To appear in INTERNATIONAL JOURNAL OF LASSO, VOLUME 34, NUMBER 1-2 (2015)

BEYOND QUESTIONNAIRES: COMMUNITY-BASED MEASURES OF BILINGUALISM*

CATHERINE E. TRAVIS
Australian National University

RENA TORRES CACOULLOS
Pennsylvania State University

ABSTRACT. Bilingual speakers are often categorized according to degree of language contact on the basis of proficiency tests, which may have prescriptive biases, or questionnaire responses, which may be limited in scope and open to misinterpretation. Here an alternative approach is put forward and applied to the New Mexico bilingual community. Three forms of data are triangulated to construct a sociolinguistic profile: variables derived from self-reports in questionnaire items; content analysis of sociolinguistic interviews through systematic extraction of speakers' spontaneous comments on their linguistic experiences; and operationalization of language predominance as a measure of relative level of use and activation.

Keywords: bilingualism, bilingual speech corpus, questionnaires, sociolinguistic interview, New Mexico

1. ASSESSING LANGUAGE CONTACT: COMMUNITY MATTERS. Linguistic structure is best understood by considering speakers in their speech communities, geographic units "with well-defined limits, a common structural base, and a unified set of sociolinguistic norms" (Labov 2007:3). Northern New Mexican Spanish speakers (*Nuevomexicanos*) constitute a bilingual speech community: besides a specified geographical location and ethnicity, they share the same varieties of each of their languages and conventions for combining them, arising from common sociolinguistic experiences. An illustration of the shared structural base of the community's Spanish variety is found in verb forms. For example, the past perfective is expressed via the Preterit (*canté* 'I sang') in most varieties of Spanish in the Americas and by the Present Perfect (*he cantado* 'I have sung') in Peninsular Spanish. New Mexican Spanish follows Latin American norms in using the Preterit. At the same time, for the first-person Present Perfect the more widely used *he* exists in variation with the homegrown form *ha* 'I have', as in *yo ha tenido que trabajar* 'I have had to work' (NMCOSS 190-3B2, 24) (Bills & Vigil 1999:53-54, 2008:145-151).

Community-specific bilingual conventions for combining languages are illustrated in the distribution of English-origin loanwords by semantic field. The generalization that loanwords from the basic vocabulary are less likely than "culture-specific vocabulary" (Thomason 2001:71-72) does not apply in New Mexico. Aaron reports that the semantic field most propitious to borrowing is kin terms, compared with technology, roles (e.g. *firefighter*) and everyday items (*bag*), (2015:466). Example (1) illustrates with English *grandma* embedded in Spanish discourse. (Here, we also see the local variant *los*, which is in variation with *nos*, as the first-person plural reflexive pronoun; Bills & Vigil 2008:145).

* We gratefully acknowledge support from the National Science Foundation (BCS-1019112/1019122, 1624966; http://nmcodeswitching.la.psu.edu) and the ARC Centre of Excellence for the Dynamics of Language, and also thank the editor of the volume, Daniel Villa.

(1)

Inmaculada:	... yo y la grandma los sentamos y hablamos,	'... I and the ((other)) grandma sat down and talked,' (NMSEB 14 Proper Spanish, 26:07-26:10; born 1952, social worker, Albuquerque)[1]

Even more revealing of community-specific patterns is the distribution of individual kin terms according to language of origin. As Figure 1 shows, within the field of kin terms, the likelihood of using an English-origin noun is conditioned by lexical item. The borrowed *dad(dy)* is by far the preferred term for 'father', but native *mamá* is favored for 'mother'. And while grandparents of both sexes tend to be referred to with the English-origin *grandpa* or *grandma*, the preferred terms for sons and daughters are Spanish *hijo(s) – hija(s)*.

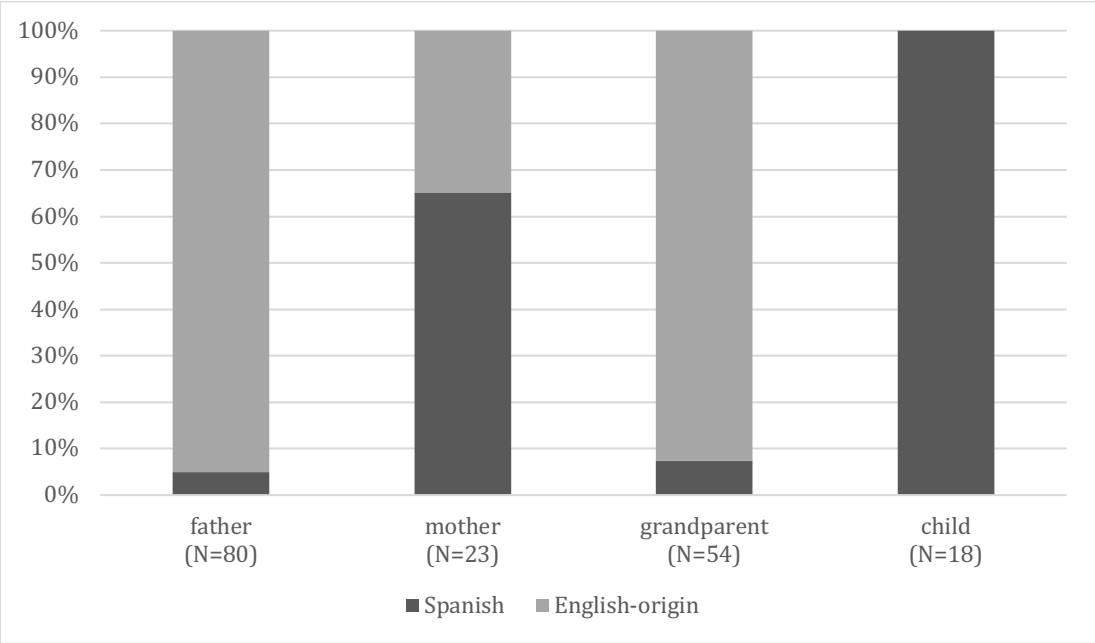

FIGURE 1. Proportions of English-origin vs. Spanish kinship terms occurring in New Mexican Spanish (NMSEB corpus) (from Aaron 2015:467).

Language contact is often blamed for language change, of the kind that is taken from other linguistic systems, or change "from above" (Labov 2007:346). Linguistic change is, however, normally from within the system, as the result of cognitive and discourse factors (e.g., Bybee 2015). We know from studies of quantitative patterns in speech communities that language change can be observed via analysis of conditions on variation according to extralinguistic characteristics of the speakers (Labov 1966). Profitable speaker groupings to demonstrate change are according to age, with young speakers leading in apparent time (Labov 1994:43-72); gender, with women

[1] Examples are reproduced verbatim from the New Mexico Spanish-English Bilingual corpus (see Torres Cacoullos & Travis 2018, Chapters 2 & 3), unless otherwise indicated. We follow transcription conventions laid out in Du Bois et al. (1993), and lines correspond to prosodic units. The translation of the Spanish original appears on the right; in order to preserve this representation in the glosses of bilingual utterances, English segments are repeated. Information given in parentheses indicates the transcript name and/or number, and the time stamp or line numbers of the excerpt, as well as the year of birth, occupation and residence of the participant at the time of the interview.

adopting innovative forms earlier than men (Labov 2001:261-293); and social class, which interacts with gender (Labov 2001:294-322) after the initial stages of the change and as the level of social awareness rises. For language change attributed to language contact, however, the scope and mechanisms remain hotly debated (Poplack & Levey 2010).

