VARIABLE YO EXPRESSION IN NEW MEXICO: ENGLISH INFLUENCE?

RENA TORRES CACOULLOS
*The Pennsylvania State University*

CATHERINE E. TRAVIS
*University of New Mexico*

**Introduction**

Spanish has been spoken in New Mexico for over 400 years, developing as a distinct dialect group over a 300-year period of relative isolation from other Spanish dialects. For the past 150 years it has been in intense contact with English. One linguistic consequence of this contact is widespread code-switching among bilingual New Mexicans, as illustrated in the following examples.

(1) O: *Yo era la que estaba* TAKING CARE OF THINGS. (117-1A3: 231)
(2) G: *Aprendí en inglés.*
    Well,
    *no,*
    *fui hasta más de* HALF OF THE EIGHTH GRADE. (47-1A2: 431)
(3) W: *Y la querían poner* WITH THE BEGINNERS o no sé qué, (190-3B2: 653)

Poplack (1993: 255) defines code-switching as the juxtaposition of multi-word sequences, each of which is internally consistent with the grammatical patterns of its respective language. Thus, in code-switching, the patterns of each language are maintained, as bilinguals switch between—not ‘mix’—their two languages (cf. Haugen 1950). Some, however, have claimed that code-switching can have an effect on the grammar of each language. In this view code-switching promotes convergence, or structural similarity due to language contact (e.g., Gumperz & Wilson 1971: 165). For example, according to Backus (2005: 334) “code-switching function[s] to model syntactic patterns which are then subsequently imitated in the base language”.

In this study, we investigate whether code-switching plays a role in convergence by examining the patterns of subject expression in the speech of New Mexican bilingual speakers. The variability in subject expression can be seen in examples (1) and (2): in (1), the subject is expressed (*yo era*) and in (2), both subjects are unexpressed (*Ø aprendí, Ø fui*). New Mexico is an ideal environment for the study of contact-induced change, as it allows us to examine the long-term effects of contact in a native, non-immigrant community in the Southwest, rather than a community undergoing loss or shift within three generations (as is the case for most immigrant communities, cf. Silva-Corvalán 1994, inter alia).

We use the comparative variationist method (Poplack & Meechan 1998, Poplack & Tagliamonte 2001) to test the hypothesis of English influence on Spanish through code-switching, comparing patterns of variable *yo* expression in two environments: (1) across speakers of differing degrees of bilingualism and (2) across contexts where code-switching is, and is not, present in the surrounding discourse. The comparisons indicate the same grammatical patterns in both speaker groups and environments. We further consider this patterning in relation to findings regarding...
subject expression in English, which although required in most contexts, does exhibit variability in others (Cote 1996, Harvie 1998, Leroux & Jarmasz 2005, Oh 2005, 2006), and find that there are some important differences in patterning between English and Spanish. We argue, then, that the prolonged contact with English in this region has not affected the speakers’ patterns of subject expression, and that this is the case regardless of whether the speakers are code-switching or not.

The study of code-switching in its social context

The study is based on the materials of the New Mexico Colorado Spanish Survey (NMCOSS), a corpus of recorded interviews of 350 speakers conducted by New Mexican graduate students in 1992–1995 under the direction of Garland D. Bills and Neddy A. Vigil of the University of New Mexico (Bills & Vigil 2008). The data consist primarily of narratives, as the interviewees tell their life stories and describe their experiences in schooling and growing up in New Mexico.

These speakers represent a well-defined bilingual speech community in which there is widespread use of smoothly flowing code-switching as a discourse mode (e.g., Gonzales 1999). This allows us to consider the patterns of members of a community, rather than the idiosyncratic behavior of individual bilinguals which may or may not be constrained by group norms. To categorize language contact phenomena we must know the particular community patterns, since these may differ even when the language pairs involved are the same (Poplack 1985, 1987). Furthermore, we draw on spontaneous discourse data, where natural patterns of language use can best be observed. This is especially relevant in the bilingual situation in New Mexico, where strong normative pressures may inhibit bilinguals’ language use, particularly in formal situations such as those defined by elicitation tasks as have been applied in previous studies (e.g., Montrul 2004, Toribio 2004).

Table 1 presents a summary of the social characteristics of the speakers on whom the analysis presented here is based. All participants maintain proficiency in Spanish and have varying degrees of English proficiency, as indicated by their ratings of fluency in Spanish and English (5 = fluent, 1 = poor). Table 1 also indicates the proportion of each speaker’s data produced in contexts of bilingual discourse: the column labeled ‘code-switching’ provides the percentage of first person singular tokens that occur in contexts of code-switching within the same Intonation Unit (IU, see below), the preceding three clauses, or the preceding 10 IUs, whichever represents the larger segment of discourse.

