Language Variation and Change, **26** (2014), 303–330.

doi:10.1017/S095439451400012X

Grammatical complexification in Spanish in New York: 3sg pronoun expression and verbal ambiguity

NAOMI LAPIDUS SHIN
University of New Mexico

ABSTRACT

This study provides evidence of grammatical complexification, operationalized as the emergence of a significant linguistic constraint on the use of a linguistic structure, in Spanish spoken in New York City. Analyses of 4276 third-person singular verbs produced in sociolinguistic interviews with first-generation Latin American immigrants and second-generation US-born Latinos demonstrate that third-person singular subject pronoun expression (*ella canta* ~ *canta* 'she sings ~ sings') is constrained by tense/mood/aspect in the second generation, but not the first. Further analysis shows that this effect reflects a strategy aimed at clear referent identification. I conclude by suggesting that the increase in attention to ambiguous verb morphology, that is, the complexification, is related to other, concomitant changes in pronoun expression patterns in Spanish in New York City.

Both simplification and complexification of grammar are well documented in situations of language contact, but studies of minority languages being displaced by majority languages usually focus on the former. This study provides evidence that complexification accompanies simplification in Spanish spoken by bilingual Latinos in the United States, where English typically displaces Spanish within three generations (Rivera-Mills, 2012; Veltman, 1990; Zentella, 1997). Comparisons of first-generation Latin American immigrants to second-generation (US-born) Latinos show complexification of third-person singular (3sg) subject pronoun use. Analyses of 4276 3sg verbs demonstrate that second-generation Latinos are sensitive to tense/mood/aspect (TMA) verb morphology as a predictor of the presence versus absence of *él/ella* ('he'/'she'), whereas newly arrived immigrants are not. More specifically, for second-generation Latinos, imperfect verbs promote *él/ella* expression, while preterit verbs promote their absence.

Why does TMA influence second-generation speakers' *él/ella* expression? I consider but reject an explanation having to do with the backgrounding/foregrounding functions associated with the imperfect and preterit. Silva-Corvalán (2001:161–163) proposed that subject expression is compatible

I am grateful to Ricardo Otheguy and Ana Celia Zentella for allowing me to use data from their corpus. I also thank the following people for their for their excellent feedback and suggestions: Bill Croft, Danny Erker, Chris Koops, Amy Lindstrom, Ricardo Otheguy, Rosa Vallejos Yopán, Jackelyn Van Buren, Damián Vergara Wilson, as well as six anonymous reviewers and the editorial staff of this journal. Any remaining errors are mine alone.

with the backgrounding function of the imperfect, but not the foregrounding function of the preterit. Analyses of *él/ella* expression lend no support to Silva-Corvalán's position. Instead this study finds evidence in favor of a functional explanation for the *TMA effect*: second-generation speakers express *él/ella* with imperfect verbs especially in contexts where reference-tracking is hindered by other factors, such as a switch in reference or competing referents in the discourse. Thus the new tendency of the second generation to use *él/ella* with the imperfect more than with the preterit is related to the greater potential for referent ambiguity with the former than with the latter and, as such, is a type of functional compensation. More broadly, this study suggests that functional explanations of linguistic variation and change, despite long-standing skepticism (e.g., Labov 1994:547–568), continue to deserve serious consideration.

Once it is clear that the TMA effect is primarily an ambiguity avoidance strategy that has arisen in the second generation, I ask why it should be these speakers who develop this strategy. I tentatively suggest that the increased attention to the imperfect (more ambiguity) versus the preterit (less ambiguity) is related to a concomitant decrease in attention to other constraints on pronoun expression in Spanish in New York City (NYC). The broad explanation pursued is that the interrelated changes are part of a larger balancing act whereby bilinguals strive to maintain communicative efficiency.

DEFINING GRAMMATICAL COMPLEXIFICATION

There are various ways in which grammars can become more or less complex. One way is through changes that affect categorical phenomena: increases or decreases in complexity can result from the addition or loss of linguistic structures or categories (e.g., Ferguson, 1982:59–63; Trudgill, 2011:15–32), as in the incorporation of case-marking into the Arawak language Tariana as a result of contact with Tucano (Aikhenvald, 2003:3). A second way in which grammars can become more or less complex is through changes that affect variable phenomena. To detect this latter type of change, patterns of structured variation are compared across groups of speakers (e.g., Poplack & Levey, 2010; Poplack & Meechan, 1998; Tagliamonte, 2012:163ff; Torres Cacoullos & Travis, 2011).[1]

How do we determine whether changes in patterns of structured variation are instances of complexification? An answer can be derived from Dahl's (2004:24) method of measuring linguistic complexity: the "length of the specification [or description] . . . of the totality of patterns it contains." Patterns that take longer to describe are more complex than patterns that are shorter to describe. Dahl's (2004:46–48) analogy of a three-course dinner elucidates this idea. Like variation in grammar, the set of dinner choices is probabilistic and predictable, as choices may be influenced by factors such as hunger and health. If I am hungry, I might choose steak as my second course, which may in turn prompt me to choose a light appetizer. Pattern complexity increases with more choices and with more factors impacting the choices (hunger, diet, etc.). As choices and factors increase

in number, so will the length of the description required to account for the dinner choices.

Returning to grammar, relative complexity in grammatical patterns can be assessed as follows: (a) the more choices of variants, the more complex the pattern; and (b) the more factors constraining the choice between variants, the more complex the pattern. Thus, I posit the following operationalization of simplification and complexification:

> The loss of a linguistic factor that constrains linguistic choice is a type of simplification, while the emergence of a new factor is a type of complexification.

For example, assume that the choice to express or omit a subject pronoun is conditioned by factors *x* and *y*, but not *z* among first-generation US Spanish speakers, while this choice is conditioned by *x, y,* and *z* among second-generation speakers. In this case, second-generation Spanish has a grammatical constraint (*z*) that was absent in the Spanish spoken by the first generation. This operationalization of complexification fits well with Dahl's (2004) definition because the gain of *z* increases the length of the description of the pattern ("conditioned by *x, y,* and *z*" as opposed to "conditioned by *x* and *y*").

Complexification, as just defined, pertains to linguistic constraints guiding variation. So, if second-generation Spanish includes constraint *x,* but first-generation Spanish does not, then this is one way that the former is more complex than the latter. But, at the same time, second-generation Spanish can be lacking a constraint that is present in first-generation Spanish. That is, second-generation Spanish can be more complex than first-generation Spanish with respect to one constraint and, at the same time, be simpler with respect to a different constraint. In other words, grammatical simplification and complexification can proceed in tandem. In the current article, I assume this to be the case, that is, that the complexification in *él/ella* expression manifested as the emergence of sensitivity to TMA accompanies (and perhaps is triggered by) simplification of other constraints on *él/ella* expression.

VARIATIONIST RESEARCH ON SPANISH SUBJECT PRONOUN USE

The extensive sociolinguistic research on Spanish subject pronoun expression provides a wealth of information regarding patterns of use across varieties spoken in Latin America, Spain, and the United States. Six variables that constrain pronoun use are: reference, priming, semantic class, clause type, reflexive, and TMA.[2]

Reference: Pronouns are expressed more often when the referent of two consecutive grammatical subjects in discourse is different (switch-reference) than when the subjects share the same referent (same-reference) (e.g., Cameron, 1993, 1995;

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

Carvalho & Bessett, in press; Holmquist, 2012; Otheguy & Zentella, 2012; Shin & Otheguy, 2009; Silva-Corvalán, 1994; Torres Cacoullos & Travis, 2011).

Priming: Subject pronouns are more likely to be expressed when the previous mention of the referent also appears as an expressed subject pronoun. The result is a clustering of the [pronoun + verb] construction, illustrated by (1).

(1) *él se hace . . . él se mete en su dieta él mismo, no . . . él no toma cerveza.* [311C][3]
'he makes himself . . . . . he puts himself on his diet, no . . . . . he doesn't drink beer.'

The finding that pronoun use triggers further pronoun use has been interpreted as a priming or perseveration effect (Cameron & Flores-Ferrán, 2004; Carvalho & Bessett, in press; Torres Cacoullos & Travis, 2011; Travis, 2007).

Semantic class of verb: Pronoun rates tend to be highest with verbs of cognition, such as *creer* 'to believe,' and least common with nonmental activity verbs (Bentivoglio, 1987:52, 60; Enríquez, 1984; Erker & Guy, 2012:541; Flores-Ferrán, 2002; Otheguy & Zentella, 2012; Otheguy, Zentella, & Livert, 2007; Posio, 2011; de Prada Pérez, in press; Silva-Corvalán, 1994; Torres Cacoullos & Travis, 2011:252; Travis, 2007).

Clause type: Several studies have found that pronoun expression is most common in main clauses, less common in subordinate clauses, and least common in coordinate clauses (Abreu, 2009:125; Enríquez, 1984:256-258, Otheguy & Zentella, 2012).

Reflexive: Pronoun omission is likelier with verbs that occur with a reflexive pronoun. For instance, *yo* is less likely to occur with a verb such as *me baño* 'I bathe myself' than with a verb such as *bailo* 'I dance' (Bayley & Pease-Álvarez, 1997; Michnowicz, in press; Otheguy & Zentella, 2012).

Tense/mood/aspect morphology: Certain verb forms, such as the imperfect (imperfective past), increase the likelihood of pronoun expression, whereas other forms, such as the preterit (perfective past) decrease it. Consider the singular forms of the three most common verb paradigms, illustrated with the verb *bailar* 'to dance'.