The general prediction is that speakers with the highest degree of contact lead in change. To ascertain change in a bilingual situation, then, speakers have to be grouped according to operationalizations of degree of contact. One measure is over real time, comparing an older stage of the community representing less contact, with a newer one representing greater contact. Where such a diachronic measure is not available, typical synchronic measures have been based on formal tests of proficiency and dominance; these, however, are clearly out of place in speech communities whose varieties are stigmatized. The questionnaire has been used as an alternative common instrument, but this, too, may give misleading results.

This article illustrates how measures of degree of contact that are meaningful and appropriate for a given speech community can be devised through triangulating different forms of data, specifically:

- variables derived from optimized questionnaire data;
- content analysis of sociolinguistic interviews;
- language distributions in spontaneous speech data.

2. BILINGUALISM IN NEW MEXICO. Spanish in northern New Mexico dates from settlement in the 16th-17th centuries from New Spain (today's Mexico). Annexed to the United States in the 1850s and admitted to the union in 1912, the region has been the site of Spanish-English contact for over 150 years (see Bills & Vigil 2008). In the schools, students were punished for speaking Spanish during the first decades of the 20th century (cf. Gonzales-Berry 2000), as recalled by several speakers recorded for the New Mexico Colorado Spanish Survey (NMCOSS) project (see Section 3).

(2)

Alicia:	Y qué pasaba si hablaba español?	'And what happened if you spoke Spanish?'
Ximena:	Nos castigaban.	'They punished us.'
Alicia:	Como --	'How --
	como qué hacían?	what would they do?'
Ximena:	Oh di- --	'Oh th- --
	y no nos dejaban salir al recess,	and they wouldn't let us go out at recess,
	nos hacían escribir,	they would make us write,
	yo no s- --	I won't --
	yo no debo espa- --	I shouldn't --
	hablar españo=l,	speak Spanish,'

(NMCOSS 47-1A2, 275-284; born 1939, housewife, Albuquerque)

(3)

Sebastián:	no tuve más que aprender a hablar inglés.	'I had no choice but to learn to speak English.
	no más --	only --
	.. No más pa' comunicar con el maestro.	.. Only to communicate with the teacher.'
Dora:	Claro.	'Sure.'
Sebastián:	porque,	'because,
	.. era como un pecado en ese tiempo,	.. it was like a sin at that time,
	.. hablar español no?	.. to speak in Spanish you know?'

(NMCOSS 102-1A1, 43-49;
born 1944, wood artist, Chamisal)

Today, Spanish is back in the schools but taught as a foreign language, to the detriment of the local variety. New Mexican Spanish is also disparaged in comparison with immigrant varieties. By way of illustration, in (4), from the New Mexico Spanish-English Bilingual (NMSEB) corpus (Section 3), Inmaculada recounts how she helped her granddaughter with her Spanish class homework only to have it marked as incorrect by the teacher (in what is likely to have been the 2000s). In (5) Trinidad tells how her daughter learned to speak nicely, because she learned Spanish with her Mexican friends rather than at home.

(4)

Inmaculada: .. they c- called it proper Spanish.
o=r,
whatever,
it was called,
but it wasn't our Spanish.
so she got everything wrong.

(NMSEB 14 Proper Spanish, 26:25-26:30;
born 1952, social worker, Albuquerque)

(5)

Trinidad:	lo hablaba muy bonito,	'she spoke it very nicely,
	como los de México.	like people from Mexico.
	... porque .. aprendió más por ella,	... because .. she learned it more on her own,
	que por nosotros.	than from us.'

(NMSEB 21 Demerits, 03:33-03:39;
born 1938, substitute teacher, Taos)

New Mexican Spanish has been propelled toward endangerment, being doubly undervalued, in comparison with English and in comparison with both immigrant and textbook varieties of Spanish (Bills & Vigil 2008:313). Though close to one half of NM's population of approximately two million are Hispanic (the largest proportion of Hispanics for any state in the U.S.), it is not the case that Spanish is flourishing. Language loss is made abundantly clear by a measure derived from Census data, which Bills, Hernández-Chávez and Hudson (1995:16) call "Loyalty". This is

the proportion of those who identify as Hispanic (or Latino) who speak only English at home—41% in NM. Even this percentage understates the loss of intergenerational language transmission among U.S.-born Hispanics: no less than 49% of this group speak only English, compared with just 6% of foreign born, as shown in Figure 2 (first and third columns).

Nevertheless, resisting the shift to English is the complementary 51% of the U.S.-born Hispanics in NM who speak "another language" (U.S. Census 2014), which we can infer is Spanish. And of these, close to 90% also report speaking English "very well", resulting in a sizable bilingual population, 45% of all native Hispanics across the state (first column, Figure 2). In the northern counties of Mora, Río Arriba, San Miguel and Taos, the proportion of native Hispanics who speak Spanish, and speak English "very well" is greater still, 58% (middle column). Moreover, while 17% of NM Hispanics are foreign-born (99% Mexican), in the northern counties only 6% are (compared with nearly 40% in the state of New York, and 33% in Texas). This long-standing, non-immigrant, bilingual community permits a compelling examination of language contact.

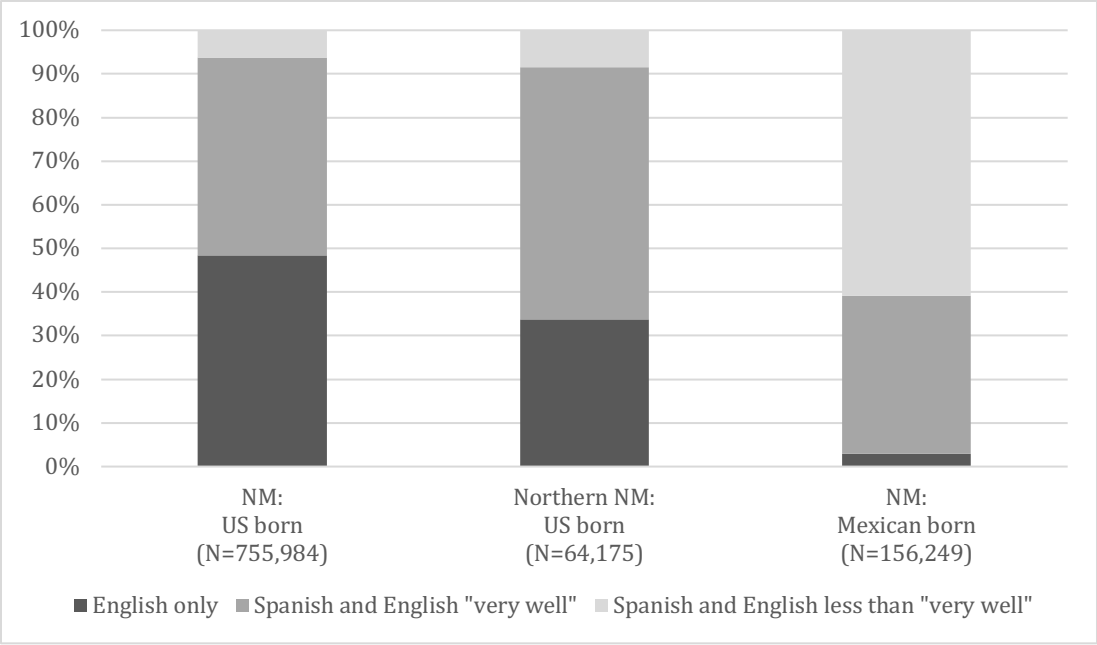

FIGURE 2. Language shift and bilingualism: Hispanic population according to language spoken in the home and level of English in New Mexico (United States Census Bureau 2014).

