood= -966.688			
p < 0.05			

Linguistic factors by data source

The comparison between the conditioning of subject pronoun expression in the interview and Facebook data can be seen in Table 6.

For the sociolinguistic interview data set, TMA, co-reference, and clause type are significant factors. These factors are in line with previous findings for interview data. The categories within the TMA factor show parallels to previous studies (eg., Lastra & Martín Butragueño, 2015 and Michnowicz, 2015) in those unambiguous forms like the periphrastic future (FW = 0.59) and the present indicative (FW = 0.53) favor expressed pronouns and ambiguous forms like the imperfect subjunctive (FW = 0.41), and the conditional (FW = 0.38) disfavors expressed pronouns. There are also deviations from the pattern in that some

unambiguous forms also favor expressed pronouns like the imperfect indicative (FW = 0.59) and the present subjunctive (FW = 0.56). However, the imperfect indicative was the verb form that most favored expressed pronouns in Veracruz (Orozco, 2016).

In terms of co-reference, oral Hermosillo Spanish again patterns like previously studied communities (Cameron, 1993; Travis, 2007; Lastra & Martín Butragueño, 2015 and Michnowicz, 2015). A complete switch with the previous subject highly favors an expressed pronoun (FW = 0.63), followed by a switch with the subject but the same referent as the object of the previous verb (FW = 0.51). In contrast the same subject disfavors expressed pronouns (FW = 0.41).

TABLE 6. COMPARISON OF THE LINGUISTIC FACTORS THAT CONDITION 1ST PERSON SINGULAR SPE IN HERMOSILLO.

SOCIOLINGUISTIC INTERVIEW					FACEBOOK				
TMA	FW	%	N	% DATA	CO-REFERENCE	FW	%	N	% DATA
Imperfect indicative	0.59	19.9	59/296	24.4	Same object	0.59	13.2	18/136	10.0
Periphrastic future	0.59	25.0	4/16	1.3	Complete switch	0.56	11.7	108/926	68.3
Present subjunctive	0.56	12.5	1/8	0.7	Same subject	0.29	4.1	12/293	21.6
Present indicative	0.53	18.9	95/503	41.5	Range = 30				
Preterit	0.41	12.1	38/314	25.9	TMA				
Imperfect subjunctive	0.41	8.3	1/12	1.0	Perfect (all)	[0.62]	15.4	6/39	2.9
Conditional	0.38	11.1	1/9	0.7	Imperfect indicative	[0.60]	14.3	11/77	5.7
Perfect (all)	0.25	5.7	3/53	4.4	Imperfect subjunctive	[0.54]	11.8	2/17	1.3
Morphological future	NA	NA	NA	NA	Present indicative	[0.50]	9.9	86/867	64.0
	Range = 34				Preterit	[0.50]	10.0	26/259	19.1
Co-reference*					Conditional	[0.50]	10.0	1/10	0.7
Complete switch	0.63	23.1	99/428	35.3	Present subjunctive	[0.43]	7.7	1/13	1.0

* Co-reference and parallelism were run in separate models since the two factor groups present a significant interaction. Both factors were found to be significant in their respective models, but since neither improved/worsened the model, co-reference was chosen as the factor to be reported in order to compare with previous studies.

SOCIOLINGUISTIC INTERVIEW					FACEBOOK				
Same object	0.51	17.0	24/141	11.6	Morphological future	[0.42]	7.5	4/53	3.9
Same subject	0.41	12.3	79/642	53.0	Periphrastic future	[0.32]	5.0	1/20	1.5
	Range = 22					Range = [30]			
Clause type					Clause type				
Main	0.56	19.9	75/377	31.1	Coordinate	[0.54]	11.8	14/119	8.8
Coordinate	0.52	16.4	98/596	49.2	Main	[0.52]	10.7	106/992	73.2
Subordinate relative	0.44	15.9	7/44	3.6	Subordinate other	[0.42]	7.4	15/203	15.0
Subordinate other	0.35	11.3	22/194	16.0	Subordinate relative	[0.41]	7.3	3/41	3.0
	Range = 21					Range = [13]			
Log like = -513.228, $p < 0.05$					Log like = -436.754, $p < 0.01$				

The last significant factor group for the sociolinguistic interview data is clause type. Previous research has shown that main clauses most favor an expressed pronoun while coordinate clauses highly disfavor expressed pronouns (Abreu, 2009; Flores-Ferrán, 2009; and Otheguy & Zentella, 2012). In line with previous studies, main clauses favor an expressed pronoun (FW = 0.56) in Hermosillo; however coordinate clauses do as well (FW = 0.52). This discrepancy may be due to the definition of what a main clause is versus a coordinate clause in informal speech data. Since it is difficult to know where one sentence begins, and another sentence ends in spontaneous speech, we consider a main clause one that is isolated, between long pauses, or one that has a subordinate clause. In comparison, coordinate clauses include any string of main clauses not separated by a long pause, even without a clearly present conjunction. In the sociolinguistic data, there is a considerable number of coordinate tokens (49.2 % of the data set). It may be that in other communities, there is a higher number of main clauses due either to coding differences or the nature of the interviews conducted.