The average amount of code-switching in preceding discourse is approximately 25% of all tokens (310/1,190), a rate which allows us to analyze the patterning of subject expression in this environment. Based on both the level of English and the amount of code-switching, we divided the speakers into two groups: a ‘bilingual’ group, of high levels of English proficiency (rated 4 or 5) and over 40% of first-person singular tokens in contexts where there is code-switching in the preceding discourse; and a ‘Spanish-dominant’ group, of low levels of English proficiency (rated 1 or 2) and under 15% of first-person singular tokens in code-switching contexts. Speakers with a high proficiency rating for English but less code-switching than that observed for the ‘bilingual’ speakers (12% to 24% vs. 40% to 52% for the ‘bilinguals’) were left unclassified.
The final column in Table 1 indicates the rate of expressed subjects for each individual speaker. While the average rate of expression overall is 34%, we have a high level of inter-speaker variability, with individual rates ranging from 13% to 64%. More importantly, we note that the rate of expression does not appear to correlate with the rating of fluency in English nor with the amount of code-switching in the data.

Table 1
Speaker characteristics (speakers from NMCOSS, Bills and Vigil (2008))

<table>
<thead>
<tr>
<th>Name</th>
<th>City or County</th>
<th>Educ</th>
<th>Born</th>
<th>Rating</th>
<th>Code-switch</th>
<th>Group</th>
<th>Rate yo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odilia</td>
<td>Bernalillo Co.</td>
<td>8</td>
<td>1930</td>
<td>5/5</td>
<td>52%</td>
<td>bilingual</td>
<td>38%</td>
</tr>
<tr>
<td>Ximena</td>
<td>Albuquerque</td>
<td>8</td>
<td>1939</td>
<td>5/5</td>
<td>50%</td>
<td>bilingual</td>
<td>22%</td>
</tr>
<tr>
<td>Tobías</td>
<td>Catron Co.</td>
<td>12</td>
<td>1907</td>
<td>4/5</td>
<td>41%</td>
<td>bilingual</td>
<td>25%</td>
</tr>
<tr>
<td>Wilmer</td>
<td>Tucumcari</td>
<td>12</td>
<td>1923</td>
<td>5/4</td>
<td>40%</td>
<td>bilingual</td>
<td>53%</td>
</tr>
<tr>
<td>María</td>
<td>Mora Co.</td>
<td>8</td>
<td>1918</td>
<td>5/2</td>
<td>15%</td>
<td>Sp-dom</td>
<td>46%</td>
</tr>
<tr>
<td>Victoria</td>
<td>Rio Arriba Co.</td>
<td>3</td>
<td>1911</td>
<td>5/1</td>
<td>11%</td>
<td>Sp-dom</td>
<td>33%</td>
</tr>
<tr>
<td>Norma</td>
<td>Española</td>
<td>2</td>
<td>1904</td>
<td>5/1</td>
<td>1%</td>
<td>Sp-dom</td>
<td>34%</td>
</tr>
<tr>
<td>Rosa</td>
<td>Taos Co.</td>
<td>5</td>
<td>1910</td>
<td>5/2</td>
<td>0</td>
<td>Sp-dom</td>
<td>33%</td>
</tr>
<tr>
<td>Pedro</td>
<td>Albuquerque</td>
<td>12</td>
<td>1947</td>
<td>4/4</td>
<td>24%</td>
<td>—</td>
<td>13%</td>
</tr>
<tr>
<td>Henry</td>
<td>Rio Arriba Co.</td>
<td>10</td>
<td>1920</td>
<td>5/2</td>
<td>22%</td>
<td>—</td>
<td>49%</td>
</tr>
<tr>
<td>Yolanda</td>
<td>Taos Co.</td>
<td>17</td>
<td>1926</td>
<td>4/5</td>
<td>14%</td>
<td>—</td>
<td>23%</td>
</tr>
<tr>
<td>Javier</td>
<td>Rio Arriba Co.</td>
<td>16</td>
<td>1909</td>
<td>5/5</td>
<td>13%</td>
<td>—</td>
<td>64%</td>
</tr>
<tr>
<td>Sebastián</td>
<td>Taos Co.</td>
<td>10+</td>
<td>1944</td>
<td>5/5</td>
<td>12%</td>
<td>—</td>
<td>13%</td>
</tr>
</tbody>
</table>

