Gauging convergence on the ground: Code-switching in the community*

Rena Torres Cacoullos
Pennsylvania State University, USA; ARC Centre of Excellence for the Dynamics of Language, Australia

Catherine E. Travis
Australian National University, Australia; ARC Centre of Excellence for the Dynamics of Language, Australia

Abstract
Is grammatical convergence between bilinguals' two languages inevitable and does code-switching inherently promote it? Despite the burgeoning of bilingualism studies, this question—and even what should count as code-switching—remains contentious. Cumulative scientific advances will depend on attention to the social context in which bilingual phenomena arise, proper handling of spontaneous speech data, and consideration of the probabilistic constraints underlying occurrence rates of linguistic forms. We put forward this program of study as implemented in systematic quantitative analysis of linguistic structures in the New Mexico Spanish-English Bilingual (NMSEB) corpus. This unique compilation of bilingual speech by members of the Hispanic northern New Mexican community in the United States records both borrowing and—vitally—copious multi-word code-switching. Advancing the study of bilingualism is community-based data collection and accountable analysis of the linguistic conditioning of variation in both of the languages in contact as used by the bilinguals themselves, in comparison with appropriate benchmarks, again of both languages (monolingual, or earlier, varieties). The role of code-switching in convergence is evaluated through a novel on-line measure, comparisons based on the proximity of spontaneous use of the other language. Implementation of this test of proximate code-switching confirms a disjunction between bilinguals' phonology, which is more labile, and morpho-syntax, which is stable. Variation is conditioned by intra-linguistic contextual features, the distribution of which, however, may shift under code-switching, shaping patterns in the bilingual community.

Keywords
Spontaneous code-switching, community-based data, linguistic conditioning of variation, contextual distribution hypothesis, New Mexico

*Both authors contributed equally to the introduction and this issue.

Corresponding author:
Rena Torres Cacoullos, Department of Spanish, Italian, and Portuguese, Pennsylvania State University, 31 Burrowes Building, University Park, PA 16802, USA.
Email: rena@psu.edu

The study of bilingualism in its social context

The bilingual speech community as the unit of study

Corpus- and lab-based approaches increasingly interact as interest grows in the link between language use and structure (e.g. Bybee, 2010) and between production and processing (e.g. Gennari & MacDonald, 2009). Yet mostly unaccounted for in both approaches is the priority of the community over the individual: speakers are uniformly affected by their community affiliation such that in their more careful speech styles all members of the community increase their rate of use of those linguistic forms which are favored by speakers of higher social status (Guy, 2013, p. 64); a well-known example is the social and stylistic stratification of syllable-final (r)-pronunciation in New York City (Labov, 1966). For bilinguals, "an individual's personal ability is operative but is mediated by the norms of his speech community" (Poplack, Sankoff, & Miller, 1988, p. 98).

The precedence of community norms for the study of bilingual phenomena was demonstrated in Poplack et al.'s (1988) mammoth study of borrowing, based on 20,000 tokens of single English-origin words extracted from conversations with 120 francophones in the national capital region of Canada. For overall borrowing rate, a better predictor than individual bilingual proficiency was social class membership, as measured by occupation (members of higher social classes tending to avoid loanwords). Furthermore, environmental influence, as measured by majority vs. minority status of French in the neighborhood of residence, was paramount for the proportion of instances of active borrowing (nonce words directly drawn from English vs. established borrowings transmitted via other francophones). The interpretation of this remarkable finding is that "behavior with respect to use of borrowings is *acquired*" (Poplack et al., 1988, p. 98, emphasis in original).

The priority of community norms is verified in this volume through systematic quantitative analysis of the New Mexico Spanish-English Bilingual (NMSEB) corpus. For example, single English-origin words incorporated in the Spanish of bilinguals in New Mexico include kinship terms *grandma* and *dad* as the preferred form whereas native *mamá* is more recurrent and widespread than *mom* (Aaron). As to single English verbs, an *hacer* 'to do'+VERB$_{Eng}$ construction (e.g. *yo hacía draw mejor* 'I drew better', lit. 'I did-SPAN draw-ENG') is productively employed (Wilson & Dumont). Such patterns cannot be detected in experiments of random university students nor in corpora constituted of amalgamated texts culled from a range of dialects or contact situations, which lack a basis in the speech community.

An ideal contact site for the observation of bilingual phenomena

Studies of communities formed by recent immigrants widely report changes in the immigrant language which occur abruptly, between generations. But to the extent that the population is transitioning to the majority language, the "changes" to the immigrant variety will necessarily be transitional, and hence such settings are not instructive as to contact-induced change. For example, differences observed between first-generation immigrants, who are taken as benchmarks, and second- or even third-generation speakers in Los Angeles (Silva-Corvalán, 1994) or New York (Otheguy & Zentella, 2012), may turn out to be transient characteristics, not destined to be transmitted if there is shift to English within three generations. Or, the contact situation may be interpreted as being "too young" for changes to have occurred, as with Turkish in the Netherlands (Doğruöz & Backus, 2007, p. 218).

In contrast, the long-standing nature of the contact situation in New Mexico, where Spanish and English have existed as the main competing languages for over 150 years, allows us to examine the outcomes of long-term contact. The contact site for the NMSEB corpus, the northern part of the US

southwestern state of New Mexico, colonized by the Spanish at the end of the 16th century, is home to "arguably the oldest continually spoken variety of Spanish anywhere in the Americas that has not been updated by more recent immigration" (Lipski, 2008, p. 193). This is known as Traditional New Mexican Spanish, traits of which have been documented as early as 100 years ago by Espinosa (1909) and more recently in a linguistic atlas by Bills & Vigil (2008).

Since 1850, when New Mexico was pronounced a Territory of the United States, the proportion of non-Spanish-speaking English speakers has grown, with precipitating events most likely including the arrival of the railroad in 1878. In 1890, 70% of the New Mexican population reported on the US Census that they "could not" speak English, a figure which dropped to 51% in 1900 and to 33% in 1910 (Fernández-Gibert, 2010, p. 48). Following statehood in 1912, English was increasingly imposed, supplanting Spanish in the school system even in northern rural communities by the mid 1940s (Gonzales, 1999, p. 20; Lipski, 2008, p. 203). Where Spanish is taught in the public schools today, it is primarily as a foreign or second language and to the detriment of the local variety (cf. Gonzales-Berry, 2000). Shift to English is underway in this community, evidenced partly in the fact that in the 2010 census 40% of those who identified themselves as Hispanic (or Latino) reported speaking no Spanish (American Community Survey one-year estimates). Thus, New Mexican Spanish-English bilinguals represented in the NMSEB corpus are of precious value.

Given the manifold factors involved in change, despite the attractiveness of proposed classifications of contact situations (for example, Thomason and Kaufman's (1988) distinction between borrowing and substratum interference), there is no model by which to predict whether change will occur, much less to determine whether it is promoted by code-switching. This can only be assessed on a case-by-case basis for the particular community and linguistic structure. The papers in this volume are able to do just that by capitalizing on the highly bilingual nature of the NMSEB corpus, which achieves the recording of copious multi-word stretches—unambiguous code-switches—in each language. For example, Balukas and Koops, in their study of voice onset time (VOT) find that the majority of the /ptk/-word tokens examined fall within a 5-second window following the code-switch (English 66%; Spanish 81%). Benevento and Dietrich find that close to one-quarter (157/699) of the tokens of Spanish expressed 1sg subjects in their study occur with multi-word English in the same or preceding clause (cf. also Torres Cacoullos & Travis (2015b, Table 1) who obtain a similar proportion of tokens in maximal proximity to code-switching (424/1494) for pre-verbal and unexpressed 1sg subjects in this corpus).

Bilinguals: the criterion of regular use

How do we select a bilingual speaker sample? Because individual differences in learning history and current "language use and activity" impinge on (models of) bilingual processing (Kroll & Tokowicz, 2005:532), studies pooling assorted participants of unknown extralinguistic characteristics, including heterogeneous groups of university students, fail to yield interpretable results. This is not (merely) a matter of screening participants for proficiency or other individual traits as assessed by different tests, nor is it one of classifying participants according to self-reports of code-switching or domains of language use. For bilinguals just as for monolinguals, "the behavior of an individual can be understood only through the study of the social groups of which he or she is a member" (Labov, 2010, p. 7).