As shown in Table 1, *bailaba* can be first-person singular (1sg) or a 3sg, but no such ambiguity exists in the simple present or the preterit.[4] Numerous studies have found that morphologically ambiguous forms as in the imperfect promote pronoun expression, whereas unambiguous forms as in the preterit do not (Abreu, 2009; Bentivoglio, 1987:45; Carvalho & Bessett, in press; Flores-Ferrán, 2002; Hochberg, 1986; Hurtado, 2005; Lastra & Martín Butragueño, in press; Michnowicz, in press; Orozco, in press; Otheguy & Zentella, 2012; de Prada Pérez,

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

TABLE 1. *1sg, 2sg, and 3sg forms in the simple present, preterit, and imperfect indicative*

Person	Simple Present	Preterit	Imperfect
1sg	bailo	bailé	bailaba
2sg	bailas	bailaste	bailabas
3sg	baila	bailó	bailaba

in press; Travis, 2007). There is also evidence that this TMA effect is significant in switch-reference but not same-reference contexts (Cameron, 1994:34–37). Previous research, however, has not focused on one grammatical person in particular (except Travis, 2007; see also Torres Cacoullos & Travis, 2011, who study 1sg), thus it is unknown how TMA may affect the different grammatical persons.

Changes in Spanish subject pronoun expression in New York City

Previous studies based on the Otheguy-Zentella Corpus of Spanish in NYC[5] have found evidence of intergenerational change in subject pronoun use (e.g., Otheguy & Zentella, 2012; Otheguy et al., 2007; Shin & Otheguy, 2013). The corpus consists of sociolinguistic interviews conducted with 140 adult Spanish speakers in NYC, selected based on various criteria including regional origin (Mainland Latin America vs. Caribbean), gender, and social class. To study change, two groups of speakers are compared: (i) newly arrived immigrants from Latin America, called "newcomers," and (ii) Latinos raised in NYC, called "New-York Raised" or NYRs. Comparisons of newcomers and NYRs show increasing use of pronouns reflected in overall rates of occurrence, as well as weakening of grammatical constraints on pronoun use. Otheguy and Zentella (2012:151–177, 185–190) find that reference, clause type, and semantic class exert a weaker effect among NYRs than among newcomers, leading them to conclude that "the picture that emerges is one where newcomers have transmitted intact large portions of their grammar to the NYRs, who have nevertheless changed some of it by keeping but weakening some of the principles that guided usage in the previous generation" (2012:177).

While some variables guiding pronoun use undergo weakening, Otheguy and Zentella's (2012:164, 186) study suggested a strengthening of TMA, as the *range*[6] between the imperfect and the imperative (the top and bottom TMA factors) is greater among NYRs than among newcomers. However, Otheguy and Zentella's (2012) approach of grouping together pronouns that differ in person/number obscures differences between newcomers and NYRs in terms of significance of predictor variables. This establishes the need for additional studies focusing on each of the grammatical persons (see also Travis & Torres Cacoullos, 2012:713). By analyzing 3sg pronouns separately from the other pronouns, the current study uncovers new evidence showing that NYRs, but not newcomers, are sensitive to TMA verb morphology as a predictor of *él/ella* expression.

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

METHODOLOGY

The data for the present study come from the Otheguy-Zentella corpus and focus on the two polar opposite types of participants with respect to amount of exposure to life in NYC: recent arrivals to NYC, the newcomers; and US-born and raised Latinos, NYRs. The specific criteria for being considered a newcomer or NYR are as follows. Newcomers are immigrants who arrived at a linguistically mature age (17 or older), who had been in the United States for a short time when the interview was conducted, and whose speech, therefore, most closely resembles that of the country of origin. NYRs, on the other hand, were either born in New York or arrived by age 3. In the corpus there are 39 newcomers and 26 NYRs, resulting in a total of 65 participants,[7] ages 17 to 73. Differences between newcomers and NYRs may be considered "apparent-time" differences, which are indicators of diachronic change (Chambers, 2004:355–363). In the current article, intergenerational change refers to apparent-time changes that occur over the course of one generation, that is, between the newcomer and NYR stages.

As mentioned earlier, the corpus is a balanced sample in terms of both dialect and NYC generation. Of the 39 newcomers, 19 hail from Caribbean countries and 20 from Mainland Latin America. Of the 26 NYRs, exactly half of them have roots in the Caribbean; the other half have roots in the Mainland. As I will demonstrate later, regional differences do not play a role in the primary change explored in the current study.

The envelope of variation for this study includes tokens of all tensed 3sg verbs in the corpus that occur with a subject pronoun (either *él* or *ella*) but could have occurred without one, or that are found without *él* or *ella* but could have been found with one of these pronouns. In other words, only *variable* contexts are studied, that is, contexts where both presence and absence of a pronoun are possible. To illustrate, consider example (2), produced by a Colombian newcomer. The relevant contexts are underlined. The instance where a pronoun could have occurred but did not is signaled by the symbol Ø in the Spanish original and by parentheses in the English translation.

(2) *Aquí en Estados Unidos en un tren yo noté que una persona, un señor se acercaba mucho a mí, Ø me empujaba, Ø me empujaba, cuando yo sentí el ruido de la cremallera en mi cartera que Ø estaba corriendo, Ø volteé a mirar y claro el señor estaba abriendo mi cartera.* [021C]
'Here in the United States on a train I noticed that a person, a man got really close to me, (he) pushed me, (he) pushed me, when I heard the sound of the zipper of my purse that (he) was pulling back, (I) turned around to look and of course the man was opening my purse.

All underlined verbs in example (2) are sites of variation. For example, the speaker produces the 1sg pronoun *yo* with *noté*, but she could have omitted *yo*, saying instead *Ø noté que una persona.* Similarly, when the speaker says *Ø me empujaba* '(he) was pushing me,' she does not insert the pronoun *él*, but could

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

have done so. When repeating *me empujaba*, the speaker could have opted to include *él*. While *él me empujaba, él me empujaba* may strike some readers as redundant, there is no categorical rule prohibiting the use of *él* in either occurrence of *me empujaba* (Torres Cacoullos & Travis, 2011; Travis, 2007). In all contexts underlined in example (2), the speaker could have opted to include or omit the subject pronoun.[8]

The six predictors of *él/ella* expression investigated in this study are: TMA, reference, priming, clause type, semantic class of verb, and reflexive.[9] *TMA* includes three factors—simple present, preterit, and imperfect—all in the indicative mood, because 90% of the 3sg verbs in the corpus are conjugated in these three forms. *Reference* is defined here as whether the referent of two consecutive grammatical subjects is the same (same-reference) or different (switch-reference). *Priming* is operationalized following Travis (2007) and Torres Cacoullos and Travis (2011): For each verb included in the study, I looked for the nearest previous mention of that verb's subject. If the previous mention occurred (a) in subject position, (b) as present or absent *él/ella*, (c) within a context of variation, and (d) within the previous six clauses, then the verb was coded as having a previous mention that was either present or absent *él/ella*. Note that all the criteria (a to d) had to obtain for a previous mention to be coded as eligible for priming. In contrast, if there was no previous mention of the referent or there was a previous mention of the referent but it occurred as a lexical noun phrase, in object position, or in any context in which there is little to no variation between expression and omission of pronouns, then the verb under study was coded as "other." In summary, the priming variable included three factors: previous mention = present *él/ella*, absent *él/ella*, or other.[10]

Clause type includes three broad categories: main, subordinate, and coordinate. *Semantic class* includes (a) verbs that describe mental activities, such as, *pensar* 'to think'; (b) stative verbs, such as *estar* 'to be'; and (c) verbs that describe nonmental or "external" activities, such as, *escribir* 'to write.' The six predictors are summarized in Table 2.

Data and statistical methodology

The dataset consists of 4276 present, preterit, or imperfect 3sg verbs that occur with or without *él/ella*, 2275 of which were produced by newcomers and 2001 by NYRs.

TABLE 2. *Predictor variables included in the present study, along with their factors*

Variables	Factors
1. TMA	(3) Present, preterit, imperfect (all indicative mood)
2. Reference	(2) Same, switch
3. Priming	(3) Previous mention = present *él/ella*, previous mention = absent *él/ella*, other
4. Clause	(3) Main, subordinate, coordinate
5. Semantic	(3) Mental/estimative, stative, or external activity
6. Reflexive	(2) Verb occurs with or without reflexive pronoun

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

TABLE 3. *Variables constraining* él/ella *expression, newcomers and NYRs, fixed effects*

Newcomers			NYRs		
Variable	**Wald**	***p***	**Variable**	**Wald**	***p***
Reference	112.48	<.0001	Priming	93.24	<.0001
Priming	94.86	<.0001	Reference	81.93	<.0001
Clause	15.17	<.0001	TMA	13.58	.001
[Reflexive]	5.57	.02	Clause	11.79	.003
[Semantic]	2.94	.23	[Reflexive]	1.23	.27
[TMA]	.11	.97	[Semantic]	.13	.94
Verbs, *N*	2275		Verbs, *N*	2001	
Nagelkerke R^2	.15		Nagelkerke R^2	.15	
% Correct (% baseline)	69.0 (66.8)		% Correct (% baseline)	64.1 (53.5)	
Deviance	2635.94		Deviance	2531.35	
Intercept	−.77		Intercept	−.15	
Centered input probability	.32		Centered input probability	.46	

Notes: Variables included in regressions: priming, reference, TMA, clause, semantic class, reflexive, reference by TMA, reference by clause.
Variables for which the *p*-value is greater than .01 appear in brackets.