The data have been transcribed in accordance with the approach developed at the University of California, Santa Barbara (cf. Du Bois et al. 1993). This transcription method includes the annotation of features such as pauses, speaker overlaps, false starts and so on, all of which play a key role in the interaction between the interlocutors. Fundamental to this approach is the notion of the Intonation Unit (IU), that is, “a stretch of speech uttered under a single coherent intonation contour” (Du Bois et al. 1993: 47). Though defined prosodically, IUs show a strong correlation with the clause, demonstrating that they are functionally relevant units (Chafe 1994: 64). The IU provides us with a unit of measure for spontaneous discourse data, where the notion of ‘sentence’ is extremely difficult (if not impossible) to apply (cf. Crystal 1979, Chafe & Danielewicz 1987, Harvie 1998: 24). One study conducted on the same community of speakers under consideration here found that over 10% of code-switches occurred internally to the IU (Durán Urrea 2006), illustrated in (1) – (3), compared with just 4% found for four speakers of Mexican descent in California in a study conducted by Shenk (2006).
Spanish subject expression: Patterns vs. rates

Subject expression is one of the most studied variables in investigating the linguistic consequences of Spanish-English contact. Findings in this regard have been conflicting; while some propose that increased rate of use of expressed subjects does correlate with contact with English (Lipski 1996, Montrul 2004, Toribio 2004, Lapidus & Otheguy 2005a, b, Otheguy & Zentella 2007, Otheguy, Zentella & Livert 2007), others argue that there is no such correlation (Morales 1986, 2000, Silva-Corvalán 1994, Bayley & Pease-Alvarez 1997, Travis 2007). Indeed, some studies have found a lower rate of expression in speakers or varieties with greater (or more prolonged) contact with English (Silva-Corvalán 1994, Bayley & Pease-Alvarez 1997, Travis 2007).

Reliance on overall rates of use to determine contact-induced change is problematic because it is well known that regional dialects vary enormously. In particular, Caribbean dialects are reported to demonstrate higher rates of subject expression than other varieties (cf. Cameron 1992, 1993, Lipski 1994, Otheguy & Zentella 2007, Otheguy et al. 2007). However, despite divergent rates of expression, no study has reported different grammatical patterning across those dialects, and instead those studies which have compared patterning across different dialects have found remarkably similar results (e.g., Cameron 1993, 1994 for speakers from San Juan, Puerto Rico, and Madrid, Spain, Flores-Ferrán 2002, 2004 for Puerto Ricans living in Puerto Rico and in New York City, and Travis 2007 for speakers from New Mexico and Colombia).

Another extra-linguistic feature that has been found to affect rates of subject expression is genre (cf. Silva-Corvalán 2001: 163). Travis (2007: 130), for example, found a much higher rate of expressed first-person singular subjects in conversation than in personal narrative. The narratives examined were found to have a very high degree of continuity of subject, while the interactive conversations were found to have more shifting of topics, and therefore less continuity of subject. Given that continuity of subject is one of the strongest factors that has been found to affect subject expression in Spanish, the distribution of the data in terms of subject continuity has a major impact on rates of expression. That is, diverging rates of expression may be due not to grammatical differences, but to genre-driven differences in distribution.

In examining constraints on variable subject expression, Silva-Corvalán (1994: 159) has observed loss of pragmatic constraints in varieties in contact with English in Los Angeles, specifically loss of the clarifying use of expressed subjects with ambiguous verb forms, which she attributes to discourse-pragmatic transfer from English. Otheguy et al. (2007), who consider subject expression in speakers from Caribbean and mainland Latin America as they come together in New York City, examine the effect of grammatical person, finding some shifts in patterning, which they attribute to the dialect contact.

In studies of several non-contact varieties of Spanish, the linguistic factors that have been found to uniformly affect patterns of subject expression include subject continuity and discourse connectedness, semantic class of verb, Tense-Mood-Aspect, and person (for reviews see Silva-Corvalán 2001: 154-169). As we will see below, similar factors are identified as significant in this research.
A robust constraint uncovered in sociolinguistic and psycholinguistic research is a priming, or parallel structure, effect, whereby the use of a certain structure in one utterance functions as a prime on a subsequent utterance, such that that same structure is repeated (Poplack 1980, Weiner & Labov 1983, Bock 1986, Scherre & Naro 1991, Labov 1994, Cameron & Flores-Ferrán 2003, inter alia). This has been observed for subject expression in several dialects of Spanish (cf. Cameron 1994, Flores-Ferrán 2002, Travis 2007). Priming in monolingual Colombian Spanish is illustrated in example (4) from Travis (2007: 107), in which there is continuity of subjects across two clauses, yet both are expressed.

(4)  *Yo no sabía cuánto yo quiero a mi hermano.*

In recent psycholinguistic research, priming has been observed to occur across languages, specifically for passive constructions in Spanish and English (Hartsuiker, Pickering & Veltkamp, 2004), but this factor has not yet been considered in spontaneous discourse.