The speakers comprising the NMSEB corpus are at least third-generation *Nuevomexicanos* 'New Mexicans', all from northern New Mexico, the heart of Traditional New Mexican Spanish, where contact with recent Mexican immigrants has been minimal. Most of the 40 participants hail from counties that have a high proportion of Hispanics (47–80%), indicated by the darker shades in the map in Figure 1. In particular, NMSEB participants were born in Rio Arriba (11 of 40
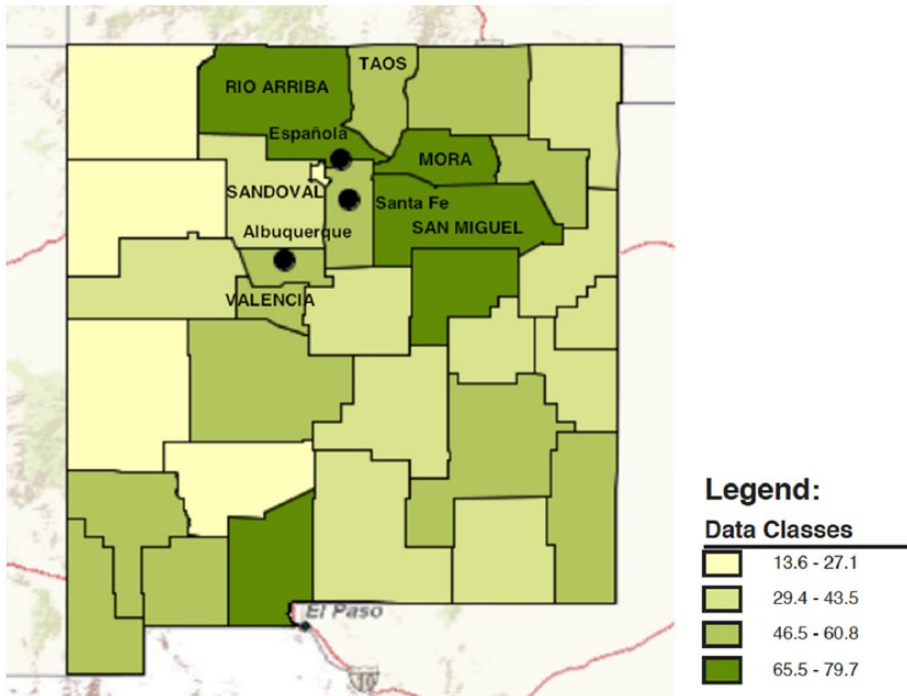
Figure 1. Major birthplaces of NMSEB participants (counties and Cities) and percent of people who are Hispanic or Latino in the state of New Mexico (2007–2011 American Community Survey 5-Year Estimates).

participants) and Taos (10/40), as well as the city of Española (8/40) (and most, at the time of the interview, resided either in the county of their birth or in a neighboring county, 33/40). NMSEB speakers, born between 1923 and 1989, are 23 women and 17 men whose occupations include miners, ranchers, schoolteachers, and financial administrators (see Appendix 1 and Travis & Torres Cacoullos (2013) for more details).

Choosing a sample of bilingual speakers by subjecting them to a battery of proficiency tests (such as those administered for experimental studies) would have been inappropriate in the New Mexican context, given the association of such tests with the very kinds of educational institutions that have demoted New Mexican speech varieties (cf. Sankoff, 1988, p. 145). It would also be of dubious value, given the prescriptive bias of (many) such tests (see Dąbrowska, 2012, on the effect of education on performance in experimental tasks).

The criterion for inclusion adopted in NMSEB is that of regular *use* of both languages (cf. Poplack, 1993, p. 254), as observed over time by the fieldworkers or as reported to the fieldworkers by other acquaintances. The speakers comprising NMSEB are bilingual in that they regularly use both languages with the same interlocutor in the same domain. Speaker bilingualism as per the use criterion is verified by triangulating data from:

(1) *Self-reports in response to questionnaire items.* NMSEB participants can be considered early bilinguals (though this may be an idealized notion (cf. Grosjean, 1998:133)): their first language was Spanish (35 of 40 participants), they learned English at school (31/40), and they prefer

to speak English (20/40), Spanish (12/40) or both (8/40). In the aggregate, they use both languages not just in the home (31/40), but also with friends (31/40) and at work (24/40) (Travis & Torres Cacoullos, 2013:182). It should be noted that self-ratings cannot be taken as mathematically exact figures. One participant (Inmaculada) who was interviewed in 2010 and again 1 year later rated her Spanish as a 5 on a 5-point scale both times, but her English as a 4 the first time and a 5 the second; and while she gave English as her preferred language in 2010, she described both as equally preferred in 2011.

(2) *Content analysis of the recordings.* Systematic extraction of socio-demographic, linguistic-history and language-attitude information arising naturally in the course of the recordings complements and in some cases even corrects information from self-reports. For example, one speaker told of learning to read English from her brother, seen in (1). Classifying her as a bilingual who learned one of her languages in school would fail to capture the role of her family in her acquisition of English.

(3) *The production data themselves*, which comprise stretches of both Spanish and English (see section on code-switching as an in-group discourse mode below).

(1)

Sandra   *you know my brother was my teacher.*
          ... (H) we would --
          .. coming back from the school?
          there in the arroyo?
((10 intervening lines of transcription))
          and he would make --
          a blackboard out of the sand.

[03 Dos Comadres, 0:58:01-0:58:17][1]

Of course, speakers may also display different *degrees* of bilingualism, operationalized by tabulating individual characteristics. Again proficiency tests are of little use here; even if they were reliable for speakers of stigmatized varieties, they provide no window on bilinguals' actual experience with their two languages. Sufficient extralinguistic characterization of the bilingual speakers is indispensable, because if convergence is in progress, linguistic behavior should show correlations with some external measure of contact.

Measures of contact enlisted have been the proportion of speakers of each language in the neighborhood (Poplack et al., 1988, p. 51), age with respect to a relevant milestone (such as Quebec Bill 101 which ratified French as the official language in 1977) (Poplack, Walker, & Malcolmson, 2006, p. 187) and immigrant generation (Silva-Corvalán, 1994, p. 15-16). If the hypothesis is one of inter-dialect influence, as for Spanish in New York (Otheguy & Zentella, 2012), pertinent measures would be demographic makeup of the neighborhood of residence by national-origin group (Hernández, 2011) or more direct indicators of the extent of a speaker's interactions with speakers of other varieties.

Once speakers are adequately characterized with respect to their social roles, phenomena surmised to be of a bilingual nature may turn out to be conditioned by social class (occupation or education). For example, use of the French subjunctive mood, presumed to be undergoing attrition in Canada due to contact with English, is explained by neither speaker age nor contact with English, whether at the level of the individual (English proficiency) or the community (proportion of francophones in neighborhood). Poplack (1997, p. 292) shows that the explanation is class based: a

genuine predilection of professionals for the subjunctive is revealed (once the lexical effect of the association of the subjunctive with *falloir* 'have to' is accounted for).

A class-based effect is also proposed by Shin and Otheguy (2013, p. 436, 441-442) for Spanish subject pronoun use in New York, based on a higher rate in Colombians and possibly Cubans who have lived in the city more than five years compared with newcomers; these are precisely those Latino national-origin groups whom census data indicate to have high affluence rankings. High-status occupations may be associated with shift to English, as Bills, Hernández Chávez and Hudson (1995, p. 23) show for Southwestern states. Alternatively, if affluent Latinos are susceptible to influence from English and this is due to looser social networks and more interaction with anglophones, as Shin & Otheguy (2013, p. 442-443) hypothesize, this could be fruitfully explored by operationalizations of density of interactions.

Degree of bilingualism can also be measured through distributions of linguistic features, for example the proportion of nonce vs. established loans in a speaker's recorded speech (Aaron) (though care must be taken that frequencies are not epiphenomenal of, for example, topic). In their study of the bilingual *hacer* 'to do'+VERB$_{Eng}$ device, Wilson and Dumont classify speakers according to proportion of bilingual Intonation Units (see the section on prosody-based transcription, below).