As the focus here is grammatical patterning, rather than overall rates of pronoun use, the primary methodology is multivariate analysis from which variable and constraint hierarchies are constructed. Those hierarchies are compared across groups of people, in this case, newcomers and NYRs in NYC. Logistic regression models that included fixed factors only (the six independent variables), run in both SPSS and Rbrul (Johnson, 2009a, 2009b), were compared with mixed effects models run in Rbrul including two random factors—individual speaker and verb lexeme. The comparison between the fixed-effects and mixed-effects models, which is reported in Appendix A, shows that there were no major differences between the two models. Results from the regressions that included fixed factors only are presented in Tables 3 and 4.

RESULTS: THE IMPACT OF SIX PREDICTORS ON *ÉL/ELLA* EXPRESSION

Tables 3 and 4 present variable and constraint hierarchies based on two logistic regressions[11]—one for newcomers and one for NYRs—investigating the influence of reference, priming, clause, semantic class, reflexive, and TMA on 3sg pronoun expression. I use *variable hierarchies* to refer to the rankings of predictor variables such as reference or TMA, and *constraint hierarchies* to refer to the rankings of factors within each variable, such as main clause or preterit tense.[12]

As to interactions between the predictor variables, for both groups of speakers, reference by clause was significant (newcomers: $p < .0001$; NYRs: $p = .001$) and reference by TMA approached significance among NYRs for preterit ($p = .066$) and imperfect verbs ($p = .066$). Because previous research indicates that the

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

TABLE 4. *Factors constraining* él/ella *expression, newcomers and NYRs, fixed effects only*

	Newcomers					NYRs				
Variable	FW	Log Odds	*p*	*n*	*% él/ella*	FW	Log Odds	*p*	*n*	*% él/ella*
Priming						**Priming**				
él/ella >	.65	.62	<.0001	507	47	.64	.59	<.0001	642	59
Other	.49	−.05	.47	741	38	.49	−.04	.62	647	50
Ø >	.36	−.57	<.0001	1027	23	.37	−.55	<.0001	712	32
Range 29						*27*				
Reference						**Reference**				
Switch	.64	.58	<.0001	918	46	.63	.52	<.0001	886	58
Same	.36	−.58	<.0001	1357	24	.37	−.52	<.0001	1115	37
Range 28						*26*				
Clause						**Clause**				
Main	.56	.25	<.0001	1198	38	.56	.23	.001	1154	50
Subordinate	.47	−.13	.11	469	31	.46	−.15	.07	392	38
Coordinate	.47	−.12	.13	608	26	.48	−.07	.39	455	44
Range 9						*10*				
[Reflexive]						**[Reflexive]**				
Nonreflex	.55	.19	.02	2046	34	.52	.10	.28	1819	47
Reflexive	.45	−.19	.02	229	28	.48	−.10	.28	182	40
Range 10						*4*				
[TMA]						**TMA**				
Imperfect	.50	−.02	.97	511	34	.56	.25	.001	494	51
Present	.50	−.03	.75	1113	33	.50	−.02	.91	990	47
Preterit	.50	−.01	.97	651	33	.44	−.23	.001	517	41
Range 0						*12*				
[Semantic class]						**[Semantic class]**				
Mental	.55	.20	.19	154	38	.49	−.04	.95	153	45
Stative	.48	−.09	.71	858	34	.51	.03	.85	807	48
External	.47	−.12	.19	1263	32	.50	.01	.97	1041	45
Range 8						*2*				

Note: Variables for which the *p*-value is greater than .01 appear in brackets.

TMA effect impacts switch-reference but not same-reference contexts (Cameron, 1994), the reference by TMA variable was included in the regressions. Another issue that was taken into consideration was the possibility of multicollinearity, which does not appear to be a problem in the current study, as demonstrated by the measures in Appendix B.

The two variable hierarchies presented in Table 3 are constructed based on logistic regressions run in SPSS and Rbrul. SPSS provides a Wald statistic for each variable, which tests the significance of individual coefficients (Menard, 2002:43, Otheguy & Zentella, 2012:158–161). In Table 3, the variables are ranked in order of highest to lowest Wald value. Statistical significance values are listed under the column heading *p*. The significance cutoff value for these analyses is .01 due to the Bonferroni adjustment needed when including six predictor variables (Johnson, 2009a:363). Reported at the bottom of the table are (a) Nagelkerke R^2, which indicates the amount of variability explained by the logistic regression model as a whole; (b) % correct (baseline), which is the percentage of correctly predicted cases as compared with the baseline prediction and indicates the predictive efficiency of the model; (c) deviance, which measures how well the model fits the data (the smaller the deviance, the better the fit); (d) intercept, which provides a baseline for building the model predictions and is similar to the input value provided by Goldvarb; and (e) centered input probability, which is the inverse logit of the model intercept and reports the overall prediction of the model. (For information regarding measures a and b, see Menard, 2002:18–41, and Szmrecsanyi, 2005:119; for measures c, d, and e, see Johnson, 2009b, and Tagliamonte, 2012:140–143).

The primary point to take away from Table 3 is that TMA predicts *él/ella* expression among NYRs but not newcomers. At the same time, the speaker groups are similar with respect to the other five variables: *él/ella* is conditioned most strongly by priming/reference, followed by clause (and reference interacts with clause for both groups). Reflexive does not reach significance, and semantic class does not condition *él/ella* for either group.[13] The speaker groups are also similar in terms of how much variability is predicted by the regression model, approximately 15% for each group, per the Nagelkerke R^2 values. The deviance and centered input probability values indicate that the model is a better fit for the NYRs than for the newcomers. To summarize, the hierarchies in Table 3 tell us that newcomers and NYRs are similar, because several variables are significant for both groups. At the same time, there is also evidence for complexification because TMA is significant among NYRs but not among newcomers.

Table 4 presents the results from the same regressions reported in Table 3, but this time with more detailed information for each factor. Each variable is listed in boldface with its factors underneath. For example, priming includes three factors: (i) previous mention of referent was subject pronoun *él* or *ella,* which is denoted by 'él/ella >'; (ii) previous mention was an omitted subject pronoun, denoted by 'Ø >'; and (iii) other, which were cases not considered relevant for priming. Also included in Table 4 are factor weights, log odds, *p*-values, and range. *Factor weights* (FW) indicate the probability that the dependent variable

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

will occur. The closer to 1.0, the greater the probability that *él/ella* will be expressed; the closer to 0, the greater the probability that *él/ella* will be omitted. *Log odds* are coefficients that, like FWs, express the probability that the dependent variable will occur. Positive log odds indicate that a factor favors *él/ella* expression; negative log odds indicate that a factor favors omission. The *p*-value(s) indicate the statistical significance associated with each factor. *Range* is calculated by subtracting the lowest factor weight from the highest. Generally speaking, the greater the range, the stronger the variable (Tagliamonte, 2012:123).

The constraint hierarchies in Table 4, like the variable hierarchies in Table 3, show similarities between newcomers and NYRs in terms of how priming, reference, and clause impact *él/ella* expression. For both groups, the same factors are significant and the direction of the effects is the same. Expressed *él/ella* promotes a subsequent expressed *él/ella*, while omitted *él/ella* promotes a subsequent omitted *él/ella*, which can be interpreted as evidence of priming. Also, for both groups, switch-reference promotes *él/ella* expression, same-reference promotes *él/ella* omission, and main clauses promote *él/ella* expression. Nevertheless, and of primary importance for the current article, the results in Table 4 underscore the evidence for complexification: NYRs are sensitive to TMA, while newcomers are not. Moreover, the results pinpoint the complexification: NYRs' *él/ella* expression is constrained by TMA-imperfect and TMA-preterit, while for newcomers, neither imperfect nor preterit verbs significantly predict *él/ella* expression. This difference between newcomers and NYRs is also clear in the rates of *él/ella* expression in each TMA category. Newcomers' rates are nearly identical across TMA categories (34%, 33%, 33% for imperfect, present, and preterit, respectively; [X^2 (1,2) = .12, p = .94 (ns)]), whereas NYRs express significantly higher *él/ella* rates with imperfect (51%) than with present (47%) or preterit (41%) verbs (X^2 (1,2) = 10.67, p = .005).

A note on regional differences and Spanish subject pronoun expression

In order to investigate the possibility that the differences between newcomers and NYRs reported herein are due to dialect differences instead of intergenerational changes, the participant groups were divided into two dialect regions—the Caribbean (Cuban, Dominican, Puerto Rican) and the Latin American Mainland (Colombian, Ecuadorian, Mexican). This division is based on the well-known fact that subject pronoun expression rates are higher in the Caribbean than elsewhere, a trend that remains significant in NYC (Otheguy & Zentella, 2012:178–199). The same regressions were run as before, one for newcomers and one for NYRs, but this time with region added as a seventh independent variable. Region was not significant for either group (newcomers: p = .64; NYRs: p = .24). Next the data were divided into four sets: Caribbean newcomers, Mainlander newcomers, Caribbean NYRs, and Mainlander NYRs. Bivariate analyses show that neither Caribbean nor Mainlander newcomers differentiated between imperfect and preterit verbs with respect to *él/ella* expression (Caribbean newcomers: imperfect

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

31%, preterit 31%, X^2 (1,1) = .001, p = .97; Mainlander newcomers: imperfect 36%, preterit 34%, X^2 (1,1) = .26, p = .61). In contrast, both Caribbean and Mainlander NYRs expressed significantly higher rates of *él/ella* in the imperfect than in the preterit (Caribbean NYRs: imperfect 52%, preterit 42%, X^2 (1,1) = 5.10, p = .02; Mainlander NYRs: imperfect 51%, preterit: 41%, X^2 (1,1) = 5.36, p = .02). Thus the differences between newcomers and NYRs outlined in the current article are not an artifact of regional differences and instead reflect intergenerational change in Spanish in NYC.