3. REAL-TIME SPEAKER GROUPINGS. The bilingual speech of northern NM community members provides a precious window into grammars in contact. The time-depth of research in this area is substantial, dating back to the early 20th century (e.g. Espinosa 1911). Real-time comparisons can be drawn thanks to the large-scale survey for a linguistic atlas of Spanish as spoken in New Mexico and southern Colorado, carried out by Garland Bills and Neddy Vigil in the 1990s, from which they produced the New Mexico Colorado Spanish Survey (NMCOSS) (Bills & Vigil 2008). As well as a lexical survey, NMCOSS includes stretches of spontaneous speech recorded by community members. The older speakers in this corpus, in whose recordings there are no multi-word English code-switches and only established borrowings, provide a benchmark of an earlier—

less bilingual—variety of Spanish against which to compare bilingual speakers today. Here we draw on a sample of 11 such speakers from Northern NM, born between 1897 and 1918.[2]

The New Mexico Spanish-English Bilingual corpus (NMSEB), collected in the years 2010-2011, provides a contemporary comparison. NMSEB records 40 *Nuevomexicano* speakers from northern NM counties born between 1922 and 1989 (Torres Cacoullos & Travis 2018: Chapters 2 & 3). NMSEB consists of spontaneously produced speech data from sociolinguistic interviews conducted by extended family members or acquaintances of the participants (cf. Labov 1984, Poplack 1993), totaling 29 hours, or 300,000 words. All speakers make regular use of both languages, as is evident in the interviews, where participants seamlessly switch back and forth between languages. This results in a uniquely bilingual corpus with even amounts of Spanish and English (see Section 4). Together, these corpora allow for examination of change over time and, crucially, according to degree of contact with English, from more Spanish-dominant (selected from NMCOSS) to more bilingual (NMSEB) speakers. Our first speaker groupings, then, are based on two corpora, for a real-time comparison between the linguistic behavior of NMSEB and NMCOSS speakers.

Subject pronoun expression has been described as the "showcase variable in variationist sociolinguistics" for Spanish, in particular in the study of Spanish in contact with English in the U.S. (Bayley et al. 2012:50). The prevalent hypothesis is that the near obligatory nature of subject expression in English should drive up the rate of subject expression in Spanish, and more so in speakers with greater contact with English (e.g. Otheguy & Zentella 2012). If contact-induced change has taken place, we would expect to find, first, higher rates of pronoun expression among NMCOSS speakers than other monolingual varieties due to the long-term contact with English in NM, and second, even higher rates among the NMSEB speakers, due to their greater contact with English.

Figures 3 and 4 compare rates of subject pronoun expression across different varieties, for 1sg and for 3sg human specific subjects, the two most frequent grammatical persons to occur in speech data (Torres Cacoullos & Travis 2018:127). As seen here, 1sg expression ranges from 25% to 50%, and 3sg from 14% to 39%. The prediction for the NM varieties is not upheld. NMCOSS speakers do not have higher rates of expression but sit somewhere in the middle (34% for 1sg and 21% for 3sg); nor do NMSEB speakers have higher rates of expression than NMCOSS (28% for 1sg and 17% for 3sg).

More decisive for ascertaining change is the linguistic conditioning of subject pronouns according to contextual features (cf. Poplack et al. 2012). One such feature is grammatical person. As can be seen in Figures 3 and 4, a general tendency is a greater favoring of pronominal subjects with 1sg than 3sg, and this is also the case for both NMCOSS and NMSEB. In fact, the linguistic conditioning of Spanish subject pronouns is similar across varieties, including in NM (Carvalho et al. 2015:xiv–xv, Silva-Corvalán & Enrique-Arias 2017:172–187, Torres Cacoullos & Travis 2018:74-110).

[2] Thanks to Garland Bills and Neddy Vigil for access to NMCOSS recordings to create a comparison sub-corpus. In the transcriptions for this sub-corpus, we have used pseudonyms for the participants and the interviewers. The eleven older, Spanish dominant speakers are participants number 4, 10, 20, 76, 219, 236, 246, 272, 310, 313 and 316.

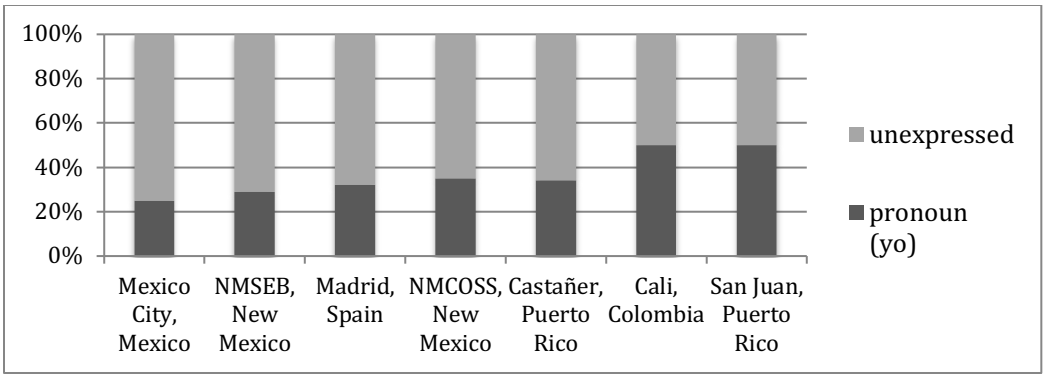

FIGURE 3. Rate of subject pronoun expression vs. Ø across different varieties: 1sg.