Code-switching as an in-group discourse mode: Community-based field methods

Multi-word code-switching may (or may not) be a community discourse mode (Poplack, 1998, 2000 [1980]). Based on ethnographic participant-observation, code-switching has been described as "the appropriate code for the Hispano community" of northern New Mexico (Gonzales, 1999, p. 29). That is, *Nuevomexicano* Spanish-English bilinguals regularly code-switch without particular rhetorical motivations. As one participant put it,

(2)

Miguel:	…*(1.1) cuando estoy con mexicanos*,	'…(1.1) when I'm with Mexicans,
	o algo que --	or something that --
	.. son de México?	.. they're from Mexico?
	.. como que entonces quiero hablar es- --	.. like then I want to speak Span- --
	.. <@ inglés @>.	.. <@ English @>.
	.. como que se me sale más easy.	.. like it comes out more easy.
	((6 intervening lines of transcription))	
	... y en inglés,	… and in English,
	when I'm speaking,	*when I'm speaking*,
	… hay veces que quiero poner una ..	… there are times when I want to put a ..
	Spanish word in there.	*Spanish word in there.'*

[04 Piedras y gallinas, 1:11:03-1:11:24]

Who the fieldworker is and where the recordings are made are critical to obtaining samples of multi-word code-switching (rather than merely single-word items or tags). As the use of code-switching as a discourse mode in New Mexico depends on whether the interlocutors belong to the speech community (Gonzales, 1999, p. 30), code-switching data can only be collected by in-group members (cf. Clyne, Eisikovits, & Tollfree, 2001, p. 235-236; Poplack, 1993, p. 260). Thus, the fieldworkers for NMSEB were *Nuevomexicano* students of the University of New Mexico who were community insiders by virtue of their ethnicity and relationship with the participants, who were family or family acquaintances. The recordings were made in the home of the participant or that of a family member or friend.

To circumvent observer effects, NMSEB is based on the technique of the "sociolinguistic interview" (Labov, 1984). Fieldworkers were instructed to speak as they naturally would and to formulate questions from an insider's point of view, for example, about childhood games in the acequias ("irrigation ditches"). A primary goal is to gather narratives of personal experience, where "style is regularly shifted towards the vernacular" (Labov, 1984, p. 32), which is least affected by self-monitoring or hyper-correction and thus provides the most systematic data for linguistic analysis (Labov, 1972, p. 208; 1984, p. 29). Also gathered are dialogic conversations between speakers of the same vernacular. Example (3) illustrates the kind of vernacular features we are able to observe in these data, including

- single-word items: English-origin *misstiste* 'miss-Preterite-2sg' surrounded by Spanish (line 6), Spanish-origin *entonces* 'so, then' surrounded by English (lines 7-8);
- spontaneous switching between speakers (lines 1 and 2);
- and multi-word code-switching within the clause (and prosodic unit) (line 4).

(3)

1.	Fabiola	*.. who's going to the party?*	'.. who's going to the party?'
2.	Molly	*...(0.7) no sé no,*	'…(0.7) I don't know no.'
3.		*.. no más me dijo que,*	.. she just said that,
4.		*.. (H) .. que le iban a hacer algo at five or six something.*	.. (H) .. that they were going to do something *at five or six something.'*
5.	Fabiola	*.. oh this evening?*	*'.. oh this evening?*
6.		*… oh so entonces no misstiste nada.*	… oh so you didn't miss anything.
7.		*…(1.3) you'll be okay entonces.*	…(1.3) *you'll be okay then*.
8.		*... you'll be fine,*	… *you'll be fine,*
9.		*como decía el Chalo.*	as Chalo would say.'

[09 La salvia, 0:44:56-0:45:11]

Through the community-based method of sociolinguistic interviews conducted by in-group members, the NMSEB corpus achieves a roughly even distribution of speech produced in Spanish and English by the same speakers (though amounts of Spanish and English vary across the recordings, as noted by Wilson & Dumont). This allows comparing the language varieties in actual contact, as in the examination of convergence in VOT (Balukas & Koops) or of (dis)similarities in the marking of discourse properties of nouns (Aaron).

Critically, NMSEB records abundant smooth multi-word code-switching. This sets it apart from corpora in which material from the other language is sparse or is restricted to single-word items, whether due to data collection methods or as a reflection of the fact that in most bilingual communities single-word borrowings predominate (e.g. Poplack, 2000, p. 222). The copiousness of alternations between stretches in each language allows testing for the effect of use of two languages using on-line measures, as we describe below (see the section on code-switching in NMSEB).

Handling a corpus of spontaneous bilingual speech

Community-based transcription

A corpus that can be used for accountable analysis requires comprehensive transcription. For NMSEB, transcription was done in the program *ELAN* (Lausberg & Sloetjes, 2009), in which each line of transcript is aligned with the corresponding audio. This alignment increases accuracy at the

time of transcription and, in providing the researcher ready access to the original speech data, can also enhance subsequent analysis.

To maximize searchability, we make use of standard orthography. We make no attempt to capture phonetic realizations, because the many decisions to be made are best left to the analyst as they relate to the specific study (as in Brown's study of initial /d/ weakening and Balukas & Koops' study of Voice Onset Time). For example, despite variable aspiration and elision of /s/ in Traditional New Mexican Spanish (cf. Brown, 2005), we transcribe an *s* in all cases where plural meaning was clear from the context (thus, always *los muchachos* 'the boys', and never *loh muchachoh* or *lo muchacho*, for example).

Consistent with this, we make no attempt to capture the more English- or Spanish-like realization of English-origin forms; for example we use the English spelling of *grandpa* and *grandma*, despite their varied pronunciation in the corpus. This enables us to systematically extract such items and calculate their recurrence (overall frequency of use) and diffusion (use across different speakers in the community), rather than attempting to tag them at the time of transcription based on phonetics, which, as is well known, is a poor predictor of the status of an item as a loanword or a code-switch (cf. Poplack & Meechan, 1998, p. 134). Aaron demonstrates that such kinship terms are established in this community (as noted above), and in applying morpho-syntactic tests, is able to determine that single English-origin nouns are mostly borrowings into Spanish.

Given that variation is a significant linguistic fact to be confronted, following Poplack (1993, p. 265-266), we do represent morphological and lexical variants of the community. Thus, we capture forms such as the first-person plural object pronoun *los* (as in the third line of example (13), vs. standard *nos*) (Bills & Vigil, 2008, p. 145), first-person singular Perfect auxiliary *ha* (standard *he* e.g. *ha visto* vs. *he visto* 'I have seen'), and /b/ in Imperfect forms (e.g. *traíba* 'I brought' vs. *traía*) (cf. Bills & Vigil, 2008, p. 74). We use non-standard orthography for the few cases of accepted abbreviations such as *cause* (for *because*) in English and *pa'* (for *para* 'for/toward') in Spanish. To ensure uniformity, all transcribers followed a Transcribing Protocol developed specifically for this project.[2]

Just as the data collection is community based, so too must be the transcription. Speakers of other varieties of Spanish and/or English unfamiliar with the community display misunderstandings. For example, in an earlier project a linguistics graduate student from Puerto Rico, otherwise an excellent transcriber, transcribed "Tierra María" instead of "Tierra Amarilla" (realized with the weakened /y/ variant).

Prosodic units and the structure of spoken discourse

For the analyst to accurately understand what is being said from the transcription, detail beyond the individual words needs to be included. Particularly important is prosody, which works together with syntax to delineate relevant units of spoken interaction. As an illustration, consider example (4). As transcribed here, there are two possible interpretations of the 1sg pronoun *yo* (in bold): it could be a post-verbal subject on *dije* ('said I'), or a preverbal subject on *no puedo* ('I cannot'). It is only once prosodic units are marked, as in (5) (appearing on different lines)[3], that these clauses can be demarcated: *yo* is post-verbal subject on *dije* (as there is no prosodic break here), while *puedo* has an unexpressed subject.