While it is clear that oral Hermosillo Spanish, in general, patterns like previously studied communities, especially other Mexican varieties, there are stark differences between the sociolinguistic interview data set and the data collected from Facebook posts. First, only one factor is significant for the Facebook data set, co-reference, while TMA, co-reference, and clause type are significant for the interview data. This may be due to the distribution of tokens within the two data sets. For TMA, for example, in the oral data, there are three factors that have a

significant percentage of the data, the present indicative (41.5 %) imperfect, indicative (24.4 %), the preterit (25.9 %). While the present indicative is the category with the highest percent of the data, it is still less than half. On the other hand, in the Facebook data, only two categories have a significant percentage of the data, present indicative (64 %) and the preterit (19.1 %). The present indicative is used in well over half of the tokens. A similar discrepancy between the two data sources can be found for clause type, but to a higher degree. For the interview data, coordinate clauses have 49.2 % of the data, main clauses have 31.1 %, and other subordinate clauses have 16.0 %. However, for the Facebook data set, main clauses consist of 73.2 % of the data set, subordinate other only 15.0 %, and coordinate 8.8 %. While the data is more evenly distributed for the conversational, sociolinguistic interview, the nature of posts in Facebook lend for a very slanted distribution for the social media group. A large majority of the tokens come from original posts consisting of a sole main clause. It is not surprising, then, that a large portion of tokens for the Facebook data set are coded as main clauses and in the present indicative. This distribution alone is a significant difference between the interview data and the social media data on Facebook.

After discussing the differences in the factor groups that condition pronoun expression in the two groups, we now focus on each factor separately. First, co-reference, the only significant factor for the Facebook group, patterns similarly for the two data sets. For both the interview and the Facebook groups, a switch in subject favors expressed pronouns (interview, FW = 0.63; Facebook, FW = 0.56), while the same subject disfavors them (interview, FW = 0.41; Facebook, FW = 0.29). However, while in the sociolinguistic interview data, a complete switch most favors an expressed pronoun, a switch in the subject but same object is the category that most favors expressed pronouns (FW = 0.59). While TMA was not significant in the Facebook data, there are a couple of similarities with the interview data in terms of the hierarchy of the within-group factors. The imperfect indicative is the second-highest category favoring expressed pronouns (FW = 0.60) in the Facebook group and the highest (FW = 0.59) in the sociolinguistic interview group. The present indicative also slightly favors expressed pronouns in both data sets (interview, FW = 0.53; Facebook, FW = 0.50). The rest of the factors within TMA shows opposite trends. For example, where periphrastic future highly favors expressed pronouns in the interview data set (FW = 0.59), it highly disfavors them in the Facebook data set (FW = 0.32). Additionally, the imperfect subjunctive, conditional, and perfect TMAs all disfavor expressed pronouns for the sociolinguistic data but

favor them for the Facebook posts. In terms of clause type, on the other hand, the pattern is relatively similar for the two groups. Main and coordinate clauses favor expressed pronouns in both groups while subordinate relative and subordinate other disfavor expressed pronouns.

Based on the discussion above, it is seen that the prediction in Hypothesis 3 that there will be differences in the constraints that condition subject pronoun expression between the sociolinguistic interviewers and the Facebook posts is borne out in these data sets. It is also clear that the genre effects previously reported to affect the overall frequency of expressed pronouns (Solomon, 1999; Flores-Ferrán, 2002, 2010; and Travis, 2007) can also affect the factors that condition pronoun expression. The nature of social media posts, represented here in the Facebook data, also plays a key role, in addition to being a different genre, in that the majority of occurrences happen in an original post consisting of one main clause. While there are responses to posts that mimic conversational interactions, these are rarer (at least in this data set).