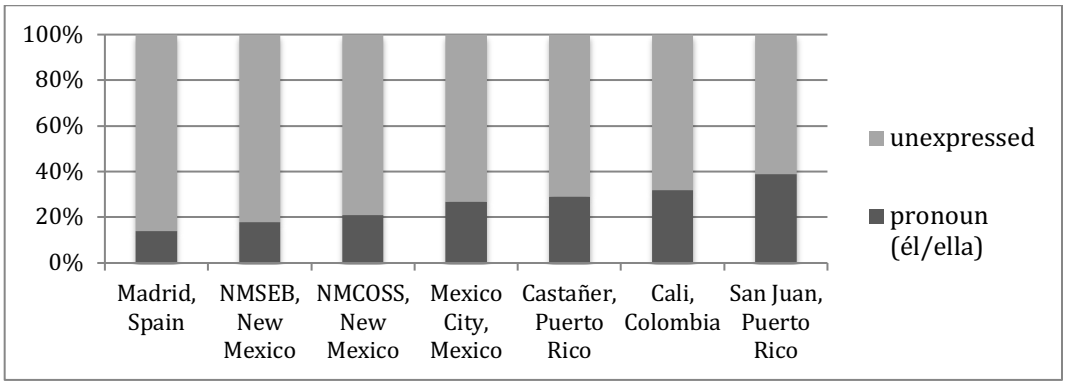

FIGURE 4. Rate of subject pronoun expression vs. Ø across different varieties: 3sg human specific.[3]

Thus, this first comparison reveals no change over time in either rates or conditioning of pronominal subject expression, between the more Spanish dominant NMCOSS speakers and the bilingual NMSEB speakers. We may still, ask, however, whether there are differences within NM bilingual speakers according to the degree of contact with English that correlate with subject expression rates. For this, we turn to the questionnaire responses, the content of the interviews, and the production data.

4. VARIABLES DERIVED FROM OPTIMIZED QUESTIONNAIRE DATA. Questionnaires are widely used in language contact studies, with tabulated responses being employed as a measure of degree of contact, or of bilingualism, against which to compare linguistic behavior. For NMSEB, information about the participants' sociolinguistic history was collected in the course of the sociolinguistic interview (see Section 5). Following the interview, fieldworkers filled in responses to a short questionnaire, which, besides basic social information about the participants (age, place of residence, etc.), contained questions related to their experience with Spanish and English. These included self-reports of "first" and "preferred" language, as well as ratings and domains of use of

[3] Sources: NMSEB (N = 1sg 3,296, 3sg 2,275) and NMCOSS (N = 1sg 1,009, 3sg 685) (Torres Cacoullos & Travis 2018:142), Mexico City (N = 1sg 939, 3sg 450; Lastra & Butragueño 2015:43), Madrid, Spain (N = 1sg 10,185, 3sg 2,501; Enríquez 1984:348), Castañer, Puerto Rico (N = 1sg 1,527, 3sg 603; Holmquist 2012:211), Cali, Colombia (N = 1sg 1,389, 3sg 1,413; Torres Cacoullos & Travis 2018:124), San Juan, Puerto Rico (N = 1sg 949, 3sg 443; Cameron 1992:233).

each language. Responses to ten items (listed in the left-hand column of Table 1) were assigned a score from 1 to 3, 3 indicating more English, 1 more Spanish, and 2 somewhere in between (such as when a response of "both" was given).

A simple averaging of scores may be deceptive, as responses may not be independent of each other, and they may not contribute equally to a mean score. To address this, we carried out data optimization using principal component analysis (PCA), which reduced the multidimensional space made up of the 10 questionnaire items to a smaller number of orthogonal components (cf. Hoffman & Walker 2010:47-48). Four principal components accounted for 73% of the total variance in the dataset. As listed in the top row in Table 1, we interpret these as: preferred language (PC1), preferred language for TV (PC2), preferred language for radio (PC3), and self-rating of one language relative to the other (PC4). The first principal component (PC1), preferred language, alone accounts for 32% of the variance (second row).

	PC1 PREFERRED LANGUAGE	PC2 TV	PC3 RADIO	PC4 SELF-RATING
Variance accounted for	32%	16%	14%	11%
QUESTIONNAIRE ITEM				
Preferred language	**0.64**	-0.37	0.17	0.32
First language	0.23	-0.07	0.34	**-0.52**
Self-rating (relative English-Spanish)	0.17	-0.11	0.34	**0.53**
How English learned	0.24	-0.07	0.39	**-0.56**
Preferred language for				
... radio	**0.53**	0.37	**-0.49**	-0.09
... TV	0.13	**0.81**	**0.42**	0.17
... reading books	0.09	-0.05	0.11	0.05
Language spoken				
... with family	0.005	-0.21	0.07	-0.004
... with friends	0.23	-0.02	-0.32	-0.09
... at work	0.30	-0.01	-0.22	-0.02

TABLE 1. Loadings of contact-with-English questionnaire items on principal components.

Table 1 shows the associations (loadings) of each of the 10 questions with these principal components; bolded text in shaded cells indicates questions with relatively stronger associations with a particular component (with a magnitude greater than 0.4) (cf. Horvath & Sankoff 1987:194). While we might predict preferred language, first language and relative self-rating all to load onto the same principal component, this is not the case. And though first language and self-rating do load on to the same principal component (PC4), they do so with opposite signs. In fact, preferred language scores do not correlate strongly with self-rating scores nor with reported "first" language (Pearson correlation test $r = 0.32$, $p = 0.045$; $r = 0.33$, $p = 0.038$, respectively).[4]

[4] Statistics done in R (R Core Team 2015) using the base package or the package Hmisc (Harrell 2015).

For an initial interpretation of these results, we may consider the distribution of participants according to their responses to these items. For preferred language, one half of the speakers report that they prefer English (20/40), close to one-third Spanish (12/40), and the remainder state "both" (8/40). For self-rating, on the other hand, most rate their English and Spanish as equal (25/40), and the rest are evenly divided (7/40 rating their Spanish, and 8/40 their English, higher). Asked about their first language, most participants list Spanish (35/40) and state that they learned English when they started school (32/40). However, from the content of the interviews it is clear that these responses cannot be taken at face value, as we will see below.

As they load on to different components, preferred language and relative self-rating can be used as two independent measures of degree of contact for these bilingual speakers. The hypothesis of contact-induced change would predict higher subject pronoun expression rates for speakers who state a preference and/or a higher rating for English. Nevertheless, subject pronoun expression rate is affected neither by preferred language score, as shown in Figure 5, nor by self-rating, seen in Figure 6.

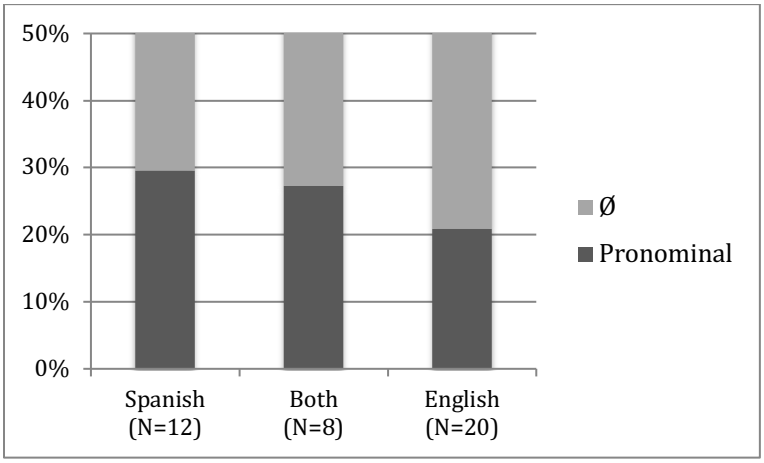

FIGURE 5. Rates of subject pronouns according to reported preferred language in NMSEB ($F = 2.06$, $p = 0.141$). For speakers preferring Spanish $M = 30\%$, $SD = 17.01$; both the same $M = 27\%$, $SD = 12.45$; English $M = 21\%$, $SD = 8.60$.