(4)

   Ivette: <u>*dije* **<u>yo</u>** *no no* <u>*puedo*</u> *estar yendo pa' atrás y pa' adelante.*</u>
                  '.. I said no I can't be going backwards and forwards.'

(5)

Ivette	*.. dije yo*,	'.. I said,
	No.	No.
	no puedo estar yendo pa' atrás y pa' adelante.	(I) can't be going backwards and forwards.'

[06 El túnico, 0:43:54-0:43:57]

The prosodic unit has been a construct of linguistic analysis at least since Halliday (1967); in NMSEB, we follow the transcription method outlined in Du Bois, Schuetze-Coburn, Cumming and Paolino (1993), which employs the Intonation Unit (IU), "a stretch of speech uttered under a single coherent intonation contour" (Du Bois et al., 1993:47).

The IU makes a difference in interpretation of examples such as (5) because, in accordance with observed correlations between syntax and prosody, there is a very strong tendency for pronouns to occur in the same IU as the verb. In a sample extracted from NMSEB, 95% (664/696) of pre- and post-verbal expressed 1sg subject pronouns occur in the same IU as the verb; in monologic Pear Story narrative data of American English, this was categorical (Croft, 1995, p. 859). Indeed, when a subject pronoun occurs in a different IU from the verb, it is not always straightforwardly a subject, as for example for the *yo* in line 3 in (6) (where (I) in parentheses in the translation indicates the possibility of *die* having an unexpressed subject). Hence, Benevento and Dietrich exclude such tokens from their analyses of word order, as have we in our studies of subject expression (Torres Cacoullos & Travis, 2015a, 2015b; Travis & Torres Cacoullos, 2012).

(6)

1.	Rocío	*.. yo creo que ahí me voy a quedar hasta que=*,	'.. I think that I am going to stay there until,
2.		*.. me muera*,	.. (I) die,
3.		**yo**.	I.
4.		*o ella*.	or her.'

[05 Las tortillas, 0:46:32-0:46:36]

The IU thus aids in delimiting the clause. It also proves to be a more profitable unit of analysis in spoken discourse than the highly nebulous "sentence" (e.g., Miller, 1995, p. 132). A particularly clear example of this can be seen in attempts to account for unexpressed subjects in English as occurring only in "initial position", based on an undefined notion of the sentence (cf. Napoli, 1982, p. 99; Roberts & Holmberg, 2010, p. 5). Torres Cacoullos and Travis (2013, p. 11) are able to pin down this constraint and demonstrate that it is prosodic-initial position that constrains expression: outside of coordinated contexts, unexpressed subjects in English virtually only occur in IU-initial position, illustrated in (7). We will see below (see the section on conflict sites) that this prosodic constraint provides a rigorous test of convergence in subject expression.

(7)

Manuel: *.. I was able to scramble and*,
      .. find the --
      my other flashlight,
      ... Ø turned it on,
      Ø (H) worked on the one I had just broke.

[16 Trip to Africa pt.1, 0:26:06-0:26:12]

Prosodic segmentation is no less relevant for the study of phonetics than it is for the study of morphosyntax. In this volume, Balukas and Koops call on the IU to delimit their tokens for the study of VOT, and Brown does so to identify the preceding phonological segment conditioning reduction of /d/.

The difference transcription makes to the analysis of bilingual discourse

IU-based transcription is of particular value for bilingual discourse. Although an early distinction was made between 'intra-' and 'inter' sentential code-switching (Myers-Scotton, 1993; Poplack, 2000 [1980]), the points at which languages can be switched are insightfully examined through the IU. Most code-switches occur across rather than within IUs. That is, cases such as that within line 1 in example (8) are less common than cases such as that between lines 2 and 3. Wilson and Dumont report that just under 10% (2587/28,500) of multi-word IUs contain Spanish and English words (or combinations of words). Thus, despite frequent code-switching in NMSEB, the speakers tend to keep the two languages separate prosodically.

(8)

1.	Victoria	*…(0.8) porque si no lo hago while it's in my head*,	'…(0.8) because if I don't do it *while it's in my head*,
2.		*well then*,	*well then*,
3.		*no se hace*.	it doesn't get done.'

[12 Juego de scrabble, 0:09:47-0:09:51]

Delimiting syntactic units is important for consideration of the ways in which languages are combined. Benevento and Dietrich, following Torres Cacoullos and Travis (2015b), seek to identify cases of code-switching within the one constituent. Completion of the constituent is defined as the point at which nothing further is 'projected' neither syntactically nor prosodically (Hopper & Thompson, 2008). Thus, included as a switch within the one constituent are examples such as (9), where multi-word Spanish and English occur in the same IU, and (10), where multi-word Spanish and English occur in different IUs but the first IU projects subsequent material both syntactically (e.g. a verb that projects a direct object) and prosodically, with continuing intonation (marked by a comma at the end of the IU, which represents a slight rise in pitch and gives an indication that the speaker will continue speaking (Du Bois et al., 1993, p. 53)). However, English is not counted as having occurred in the same constituent if it occurs when the material produced in the following IU is not projected either syntactically or prosodically (such as in cases of period intonation), as in (11), where *con petticoats* 'with petticoats' is an adjunct.

(9)

| Miguel | *… pues le dije que I was gonna go= y*, | '… well I told him that *I was gonna go=* and,' |

[04 Piedras y gallinas, 1:07:02-1:07:04]

(10)

| Monica | *.. yo creo que tendríanos*, | '… I think that (we) would have been, |
| | *.. maybe twelve or thirteen years*. | *.. maybe twelve or thirteen years.'* |

[11 El trabajo, 0:43:56-0:43:59]

(11)

| Ivette | *.. vide .. todas las girls.* | '.. I saw .. all the girls. |
| | *...(1.7) con .. petticoats.* | …(1.7) with .. petticoats.' |

[06 El túnico, 0:13:20-0: 13:24]

The IU can also inform the identification of switch points. In the following example, the speaker ends line 2 with the English-origin item *taite* 'tight', and then continues in a new IU in English. On the basis of the presence of the prosodic break here, it could be argued that *taite* should be treated as a single English-origin item, to be examined as a candidate borrowing (cf. Aaron), and that the switch takes place in line 4.

(12)

1.	Inmaculada	*.. te castigó Dios,*	'.. God punished you,
2.		*por an- andar de taite.*	for being tight ((with money)).'
3.	Lucy	*[mhm]*.	'[mhm].'
4.	Inmaculada	*[(H)] so that was funny*.	'[(H)] *so that was funny.'*

[08 Graduación Familiar Pt.2, 0:05:11-0:05:15]

Finally, as well as prosodic units, NMSEB transcription annotates a range of prosodic features that can be drawn on for testing proposed factors relevant to bilingual speech. In their examination of *hacer* 'to do'+VERB$_{Eng}$, Wilson and Dumont operationalize truncation and IU-internal pauses as diagnostics of cognitive effort to test the hypothesis that this bilingual compound verb is associated with a heavier cognitive load than monolingual speech.

In sum, there are new insights to be gained from incorporating the prosodic feature of the IU into analyses of bilingualism, including repercussions for assessing the sites and smoothness of code-switching. Paramount is that the transcription of the corpus be responsible (faithful to the speech community) and exhaustive (rather than opportunistic), to enable systematic quantitative analysis of a range of phenomena.

Quantitative reasoning

For linguistic change to take place, there must be diffusion of a new pattern across a community (Weinreich, Labov, & Herzog, 1968), i.e. quantitative regularities that differ from pertinent benchmarks.

Linguistic conditioning: Constraints beneath (equivocal) rates

Convergence is construed as an increase in similarity between bilinguals' two languages, raising the question of how similarity of linguistic structure may be evaluated. What kind of quantitative analysis is capable of substantiating change by pinpointing linguistic (dis)similarity? Pronouncements of convergence fabricated from intuited deviations from a native-speaker norm overlook the fact that speech is characterized by variability, in bilingual just as in monolingual settings. Once language variation is acknowledged, the next pitfall to overcome is the conflation of variation with change in progress. How can these be distinguished?