Summary of results: The influence six predictors of él/ella expression

By focusing on 3sg pronoun expression, the results uncover a heretofore unreported apparent time change: TMA significantly constrains NYRs', but not newcomers' 3sg pronoun use. In particular, NYRs are sensitive to the imperfect as a predictor of *él/ella* expression and to the preterit as a predictor of *él/ella* omission. I will further explore these results in an attempt to gain purchase on this emerging sensitivity to TMA.

EXPLAINING THE EMERGING TMA EFFECT IN SPANISH IN NYC

Before exploring reasons for why complexification arises in the grammar of second-generation speakers, we must first understand the TMA effect itself. Why do speakers choose to express pronouns with verbs in the imperfect and to omit pronouns with verbs in the preterit? There are currently two explanations in the literature, one focusing on the avoidance of ambiguity and the other on the foregrounding/backgrounding functions of the verb tenses. In the following sections, I discuss each of these explanations.

TMA effect as a strategy to avoid ambiguous reference

The most common explanation for the TMA effect on pronoun usage focuses on verb ambiguity; pronouns are expressed more often with the imperfect than with the preterit due to the greater amount of morphological ambiguity in the former than in the latter (see Table 1). There are two lines of evidence indicating that morphological ambiguity triggers pronoun expression primarily in contexts where reference-tracking is hindered by other variables. The first is that the TMA effect applies in contexts of switch- reference but not same-reference. The second is that the tendency to express *él/ella* is related to contextual ambiguity and the potential for such ambiguity is greater with imperfect verbs.

The first line of evidence that the TMA effect serves to disambiguate reference comes from the link between TMA and reference. As mentioned earlier, Cameron (1994:34–37) found that the TMA effect was significant in switch-reference but not same-reference contexts. In the current study, TMA interacts with reference

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

TABLE 5. *Rates expressed* él/ella *by TMA, same-reference and switch-reference contexts*

TMA	Newcomers				NYRs			
	Same		Switch		Same		Switch	
	n	%	*n*	%	*n*	%	*n*	%
Preterit	374	23	277	46	271	33	246	51
Present	682	25	431	46	559	39	431	57
Imperfect	301	23	210	49	285	39	209	69
Total	1357		918		1115	37	886	58

among the NYRs in the same way: in the NYRs' regression analysis, the TMA*reference variable approached significance for preterit ($p = .066$) and imperfect verbs ($p = .066$). The interaction is more clearly illustrated by bivariate analyses comparing same-reference and switch-reference contexts, which are presented in Table 5.

Table 5 shows that NYRs are highly sensitive to TMA for switch-reference (X^2 (1,2) = 15.57, $p < .0001$), but not same-reference (X^2 (1,2) = 3.46, $p = .18$ [ns]). In switch-reference contexts, NYRs' TMA effect is robust: the *él/ella* expression rate reaches 69% with imperfect verbs, a rate that is 18 percentage points higher than the rate with preterit verbs. In contrast, TMA did not impact newcomers' use of *él/ella*, even when separating same-reference (X^2 (1,2) = .82, $p = .66$ [ns]) from switch-reference (X^2 (1,2) = .53, $p = .77$) contexts. The finding that TMA matters especially when switching reference tells us that NYRs pay attention to TMA when it is helpful to do so; when speakers switch reference, the referent becomes less accessible and more information must be encoded in the subject noun phrase in order to identify the referent (e.g., Ariel, 1990; Diver, 2011:257).

The second line of evidence showing that the TMA effect is related to avoiding ambiguity comes from research showing that speakers increase their use of pronouns in contexts of "contextual ambiguity," that is, when the discourse context hinders reference-tracking due to, for example, competing referents (Hurtado, 2005:344–346; Ranson, 1991:140–148). To illustrate contextual ambiguity, consider example (3), in which a Colombian woman describes how she learned English upon arriving in the United States at the age of 3. Example (3) has been altered from the original version; all subject pronouns have been removed. The original version, in which the subject pronouns have been reinserted, is presented in (4). Relevant contexts are underlined and the symbol "?" is used in the English version to denote ambiguous reference.

(3) *No . . . [a]Ø no tuve mucho . . . ah . . . problema con el inglés, porque como mi hermano . . . [b]Ø tenía un hermano mayor, y eh . . . E [c]Ø iba a la casa y [d]Ø hablaba y así mismo [e]Ø me enseñaba el inglés. Más, [f]Ø hablaba con el otro hermano mío, entonces [g]Ø aprendí el inglés, más o menos.*
No . . . [a](I) didn't have much . . . ah . . . problem with English because my brother . . . [b](I) had an older brother, and uh, . . . And [c](?) went to the house, and [d](?) talked and

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

just like that [e](?) taught me/myself English. Actually, [f](?) talked with my other brother, so [g](I) learned English, more or less.

The morphology of some pronounless verbs in (3) uniquely marks the person and number of the referent. The verbs in contexts a and g are unambiguous because they are morphologically marked for 1sg. But c, d, e, and f are both morphologically and contextually ambiguous.[14] Who "went to the house, and talked"? Who "taught" the participant English? Who "talked with [her] other brother"? Ambiguity arises primarily because there are competing referents (the speaker and her older brother) for the imperfect verbs in c, d, e, and f. Now consider (4), which is what the speaker really said.

(4) *No . . . [a]Yo no tuve mucho . . . ah . . . problema con el inglés, porque como mi hermano . . . [b]yo tenía un hermano mayor, y eh . . . . Y [c]él iba a la casa y [d]él hablaba y así mismo [e]me enseñaba el inglés. Más, [f]él hablaba con el otro hermano mío, entonces [g]yo aprendí el inglés, más o menos.*
No . . . [a]I didn't have much . . . ah . . . problem with English because my brother . . . [b]I had an older brother, and uh, . . . And [c]he went to the house, and [d]he talked and just like that [e](I? he?) taught me/myself English. Actually, [f]he talked with my other brother, so [g](I) learned English, more or less.

In (4), the inclusion of *él* identifies the older brother as the subject of contexts c, d, and f.[15] By saying *él*, the speaker increases the likelihood of communicative efficacy.

Another example of the greater potential for ambiguous reference with verbs in the imperfect is presented in (5). "P" indicates the turn of the participant; "I" indicates that of the interviewer.'

(5) P: . . . *el día que yo me gradúe y le enseñé a mi mamá que . . . Por el día del orgullo de mi mamá de verme graduar de xxxx.*
I: *Sí.*
P: *Ella siempre quería eso.*
P: . . . the day that I graduated and I showed my mom that . . . For the day of my mom's pride from seeing me graduate from xxxx.
I: Yes.
P: She always wanted that.

Had the speaker not included *ella* 'she', the likeliest subject referent of *quería* would have been *yo* 'I'. Both (4) and (5) show that communication can be hindered due to the fact that imperfect morphology does not distinguish between 1sg and 3sg. Expressed pronouns, then, help disambiguate reference. In contrast, potential ambiguity between 1sg and 3sg does not arise with preterit verbs due to distinctive person morphology.

To investigate the link between the TMA effect and contextual ambiguity, I coded 3sg verbs in the preterit or imperfect for *potential contextual ambiguity*. Drawing from Ranson (1991:142–145), potential contextual ambiguity was operationalized as *no potential ambiguity* and *potentially ambiguous*.

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

No potential ambiguity. The 3sg verbs are *not* potentially ambiguous if, even when expressed subject pronouns are removed, reference can be clearly established because the referent was identified (anywhere)[16] in the preceding discourse and there were no pragmatically viable competing referents,[17] as in (6) and (7).

(6) P: *Estuvo casi, así como seis meses, por un tiempo internada, y por otro tiempo, como otros cuatro meses, así. . . .*
I: *¿Qué tenía?*
P: *Ø Tenía . . . um . . . en realidad no saben bien que sea . . . que sea la causa. Pero lo que Ø tenía, eran convulsiones, como si fuera epilepsia.* [315M]
P: (She) was almost, like that for almost six months, admitted for a time, and another time, like another four months, like that . . .
I: What did (she) have?
P: (She) had . . . um . . . in reality (they) don't really know what is . . . what is the cause. But what (she) had, were convulsions, as if it were epilepsy.

In (6), the participant is talking about his sister who had been hospitalized. There is no potential ambiguity upon encountering the imperfect verb *tenía* 'had', even though morphologically it could be 1sg or 3sg. The establishment of the sister as the main referent in the discourse, coupled with the general knowledge that a person who was hospitalized is the person who 'had something', renders reference unambiguous.

Example (7) is also unambiguous; had the speaker omitted *ella*, *nació* would still refer to the daughter, not the father.

(7) . . . *el papá de ella murió como a los ocho meses que ella nació.* [311C]
' . . . her father died about eight months after she was born.'

The lack of potential ambiguity in (7) is due to general knowledge; if the father had died at the age of eight months, he would not have been a father. Notice that there were always two steps involved in coding for potential ambiguity in cases with expressed pronouns: first, the pronoun was omitted, and second, upon rereading the context, it was determined whether the referent was still clear. If the referent was still clear, the context was coded as *not* potentially ambiguous.

Potentially ambiguous. The 3sg verbs were coded as potentially ambiguous if reference could not be established via contextual cues. In these cases, reference could not be established after removing subject pronouns. Potential contextual ambiguity generally resulted from competition between 1sg and 3sg referents, as in (5) and (8).