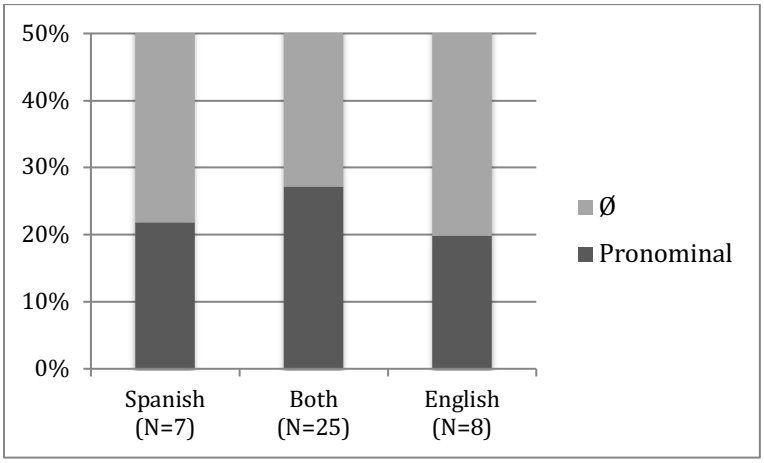

FIGURE 6. Rates of subject pronouns according to relative self-rating in NMSEB ($F = 1.32$, $p = 0.28$). For speakers rating their Spanish higher than their English $M = 22\%$, $SD = 11.87$; both the same $M = 27\%$, $SD = 13.66$; English higher than Spanish $M = 20\%$, $SD = 8.9$.

Two other factors that emerge in the PCA are language preferred for radio and TV, associated with two different principal components (PC1 and PC2 respectively). Where they are associated with the same component (PC3), they have opposite signs. This captures the preference by NMSEB participants to listen to the radio in Spanish (27/40) but to watch TV in English (30/40). Why might this be the case?

There is little local content on the television in NM, beyond some news coverage, and even that is not delivered in the local variety; instead, programming comes from Los Angeles, Miami or other national and international sources. On the other hand, there are local radio stations that play New Mexican music, and we infer positive affect toward New Mexican language and culture from speakers' references to local stations such as KDCE (Española) and KDNM (Reserve) and New Mexican bands such as *Cuarenta Y Cinco*. Thus, the bilingualism measures suggested by the PCA of questionnaire responses can be properly interpreted only with intimate knowledge of the community, together with consideration of the participants' own comments, the topic of the next section.

5. CONTENT ANALYSIS OF SOCIOLINGUISTIC INTERVIEWS. The sociolinguistic interview, especially when conducted by a community-member, records unmonitored speech that is as close as possible to the vernacular, the "most systematic data for linguistic analysis" (Labov 1984:29). As well as producing appropriate linguistic data, this data collection method supplies information on the linguistic experiences of the participants, which often arises naturally in the course of the conversations. Content analysis involves systematic extraction of demographic, linguistic history and language attitude information from transcriptions of the recordings (Poplack et al. 2006:196-207). The content—what people say in their own words— contextualizes questionnaire responses and also brings to the analyst's attention issues and attitudes relevant to the community beyond the pre-determined categories imposed by the questionnaire. In this way, it illuminates the nature of bilingualism in the community and the sociolinguistic profile of the speakers.

As we saw above, responses to questionnaire items concerning first language, language preference and self-ratings are not correlated in the predicted way. This is an indication that responses even to direct, apparently simple, questions like "What was your first language?" or "What language do you speak better?" may not be reliable, as the questions themselves may not be applicable to members of bilingual communities. This is particularly so for minority community members who speak stigmatized varieties, where speakers' judgments of their own linguistic abilities may be affected by linguistic insecurity (Sankoff 1988:145).

There is ample evidence in the NMSEB corpus of what Bills and Vigil describe as an "inferiority complex" by northern New Mexicans about their speech (2008:13). For example, Leandro reported on the questionnaire that Spanish was his first language and that he learned English at school and in the military; but in the course of the interview, he said that when he first attended school in the south of the state and had contact with Mexicans, he didn't know *mexicano* 'Spanish' either (6).

(6)

Leandro:	los mexicanitos esos,	'those Mexican guys,
	they --	they --
	.. hablaban muy diferente el mexicano que nosotros,	.. spoke Spanish very differently from us,'
Ricardo:	Yeah.	'Yeah.'
Leandro:	y nosotros hablábanos otra class de language sandwich <@ que [le dicen @>,	'and we spoke another kind of language sandwich <@ as [they say @>,
	@@@	@@@
	de Nuevo México].	of New Mexico].'
Ricardo:	[@@@@]	'[@@@@]'
Leandro:	pero cuando fuimos allá,	'but when we went there,'
Ricardo:	uh [huh].	'uh [huh].'
Leandro:	[pues] estábanos como mudos ahí sí ve.	'[well] it was like we were mute you know.
	@ no sabíanos ni inglés,	@ we didn't know English,
	y no sabíanos el mexicano !bien.	and we didn't know Spanish well.'

(NMSEB 25 El Servicio, 05:07-05:31; born 1933, miner, Río Arriba)

Some questionnaire responses from NMSEB participants challenge the very legitimacy of the questions asked. For example, one speaker, Rocío, rated both her languages a 4 out of 5, clearly conducting her self-assessment under a normative lens. Comments in her recording starkly evince linguistic insecurity: despite describing herself as someone who entered school knowing only Spanish, she claims that she learned Spanish by translating letters from Mexico for her aunt's boss (7). As an adult, Rocío sees herself as having improved her Spanish while hosting Mexican visitors whom she helped learn English (8), and describes the Spanish in her own family as *mocho* (broken) (9). Her use of this term encapsulates the negative view of the local variety, as does her attribution of the label *Spanglish* (cf. Lipski 2008:38-74, Otheguy & Stern 2010).

(7)

Rocío: ... and he used to write letters.
...(1.4) to my tía's boss. ((my aunt's boss'))
... a=nd,
...(1.5) eh he .. decided,
... that .. I was going to learn how to read them,
... and that I was going to have to .. learn how to write.
... and I was going to translate them.
...(0.9) and that's how I learned how to speak Spanish.

(NMSEB 05 Las Tortillas, 57:45-58:05; born 1945, teacher aid, Santa Fe)

(8)

Rocío: ... your boys learned how to speak English,
.. the same way that they're teaching me,
the proper way.
... of speaking Spanish.