In this study we compare patterns of variability to examine contact-induced change. Since, as we have seen, rates of subject expression fluctuate due to extra-linguistic factors, such as dialect or genre, they are of limited value as diagnostics of grammatical similarity or difference. Grammatical differences are instead evidenced in the linguistic conditioning of variability, or the patterns of co-occurrence observed in the data (Poplack & Tagliamonte 2001: 92). Though in previous language-contact studies subject expression is often assumed to be obligatory in English, unexpressed subjects do occur (for example, (7) below). In the present study, we address this lack of attention to variability in English by comparing New Mexican patterns not only with those found in non-contact Spanish varieties but also with the linguistic conditioning of variability reported for varieties of English (Cote 1996, Harvie 1998, Leroux & Jarmasz 2005, Oh 2005, 2006).

**Variable subject expression: New Mexican Spanish vs. English**

This study is based on 13 interviews from the NMCOSS (see Table 1), representing a total of five and a half hours of speech (approximately 55,000 words). All tokens of first-person singular verbs were extracted, excluding truncated tokens where the speaker cut off before completing the verb and tokens produced by the interviewers, as there was insufficient data to identify their patterns of code-switching. Although ‘contrast’ has been described as one context where subjects are obligatorily marked (e.g., Silva-Corvalán 1994, Toribio 2004, inter alia), we excluded such tokens only where this was morphologically marked, such as through the use of *yo mismo*.\(^5\) We did not exclude unmarked tokens of contrast because, as noted by Otheguy et al. (2007: 775-776), unexpressed subjects do occur in this environment, and furthermore it is not possible to establish objective criteria for identifying such tokens in natural data (cf. Amaral & Schwenter 2005), making replicability difficult if not impossible.\(^6\)

These protocols resulted in 1,190 tokens, which were analyzed for their patterning in relation to the following constraints, or factor groups: previous realization (priming), subject continuity, semantic class of verb, ambiguity of verb morphology, clause type, and position in the turn. With the aim of discovering the set of factor groups which jointly account for the largest amount of
variation in a statistically significant way (Sankoff 1988), all factor groups were considered individually and together in multivariate analysis using GoldVarb X (Sankoff, Tagliamonte & Smith, 2005). We now turn to the discussion of these results.

Table 2 shows the results of a Variable-rule analysis of the contribution of environmental factors selected as significant to the choice of expressed yo in all speakers. In this and following tables, the 'input' (here .32) indicates the overall likelihood that the variant—expressed yo—will occur. In the first column, the numbers represent the probability (or factor weight) that each factor contributes to the occurrence of the variant: the closer to 0, the less likely that yo will be expressed and the closer to 1, the more likely that it will be. The range provides an indication of the relative strength of each group of factors in the analysis. In these results, previous realization has the strongest effect, with strong effects also for subject continuity and semantic class of verb, while ambiguity, though significant, is relatively weak. Clause type and position in turn were not significant in this multivariate analysis (as in other studies, e.g., Bentivoglio 1987, Silva-Corvalán 1994, Travis 2007). The second column shows the percentage of expressed yo, the third column the total number of tokens in each factor, and the fourth column the percentage of the data each factor makes up. We will be focusing on the factor weights of the first column, which indicate the ‘constraint hierarchy’, or direction of effect.

Table 2
Variable-rule analysis of the contribution of factors selected as significant to the choice of expressed yo in 13 New Mexican Spanish speakers

<table>
<thead>
<tr>
<th></th>
<th>Prob</th>
<th>% yo</th>
<th>N</th>
<th>% data</th>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 1190; Input: .32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Overall rate: 34%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous realization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressed</td>
<td>.69</td>
<td>51%</td>
<td>389</td>
<td>36%</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Not expressed</td>
<td>.39</td>
<td>23%</td>
<td>680</td>
<td>64%</td>
<td>#</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subject continuity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different subject</td>
<td>.58</td>
<td>41%</td>
<td>659</td>
<td>55%</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Same subject</td>
<td>.41</td>
<td>26%</td>
<td>531</td>
<td>45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic class of verb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological</td>
<td>.67</td>
<td>52%</td>
<td>227</td>
<td>19%</td>
<td>*</td>
<td>?</td>
</tr>
<tr>
<td>All other verbs</td>
<td>.46</td>
<td>30%</td>
<td>963</td>
<td>81%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguity of verb morphology (person)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>.57</td>
<td>36%</td>
<td>258</td>
<td>22%</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Not ambiguous</td>
<td>.48</td>
<td>34%</td>
<td>932</td>
<td>78%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factors not selected as significant:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clause type (main vs. subordinate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>?</td>
</tr>
<tr>
<td>Position in turn (initial vs. medial)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

*#/ indicate same/different direction of effect in other studies.
First, *yo* is more likely to be expressed when the preceding coreferential subject is also expressed, as in (5) (factor weight .69), than when it is unexpressed, as in (6) (.39). As reviewed in the previous section, such repetition—structural priming or parallelism—is a major effect in linguistic variation (Labov 1994: 547-568) and has been observed for subject expression in several varieties of Spanish (Cameron 1994, Flores-Ferrán, 2002, Travis 2007).

