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 (cánté ‘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 homogrown 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.
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)

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).](image)

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.
adapting 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 (NMC OSS) 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-- and they wouldn’t let us go out at recess,
yo no nos dejaban salir al recess, they would make us write,
nos hacían escribir, I won’t --
yo no s-- I shouldn’t --
o no debo espa-- speak Spanish,’
hablar español=I,

(NMC OSS 47-1A2, 275-284; born 1939, housewife, Albuquerque)
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 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, como los de México. ... porque .. aprendió más por ella, que por nosotros.
‘she spoke it very nicely, like people from Mexico. ... because .. she learned it more on her own, 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).](image)

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

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.
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 Cacoulls & Travis 2018:142), Mexico City (N = 1sg 939, 3sg 450; Lastra & Butragueño 2015:43), Madrid, Spain (N = 1sg 10,185, 3sg 2,501; Enriquez 1984:348), Castanher, Puerto Rico (N = 1sg 1,527, 3sg 603; Holmquist 2012:211), Cali, Colombia (N = 1sg 1,389, 3sg 1,413; Torres Cacoulls & 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).

<table>
<thead>
<tr>
<th>QUESTIONNAIRE ITEM</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preferred language</td>
<td>0.64</td>
<td>-0.37</td>
<td>0.17</td>
<td>0.32</td>
</tr>
<tr>
<td>First language</td>
<td>0.23</td>
<td>-0.07</td>
<td>0.34</td>
<td>-0.52</td>
</tr>
<tr>
<td>Self-rating (relative English-Spanish)</td>
<td>0.17</td>
<td>-0.11</td>
<td>0.34</td>
<td>0.53</td>
</tr>
<tr>
<td>How English learned</td>
<td>0.24</td>
<td>-0.07</td>
<td>0.39</td>
<td>-0.56</td>
</tr>
<tr>
<td>Preferred language for</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>... radio</td>
<td>0.53</td>
<td>0.37</td>
<td>-0.49</td>
<td>-0.09</td>
</tr>
<tr>
<td>... TV</td>
<td>0.13</td>
<td>0.81</td>
<td>0.42</td>
<td>0.17</td>
</tr>
<tr>
<td>... reading books</td>
<td>0.09</td>
<td>-0.05</td>
<td>0.11</td>
<td>0.05</td>
</tr>
<tr>
<td>Language spoken</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>... with family</td>
<td>0.005</td>
<td>-0.21</td>
<td>0.07</td>
<td>-0.004</td>
</tr>
<tr>
<td>... with friends</td>
<td>0.23</td>
<td>-0.02</td>
<td>-0.32</td>
<td>-0.09</td>
</tr>
<tr>
<td>... at work</td>
<td>0.30</td>
<td>-0.01</td>
<td>-0.22</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

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 $M = 20\%$, $SD = 8.9$. 

8
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).
Leandro: Los mexicanitos esos, they -- .. hablaban muy diferente el mexicano que nosotros, ‘those Mexican guys, they -- .. spoke Spanish very differently from us,’

Ricardo: Yeah. ‘Yeah.’

Leandro: Y nosotros hablábanos otra class de language sandwich <@ que [le dicen @>, @@@ de Nuevo México]. ‘and we spoke another kind of language sandwich <@ as [they say @>, @@@ of New Mexico].’

Ricardo: [@@@] ‘[@@@@]’

Leandro: Pero cuando fuimos allá, ‘but when we went there,’

Ricardo: Uh [huh]. ‘uh [huh].’

Leandro: [pues] estabamos como mudos ahí si ve. @ no sabíamos ni inglés, y no sabíamos el mexicano !bien. ‘[well] it was like we were mute you know. @ we didn’t know English, and we didn’t know Spanish well.’

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).

Rocío: … and he used to write letters.

…(1.4) to my tía’s boss. (((‘my aunt’s boss’)))

… and,

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

Adriana: .. pero muy <@ mocho @>. .. but very <@ broken @>.’
Rocío: .. [mocho]? ‘.. [broken]?’
Adriana: ‘[@@]’
Rocío: mhm. ‘mhm.’
Adriana: ‘[yeah.]’
Rocío: [hm lo] hablaban. ‘[hm they] spoke it. ...(0.7) the Spanglish.’

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).

Sandra: you know my brother was my teacher. ... we would -- .. coming back from the school? there in the arroyo?

and he would make -- a blackboard out of the sand. ...(0.7) mira, me decía, ...(0.8) lo=s -- .. el -- los ay bee cees, ((ABCs)) ... son .. puros palitos, .. y bolitas. ‘you know my brother was my teacher. ... we would -- .. coming back from the school? there in the arroyo? and he would make -- a blackboard out of the sand. ...(0.7) look, he would say to me, ...(0.8) the -- .. the -- the ABCs, ... are .. just little sticks, .. and little balls.’
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).
Monica: ‘we were always speaking, ... half and half Spanish, and half English.’
(NMSEB 11 El Trabajo, 00:44-00:48; born 1941, factory worker, Albuquerque)

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 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” (Myslin & Levy 2015:871).

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.
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: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


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


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]