Often invoked as evidence of change is a difference in rate of use, such that a rise in the frequency of a variant with an apparent counterpart in the contact language, at the expense of

the variant without such a counterpart, is interpreted as contact-induced change. An example would be an increased rate of expressed (vs. unexpressed) subject pronouns in Spanish in contact with English, attributed to the overwhelming preference for expressed subjects in the latter. A difference in overall rates alone, however, is an equivocal test of convergence for three reasons.

First, vast differences in rates of expression are found within non-contact varieties of Spanish. For instance, the reported rate of first-person singular *yo* 'I' is double in a Cali, Colombia vs. a Mexico City study (respectively, 49% (Travis & Torres Cacoullos, 2012, p. 726) and 25% (Lastra & Butragueño, 2015, p. 43)). Though a difference in mean values may be statistically significant, this leaves unanswered the linguistic import of such a difference.

Second, a more certain characterization of linguistic structure is the linguistic conditioning of the variation—the magnitude and especially the direction of effect of probabilistic constraints on choice of one over another variant of a *linguistic variable* (Labov, 1972a, p. 94). Whereas the rate of a variant may differ according to interviewer (Hernández, 2011), dialect (Cameron, 1994), genre (Travis, 2007) or other considerations extraneous to linguistic structure, the quantitative effects of factors (contextual features) on the choice of an expressed subject are impervious to the vicissitudes of such rate fluctuations. Travis (2007, p. 130) demonstrates that the lower rate of 1sg subject pronouns in monologic narrative data as opposed to interactive conversation is not due to different constraints in the two genres (in both, operative is the widely found effect of subject continuity whereby coreferential contexts disfavor subject expression), but that it is the contextual distribution (see section on contextual distribution below) that gives rise to this particular genre difference, as subject continuity tends to be greater in narratives.

Third, even where overall rate differences may correspond to change, they are uninformative as to the direction of that change (Poplack, Zentz, & Dion, 2012b, p. 250). Illustrating again with Spanish subject pronouns, while convergence with English has been predicted to result in a 'higher' rate of expression, a 'lower' rate could also be interpreted as a contact-induced change, if it could be shown to involve erosion of bilinguals' sensitivity to factors favoring expressed subjects (Silva-Corvalán, 1994, p. 163, cf. Sorace, 2004). This brings us back to the linguistic conditioning: while overall rates are equivocal, (the existence and nature of) change is evinced through change in the linguistic conditioning of variation, or the probabilistic grammar(s) of the speakers in contact.

This is borne out in Benevento and Dietrich's study of subject-verb order in NMSEB. They find that the rate of post-verbal first singular subject pronoun *yo* in New Mexican bilinguals is higher than in some monolingual varieties, which would seem to counter a convergence interpretation, but lower than in an earlier variety of New Mexican Spanish, which might sustain such an interpretation. The superior test of linguistic conditioning of variable postposing, which they reveal to be parallel to that reported for monolingual varieties, fails to support convergence under code-switching, and demonstrates no erosion of lexico-pragmatic factors.

Benchmarks for comparisons: When is 'change' change?

It is generally acknowledged that integral to tests of convergence are comparisons of bilingual with benchmark varieties. But just what varieties serve as benchmarks? It cannot be the educated standard, as the NMSEB speaker in example (13) admonishes, where she recounts how a teacher marked as wrong the Spanish homework she helped her granddaughter with, stressing that "proper Spanish" is not "our Spanish".

(13)

Inmaculada	*(H) so me senté áhi yo,*	'(H) so I sat there,
	yo y mi suegra=,	me and my mother-in-law,
	(SNIFF) los sentamos y c- --	(SNIFF) we sat and --
	y y=,	and and,
	y= le ayudamos.	and we helped her.
	(H) pues agarró todo mal.	(H) well she got everything wrong.'

((21 intervening lines of transcription))

	.. they c- called it proper Spanish.	*.. they c- called it proper Spanish.'*
Lucy	*mhm.*	'*mhm.*'
Inmaculada	*o=r,*	'*o=r,*
	whatever,	*whatever,*
	it was called,	*it was called,*
	(H) but it wasn't our Spanish.	*(H) but it wasn't our Spanish.*'
Lucy	*hm.*	'*hm.*'
Inmaculada	*so she got everything wrong.*	*'so she got everything wrong.*
	(H) so I went to the school,	*(H) so I went to the school,*
	and I complained.	*and I complained.*
	(H) and I said wait you can't,	*(H) and I said wait you can't,*
	X,	*X,*
	.. mark her wrong,	*.. mark her wrong,*
	because that's how .. [the] community	*because that's how .. [the] community,'*
Anita	*[X].*	*[X].*
Inmaculada	*communicates.*	*communicates.*'

[14 Proper Spanish, 0:25:55-0:26:32]

The points of comparison that provide stringent tests of convergence are depicted in Figure 2, at the center of which are the Spanish and English of the New Mexican bilinguals represented in NMSEB (cf. Poplack & Levey, 2010).

Strict tests of contact-induced change must begin with comparisons with a pre-contact variety or at least an earlier stage of the same variety, in order to ascertain whether change has actually occurred. An example is the corpus of 19th century vernacular Quebec French used to compare patterns of preposition placement in the capital region of Canada (Poplack, Zentz, & Dion, 2012a). For immigrant contact varieties, the benchmark has been the generation of speakers who arrived as adults, corresponding to the bilinguals usually by national origin (Otheguy & Zentella, 2012; Silva-Corvalán, 1994). For New Mexican bilinguals an earlier-stage benchmark is found in the older speakers recorded for the New Mexico Colorado Spanish Survey (cf. Bills & Vigil, 2008) which include people born as early as 1897 and who are evidently Spanish dominant, with few multi-word English strings and minimal single-word English items in their recordings (largely limited to discourse markers *yeah* and *you know*). This comparison group is depicted at the top of Figure 2.

If comparisons with an earlier stage confirm a change among bilinguals, then it remains to be ascertained whether it is contact induced. This is accomplished by comparisons between varieties of both of the bilinguals' languages, pivoted on conflict sites.

Conflict sites: When is change contact induced?

Having determined that change has taken place, to demonstrate that it is one of convergence requires pinpointing *conflict sites*, "areas at which the structures of the language pair do not match" (Poplack & Meechan, 1998, p. 132). This is because despite apparent gross differences there may be
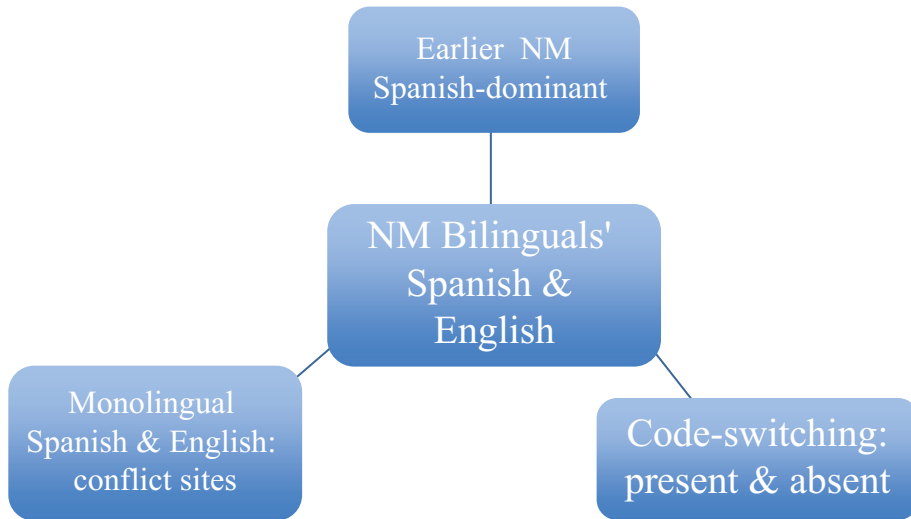
Figure 2. Linguistic comparisons for a test of convergence in New Mexican bilinguals: a pre-contact variety (top) (see the section on benchmarks for comparisons); determination of conflict sites through cross-language comparisons across monolingual Spanish and English (left) and of the Spanish and English of the bilinguals themselves (center) (see the section on conflict sites); comparisons of patterns when the speakers have recently code-switched to English vs. when they have not (see the section on code-switching in NMSEB).

commonalities between the languages in contact (for example, due to cross-linguistic tendencies). Conflict sites are verified by the comparisons between monolingual varieties indicated at the left of Figure 2.