(5)  

\[ \begin{align*}
H: & \quad \text{ahi me levanté yo,} \\
& \quad \text{y le dije yo,} \\
& \quad \text{cuando te den un nombre no importa,} \\
& \quad \text{le dije yo,} \\
& \quad \text{.. cómo suena el nombre,} \\
& \quad \text{le dije yo.} \\
& \quad \text{Es el nombre.} \\
\end{align*} \]

(6)  

\[ \begin{align*}
S: & \quad \text{Ahora } \emptyset \text{ estoy --} \\
& \quad \emptyset \text{ hago muebles.} \\
D: & \quad \text{Mhm.} \\
S: & \quad \text{.. } \emptyset \text{ Hago escultura.} \\
D: & \quad \text{Mhm.} \\
S: & \quad \text{.. } \emptyset \text{ Trabajo madera.} \\
\end{align*} \]

Second, expressed *yo* is less likely to occur when it is coreferential with the immediately preceding subject, as in (6) (factor weight .41), than when it is not, as in (5), *cómo suena el nombre, le dije yo* (.58). This effect, sometimes termed ‘switch reference’, is the most robust effect that has been found in quantitative studies of variable subject expression in Spanish (cf. Silva-Corvalán 1982, 1994, Enríquez 1984, Hochberg 1986, Morales 1986, Bentivoglio 1987, Cameron 1994, 1995, Bayley & Pease-Alvarez 1997, Ávila-Shah 2000, Flores-Ferrán 2002, 2004, Travis 2007, inter alia).

Finally, *yo* appears to be most favored by psychological (.67) as opposed to other semantic classes of verbs (.46), a result also reported in numerous previous studies (e.g., Enríquez 1984: 240, Bentivoglio 1987: 60, Silva-Corvalán 1994: 162, Travis 2007: 115). However, close analysis of this class reveals that one half of the tokens (50%, 114/227) represent just two verbs that tend to occur in set constructions, namely *yo* (no) sé ‘I (don’t) know’ and *yo* creo ‘I think’, both of which show very high rates of expression (sé at 51%, 37/73, and creo at 85%, 35/41). The behavior of these two highly frequent verbs seems to be pulling up the rate for the class as a whole, suggesting that the effect is largely one of particular constructions rather than of a general class of psychological verbs (cf. Travis 2006).

Although morphological ambiguity is selected as significant in these data, it is the weakest effect as indicated by the small *range* (9), which is two to three times lower than that of the other significant factor groups. Prior results regarding the effect of ambiguity of verb form (for example in first- and third-person singular Imperfect forms such as *tenía* ‘I - he/she/it had’) have been inconclusive. A number of studies have found a correlation between morphological ambiguity and expressed subjects (Hochberg 1986, Cameron 1994, Silva-Corvalán 1994, Bayley & Pease-Alvarez 1997), while others have found no such correlation (Enríquez 1984,
Bentivoglio 1987, Ranson 1991). It has also been noted that cases of true ambiguity are rare in natural discourse, as even with unexpressed subjects the morphological ambiguity is generally resolved by context (Bentivoglio 1987: 45, Silva-Corvalán 1994: 154, Ávila-Shah 2000: 242). If this is the case, then expressed subjects are not required to disambiguate meaning, and this correlation is worth investigating further. In these data, we observe a skewed distribution of semantic classes across Tense-Mood-Aspect: while just 7% (18/258) of the tokens of ambiguous verbs involve psychological verbs, three times as many non-ambiguous verbs involve psychological verbs (22%, 209/932), because psychological verbs tend to occur overwhelmingly in the (non-ambiguous) Present Indicative (e.g., in the constructions *yo (no) sé* and *yo creo* we saw above). In contrast, the verb class that least favors subject expression, motion verbs (not shown separately in Table 2) (cf. Travis 2007), makes up 19% (49/258) of the ambiguous contexts but only 9% (81/932) of non-ambiguous ones. Regular patterning of verb classes with certain tenses has been observed for several languages, including Spanish (Weber & Bentivoglio 1991, Travis 2006), Brazilian Portuguese (Silveira 2007) and English (Scheibman 2002), and such a skewing (which has been largely overlooked in the literature on subject expression) may account for the conflicting results in previous research.