(NMSEB 05 Las Tortillas, 059:32-059:38)

(9)

Rocío:	.. mis= hermanos hablaban poquito,	'.. my brothers and sisters spoke some,
Adriana:	.. pero muy <@ mocho @>.	.. but very <@ broken @>.'
Rocío:	.. [mocho]?	'.. [broken]?'
Adriana:	[@@]	'[@@]'
Rocío:	mhm.	'mhm.'
Adriana:	[yeah].	'[yeah.]'
Rocío:	[hm lo] hablaban.	'[hm they] spoke it.
	...(0.7) the Spanglish.	…(0.7) the Spanglish.'

(NMSEB 05 Las Tortillas, 01:00:39-01:00:49)

Notions such as order of acquisition or early vs. late bilinguals and even the L1-L2 distinction are blurred where use of two languages is commonplace. Although most NMSEB speakers responded on the questionnaire that Spanish is their first language and that they learned English at school, it is clear from the content analysis that there was English in the community and even in the home. It also transpires that they did not necessarily start school knowing no English. Sandra, for example, reports having learned English at school, but she must have had some English at home and in the neighborhood, since she recalls that her brother taught her how to write the alphabet (10), her sister listened to radio series in English (11), and her mother spoke to English-speaking neighbors (12).

(10)

Sandra:	you know my brother was my teacher.	'you know my brother was my teacher.
	... we would --	... we would --
	.. coming back from the school?	.. coming back from the school?
	there in the arroyo?	there in the arroyo?
((10 intervening lines))		
	and he would make --	and he would make --
	a blackboard out of the sand.	a blackboard out of the sand.
	...(0.7) mira,	...(0.7) look,
	me decía,	he would say to me,
	...(0.8) lo=s --	...(0.8) the --
	.. el --	.. the --
	los ay bee cees, ((ABCs))	the ABCs,
	... son .. puros palitos,	... are .. just little sticks,
	.. y bolitas.	.. and little balls.'

(NMSEB 03 Dos Comadres, 58:01-58:25;
born 1943, administration, Española)

(11)

Sandra:	.. my sister Rita was into the soap operas.	'.. my sister Rita was into the soap operas.
	.. on the radio.	.. on the radio.'
Adriana:	.. mhm.	'.. mhm.'
Sandra:	y ahí estaba The Romance of Helen Trent.	'and there was The Romance of Helen Trent.
	... [you=ng Doctor Malone].	... [you=ng Doctor Malone].'
Adriana:	[@@@@]	'[@@@@]'
Sandra:	.. The Edge of Night.	'.. The Edge of Night.
	.. y quién sabe qué tanto.	.. and who knows what else.'

(NMSEB 03 Dos Comadres, 15:04-15:11)

(12)

Sandra:	.. so cuando entramos,	'.. so when we came in,
	oh muy gente la Mrs. Johnson,	oh she was very nice Mrs. Johnson,
	<VOX come on in,	<VOX come on in,
	come on in VOX>,	come on in VOX>,
	y entró mamá,	and mom went in,
	and she could --	and she could --
	...(0.7) she could explain herself,	...(0.7) she could explain herself,
	you know,	you know,
	she could speak English.	she could speak English.'

(NMSEB 03 Dos Comadres, 01:03:55-01:04:03)

Rather than a regime of ordered first and second language acquisition with a sharp L1-L2 distinction, content analysis of the sociolinguistic interviews underscores the pervasiveness of both languages in daily life, where NMSEB participants readily adapt their language choice to the situation, (13), or, as an in-group discourse mode, spontaneously switch between languages (14) and (15) (Gonzales 1999:29).

(13)

Bartolomé:	...(2.0) pues sabes que yo nunca ha ponido at- --	'…(2.0) well you know I have never paid at- --
	...(0.7) atención,	…(0.7) attention,
	...(0.8) si me hablan en español les hablo .. les respondo en español.	…(0.8) if they speak to me in Spanish I speak .. answer them in Spanish.
	si me hablan en inglés les respondo en inglés.	if they speak to me in English I answer them in English.'

(NMSEB 02 La Marina, 57:17-57:36;
born 1928, firefighter, Albuquerque)

(14)

Monica:	siempre estábanos hablando,	'we were always speaking,
	... mitad y mitad español,	… half and half Spanish,
	y mitad inglés=.	and half English.'

(NMSEB 11 El Trabajo, 00:44-00:48; born 1941, factory worker, Albuquerque)

(15)

Ivette:	a lot of people say that you think in Spanish.
	...(1.5) or you think in English.
	and to me I don't know what I'm thinking.

(NMSEB 06 El Túnico, 24:53-24:59; born 1946, factory worker, Albuquerque)

6. LANGUAGE DISTRIBUTIONS IN SPONTANEOUS SPEECH DATA. Language dominance (e.g. Silva-Corvalán & Treffers-Daller 2015), though an intuitive notion, is not straightforwardly operationalizable outside the lab setting, given the standard-language bias of most (proficiency) tests (cf. Dąbrowska 2012). For NMSEB, two potential measures of this, preference and self-rating, are not correlated, as we saw. Furthermore, were we able to arrive at language dominance scores, these still provide no window on bilinguals' actual experience with their two languages. Here we use the production data themselves to devise a measure that we may call LANGUAGE PREDOMINANCE.

The corpus records data from both languages by the same speakers. In (16), for example, considering the language of the clause (based on the verb), the speaker uses Spanish in lines (b), (e), (f), (h), (i) and English in (a), (c), (d), (g).

(16)

a.	Mónica:	...(1.0) he wants to go to mechanic school y --	'...(1.0) he wants to go to mechanic school and --
b.		.. todavía estamos peleando con él pero,	.. we're still arguing with him but,
c.		.. can't change his mind,	.. can't change his mind,
d.		I don't know what we're going to do.	I don't know what we're going to do.
e.		...(1.3) qué= --	…(1.3) what --
f.		.. qué querrá hacer y,	.. what he'll want to do and,
g.		y and he wants to graduate early,	and and he wants to graduate early,
h.		que porque por eso va al night school,	it's why he's going to night school,
i.		.. pa' agarrar diferentes credits,	to get different credits,'

(NMSEB 11 El Trabajo, 15:26-15:41; born 1941, factory worker/school custodian, Albuquerque)

Figure 7 shows that in NMSEB Spanish and English clauses are used with equal frequency. This even distribution of both languages renders inapplicable notions such as that of an overall "matrix language" (Myers-Scotton, 1993) or its obverse, a less-frequently-used "more salient language" (Myslín & Levy 2015:871).