For example, even though Spanish and English are classified, respectively, as a null- and non-null-subject language, there are both shared and conflicting constraints on subject expression (Torres Cacoullos & Travis, 2015a, pp. 91–97, 2015b). On the one hand, a shared constraint is coreferential structural priming such that speakers tend to repeat the same form: they are more likely to use an unexpressed subject when the previous mention of the coreferential subject was unexpressed. On the other, English shows the prosodic constraint whereby, outside of coordinated contexts, unexpressed subjects are only found in absolute initial prosodic position, as in (7) above. In English, then, (already high) expression rates are highest in non-IU-initial position; in Spanish, in contrast, rates of expressed subjects are higher in IU-initial position.

Spanish-English bilinguals' behavior at the conflict site of position in the IU serves to diagnose convergence in subject expression. If we observe a higher rate of Spanish pronoun expression in non-IU-initial vs. IU-initial position (as per the English but contrary to the Spanish pattern) this would constitute solid evidence for English influence. In NMSEB there is no such similarity in direction of effect (Torres Cacoullos & Travis, 2015a, p.93; 2015b).

Further strong evidence for or against convergence is found in comparisons of the two languages as they occur in the speech of the bilinguals themselves, depicted within the center rectangle in Figure 2. The nature of NMSEB, as a corpus in which all participants freely switch between Spanish and English throughout, is ideal for just such tests. Aaron undertakes this kind of comparison to consider the distribution of nouns across determiners in the same speakers' Spanish and English, and finds differences between them. One conflict site lies in the ratio of definite to indefinite articles, six-to-one in their Spanish but two-to-one in their English. The same kind of finding

is reported by Benevento and Dietrich for subject pronoun placement, which is robustly variable in bilinguals' Spanish but categorically preverbal in their English sample. Comparing the linguistic conditioning of variation, Torres Cacoullos and Aaron (2003) identified differences in the direction of effect of constraints on bare nouns in the New Mexican Spanish vs. English produced by the same speakers. Similarly, Poplack et al. (2012a) showed that French-English bilinguals in the national capital region in Canada adhere to the French model of preposition placement while speaking French, and to the English model while speaking English.

If a change is shown and if, in addition, it is shown to be contact induced, whether code-switching promotes that change must also be empirically tested. In the papers in this special issue, this is achieved via the comparison depicted at the right of Figure 2.

Code-switching in NMSEB: Does it promote contact-induced change?

Since code-switching is an on-line phenomenon, synchronic tests of its role in change must be devised. One way is to compare groups of code-switchers. In their study of preposition stranding as a variant of preposition placement in Canadian French (e.g. *J'avais pas personne à parler avec* 'I had no one to talk to'), Poplack et al. (2012a) compared French-English bilinguals whom they classified as "copious" code-switchers (those with 20 or more switches per recording) with "sparse" code-switchers (those with fewer than 20 code-switches). They found no difference between the two groups. Instead, the mechanism giving rise to a construction superficially similar to English preposition stranding is French-internal analogy, operationalized in terms of parallel contextual effects on phrase-final prepositions in relative clauses and bare prepositions in a native (non-contact) French transitive construction.

A novel synchronic test implemented in the papers in this volume is to make comparisons based on the proximity of spontaneous code-switching (Balukas & Koops, Benevento & Dietrich). Torres Cacoullos and Travis (2010, 2011) classified tokens of variably expressed Spanish subject pronouns according to their context of occurrence, whether in the absence or presence of code-switching by the same speaker (thus providing an operationalization for spontaneous speech of "language mode" (Grosjean, 1998, p. 136)). In NMSEB, the presence of an English multi-word sequence in the same or immediately preceding clause as a measure does not raise the rate of preverbal *yo* (Benevento & Dietrich), nor does it raise the rate of subject expression (Torres Cacoullos & Travis, 2015b, Figure 2). The evidence that Spanish patterns of subject realization (expression and position) do not converge with those of English in this long-standing bilingual community, even in the presence of maximally proximate English strings, constitutes a strong refutation of the hypothesis that code-switching inexorably stimulates convergence between bilinguals' grammars.

The stability of morphosyntax that has been demonstrated in tests of contact-induced change-in-progress is distinguished from the comparative lability of phonetics (cf. Sankoff, 2002). Employing the test of proximity to code-switching to probe phonetic convergence in NMSEB, Balukas and Koops find that there is no effect on bilinguals' Spanish VOT but that average VOT duration values in their English are lower (i.e. in the direction of Spanish) following a code-switch. In these same materials, Brown reports, for Spanish, a higher rate of the stop variant of word-initial /d/ (as opposed to the reduced [ð]) in "cognate" words (with phonological and semantic overlap) with English. Note that a parallel disjunction between the morphosyntactic (e.g. inflections, word order, determiner distribution) and phonetic adaptation of borrowings has been observed. Thus, there is morphosyntactic integration of lone other-language items regardless of their frequency or status as a dictionary-established loanword, whereas phonological integration is variable and increases with frequency and dictionary attestation (e.g. Aaron, Poplack, 2012, p. 647).

Contextual distribution in bilingual speech

Finally, variationist analysis of spontaneous bilingual speech suggests an alternative to the convergence via code-switching hypothesis. As per this *contextual distribution-via-code-switching* hypothesis, the workings of code-switching or using two languages impinge on the distribution of contexts of occurrence (Torres Cacoullos & Travis, 2015b). For Spanish subject expression, we have found that rather than code-switching intrinsically inducing grammatical alteration, what is at work is associated shifts in the frequency of contextual features contributing to variant choice, in particular, those relevant to priming. When speakers have recently used English, the distribution of previous coreferential subjects is altered, with, in the case of subject expression, proportionally fewer unexpressed primes (Torres Cacoullos & Travis, 2015b), and in the case of subject position, fewer post-verbal primes (Benevento & Dietrich).

For Spanish /d/ reduction, Brown shows that the greater a Spanish /d/-initial word's prior exposure to on-line contexts promoting reduction, the greater the likelihood of reduced articulations (an effect she terms Frequency in a Favorable Context). The effect of cognate status of the word on /d/ realization follows from this cumulative effect of different patterns of use in discourse, as, when we consider speakers' English as well as Spanish, cognate /d/ words are used less often in phonetic contexts that promote reduction than are non-cognate words.

In both cases what we observe are the consequences of distribution of contextual features on the variation, either as an on-line or a cumulative effect. In both cases, it is not code-switching or use of the other language per se that has an effect, but mechanisms operative in language variation and change more generally (mechanistic priming, phonetic environment, usage effects). In other cases, familiar mechanisms may operate entirely independently of the other language, as for the effect of lexically particular constructions on post-verbal 1sg subject pronouns (Benevento & Dietrich).

Conclusion

Insights into the workings of spontaneous code-switching can only be obtained from analyses of actual language use, and not reflections about language use (whether the researcher's or the participant's) (cf. Sankoff, 1988, p. 146). Further, analyses of speech must be accountable, that is, free from selective reporting (whether driven by prescriptive bias, stereotypes or categorical perception). Finally, grammatical (dis)similarity and hence change under language contact is most penetratingly assessed by the linguistic conditioning of variation in the bilingual community in comparison with appropriate benchmarks. Through just such analyses, which exploit the highly bilingual nature of NMSEB, the papers comprising this volume arrive at the conclusion that while single English nouns and verbs are largely integrated into Spanish conforming to community-specific norms (Aaron; Wilson & Dumont), multi-word code-switching has disjunctive outcomes, affecting phonology (Balukas & Koops; Brown) but leaving syntax unperturbed (Benevento & Dietrich). As these papers exemplify, the study of bilingualism will profit from community-based spontaneous speech data that allow squarely situating bilingual phenomena in their social and linguistic context.