(8) I: *¿Y tú eras la única americana en la escuela?*
P: *Había otra muchacha no más.*
I: *¿Tú te relacionabas bastante con esa muchacha?*
P: *No, porque ella era . . . ya mayor, ya era en octavo.* [181C]
I: and you were the only American in the school?

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

P: There was only one other girl.
I: And you hung out a lot with that girl?
P: No, because she was . . . already older, (she) was already in eighth.

Removal of the pronoun *ella* in (8) obscures the intended referent. Had the speaker said *No porque era . . . ya mayor* 'No, because Ø was . . . already older', *era* 'was' would most likely have been interpreted as having a 1sg subject (*yo era* 'I was'), and there is nothing in the preceding discourse to contradict that interpretation. It is the combination of ambiguous morphology (*era* can be 1sg or 3sg) and the existence of competing referents (also 1sg and 3sg) that render *era* potentially ambiguous. To review, cases (5) and (8) were coded as potentially ambiguous; cases like (6) and (7) were coded as not potentially ambiguous. Because we know that the difference between the imperfect and the preterit affects cases of switch-reference but not same-reference among NYRs (see Table 5), I focused the analyses of potential contextual ambiguity on 3sg preterit or imperfect verbs in contexts of switch-reference only.

Table 6 illustrates how potential contextual ambiguity impacts 3sg pronoun expression with preterit and imperfect verbs. Looking across the row labeled "Ambiguous," we see that there were 63 and 104 cases of potential contextual ambiguity among newcomers and NYRs, respectively. The majority of these potentially ambiguous cases were the result of competing referents for both newcomers and NYRs alike ($n = 97\%$ and 91%, respectively).

For both groups, imperfect verbs occurred with expressed *él/ella* at significantly higher rates in contextually ambiguous contexts than in unambiguous contexts (90% vs. 39% for newcomers [X^2 (1,1) = 35.42, $p < .0001$]; and 91% vs. 51% for NYRs [X^2 (1,1) = 28.54, $p < .0001$]). Newcomers also demonstrated sensitivity to contextual ambiguity with verbs in the preterit (X^2 (1,1) = 24.06, $p < .0001$), but this effect was not significant among NYRs (X^2 (1,1) = 3.05, $p = .08$). Overall the results indicate that newcomers express *él/ella* in order to avoid ambiguous reference regardless of verb tense, whereas NYRs' use of *él/ella* to avoid ambiguity manifests more strongly with imperfect than with preterit verbs. These

TABLE 6. *Potential contextual ambiguity by* él/ella *expression, 3sg preterit and imperfect verbs in switch-reference contexts*

	Newcomers				NYRs			
	Preterit		Imperfect		Preterit		Imperfect	
	n	*% él/ella*	*n*	*% él/ella*	*n*	*% él/ella*	*n*	*% él/ella*
Verbs	277	46	210	49	246	51	209	69
Unambiguous	255	41	169	39	219	49	132	55
Ambiguous	22	96	41	90	27	67	77	91
Pct-pt difference[a]		55		51		18		36

Note: [a]The percentage-point difference is calculated by subtracting *% él/ella* in unambiguous contexts from *% él/ella* in ambiguous contexts.

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

results suggest that newcomers attune to discursive ambiguity, whereas NYRs attune to both discursive and morphological ambiguity. NYRs' higher rate of pronouns with imperfect verbs thus reflects a routinized, but not automatic or obligatory, strategy aimed at clear communication of the subject referent.

One objection to the ambiguity account, raised by Bayley and Pease-Álvarez (1997:363–364), is that first-person plural *nosotros* 'we' and third-person plural *ellos/ellas* 'they' are also expressed more often with imperfect than with preterit verbs. Expressing these pronouns does not serve to avoid ambiguity, as there is distinctive person morphology for plural verb forms, including in the imperfect. But plural pronouns are rarely expressed in general. Diver (2011:252) argued that "arbitrary [patterns] are usually dependent on non-arbitrary [patterns]." It is possible that patterns of usage that apply to singular pronouns, which are nonarbitrary in terms of function, are mapped onto their plural counterparts. The TMA effect with plural pronouns, then, may be an arbitrary pattern that arises based on the nonarbitrary pattern involving singular pronouns.

TMA and pronoun expression: Backgrounding and foregrounding events

There is robust evidence that in narratives perfective forms tend to foreground events, while imperfect forms signal background information (e.g., Bardovi-Harlig, 1985; Comajoan, 2013; Dowty, 1986; Hopper, 1979; Longacre, 1996; Reid, 1977; Schiffrin, 1981; Tomlin, 1985). This relationship between perfectivity and foregrounding on the one hand and imperfectivity and backgrounding on the other has been documented in Spanish (Flores-Ferrán, 2007:303; Ozete, 1988; Silva-Corvalán, 1983:765).

How might the foregrounding/backgrounding functions of verbs be related to subject pronoun expression? Silva-Corvalán (1997:127, 2001:161–163) proposed that the backgrounding function of the imperfect is more compatible with expressed subjects, whereas the foregrounding function of the preterit is more compatible with subject omission. Her idea is that pronoun expression draws attention to the referent of the grammatical subject and, by so doing, draws attention away from the event (the verb). If a speaker foregrounds an event (in the preterit), she might omit the subject so as not to take away attention from that foregrounded event. In contrast, a backgrounded event (in the imperfect) lends itself to the inclusion of an overt subject because there is no competition between the subject and event. Example (9) illustrates this idea.

(9) *Yo venía de . . . no sé exactamente de dónde yo venía, yo sé que era en el tren E, . . . yo vivía aquí en la 47, me quedé en la estación de la 50, iba caminando así, como a las 10 de la mañana era y sentí una . . . un grito.* [042]
I was coming from . . . (I) don't know exactly where I was coming from, I know that it was on the E train, . . . I lived here on 47th, (I) got off at the 50th (street) station, (I) was walking like this, at about 10 in the morning it was and (I) heard a . . . a scream.

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

In (9), imperfect verbs *venía* 'was coming', *vivía* 'lived', *iba caminando* 'was walking' all describe the scene and denote background information. These imperfect verbs all occur with subject pronoun *yo*: *yo venía, yo venía, yo vivía* 'I was coming', 'I was coming', 'I lived'. Foregrounded events are those that make up the skeleton of the story: *me quedé en la estación de la 50* 'I got off at the 50th street station' and *sentí un grito* 'I heard a scream'. The two foregrounded events, 'got off at the station' and 'heard a scream', are denoted by preterit verbs that occur without a subject pronoun.

While Silva-Corvalán's foregrounding/backgrounding explanation of the TMA effect on pronoun expression cleverly draws on the well-established tendency to manipulate TMA morphology in order to highlight crucial parts of a narrative, it has not previously been tested empirically. Therefore, I coded 3sg preterit and imperfect verbs in contexts of switch-reference to test the hypothesis that pronouns are less likely to occur with verbs that denote foregrounded events and more likely to occur with verbs denoting backgrounded information. First, narrative discourse had to be distinguished from non-narrative discourse. Following Comajoan (2013:311), the main criterion used to distinguish between the two types of discourse was temporal sequencing. Narratives tell a story where the main events unfold in a particular order. All 3sg preterit and imperfect verbs that occurred within narratives were coded for whether they denoted foreground or background events. In order to avoid circularity, and again drawing on temporal sequencing, an objective criterion was adopted to differentiate between the two event types: foregrounded events are temporally sequential; if the timing sequence is changed, the story does not make sense. In contrast, backgrounded events are not temporally sequenced, even though they are part of narrative discourse (Bardovi-Harlig, 1985:265; Comajoan, 2001, 2005, 2013; Dry, 1983; Schiffrin, 1981:47). An example is given in (10), in which the underlined verbs are the foreground of a narrative about being robbed.

(10) *Entonces, yo veía que estaba así, así, y yo le dije, "¿Qué te pasa?" Y cuando miró así, me metió la mano en el bolsillo y salió corriendo con la cartera.* [005U]
So, I saw that (he) was like this, like this, and I said to him, "What's up with you?" And when he looked that way, he put his hand in my pocket and (he) went out running with my wallet.

In (10), the verbs *dije* 'said', *miró* 'looked',[18] *metió* 'put', *salió* 'went out' are temporally sequenced. The speaker says something to the thief, then the thief gives a look, then he puts his hand in the speaker's pocket, and finally he runs away with the wallet. The story would not make sense if the thief ran away with the wallet before he put his hand in the speaker's pocket. In contrast, the two verbs in *yo veía que estaba así* 'I saw that (he) was like that' are not temporally sequenced. The story would still make sense if the order was 'he was like that and I saw him'. Verbs that were not temporally sequenced but were in narratives were coded as backgrounded.

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

TABLE 7. *The impact of foregrounding/backgrounding on* él/ella *expression, with 3sg preterit and imperfect verbs in switch-reference contexts*

	Newcomers				NYRs			
	Preterit		Imperfect		Preterit		Imperfect	
	n	*% él/ella*	*n*	*% él/ella*	*n*	*% él/ella*	*n*	*% él/ella*
Not narrative	79	49	66	47	47	55	40	65
Backgrounded	31	26	137	47	58	47	160	69
Foregrounded	167	47	7	100	141	51	9	67
Total	277		210		246		209	

Rates of expressed *él/ella* with foregrounded and backgrounded preterit and imperfect verbs are presented in Table 7. There were 232 contexts (25% of the data) that were not in narratives; these could not be coded for backgrounding or foregrounding. Among the imperfect verbs that were in narratives, 297 (95%) denoted backgrounded events, which supports previous research linking morphological imperfectivity with the discursive function of backgrounding information. The preterit, on the other hand, appears to be more flexible: The majority of preterits in narratives were used for foregrounding, but 16% of newcomers' preterits and 30% of NYRs' preterits denoted background events.