Overall, the multivariate analysis shows that variability in the expression of *yo* is conditioned the same way among these New Mexican speakers as in other non-contact varieties of Spanish, in particular in terms of the strong effects for priming, coreferentiality, and semantic class observed (as can be seen by the series of asterisks in the column labeled ‘Spanish’ in Table 2, indicating the same direction of effect in this and in other studies). This shared linguistic conditioning suggests that the patterns of subject expression in New Mexican Spanish have not been affected by the extensive contact it has had, and continues to have, with English. If, however, these linguistic factors condition subject expression in the same way in English, this finding cannot serve as evidence either for or against English influence, since the similarities may be coincidental or they may be cross-linguistic tendencies. If, on the other hand, we can identify a locus of different patterning, or a ‘conflict site’, between the two languages in contact (Poplack & Meechan 1998) then we have a point of comparison to allow us to determine whether the patterns in New Mexican Spanish have been affected by English or not. According to Poplack & Tagliamonte (2001: 101), “the conflict is typically instantiated as a difference in constraint hierarchy”, that is, as a difference in direction of effect, shown in the ordering of factors within each group by probability or factor weight from more to less.

Table 2 (rightmost columns) presents a comparison of the linguistic constraints on variable subject expression in Spanish with those of English, based on Leroux and Jarmasz (2005). These results indicate a difference in the direction of the previous realization effect. In English, although unexpressed subjects are highly infrequent, they are more likely when the previous subject is expressed, as in (7) (Oh 2005: 273).

\[(7)\]  
\[A: \quad \text{And I never saw a thing about it,}\]  
\[\text{and I [looked in the next couple of evenings.}\]  
\[B: \quad [\text{Mm hm.}\]  
\[(1.0)\]  
\[A: \quad (\text{Ø}) \text{ Never saw a th- a mention of it.}\]
In Spanish, in contrast, preceding expressed subjects favor subsequent expressed subjects, as in (5) (Cameron 1994, Flores-Ferrán 2002, Travis 2007). In the strong structural priming effect, whereby previous coreferential expressed first-person singular subjects favor expression (yo...yo), the New Mexican Spanish pattern differs from the reported English pattern which does not show a priming effect. By this diagnostic difference, we fail to find evidence for English influence.

Nevertheless, might these aggregate results be masking English influence on those New Mexicans with a higher degree of bilingualism? We will now compare the linguistic conditioning of variable first-person singular subject expression across our two subgroups of ‘bilingual’ and ‘Spanish-dominant’ speakers. If there is English influence, ‘bilingual speakers’ will be distinct from ‘Spanish-dominant’ speakers, and should show patterns closer to those of English.

**Bilingualism and code-switching**

As was discussed above, speakers were assigned to two groups based on self- and interviewer ratings of fluency in English in conjunction with the amount of code-switching preceding first-person singular tokens in the discourse. As can be seen in Table 1, there is a marked difference in the average amount of code-switching between the two groups: just 6% (17/265) in the ‘Spanish-dominant’ group as opposed to 47% (214/453) in the ‘bilingual’ group. This high incidence of code-switching in the preceding discourse in the ‘bilingual’ group allows us to directly consider the effect this has on rates of subject expression for these speakers, and thus we include this as a factor group in our multivariate analysis. The prediction of the hypothesis that code-switching promotes greater structural similarity with English in the Spanish of bilinguals is that expressed yo will be favored in the presence of code-switching (e.g., Backus 2004, Toribio 2004).

We compare the linguistic conditioning of variable yo expression across the two levels of bilingualism in Table 3. For the ‘Spanish-dominant’ speakers, previous realization, subject continuity, and semantic class are significant, and for the ‘bilinguals’, previous realization and subject continuity are significant. Though semantic class does not achieve significance in the bilingual data, the direction of effect for all three factor groups is the same, with choice of yo favored by a previous yo, a change in subject from the preceding clause, and psychological verbs. The parallel results, then, between these two groups of speakers again fails to support the hypothesis of English influence. Furthermore, the presence of code-switching was not significant in the multivariate analysis.
So far we have established that the patterns of variable \( yo \) expression generally among New Mexicans are the same as in non-contact varieties of Spanish, and are different from English in the strong priming effect on \( yo \). We have also shown the same patterning among New Mexicans with a higher and lower degree of bilingualism. We now conduct a third comparison of linguistic conditioning within the group of ‘bilingual’ speakers. Toribio (2004) proposes that speech in ‘bilingual mode’ exhibits a higher level of convergence than speech in ‘monolingual mode’, and finds support for this in a higher rate of use of subject pronouns in retellings of fairy tales when speakers are asked to code-switch compared with when they are asked not to. Here we operationalize the notions of ‘bilingual’ and ‘monolingual’ mode as the use of English in the recent discourse (within the same IU preceding or following the first-person singular subject, or within the preceding three clauses or 10 IUs) and ask the question: is the linguistic conditioning of \( yo \) variability different when these speakers have recently used English than when they have been speaking Spanish without recent code-switching? If code-switching is a mechanism of linguistic change in contact situations, promoting greater similarity of Spanish with English, the prediction is that we will find differences.