Gordillo, Tim Poepsel, Miguel Ramos, Jonathan Steuck, Everardo Tapia). We thank Garland Bills and Neddy Vigil for their own inspiring work on New Mexican Spanish and other colleagues for their collaboration (Sonia Balasch, Evan Kidd, Chris Koops, Enrique Lamadrid and Damian Vergara Wilson).

Notes

1. All examples given are from the New Mexico Spanish-English Bilingual corpus (NMSEB, Torres Cacoullos & Travis, In preparation) and are reproduced verbatim from the transcripts (see transcription conventions in Appendix 2). Within brackets is the recording number, name and time stamp. In examples where Spanish and English are used, the original appears on the left, and the translation on the right, with stretches of speech originally produced in English appearing in italics.
2. Viewable at the project website (http://nmcode-switching.la.psu.edu/) under the Tools tab.
3. When a single prosodic unit cannot fit on one line, subsequent lines are indented.

References

Aaron, J. E. (2015). Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism*, *19*(4), 459–480.

Balukas, C., & Koops, C. (2015). Spanish-English bilingual VOT in spontaneous code-switching. *International Journal of Bilingualism. 19*(4), 423–443.

Benevento, N. M., & Dietrich, A. J. (2015). I think, therefore *digo yo*: Variable position of the 1sg subject pronoun in New Mexican Spanish-English code-switching. *International Journal of Bilingualism*, *19*(4), 407–422.

Bills, G.D., Hernández Chávez, E., & Hudson, A. (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, G.D., & Vigil, N.A. (2008). *The Spanish language of New Mexico and southern Colorado: A linguistic atlas*, Albuquerque, NM: University of New Mexico Press.

Brown, E.L. (2005). New Mexican Spanish: Insight into the variable reduction of "la ehe inihial" (/s-/). *Hispania*, *88*(4), 813–824.

Brown, E. L. (2015). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. *International Journal of Bilingualism*, *19*(4), 387–406.

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

Cameron, R. (1994). Switch reference, verb class and priming in a variable syntax. *Papers from the Regional Meeting of the Chicago Linguistic Society: Parasession on variation in linguistic theory*, *30*(2), 27–45.

Clyne, M., Eisikovits, E., & Tollfree, L. (2001). Ethnic varieties of Australian English. In D. Blair, & P. Collins (eds.), *English in Australia*, pp. 223–238. Amsterdam: John Benjamins.

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, *33*, 839–882.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, *2*(3), 219–253.

Doğruöz, A.S., & Backus, A. (2007). Postverbal elements in immigrant Turkish: Evidence of change? *International Journal of Bilingualism*, *11*(2), 185–220.

Du Bois, J.W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. Edwards, & M. Lampert (eds.), *Talking data: Transcription and coding in discourse*, pp. 45–89. Hillsdale, NJ: Lawrence Erlbaum Associates.

Espinosa, A.M. (1909). Studies in New Mexican Spanish, part I: Phonology. *University of New Mexico bulletin / Language series*, *1*(2), 47–162.

Fernández-Gibert, A. (2010). From voice to print: Language and social change in New Mexico, 1880–1912. In S. Rivera-Mills, & D. J. Villa (eds.), *Spanish of the U.S. Southwest: A language in transition*, pp. 45–62. Madrid: Iberoamericana.

Gennari, S.P., & MacDonald, M.C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*, 1–23.

Gonzales, M.D. (1999). Crossing social and cultural borders: The road to language hybridity. In L. Galindo, & M. D. Gonzales (eds.), *Speaking Chicana: Voice, power and identity*, pp. 13–38. Tucson: University of Arizona Press.

Gonzales-Berry, E. (2000). Which language will our children speak? The Spanish language and public education policy in New Mexico, 1890–1930. In E. Gonzales Berry, & D. R. Maciel (eds.), *The contested homeland: A Chicano history of New Mexico*, pp. 169–189. Albuquerque: University of New Mexico Press.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149.

Guy, G.R. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics*, *52*, 63–71.

Halliday, M.A.K. (1967). *Intonation and grammar in British English*, The Hague: Mouton.

Hernández, J.E. (2011). Measuring rates and constraints of word-final nasal velarization in dialect contact. In L. A. Ortíz-López (ed.), *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, pp. 54–69. Somerville, MA: Cascadilla Proceedings Project.

Hopper, P.J., & Thompson, S.A. (2008). Projectability and clause combining in interaction. In R. Laury (ed.), *Crosslinguistic Studies of Clause Combining: The multifunctionality of conjunctions* (TSL 80), pp. 99–123. Amsterdam: John Benjamins.

Kroll, J.F., & Tokowicz, N. (2005). Models of bilingual representation and processing: Looking back and to the future. In J. F. Kroll, & A. M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistics approaches*, pp. 531–553. New York: Oxford University Press.

Labov, W. (1966). *The social stratification of English in New York City*, Washington DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic patterns*, Oxford: Basil Blackwell.

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh, & J. Sherzer (eds.), *Language in use: Readings in sociolinguistics*, pp. 28–53. Englewood Cliffs, NJ: Prentice Hall.

Labov, W. (2010). *Principles of linguistic change: Cognitive and cultural factors*, vol. *3*, 3 vols, Oxford: Blackwell.

Lastra, Y., & Butragueño, P.M. (2014). Subject pronoun expression in oral Mexican Spanish. In A. M. Carvalho, R. Orozco, & N. L. Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* Georgetown: Georgetown University Press.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers*, *41*(3), 841–849 (Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. http://tla.mpi.nl/tools/tla-tools/elan/).

Lipski, J.M. (2008). *Varieties of Spanish in the United States*, Washington, DC: Georgetown University Press.

Miller, J. (1995). Does spoken language have sentences? In F. R. Palmer (ed.), *Grammar and meaning: Essays in honour of Sir John Lyons*, pp. 116–135. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1993). *Duelling languages*, Oxford: Oxford University Press.

Napoli, D.J. (1982). Initial material deletion in English. *Glossa*, *16*, 85–111.

Otheguy, R., & Zentella, A.C. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity* Oxford: Oxford University Press.

Poplack, S. (1993). Variation theory and language contact: Concepts, methods and data. In D. R. Preston (ed.), *American dialect research*, pp. 251–286. Amsterdam: John Benjamins.

Poplack, S. (1997). The sociolinguistic dynamics of apparent convergence. In G. R. Guy, C. Feagin, D. Schiffrin, & J. Baugh (eds.), *Towards a social science of language. Papers in honor of William Labov: Social interaction and discourse structures*, vol. 2, pp. 285–309. Amsterdam: John Benjamins.

Poplack, S. (1998). Contrasting patterns of code-switching in two communities. In P. Trudgill, & J. Cheshire (eds.), *The sociolinguistics reader: Multilingualism and variation*, vol. 1, pp. 44–65. London: Arnold Publishers.

Poplack, S. (2000 [1980]). Sometimes I'll start a sentence in Spanish *y termino en español*: Toward a typology of code-switching. In L. Wei (ed.), *The bilingualism reader*, pp. 221–256. London / New York: Routledge.

Poplack, S. (2012). What does the Nonce Borrowing Hypothesis hypothesize? *Bilingualism: Language and Cognition*, *15*(3), 644–648.

Poplack, S., & Levey, S. (2010). Contact-induced grammatical change: A cautionary tale. In P. Auer, & J. E. Schmidt (eds.), *Language and Space: An international handbook of linguistic variation*, vol. 1: Theories and methods, pp. 391–419. Berlin: Mouton de Gruyter.

Poplack, S., & Meechan, M. (1998). Introduction: How languages fit together in codemixing. *International Journal of Bilingualism*, *2*(2), 127–138.

Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, *26*(1), 47–104.

Poplack, S., Walker, J.A., & Malcolmson, R. (2006). An English "like no other"?: Language contact and change in Quebec. *Canadian Journal of Linguistics / Revue canadienne de Linguistique*, *51*(2), 185–213.

Poplack, S., Zentz, L., & Dion, N. (2012a). Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition*, *15*(2), 203–225.

Poplack, S., Zentz, L., & Dion, N. (2012b). What counts as (contact-induced) change. *Bilingualism: Language and Cognition*, *15*(2), 247–254.