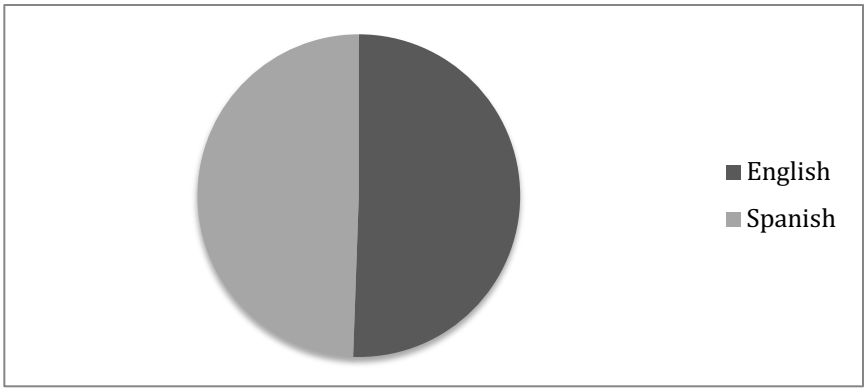

FIGURE 7. Distribution of clauses between bilinguals' two languages (NMSEB, Spanish and English, N = 33,092)

Though equal in the corpus in the aggregate, amounts of Spanish and English clauses do vary across the recordings, allowing individuals to be grouped according to their use of each of the two languages. Speakers were categorized as "English-predominant" if fewer than one-third of their clauses were in Spanish (N = 13 speakers), "Spanish-predominant" if more than two-thirds of their clauses were Spanish (N = 12) and "both" if the proportion of Spanish clauses was between one- and two-thirds of their total clause count (N = 15). Language predominance thus defined does not correlate with participants' stated language preference ($r = -0.01$, $p = 0.96$), nor strongly with self-rating ($r = 0.31$, $p = 0.05$). For example, Mónica (the speaker of example (16)) reports that she prefers to speak Spanish, yet she rates her English as higher, and she uses the two languages equally in the recording.

Language predominance could be taken as a proxy for the speaker's relative amounts of use of the two languages in their daily lives. Alternatively, it could be seen as an operationalization of language modes in a situational continuum (Grosjean 1998:136) or of activation level of languages (e.g. Kroll et al. 2006) according to which a greater proportion of clauses in English would be consistent with greater activation of that language. Either way, for subject expression, the hypothesis of contact-induced grammatical change would predict a higher rate of Spanish subject pronouns with English predominance.

Figure 8 depicts the rate of subject expression according to language predominance. Contrary to prediction, there are no significant differences in rates of subject expression across these three groupings (and, though not significantly so, the rate tends to be lowest among those speakers who are predominant in English, the opposite of the predicted direction). Thus, New Mexican bilinguals' subject pronoun rates are not influenced by language predominance in their recordings, just as they were not found to be affected by self-reported language preference or self-rating, all three measures of degree of contact failing to support a hypothesis of contact-induced grammatical change.

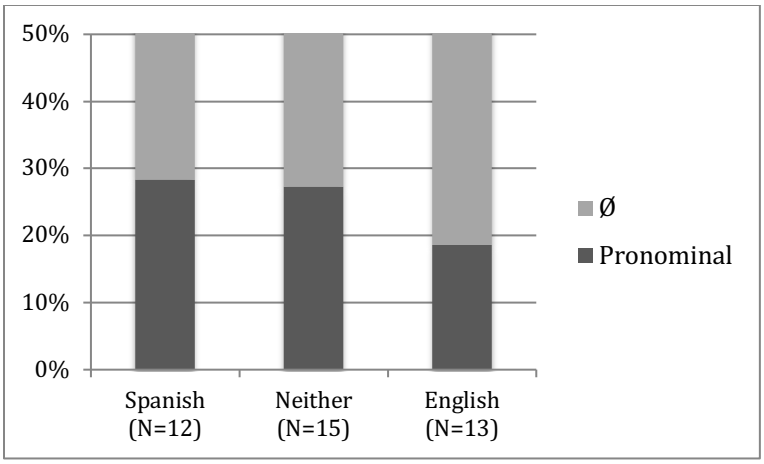

FIGURE 8. Rates of subject pronouns according to language predominance in NMSEB ($F = 2.45$, $p = 0.099$). For speakers producing more than two thirds of their clauses in Spanish $M = 28\%$, $SD = 10.92$; between one and two-thirds $M = 27\%$, $SD = 16.40$; fewer than one-third $M = 19\%$, $SD = 6.40$.

7. FINDING COMMUNITY MEASURES OF BILINGUALISM. Nearly half a century ago, Garland Bills noted the limitations of language study interested "in the accumulation of speech fragments with little concern for linguistic or sociological context" (1975:vi-vii). How can appropriate social information be collected and meaningfully interpreted in a bilingual community such as that in northern New Mexico? Bilingual speakers are often categorized according to measures such as L1, proficiency, and dominance. The analyses presented here starkly expose how it is misleading to transpose constructs such as language dominance or dichotomies such as first vs. second language (L1 vs L2) into the community setting, where using two languages is a fact of daily life.

Evidence comes from three forms of data in a bilingual speech corpus. The first is optimization of questionnaire data through a statistical technique which condenses responses (for example, on first and preferred language, learning of English, self-rating of English relative to Spanish, and domains of use) to a smaller set of factors. Second is content analysis of sociolinguistic interviews, through systematic extraction and categorization of all relevant comments arising in the course of the conversations. Such content analysis allows interpretation of questionnaire responses in the context of the community. Above all, what people say in their own words brings out issues and attitudes beyond the categories imposed by the questionnaire. Finally, a third kind of data afforded by a bilingual corpus is the distribution of languages in the transcriptions, which provides a gauge of language predominance.

The approach illustrated here shows how, beyond questionnaire data, speakers' production data from an appropriate corpus provide meaningful social groupings to test linguistic hypotheses.

APPENDIX: TRANSCRIPTION CONVENTIONS (DU BOIS ET AL. 1993)

.	final intonation contour	..	short pause (about 0.5 seconds)
,	continuing intonation contour	…	medium pause (> 0.7 seconds)
?	appeal intonation contour	…(N.0)	timed pause
--	truncated intonation contour	@	one syllable of laughter
-	truncated word	<@ @>	speech uttered while laughing
=	lengthening	(( ))	researcher's comment
[ ]	speech overlap	<VOX VOX>	speech uttered with marked voice quality

REFERENCES

AARON, JESSI ELANA. 2015. Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism* 19.459-80.

BAYLEY, ROBERT; NORMA L. CÁRDENAS; BELINDA TREVIÑO SCHOUTEN; and CARLOS MARTÍN VÉLEZ SALAS. 2012. Spanish dialect contact in San Antonio, Texas: An exploratory study. *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, ed. by Kimberly Geeslin and Manuel Díaz-Campos, 48-60. Somerville, MA: Cascadilla.

BILLS, GARLAND D. 1975. Linguistic research on United States Hispanics: State of the art. *Spanish and English of United States Hispanos: A critical, annotated, linguistic bibliography*, ed. by Richard V. Teschner, Garland D. Bills, and Jerry R. Craddock, v-xxii. Arlington, VA: Center for Applied Linguistics.

BILLS, GARLAND D.; EDUARDO HERNÁNDEZ CHÁVEZ; and ALAN HUDSON. 1995. The geography of language shift: Distance from the Mexican border and Spanish language claiming in the Southwestern U.S. *International Journal of the Sociology of Language* 114.9-27.

BILLS, GARLAND D., and NEDDY A. VIGIL. 1999. El cambio lingüístico en el español nuevomexicano: los factores de edad y educación. *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina*, ed. by José Antonio Samper Padilla and Magnolia Troya Déniz, 877-85. Las Palmas: Universidad de Las Palmas de Gran Canaria, Servicio de Publicaciones, Librería Nogal.