We compare the linguistic conditioning of variable \( yo \) expression within the same speakers in the absence vs. the presence of code-switching in Table 4. Firstly, we note that the overall rate of use of \( yo \) is higher by 12 percentage points, or by more than one third (37%), in the presence than in the absence of code-switching (44% vs. 32%). This higher probability of use of expressed \( yo \) in the presence of code-switching is seen in the input value (.44 vs. .31 in the absence of code-
switching). Since the data are from the same speakers in the same recording sessions, the rate difference is not easily attributable to extra-linguistic speaker or genre differences. Is there a linguistic constraint that we have failed to identify which would show English influence? The constraint hierarchies indicate otherwise: regardless of the presence of code-switching, the linguistic conditioning of the variation is the same.

The factor responsible, at least partially, for the increased rate of *yo* is part of the robust priming effect we have encountered throughout this study. In order to investigate in more detail this priming effect, we included a further factor in this group, namely a preceding coreferential use of English *I*, illustrated in (8).

(8) 

\[ W: \text{hijo,} \] 
\[ \text{sure I'm glad que yo no estoy --} \] 
\[ (H): \text{.. PUTTING ANY KIDS TO SCHOOL NOW. (190-3B2: 848)} \]

The same priming effect that we observed in the overall results is fully in place in the code-switching data. The choice of *yo* is most favored by previous realization as *yo* (with a factor weight of .63), less so by a previous realization as *I* (.50), and least favored when the preceding realization is unexpressed (.37). The greater favoring of *yo* by preceding *I* than unexpressed realization demonstrates that priming occurs across English and Spanish in these data, consistent with Hartsuiker et al.’s (2004) findings regarding priming across languages, as noted earlier.

As can be seen in Table 4, a preceding coreferential English *I* makes up a full 20% of the code-switching data. This reduces the number of preceding unexpressed contexts: whereas in the absence of code-switching two-thirds (66%) of the tokens have a previously unexpressed realization, in the presence of code-switching fewer than half (41%) of the tokens occur in this environment (note that this is 38% fewer than the number of preceding unexpressed contexts obtained in its absence). Since subject expression is least favored when the previous realization is unexpressed, this distribution results in a higher rate of expression overall in the bilinguals’ code-switching data. It is this difference in distribution that can then account for the different rates of subject expression in the presence and absence of code-switching. That is, it is not code-switching into English per se that results in a higher rate of expressed *yo* in Spanish, but specifically use of a preceding English *I*, because this raises the proportion of preceding expressed subjects, and thereby, through the priming effect, the rate of *yo*. Since English influence should be manifested in different patterning, these results again fail to provide evidence for contact-induced change.13
Two separate Variable-rule analyses of the contribution of factors selected as significant to the choice of expressed yo in bilingual New Mexican Spanish speakers, in absence and presence of code-switching

Table 4

<table>
<thead>
<tr>
<th></th>
<th>NO CODE-SWITCHING</th>
<th></th>
<th>CODE-SWITCHING</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 239; Input: .31</td>
<td>(Overall rate: 32%)</td>
<td>N = 211; Input: .44</td>
<td>(Overall rate: 44%)</td>
</tr>
<tr>
<td>Previous realization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expressed</td>
<td>.67</td>
<td>47%</td>
<td>75</td>
<td>.63</td>
</tr>
<tr>
<td>English</td>
<td>7</td>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td>Not expressed</td>
<td>.41</td>
<td>25%</td>
<td>146</td>
<td>.37</td>
</tr>
<tr>
<td>Range</td>
<td>26</td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Switch reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different subject</td>
<td>.62</td>
<td>42%</td>
<td>113</td>
<td>.59</td>
</tr>
<tr>
<td>Same subject</td>
<td>.39</td>
<td>24%</td>
<td>126</td>
<td>.41</td>
</tr>
<tr>
<td>Range</td>
<td>23</td>
<td></td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

Factors not selected as significant:

- Semantic class
- Clause type
- Position in turn
- Ambiguity

Conclusion

In using multivariate analysis to discover the configuration of factors accounting for variable subject expression, and drawing attention to the structural priming effect, the results we have presented are compelling for studies of subject expression and convergence. Clearly, larger scale variationist studies are called for, with more tokens and especially more speakers (cf. Otheguy et al. 2007), to enable deeper examination of patterns of variability across comparison groups and discourse modes.