Roberts, I., & Holmberg, A. (2010). Introduction: Parameters in minimalist theory. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (eds.), *Parametric variation: Null subjects in minimalist theory*, pp. 1–57. Cambridge / New York: Cambridge University Press.

Sankoff, D. (1988). Sociolinguistics and syntactic variation. In F. Newmeyer (ed.), *Linguistics: The Cambridge survey (Vol. 4, Language: The socio-cultural context)*, pp. 140–161. Cambridge: Cambridge University Press.

Sankoff, G. (2002). Linguistic outcomes of language contact. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (eds.), *The handbook of language variation and change*, pp. 638–668. Oxford: Blackwell.

Shin, N.L., & Otheguy, R. (2013). Social class and gender impacting change in bilingual settings: Spanish subject pronoun use in New York. *Language in Society*, *42*, 429–452.

Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*, Oxford: Clarendon Press.

Sorace, A. (2004). Native language attrition and developmental instability at the syntax–discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, *7*(2), 143–145.

Thomason, S.G., & Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*, Berkeley, CA: University of California Press.

Torres Cacoullos, R., & Aaron, J.E. (2003). Bare English-origin nouns in Spanish: Rates, constraints and discourse functions. *Language Variation and Change*, *15*(3), 289–328.

Torres Cacoullos, R., & Travis, C.E. (2010). Variable *yo* expression in New Mexico: English influence? In S. Rivera-Mills, & D. J. Villa (eds.), *Spanish of the U.S. Southwest: A language in transition*, pp. 185–206. Madrid: Iberoamericana.

Torres Cacoullos, R., & Travis, C.E. (2011). Using structural variability to evaluate convergence via code-switching. *International Journal of Bilingualism*, *15*(3), 241–267.

Torres Cacoullos, R., & Travis, C.E. (2013). Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics*, 63, 19–34 http://dx.doi.org/10.1016/j.pragma.2013.*08*.003.

Torres Cacoullos, R., & Travis, C.E. (In preparation). New Mexico Spanish-English Bilingual (NMSEB) corpus, National Science Foundation 1019112/1019122. http://nmcode-switching.la.psu.edu/

Torres Cacoullos, R., & Travis, C.E. (2015a). Foundations for the study of subject pronoun expression in Spanish in contact with English: Assessing inter-linguistic (dis)similarity via intra-linguistic variability.

In A. M. Carvalho, R. Orozco, & N. L. Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, pp. 83–102. Georgetown: Georgetown university Press.

Torres Cacoullos, R., & Travis, C.E. (2015b). Two languages, one effect: Structural priming in code-switching. *Bilingualism: Language and Cognition*. DOI: http://dx.doi.org/10.1017/S1366728914000406

Travis, C.E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, *19*(2), 101–135.

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

Travis, C.E., & Torres Cacoullos, R. (2013). Making voices count: Corpus compilation in bilingual communities. *Australian Journal of Linguistics*, *33*(2), 170–194.

Weinreich, U., Labov, W., & Herzog, M.I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann, & Y. Malkiel (eds.), *Directions for historical linguistics: A symposium*, pp. 95–188. Austin, TX: University of Texas Press.

Wilson, D. V., & Dumont, J. (2015). The emergent grammar of bilinguals: The Spanish verb *hacer* 'do' with a bare English infinitive. *International Journal of Bilingualism*, *19*(4), 444–458.

Author biographies

Rena Torres Cacoullos is Professor of Spanish and Linguistics at the Pennsylvania State University. Adopting a usage-based variationist approach, she studies language variation, processes of grammaticalization and the linguistic outcomes of language contact.

Catherine E. Travis is Professor of Modern European Languages at the Australian National University. Her research focuses on the way in which grammar emerges through discourse, which she investigates through the study of variation in spontaneous speech, in monolingual and bilingual contexts.

Appendix 1

NMSEB speaker characteristics

Transcript	Speaker	Year born	Sex	Educ. level	Current residence	Birthplace	Occupation
1	Susan	1934	F	high	Albuquerque	Albuquerque	Stay at home mom
2	Bartolomé	1928	M	middle	Albuquerque	Sthn Colorado	Retired fire-fighter
3	Sandra	1943	F	college	Española	Española	Retired
4	Miguel	1944	M	middle	Valencia	Valencia	Laborer
5	Rocío	1945	F	high	Santa Fe	Santa Fe	Retired school teacher aid
6	Ivette	1946	F	high	Albuquerque	Valencia	Factory worker
7	Samuel	1922	M	college	Taos	Taos	School coach
8	Inmaculada	1952	F	college	Albuquerque	San Miguel	Social worker
9	Fabiola	1954	F	college	Taos	Taos	Secretary

Appendix 1. (Continued)

Transcript	Speaker	Year born	Sex	Educ. level	Current residence	Birthplace	Occupation
9	Molly	1939	F	middle	Taos	Taos	Retired school cook
7, 10	Pedro	1953	M	college	Taos	Rio Arriba	School administrator
11	Mónica	1941	F	high	Albuquerque	Taos	Factory worker/ school custodian
12	Marta	1964	F	college	Río Arriba	Albuquerque	Guest services manager
12	Victoria	1959	F	college	Río Arriba	Española	Retired schoolteacher/ counselor
13	Betty	1925	F	high	Sandoval	Rio Arriba	Retired
14	Anita	1941	F	high	Albuquerque	San Miguel	Executive director
15, 31	Aurora	1962	F	college	Sandoval	Española	Teacher
12, 16	Manuel	1954	M	middle	Río Arriba	Rio Arriba	Electrician/rancher
17	Javier	1936	M	high	Taos	Taos	Rancher and janitor
18	Francisco	1963	M	high	Río Arriba	Rio Arriba	Miner
19	Mariana	1944	F	high	Taos	Taos	Mom/volunteer
20	Tomás	1989	M	high	Río Arriba	Rio Arriba	Unemployed
20	Dora	1953	F	unknown	Río Arriba	Rio Arriba	Housewife
21	Trinidad	1938	F	high	Taos	Española	Substitute teacher
22	Dolores	1963	F	college	Río Arriba	Española	School secretary
22	Clara	1985	F	college	Río Arriba	Española	Editor for Univision
23	Enrique	1933	M	middle	Taos	Taos	Miner, forest service
24	Diana	1941	F	high	Taos	Taos	Dry cleaner
24	Marco	1941	M	middle	Taos	Taos	Miner
25	Leandro	1931	M	middle	Río Arriba	Taos	Miner
26	Carlos	1993	M	high	Río Arriba	Española	Auctioneer
27	Eduardo	1935	M	middle	Río Arriba	Rio Arriba	Store owner/ general contractor
28	Norma	1940	F	high	Río Arriba	Rio Arriba	Retired bank employee/B&B owner
29	Rubén	1925	M	college	Valencia	Rio Arriba	Retired financial administrator
29	Víctor	1928	M	high	Valencia	Rio Arriba	Rancher
30	Cristina	1973	F	college	San Miguel	San Miguel	Self-employed
30	Neddy	1968	M	college	San Miguel	Mora	Car salesman
31	Benita	1941	F	high	Sandoval	Rio Arriba	Home maker
31	Carmela	1978	F	college	Sandoval	Española	Teacher
31	Alfredo	1941	M	high	Sandoval	Sandoval	Retired state highway department

Appendix 2

Transcription conventions (Du Bois et al., 1993)

Carriage return	new Intonation Unit	<@ @>	speech produced while laughing
.	final intonation contour	<X X>	unclear speech; transcriber's best
,	continuing intonation contour		guess at content
?	appeal intonation contour	X	unclear syllable
--	truncated intonation contour	(H)	in breath
-	truncated word	( )	vocal noises
=	lengthened syllable	(( ))	transcriber's comment
..	short pause (0.5 s)	<VOX VOX>	marked voice quality
…	medium pause (0.5–0.7 s)	%	glottal stop
…( )	timed pause (over 0.7 s)	(TSK)	click
[ ]	overlapped speech	(H)	in breath
[2 2]	used to distinguish	!	high booster (following word
	consecutive overlaps		pronounced emphatically)