BILLS, GARLAND D., and NEDDY A. VIGIL. 2008. *The Spanish language of New Mexico and southern Colorado: A linguistic atlas*. Albuquerque, NM: University of New Mexico Press.

BYBEE, JOAN. 2015. *Language Change*. Cambridge: Cambridge University Press.

CAMERON, RICHARD. 1992. Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation. Philadelphia, PA: University of Pennsylvania dissertation.

CARVALHO, ANA M.; RAFAEL OROZCO; and NAOMI LAPIDUS SHIN. 2015. Introduction. *Subject pronoun expression in Spanish: A cross-dialectal perspective*, ed. by Ana M. Carvalho, Rafael Orozco, and Naomi Lapidus Shin, xiii-xxvi. Washington DC: Georgetown University Press.

DĄBROWSKA, EVA. 2012. Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism* 2:219-53.

DU BOIS, JOHN W.; STEPHAN SCHUETZE-COBURN; SUSANNA CUMMING; and DANAE PAOLINO. 1993. Outline of discourse transcription. *Talking data: Transcription and coding in discourse*, ed. by Jane Edwards and Martin Lampert, 45-89. Hillsdale: Lawrence Erlbaum Associates.

ESPINOSA, AURELIO M. 1911. *The Spanish language in New Mexico and Southern Colorado*.

Santa Fe, NM: Santa Fe Historical Society of New Mexico 16.
GONZALES, MARÍA DOLORES. 1999. Crossing social and cultural borders: The road to language hybridity. *Speaking Chicana: Voice, power and identity*, ed. by Letticia Galindo and María Dolores Gonzales, 13-38. Tucson, AZ: University of Arizona Press.
GONZALES-BERRY, ERLINDA. 2000. Which language will our children speak? The Spanish language and public education policy in New Mexico, 1890-1930. *The contested homeland: A Chicano history of New Mexico*, ed. by Erlinda Gonzales-Berry and Donaldo R. Maciel, 169-89. Albuquerque, NM: University of New Mexico Press.
GROSJEAN, FRANÇOIS. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1.131-49.
HARRELL, FRANK E. JR. (with contributions from CHARLES DUPONT and many others). 2015. Hmisc: Harrell Miscellaneous. R package version 3.17-0. http://CRAN.R-project.org /package=Hmisc.
HOFFMAN, MICHOL F., and JAMES A. WALKER. 2010. Ethnolects and the city: Ethnic orientation and linguistic variation in Toronto English. *Language Variation and Change* 22.37-67.
HOLMQUIST, JONATHAN. 2012. Frequency rates and constraints on subject personal pronoun expression: Findings from the Puerto Rican highlands. *Language Variation and Change* 24.203-20.
HORVATH, BARBARA, and DAVID SANKOFF. 1987. Delimiting the Sydney speech community. *Language in Society* 16.179-204.
KROLL, JUDITH F.; SUSAN C. BOBB; and ZOFIA WODNIECKA. 2006. Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition* 9.119-35.
LABOV, WILLIAM. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
LABOV, WILLIAM. 1984. Field methods of the Project on Linguistic Change and Variation. *Language in use: Readings in sociolinguistics*, ed. by John Baugh and Joel Sherzer, 28-53. Englewood Cliffs, NJ: Prentice Hall.
LABOV, WILLIAM. 1994. *Principles of linguistic change: Internal factors*, vol. 1. Oxford: Basil Blackwell.
LABOV, WILLIAM. 2001. *Principles of linguistic change: Social factors,* vol. 2. Oxford: Blackwell.
LABOV, WILLIAM. 2007. Transmission and Diffusion. *Language* 83.344-87.
LASTRA, YOLANDA, and PEDRO MARTÍN BUTRAGUEÑO. 2015. Subject pronoun expression in oral Mexican Spanish. *Subject pronoun expression in Spanish: A cross-dialectal perspective*, ed. by Ana M. Carvalho, Rafael Orozco, and Naomi L. Shin, 39-57. Washington, DC: Georgetown University Press.
LIPSKI, JOHN M. 2008. *Varieties of Spanish in the United States*. Washington, DC: Georgetown University Press.
MYERS-SCOTTON, CAROL. 1993. *Dueling languages*. Oxford: Oxford University Press.
MYSLÍN, MARK, and ROGER LEVY. 2015. Code-switching and predictability of meaning in discourse. *Language* 91.871-905.
OTHEGUY, RICARDO, and NANCY STERN. 2010. On so-called Spanglish. *International Journal of Bilingualism* 15.85-100.
OTHEGUY, RICARDO, and ANA CECILIA ZENTELLA. 2012. *Spanish in New York: Language contact, dialect levelling, and structural continuity*. Oxford: Oxford University Press.

POPLACK, SHANA. 1993. Variation theory and language contact: Concepts, methods and data. *American dialect research*, ed. by Dennis R. Preston, 251-86. Amsterdam: John Benjamins.

POPLACK, SHANA, and STEPHEN LEVEY. 2010. Contact-induced grammatical change: A cautionary tale. *Language and Space: An international handbook of linguistic variation, vol. 1: Theories and methods*, ed. by Peter Auer and Jürgen E. Schmidt, 391-419. Berlin: Mouton de Gruyter.

POPLACK, SHANA; JAMES A. WALKER; and REBECCA MALCOLMSON. 2006. An English 'like no other'?: Language contact and change in Quebec. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 51.185-213.

POPLACK, SHANA; LAUREN ZENTZ; and NATHALIE DION. 2012. What counts as (contact-induced) change. *Bilingualism: Language and Cognition* 15.247-54.

R CORE TEAM. 2015. R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org.

SANKOFF, DAVID. 1988. Sociolinguistics and syntactic variation. *Linguistics: The Cambridge survey (Vol. 4, Language: The socio-cultural context)*, ed. by Frederick Newmeyer, 140-61. Cambridge: Cambridge University Press.

SILVA-CORVALÁN, CARMEN, and ANDRÉS ENRIQUE-ARIAS. 2017. *Sociolingüística y pragmática del español*. 2nd edn. Washington, DC: Georgetown University Press.

SILVA-CORVALÁN, CARMEN, and JEANINE TREFFERS-DALLER (eds.) 2015. *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press.

THOMASON, SARAH G. 2001. *Language contact: An introduction*. Washington DC: Georgetown University Press.

TORRES CACOULLOS, RENA, and CATHERINE E. TRAVIS. 2018. *Bilingualism in the community*. Cambridge: Cambridge University Press.

UNITED STATES CENSUS BUREAU. 2014. Online: *2010-2014 5-Year American Community Survey*, http://factfinder2.census.gov.

Travis
School of Literature, Languages and Linguistics / ARC Centre of Excellence for the Dynamics of Language
The Australian National University
Canberra ACT 2600
[catherine.travis@anu.edu.au]

Torres Cacoullos
Department of Spanish, Italian, and Portuguese
The Pennsylvania State University
University Park, PA 16802
[rena@psu.edu]