Analyses of *él/ella* expression in narrative sections did not support Silva-Corvalán's (1997, 2001) explanation for the TMA effect. Instead, foregrounded events occurred with higher, rather than lower, rates of pronouns. Newcomers expressed *él/ella* significantly more often with preterit verbs that were foregrounding an event (47%) than with preterits denoting background information (26%) (X^2 (1,1) = 4.91, p = .03). The same trend was found among NYRs, but the slightly higher rate of *él/ella* with foregrounded preterits (51%) than with backgrounded preterits (47%) was not significant (X^2 (1,1) = .34, p = .56). The relationship between pronoun expression and backgrounding versus foregrounding could not be tested for imperfect verbs because so few denoted foregrounded events. In summary, Silva-Corvalán's explanation of the TMA effect finds no support in the results presented here. In fact, the results point to the opposite tendency, that is, pronouns are expressed more often with foregrounded events, which suggests that participants do not compete with actions for attention. Instead, the event is construed as a unit involving the primary participant(s) and the activity. Drawing attention to the participant involved in the event via subject pronoun expression may serve to highlight the event as a whole.

To summarize, although there is strong evidence that perfective and imperfective verb morphology correspond to foregrounding and backgrounding in narratives, the results herein show no support for the idea that expressed pronouns are incompatible with foregrounded events. We are left with one coherent explanation for the TMA effect: it reflects a drive to avoid ambiguity in reference.

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

EXPLAINING COMPLEXIFICATION IN NYC

Having explained the TMA effect on pronoun expression in Spanish as an ambiguity avoidance strategy, I now discuss the reasons for the emergence of this particular strategy in Spanish in NYC. Why does TMA become a significant predictor of 3sg pronoun expression in second-generation Spanish? The answer provided in this paper is guided by a functional approach to language change, which highlights communicative needs as one reason for the adoption or retention of linguistic forms and strategies (Andersen, 1982:97; Christiansen & Chater, 2008; Kiparsky, 1982; Martinet, 1952; Nettle, 1999; Shin & Otheguy, 2009). I have shown that second-generation Spanish speakers' sensitivity to TMA as a predictor of *él/ella* expression is related to the greater amount of ambiguity in person morphology in the imperfect than in the preterit, which in turn increases the likelihood of ambiguous reference. This suggests that the differential use of *él/ella* is directly related to the need to identify clear referents, which is part of a more general drive to communicate clearly.

If the complexification in Spanish in NYC is triggered by communicative needs, we might assume that something else is changing in the system that is pulling in the opposite direction, that is, a change that is obscuring communication. The idea here is that complexification in second-generation Spanish arises in order to compensate for other, concomitant changes that introduce potential obstacles for reference-tracking. What changes occur that obscure reference? Although a direct relationship between increasing sensitivity to TMA as a predictor of *él/ella* expression and other intergenerational changes in Spanish in NYC cannot be established conclusively in the current article, I will outline several possibilities that are worth exploring.

One explanation is that increasing sensitivity to TMA is related to decreasing sensitivity to reference. Recall that Otheguy and Zentella (2012:163, 185) found evidence of intergenerational weakening of reference as a predictor of pronoun expression in analyses in which all grammatical persons were included. In the current study, reference was a strong predictor of *él/ella* expression in both the first and second generations. Nevertheless, Shin and Otheguy (2009) showed that desensitization to reference occurs more readily with 1sg and second-person singular (2sg) verbs than with 3sg verbs. Even though weakening of reference primarily affects first-person and second-person forms and the increase in sensitivity to TMA affects third person, it is possible that the two changes are related. If so, this would suggest a type of trade-off: as second-generation speakers become less sensitive to some discourse constraints, they become increasingly sensitive to morphological factors. The interpretation that morphological factors remain pertinent or increase in strength while discourse factors undergo weakening is compatible with a body of research suggesting that, compared to morphological or semantic constraints on syntactic structures, discourse-pragmatic constraints are particularly susceptible to change (for discussion, see Sorace, 2012).

Another change that might be related to increasing sensitivity to TMA has to do with coda –s deletion. Otheguy and Zentella (2012:165, 176, 186–189, 198), in their study of Spanish in New York, find that, compared to newcomers, NYRs pay less attention to coda –s as a predictor of 2sg (*tú*) pronoun expression. That is, among newcomers, 2sg verbs pronounced without coda –s (as in *tú cantaba* 'you sang') favor *tú* expression, whereas 2sg verbs pronounced with coda –s (as in *tú cantabas* 'you sang') favor *tú* omission. This tendency is weaker among NYRs.[19] Perhaps, then, as speakers move away from attending to ambiguity of tokens within a verb tense, they increase their focus on ambiguity of the verb tense as a whole. If this is correct, the complexification in the current study might represent a change from a strategy that targets specific verb tokens to one that applies at the verb paradigm level.

A third possibility can be derived from an observation about newcomers' lack of sensitivity to TMA as a predictor of *él/ella*. If the TMA effect represents an ambiguity avoidance strategy, then why is it absent from the speech of newcomers? One possible answer is that, for newcomers, imperfect verbs without a subject pronoun are assumed to refer to third-person referents by default. In other words, *bailaba* 'danced' is automatically interpreted as *él/ella bailaba* 'he/she danced'. Then, if a newcomer deviates from the default, choosing a 1sg rather than a 3sg referent, he will signal this departure by including *yo* 'I' with the imperfect verb. Analyses of *yo* lend some support to this hypothesis, as newcomers produce significantly higher rates of *yo* with verbs in the imperfect (49% of 5060 verbs) than in the preterit (30% of 7129 verbs) (X^2 (1,1) = 452.76, $p < .0001$). If newcomers' grammar includes a strategy of assigning 3sg reference as a default, then it is also possible that another intergenerational change in NYC is the loss of this default strategy. Seen this way, increasing sensitivity to TMA could be compensation for the loss of the default 3sg reference assignation.

A complete explanation for grammatical complexification in Spanish in NYC is likely to be multifaceted, potentially related to decreasing sensitivity to reference and –s deletion, and perhaps, too, to the loss of a default strategy. There are a number of other changes in Spanish in New York that are likely to occur simultaneously, some of which may hinder clear communication and, therefore, trigger other changes to increase communicative efficiency. Thus, the idea is that a decrease in attention to one factor coincides with an increase in attention to other factors. Put more generally, linguistic changes do not occur in a vacuum. Understanding the nature of how changes relate to each other is a worthy pursuit (see Erker, 2012).

A final point worth mentioning is that it is not yet clear whether the changes detected in NYC are generalizable to other bilingual settings. Some additional evidence that sensitivity to TMA increases with bilingualism comes from Torres Cacoullos and Travis (2011:254) in their study of 1sg pronoun *yo* 'I' in New Mexico, where TMA is significant among bilinguals who use both Spanish and English regularly, but not among Spanish-dominant speakers who rarely use English. On the other hand, Silva-Corvalán (1994:156) found that first-generation

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press

and second-generation, but not third-generation, Spanish speakers in Los Angeles are sensitive to the impact of verb form ambiguity on pronoun expression. Thus, while the ongoing changes in NYC appear to be the result of increasing bilingualism, it is also possible that different bilingual settings yield different linguistic changes.

CONCLUSIONS

In this article, grammatical complexification was operationalized as the emergence of a new linguistic constraint guiding the variable use of a linguistic structure. This approach to examining language change and, in particular, complexification, is especially important for research on minority language grammars in situations of language shift because the norm has been to focus on simplification and loss of grammar in these cases. Focusing on simplification implicitly highlights what is "missing" from bilinguals' grammar. But there can be elements of grammar that are present in the bilinguals' grammar, but missing from monolinguals' grammar. Ignoring this type of complexification results in a skewed view of bilingual grammar, a view that can unfortunately propagate the common misconception that bilinguals' grammar is flawed.

The current study provides evidence of complexification in Spanish by comparing first-generation Latin American immigrants and second-generation bilingual Latinos in NYC and examining patterns of Spanish subject pronoun use in natural discourse. Multivariate analyses of those patterns showed that second-generation speakers were sensitive to TMA, while first-generation speakers were not. Thus the study uncovers one area of the grammar underlying 3sg pronoun use that complexifies over the course of one generation. It is worth reiterating that the claim is *not* that the entire grammar becomes more complex; indeed, there is abundant evidence that processes of simplification accompany change in situations of language shift (Silva-Corvalán, 1994:20–55, 213–214; 2001:318-327; Sasse, 2001; Trudgill, 2011). Instead, the argument is more modest: NYRs' pattern of 3sg subject pronoun expression has a component that is absent from the newcomers' grammar, and in that sense there is evidence of complexification.

After providing evidence for second-generation Spanish speakers' newfound sensitivity to TMA as a predictor of *él/ella* expression, two explanations for this TMA effect were explored. By operationalizing contextual ambiguity, I was able to show that NYRs expressed *él/ella* with imperfect verbs especially in contexts where reference would be ambiguous if the pronoun were omitted. These results demonstrated that the TMA effect represents a drive to refer unambiguously. An alternative explanation having to do with backgrounding/foregrounding functions associated with verb tenses was ruled out. Finally, a functional explanation for the grammatical complexification was proposed: NYR, bilingual Latinos compensate for other, concomitant changes by paying more attention to ambiguous verb morphology, which suggests that one principle driving the pattern of change is a need to clearly identify referents or, more broadly put, a need to

communicate clearly. By increasing attention to some factors and decreasing attention to others, the language user creates an effective and balanced system.