Conclusion

In this study, we examined first-person singular subject pronoun expression in Hermosillo, Sonora Spanish. We explored variation in social media in the form of Facebook posts. By comparing sociolinguistic interview data to Facebook posts, we found considerable differences in subject pronoun expression within the two data sources. First, Facebook posts showed a lower overall frequency of first-person expressed pronouns (10.2 %) than the sociolinguistic interview data (16.7 %). Second, using data source as a predictor variable, we determined that the difference in frequency is statistically significant and the second-ranked factor in the conditioning of subject pronoun expression in Hermosillo. Third, when separating the data-by-data source (interview vs. Facebook), we also found crucial differences in the factors that condition the variable based on the medium in which the data was produced. Namely, while the interview data is more equally distributed to different contexts, Facebook posts produce mostly main clauses and present tense tokens. Due to this fact, TMA, co-reference, and clause type were all significant factors for the sociolinguistic interview data set, but only co-reference was significant for the Facebook group.

The results presented in this study add to our previous knowledge of genre effects in the overall frequency of subject pronoun expression (Solomon, 1999; Flores-Ferrán, 2002, 2010; and Travis, 2007) by demonstrating that the effect

is also more structural. Furthermore, this study has provided more insights into the differences in written and oral data variation. We have also presented evidence further supporting the findings of Martínez (2007) that there are differences in pronoun frequency wherein more formal writing, more pronouns are expressed than in informal. By comparing data from written narratives in an academic setting among Sonoran monolinguals (Fernández Flórez 2019) to the more informal written data in the Facebook group of this study, we found that the formal written context showed a slightly higher rate of expression (12.0 %, Fernández Flórez, 2019) as compared to the informal Facebook context (10.2 %, this study). Additionally, the representation of data from Hermosillo sociolinguistic interview data and Facebook posts from Hermosillo speakers (verified through acquaintances of each participant) allowed for the direct comparison of the differences in the variation between social media and written data versus the interview/oral data. Future studies should consider working with social media data and other written sources to further our understanding of variation in written and social media contexts.

References

- Abreu, L. (2009). *Spanish subject personal pronoun use by monolinguals, bilinguals and second language learners* (Ph.D. dissertation). University of Florida.
- Alfaraz, G. (2015). "Variation of overt and null subject pronouns in the Spanish of Santo Domingo". In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (Eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal perspective* (pp. 3-16). Washington, D.C.: Georgetown University Press.
- Bayley, R., & Pease-Alvarez, L. (1997). "Null Pronoun Variation in Mexican-Descent Children's Narrative Discourse". *Language Variation and Change*, 9(3), pp. 349-71.
- Bessett, R. M. (2018). "Testing English Influence on First Person Singular 'Yo' Subject Pronoun Expression in Sonoran Spanish". In J. E. MacDonald (Ed.), *Issues in Contemporary Trends in Hispanic and Lusophone Linguistics* (pp. 355-72). Amsterdam: John Benjamins.
- Cameron, R. (1993). "Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain". *Language Variation and Change*, 5(3), pp. 305-334.
- Carvalho, A. M. & Bessett, R. M. (2015). "Subject Pronoun Expression in Spanish in Contact with Portuguese." In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (Eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal perspective* (pp. 143-165). Washington D.C.: Georgetown University Press.

- Carvalho, A. M., & Child, M. (2011). "Subject Pronoun Expression in a Variety of Spanish in Contact with Portuguese". In J. Michnowicz & Dodsworth, R. (Eds.), *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics* (pp. 14-25). Somerville, MA: Cascadilla Proceedings Project.
- Díaz-Campos, M. (2014). *Introducción a la sociolingüística hispánica*. Mladen, Oxford: Wiley Blackwell.
- Elizaincín, A. (1995). "Personal pronouns for inanimate entities in Uruguayan Spanish in contact with Portuguese". In C. Silva-Corvalán (Ed.), *Spanish in Four Continents: Studies in Language Contact and Bilingualism* (pp. 117-131). Georgetown: Georgetown University Press.
- Fernández Flórez, C. (2019). *Written Variation among Speakers of Spanish as their First, Second and Heritage Language: The Case of the Subject Pronoun Expression* (Doctoral dissertation). Tucson: University of Arizona.
- Flores-Ferrán, N. (2002). *Subject personal pronouns in Spanish narratives of Puerto Ricans in New York City: A sociolinguistic perspective*. Munich: Lincom Europa.
- Flores-Ferrán, N. (2004). "Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact?" *Language Variation and Change*, 16(1), pp. 49-73.
- Flores-Ferrán, N. (2009). "Are you referring to me? The variable use of UNO and YO in oral Discourse". *Journal of Pragmatics*, 41(9), pp. 1810-1824.
- Flores-Ferrán, N. (2010). "¡Tú no me hables! Pronoun expression in conflict narratives". *International Journal of Sociology of Language*, 2010(203), pp. 61-82.
- Gudmestad, A., House, L., & Geeslin, K. L. (2013). "What a Bayesian analysis can do for SLA: New tools for the sociolinguistic study of subject expression in L2 Spanish". *Language Learning*, 63(3), pp. 371-399.
- Lapidus, N., & Otheguy, R. (2005). "Overt non-specific *ellos* in Spanish in New York". *Spanish in Context*, 2(2), pp. 157-174.
- Lastra, Y., & Butragueño, P. M. (2015). "Subject pronoun expression in oral Mexican Spanish". *Subject pronoun expression in Spanish: A cross-dialectal perspective*, pp. 39-57.
- Manjón-Cabeza Cruz, A., Pose Furest, F. y Sánchez García, F. (2016). "Factores determinantes en la expresión del sujeto pronominal en el corpus PRESEEA de Granada". *Boletín de Filología*, LI(2), pp. 181-207.
- Martinez Mira, M. I. (2009). "Spanish heritage speakers in the Southwest: Factors contributing to the maintenance of the subjunctive in concessive clauses". *Spanish in Context*, 6(1), pp. 105-126.
- Martínez, G. (2007). "Writing back and forth: the interplay of form and situation in heritage language composition". *Language Teaching Research*, 11(1), pp. 31-41.