We have seen that the New Mexican speakers show similar patterns of variable first-person singular subject expression as in non-contact varieties of Spanish, with use of yo favored by previous realization of coreferential subjects as yo, when the subject of the immediately preceding verb is non-coreferential, and in constructions with psychological verbs. Unlike other studies that have disregarded conditioning factors in English, we have identified one major difference in terms of the structural priming effect, which contrasts with English patterns of greater probability of subject omission following subject expression. Furthermore, we find the same linguistic conditioning across groups distinguished by level of bilingualism, as well as among ‘bilingual’ speakers in the presence and absence of code-switching. Rather than support for Spanish convergence with English in subject expression and for code-switching as a mechanism of contact-induced change, we have adduced evidence for the local process of priming, with English I favoring choice of Spanish yo more than unexpressed subjects but less than Spanish yo itself.
Notes
1 Numbers in parentheses following examples indicate NMCOSS (Bills & Vigil 1999) interview number, tape, side and segment, and the transcription line number. Cases of code-switching to English appear in SMALL CAPS. Transcription conventions are given in the appendix. We thank Garland Bills and Neddy Vigil for permission to use the NMCOSS materials, and Jenny Dumont, Damián Vergara Wilson, and James Walker for helpful comments on this chapter.
2 We restrict code-switching to multi-word sequences, since nonce borrowing of single-word English-origin nouns is prevalent in this community (see Clegg, this volume).
3 This is determined from an average of one rating given by the participant and another by the interviewer; most of these ratings coincide.
4 In most cases, the preceding three clauses occurred within 10 IUs. There are some cases where there is a lot of backchanneling, or speech not incorporated into clauses, where the preceding three clauses were expressed in more than 10 IUs.
5 This occurred just once in the data: Comencé yo misma... a leer inglés. (20-1A1: 736)
6 Contexts of contrast as indicated by a change in subject and clauses conjoined by pero ‘but’ (e.g., Me han invitado pero nunca he ido (102-1A1: 374)) show a yo rate of 38% (10/26), a rate that is not significantly higher than the overall rate of 34% (compared with 45% (51/113) in clauses with a change in subject conjoined by y ‘and’).
7 The ordering of factor groups by the range (the difference between the highest and lowest factor weight in a factor group) coincides with the order of selection (in GoldVarb’s step-up analysis) in the Variable-rule analyses shown in Table 2 – Table 4.
8 Note that in some contexts, the effect of subject continuity works against the priming effect; cf. Travis (2007: 124) for discussion.
9 This was also the case for Travis (2005, 2007).
10 The interaction with ‘Semantic class’ promotes the selection as significant of ‘Ambiguity’ despite near-identical rates within the factor group (Table 2).
11 A further explanation has been proposed by Silva-Corvalán (2001: 161-163), namely that it is not the ambiguity but the discourse function of the different TMAs that motivates their use with expressed or unexpressed subjects, a hypothesis supported by Bayley and Pease Alvarez (1997: 363). The present data follow the patterns predicted by Silva-Corvalán: Preterit forms, which have a foregrounding function, show a lower yo rate (27%, 84/309) than Imperfect, Conditional, and Subjunctive forms, which have a backgrounding function (37%, 90/245).
12 Factor weights for non-significant factors are from the first iteration of GoldVarb’s stepping down analysis (Tagliamonte 2006: 252).
13 Another test of convergence is in subject placement, as English subjects are almost categorically preposed. As with rates of expression, the rates of pre- and post-posed subjects in our data do not support the hypothesis of code-switching enhancing convergence: expressed yo occurs in postverbal position with the same relative frequency in all comparison groups: 29% (29/100) in the ‘Spanish-dominant’ and 25% (43/170) in the ‘bilinguals’; and, for the latter group, 27% (21/77) in the absence of code-switching and 24% (22/92) in the presence of code-switching (these differences are not significant in pair-wise Chi-square tests).
References


“Ashes to Ashes: The Historical Basis for Dialect Variation in New Mexican Spanish”, in: Romance Philology 53, 43-66.


Cameron, Richard/Flores-Ferrán, Nydia (2003): “Perseveration of Subject Expression Across Regional Dialects of Spanish”, in: Spanish in Context 1, 41-65.


--- (2004): “Spanish Subject Personal Pronoun Use in New York City Puerto Ricans: Can We Rest the Case of English Contact?”, in: *Language Variation and Change* 16, 49-73.


http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm (June 3, 2007).


Appendix: Transcription Conventions (Du Bois et al. 1993)

LETTER: speaker label . final intonation contour
Carriage return: new Intonation Unit , continuing intonation contour
.. short pause (about 0.5 secs) ? appeal intonation contour
(H) in breath -- truncated intonation contour