NOTES

1. The position that changes in structured variation reflect changes in grammar rests on a model of grammar that accommodates probabilistic patterns. The usage-based approach is well suited for this purpose, as it posits that grammar includes probabilistic information about the likelihood of form-form and form-meaning pairs (Bybee, 2001, 2010; Croft, 2010; Diver, 1995, 2011; Manning, 2003). In such a model, increasing complexity in patterns of usage reflects increasing complexity in grammar itself.

2. Other variables are person/number (e.g., Otheguy & Zentella, 2012), lexical frequency (e.g., Erker & Guy, 2012), and specificity of the referent (e.g., Cameron, 1993, Lapidus & Otheguy, 2005). In the current study, there were eight cases of nonspecific reference; three of which occurred with expressed *él/ella*. The eight cases were included in the study.

3. The numbers at the end of the examples refer to the number assigned to the participant.

4. *Bailaba* can also mean 'you-formal used to dance'. Furthermore, in some varieties of Spanish, final-syllable –s is not always pronounced, resulting in even more ambiguity.

5. The corpus was developed at the Graduate Center of the City University of New York (CUNY) with support from the National Science Foundation (#BCS 0004133), Professional Staff Congress of CUNY (#62666-00-31), and a CUNY Collaborative Grant (#09-91917).

6. Range is the difference between the top and bottom factors influencing pronoun expression (Otheguy & Zentella, 2012:165; Tagliamonte, 2012:127).

7. The remaining 75 participants arrived either between the ages of 3 and 17 or after age 17 and had been in the United States for more than five years at the time of their interview. Otheguy and Zentella (2012:33) classified these as "established immigrants," who have had more exposure to life in NYC than newcomers, but less exposure than NYRs.

8. Not all contexts should be regarded as variable. For example, 3sg subject pronouns can refer to animate entities (human and animals alike), but subject pronouns referring to inanimate entities are so rare that it makes sense to exclude all inanimate subjects. For an in-depth discussion of the envelope of variation, see Otheguy and Zentella (2012:45–67).

9. Most of the coding of TMA, reference, clause, semantic class, and reflexive was completed by Ricardo Otheguy and his research assistants at the CUNY Graduate Center.

10. I did not consider distance from the previous mention because it could only be calculated for tokens that were coded as having an eligible previous mention (approximately half the data). Priming effects were evident without considering distance, most likely because the previous mention occurred within the previous two clauses in about 80% of the tokens. If distance were taken into consideration, we would presumably see an even stronger priming effect (Torres Cacoullos & Travis, 2011; Travis, 2007).

11. In a regression run with data from newcomers and NYRs together, generation, which distinguished between the two groups, was significant ($p < .0001$).

12. Some sociolinguists employ the terms *factor groups* and *factors* for what I am calling *predictor variables* and *constraints/factors*, respectively (Tagliamonte, 2012:9).

13. Previous research has shown that speakers tend to use pronouns with cognitive verbs such as *creer* 'to believe' (Enríquez, 1984; Otheguy & Zentella, 2012; Travis & Torres Cacoullos, 2012). The current study suggests that this effect is not relevant for third-person referents (see also Bentivoglio, 1987:51–52).

14. Context b is morphologically ambiguous, but contextually unambiguous.

15. The subject of context e most likely refers to the older brother, but could mean '(I) taught myself English'. There is no difference between "me" and "myself" in Spanish (both are *me*), so this context remains ambiguous.

16. Although I considered previous mentions of the referent anywhere in the preceding discourse, of 775 cases where a coreferential previous mention was identified, 96% were within the previous six clauses.

17. Inanimate referents can compete for 3sg referent assignment, as in: *"Entonces, igual con mi hermano, él era igual, él tenía responsabilidades como nosotros."* [322E] 'So, the same with my brother, he was equal, he had responsibilities like we did.' Without *él*, *era igual* could mean 'it was the same'. Out of 156 cases where potential contextual ambiguity was due to competing referents, 28 had competing inanimate referents.

18. Longacre (1996:22) wrote that temporal adverbial clauses "serve for back-reference and cohesion." But some adverbials denote foregrounded events (Comajoan, 2013:332–334; Croft, 2001:334–335;

Tomlin, 1985:115). I coded when-clauses as foregrounded if they were clearly part of a temporally fixed sequence of events.

19. Caribbean speakers' decreasing attention to coda –s as a predictor of pronoun use might be related to their increasing tendency to pronounce –s (Erker, 2012; Otheguy & Zentella, 2012:186–189).

REFERENCES

Abreu, Laurel. (2009). *Spanish subject personal pronoun use by monolinguals, bilinguals, and second language learners*. Ph.D. dissertation, University of Florida.

Aikhenvald, Alexandra. (2003). *A grammar of Tariana, from northwest Amazonia*. Cambridge: Cambridge University Press.

Andersen, Roger W. (1982). Determining the linguistic attributes of language attrition. In R. D. Lambert & B. F. Freed (eds.), *The loss of language skills*. Rowley: Newbury. 83–118.

Ariel, Mira. (1990). *Accessing noun-phrase antecedents*. London: Rutledge.

Bardovi-Harlig, Kathleen. (1985). A narrative perspective on the development of the tense/aspect system in second language acquisition. *Studies in Second Language Acquisition* 17:263–291.

Bayley, Robert, & Pease-Álvarez, Lucinda. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9:349–371.

Bentivoglio, Paola. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.

Bybee, Joan. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

_____. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Cameron, Richard. (1993). Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change* 5:305–334.

_____. (1994). Switch-reference, verb class, and priming in a variable syntax. In K. Beals (ed.), *Papers from the 30th regional meeting of the Chicago Linguistics Society.* Vol. 2: *The parasession on variation in linguistic theory*. Chicago: Chicago Linguistics Society. 27–45.

_____. (1995). The scope and limits of switch-reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6/7:1–27.

Cameron, Richard, & Flores-Ferrán, Nydia. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1):41–65.

Carvalho, Ana, & Bessett, Ryan. (in press). Subject pronoun expression among Spanish-Portuguese bilinguals. In Carvalho *et al.*

Carvalho, Ana, Orozco, Rafael, & Shin, Naomi L. (eds.). (in press). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.

Chambers, J. K. (2004). Patterns of variation including change. In J. K Chambers, P. Trudgill, & N. Schilling-Estes (eds.), *The handbook of language variation and change*. Oxford: Blackwell. 349–372.

Christiansen, Morten H., & Chater, Nick. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5):489–509.

Comajoan, Llorenç. (2001). *The acquisition of past temporality and past morphology in Catalan.* Ph.D. dissertation, Indiana University.

_____. (2005). The acquisition of perfective and imperfective morphology and the marking of discourse grounding in Catalan. In D. Ayoun & R. Salaberry (eds.), *Tense and aspect in Romance languages*. Amsterdam: Benjamins. 35–78.

_____. (2013). Defining and coding data: Narrative discourse grounding in L2 studies. In M. R. Salaberry & L. Comajoan (eds.), *Research design and methodology in studies on L2 tense and aspect*. Berlin: Mouton de Gruyter. 309–350.

Croft, William. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

_____. (2010). The origins of grammaticalization in the verbalization of experience. *Linguistics* 48:1–48.

Dahl, Östen. (2004). *The growth and maintenance of linguistic complexity*. Studies in Language Companion Series 71. Amsterdam: John Benjamins.

Diver, William. (1995). Theory. In E. Contini-Morava & B. Sussman Goldberg (eds.), *Meaning as explanation: Advances in linguistic sign theory*. Berlin: Mouton de Gruyter. 43–114.

_____. (2011). Avoidance of the obvious: The pronoun as a minimax solution. In A. Huffman & J. David (eds.), *Language: Communication and human behavior. The linguistic essays of William Diver*. Leiden: Brill. 247–264.