- Martínez-Lara, J.A., Guerrero González, S. y González Vergara, C. (2021). “(Yo) opino: la variación del sujeto pronominal en el habla de Santiago de Chile”. In B. M. A. Rogers & M. A. Figueroa Candia (Eds.), *Chilean Spanish Linguistics: Studies on variation, innovation, contact, and identity* (pp. 283-308). Malaga: Vernon Press.
- Michnowicz, J. (2015). “Subject pronoun expression in Yucatan Spanis”. In A. M. Carvalho, Orozco, R., & Lapidus Shin, N. (Eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal perspective* (pp. 101-120). Washington, D.C.: Georgetown University Press.
- Orozco, R. (2016). “Subject Pronoun Expression in Mexican Spanish: ¿Qué pasa en Xalapa?” *Proceedings of the Linguistic Society of America*, 1(7), pp. 1-15.
- Otheguy, R, Zentella, A. C., & Livert, D. (2007). “Language and dialect contact in Spanish of New York: Toward the formation of a speech community”. *Language*, 83(4), pp. 770-802.
- Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialect leveling, and structural continuity*. Oxford: Oxford University Press.
- Prada Pérez de, A. (2015). “First person singular subject pronoun expression in Spanish in contact with Catalan”. In A. M. Carvalho, R. Orozco, & N. Lapidus Shin (Eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal perspective* (pp. 121-142). Washington, D.C.: Georgetown University Press.
- Repede, D. (2019). “Condicionantes sociolingüísticos de la expresión del sujeto pronominal en el sociolecto alto del corpus PRESEEA-Sevilla”. *Revista de investigación lingüística*, 22, pp. 397-423.
- Roselló Verdeguer, J. (2021). “Factores determinantes en la expresión del sujeto pronominal en el español hablado de Valencia”. *Lingüística en la red*, 18(2), pp. 1-24.
- Soares da Silva, H. (2006). *O parâmetro do sujeito nulo: Confronto entre o português e o espanhol*. (PhD dissertation). Rio de Janeiro: Universidade Federal do Rio de Janeiro.
- Solomon, J. (1999). *Phonological and syntactic variation in the Spanish of Valladolid Yucatán*. (Ph.D. dissertation). Palo Alto: Stanford University.
- Torres-Cacoulllos, R., & Travis, C. (2010a). “Variable *yo* expression in New Mexico: English influence?” In S. Rivera-Mills & D. Villa Crésap (Eds.), *Spanish of the U.S. southwest: A language in transition* (189-210). Madrid: Iberoamericana/Vervuert.
- Torres Cacoulllos, R., & Travis, C. (2010b). “Testing convergence via code-switching: priming and the structure of variable subject expression”. *International Journal of Bilingualism*, 15(3), pp. 1-27.

- Travis, C. E. (2007). "Genre effects on subject expression in Spanish: Priming in narrative and Conversation". *Language Variation and Change*, 19(2), pp. 101-135.
- Yus, F. (2002). "El chat como doble filtro comunicativo". *Revista de Investigación Lingüística*, 5(2), pp. 141-169.
- Yus, F. (2003). "La lengua de los medios en la red y su interpretación". In J. Díaz Noci & R. Salverría (Eds.), *Manual de redacción ciberperiodística* (pp. 309-352). Barcelona: Ariel.