Dowty, David R. (1986). The effects of aspectual class on the temporal structure of discourse: Semantics or pragmatics? *Linguistics and Philosophy* 9(1):37–61.
Dry, Helen. (1983). The movement of narrative time. *Journal of Literary Semantics* 12:19–53.
Enríquez, Emilia V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.
Erker, Daniel. (2012). Of categories and continua: Relating discrete and gradient properties of sociophonetic variation. *University of Pennsylvania Working Papers in Linguistics* 18(2):11–20.
Erker, Daniel, & Guy, Gregory. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3):526–557.
Ferguson, Charles A. (1982). Simplified registers and linguistic theory. In L. K. Obler & L. Menn (eds.), *Exceptional language and linguistics*. New York: Academic Press. 49–66.
Flores-Ferrán, Nydia. (2002). *A sociolinguistic perspective on the use of subject personal pronouns in Spanish narratives of Puerto Ricans in New York City*. Munich: Lincom-Europa.
_____. (2007). Is the past really the past in narrative discourse? In K. Potowski & R. Cameron (eds.), *Spanish in contact: Policy, social and linguistic inquiries*. Amsterdam: John Benjamins. 297–310.
Hochberg, Judith. (1986). Functional compensation for /-s/ deletion in Puerto Rican Spanish. *Language* 62:609–621.
Holmquist, Jonathan. (2012). Frequency rates and constraints on subject personal pronoun expression: Findings from the Puerto Rican highlands. *Language Variation and Change* 24:203–220.
Hopper, Paul. (1979). Aspect and foregrounding in discourse. *Discourse and Syntax* 12:213–241.
Hurtado, Luz Marcela. (2005). Condicionamientos sintáctico-semánticos de la expresión del sujeto en el español colombiano. *Hispania* 88(2):335–348.
Johnson, Daniel Ezra. (2009a). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3(1):359–383.
_____. (2009b). *Rbrul manual.* Retrieved from http://www.danielezrajohnson.com/Rbrul_manual.html. Accessed December 15, 2013.
Kiparsky, Paul. (1982). *Explanation in phonology*. Dordrecht: Foris.
Labov, William. (1994). *Principles of linguistic change: Internal factors*. Malden: Blackwell.
Lapidus, Naomi, & Otheguy, Ricardo. (2005). Overt nonspecific *ellos* in Spanish in New York. *Spanish in Context* 2(2):157–174.
Lastra, Yolanda, & Martín Butragueño, Pedro. (in press). Subject pronoun expression in oral Mexican Spanish. In Carvalho *et al.*
Longacre, Robert. (1996). *The grammar of discourse*. 2nd ed. New York: Plenum Press.
Manning, Christopher. (2003). Probabilistic syntax. In R. Bod, J. Hay, & S. Jannedy (eds.), *Probabilistic linguistics*. Cambridge: MIT Press. 289–341.
Martinet, André. (1952). Function, structure and sound change. *Word* 8:1–32.
Menard, Scott. (2002). *Applied logistic regression analysis*. 2nd ed. Thousand Oaks: Sage.
Michnowicz, Jim. (in press). Subject pronoun expression in contact with Maya in Yucatan Spanish. In Carvalho *et al.*
Miller, Karen. (2013). Acquisition of variable rules: /s/-lenition in the speech of Chilean Spanish-speaking children and their caregivers. *Language Variation and Change* 25:311–340.
Nettle, Daniel. (1999). Functionalism and its difficulties in biology and linguistics. In M. Darnell, E. Moravcsik, M. Noonan, & K. Wheatley (eds.), *Functionalism and formalism in linguistics*. Vol. 1. Amsterdam: John Benjamins. 445–468.
Orozco, Rafael. (in press). Pronominal variation in Costeño Spanish. In Carvalho *et al.*
Otheguy, Ricardo, & Zentella, Ana Celia. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.
Otheguy, Ricardo, Zentella, Ana Celia, & Livert, David. (2007). Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language* 83:770–802.
Ozete, Oscar. (1988). Focusing on the preterite and the imperfect. *Hispania* 71:687–691.
Poplack, Shana, & Levey, Stephen. (2010). Contact induced grammatical change: A cautionary tale. In P. Auer & J. Schmidt (eds.), *Language and space: An international handbook of linguistic variation*. Vol. 1. Mouton de Gruyter. 391–418.
Poplack, Shana, & Meechan, Marjory. (1998). How languages fit together in codemixing. *International Journal of Bilingualism* 2(2):127–138.
Posio, Pekka. (2011). Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics* 43(3):777–798.

de Prada Pérez, Ana. (in press). First person singular subject pronoun expression in Spanish in contact with Catalan. In Carvalho *et al.*

Ranson, Diana L. (1991). Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change* 3(2):133–152.

Reid, Wallis. (1977). The quantitative validation of a grammatical hypothesis: The passé simple and the imparfait. In J. A. Kegel, D. Nash, & A. Zaenen (eds.), *Proceedings of the Seventh Annual Meeting of the Northeast Linguistics Society*. Cambridge: MIT Press. 315–333.

Rivera-Mills, Susana. (2012). Spanish heritage language maintenance: Its legacy and its future. In S. Beaudrie & M. Fairclough (eds.), *Spanish as a heritage language in the US: The state of the science*. Washington, DC: Georgetown University Press. 26–64.

Sasse, Hans Jurgen. (2001). Typological changes in language obsolescence. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (eds.), *Language typology and language universals: An international handbook*. Berlin: de Gruyter. 1668–1677.

Schiffrin, Deborah. (1981). Tense variation in narrative. *Language* 57:45–62.

Shin, Naomi Lapidus, & Otheguy, Ricardo. (2009). Shifting sensitivity to Continuity of reference: Subject pronoun use in Spanish in New York City. In M. Lacorte & J. Leeman (eds.), *Español en Estados Unidos y en otros contextos de contacto: Sociolingüística, ideología y pedagogía*. Madrid: Iberoamericana. 111–136.

_____. (2013). Social class and gender impacting change in bilingual settings: Spanish subject pronoun use in New York. *Language in Society* 42:429–452.

Silva-Corvalán, Carmen. (1983). Tense and aspect in oral Spanish narrative. *Language* 59:760–780.

_____. (1994). *Language contact and change*. New York: Oxford University Press.

_____. (1997). Variación sintáctica en el discurso oral: Problemas metodológicos. In F. Moreno Fernández (ed.), *Trabajos de sociolingüística hispánica*. Alcalá de Henares: Universidad de Alcalá. 115–135.

_____. (2001). *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.

Sorace, Antonella. (2012). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism* 1(1):1–33.

Szmrecsanyi, Benedikt. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1:113–149.

Tagliamonte, Sali. (2012). Variationist sociolinguistics: Change, observation, and interpretation. Malden: Wiley-Blackwell.

Tomlin, Russell. (1985). Foreground-background information and the syntax of subordination. *Text* 5 (1-2):85–122.

Torres Cacoullos, Rena, & Travis, Catherine E. (2011). Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism* 15 (3):241–267.

Travis, Catherine. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19:101–135.

Travis, Catherine E., & Torres Cacoullos, Rena. (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4):711–748.

Trudgill, Peter. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

Veltman, Calvin. (1990). The status of the Spanish language in the United States at the beginning of the 21st century. *International Migration Review* 24(1):108–123.

Zentella, Ana Celia. (1997). *Growing up bilingual*. Malden: Blackwell.

APPENDIX A

Fixed effects only compared to mixed effects

TABLE A1. *Factors constraining* él/ella *expression, newcomers and NYRs, and fixed and mixed effects*

Deviance	2630.55		2583.39		2531.35		2490.90	
Intercept	−.76		−.82		−.15		−.27	
Centered input probability	.32		.31		.46		.43	
	Newcomers[a]				NYRs[b]			
	Fixed Effects Only		Mixed Model		Fixed Effects Only		Mixed Model	
Variable	FW	Log Odds	FW	Log Odds	FW	Log Odds	FW	Log Odds
Reference					**Reference**			
Switch	.64	.58	.65	.61	.63	.52	.63	.53
Same	.36	−.58	.35	−.61	.37	−.52	.37	−.53
Priming					**Priming**			
él/ella >	.65	.62	.63	.55	.64	.59	.63	.53
Other	.49	−.05	.50	−.02	.49	−.04	.50	−.02
Ø >	.36	−.57	.37	−.53	.37	−.55	.38	−.51
Clause					**Clause**			
Main	.56	.25	.56	.25	.56	.23	.57	.26
Subordinate	.47	−.13	.47	−.13	.46	−.15	.47	−.10
Coordinate	.47	−.12	.47	−.12	.48	−.07	.46	−.16
[TMA]					**TMA**			
Imperfect	.50	−.01	.51	−.05	.56	.25	.56	.23
Present	.50	−.02	.50	−.05	.50	−.02	.51	.05
Preterit	.50	−.01	.49	−.01	.44	−.23	.43	−.27
[Reflexive]					**[Reflexive]**			
Nonreflexive	.55	.19	.54	.16	.52	.10	.52	.09
Reflexive	.45	−.19	.46	−.16	.48	−.10	.48	−.09
[Semantic class]					**[Semantic class]**			
Mental	.55	.21	.56	.25	.49	−.04	.50	.02
Stative	.48	−.09	.46	−.16	.51	.03	.50	.01
External	.47	−.12	.48	−.09	.50	.01	.50	.02

Notes: Random factors = speaker and verb lexeme; fixed factors: priming, reference, clause, TMA, semantic, reflexive, clause*reference, clause*TMA.
Variables for which the *p*-value is greater than .01 appear in brackets.
[a]Newcomers: Verbs with highest centered factor weights (in order from highest to lowest): *hablar, trabajar, venir, ayudar, ser*; verbs with lowest centered factor weight (lowest to highest): *poner, casar, volver, ir, aprender.*
[b]NYRs: Verbs with highest centered factor weight: *trabajar, vivir, ser, hablar, criar*; verbs with lowest centered factor weight: *dar, ver, dejar, matar, estar, tener.*

APPENDIX B

Correlations between factors

Table B1 reports the *tolerance statistic* and the *variance inflation factors* (VIFs) for each predictor variable in the logistic regression models. The tolerance value indicates, for each predictor variable, the percentage of variance that cannot be accounted for by the other predictors. A tolerance of $<.20$ is considered potentially problematic; a value of $<.10$ indicates a collinearity problem (Menard, 2002:76). VIFs measure the strength of relationships among predictor variables; a value >2.5 may indicate collinearity issues (Miller, 2013:341–342; Szmrecsanyi, 2005:142). As demonstrated by the tolerance values and VIFs in Table B1, multicollinearity is not an issue in the current study.

TABLE B1. *Multicollinearity statistics for multivariate analysis of* él/ella *expression by newcomers and NYRs*

	Newcomers		NYRs	
Variable	Tolerance	VIF	Tolerance	VIF
Reference	.936	1.068	.924	1.082
Priming	.943	1.060	.929	1.077
Clause	.984	1.017	.984	1.016
Reflexive	.973	1.028	.957	1.045
Semantic	.951	1.052	.914	1.095
TMA	.942	1.062	.900	1.045

https://doi.org/10.1017/S095439451400012X Published online by Cambridge University Press