

of code-switches occurred at points where the word orders were different. The key insight of the Equivalence Constraint is that the concatenated strings constituting code-switches internally follow the grammar of their respective language, but that at the points at which languages are switched the two grammars are ‘equivalent’ (Poplack 1980, Sankoff 1998). For example, in (1), where Miguel (a pseudonym) begins in Spanish and ends in English, each string adheres to the respective monolingual norms. On the one hand, the Spanish string follows the Spanish ‘there are times that’ rather than the English preference for *there are times when over there are times that* (a search in the Corpus of Contemporary American English [Davies 2008-] yielded frequencies of 856 versus 49). On the other, the English string displays adjective noun order rather than the Spanish preference for postnominal adjectives (e.g. Delbecque 1990). However, at the juncture between languages, there is word string equivalence between the two languages, since the determiner precedes the noun in both Spanish and English (Sankoff & Poplack 1981:17).

Here we focus on the boundary of main and complement clauses, illustrated in (2) for English and (3) for Spanish. In (2), the English complementizer *that* is absent, indicated with a \emptyset between the main clause [MC] and the complement clause [CC]. The complementizer is in fact absent most of the time in English (see Shank, Plevoets and Van Bogaert [2009] for a summary of *that/zero* alternation). Though rates of complementizer presence vary according to register, they range approximately between 10% and 20% in corpora of spoken English (e.g. Thompson & Mulak 1991; Torres Cacoullós & Walker 2009). In Spanish, in contrast, the complementizer *que* is virtually always present (Silva-Corvalán 1994:137).

- (2) Pedro: ... *I thought*_[MC] \emptyset *it was a pretty big town back then.* _[CC]
(10 El Timbre Portátil, 35:55-35:57)
- (3) Alfredo: *yo pensé*_[MC] *que estaba muy alto.* _[CC]
‘*I thought*_[MC] *that* *it was very high.* _[CC]’
(31 Speed Limit, 52:11-52:12)

For English-Spanish bilinguals, the juncture between MC and CC would be a site of *variable equivalence* (Torres Cacoullós & Poplack 2016). Sites of variable equivalence display a quantitative rather than a qualitative mismatch—where the languages differ only sometimes, due to independent, but inherently variable, processes in one or both of the languages. This sometimes-applying mismatch between the languages at the main and complement clause boundary has the consequence that for code-switching there arises the problem of variable equivalence at this juncture. In this paper, we offer evidence for prosodic separation as a bilingual strategy for dealing with cross-language mismatches found at the juncture of two languages.

2. A community-based bilingual speech corpus

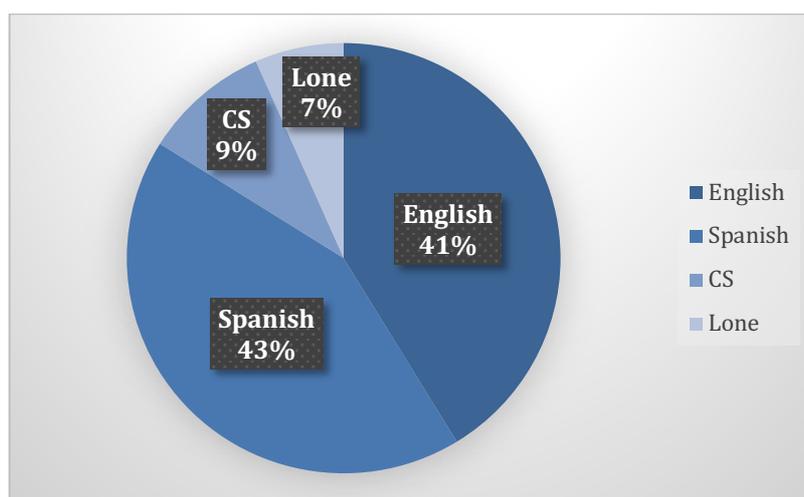
To address the prosody of code-switching, we visit northern New Mexico, where code-switching is “the appropriate code for the Hispano community” (Gonzales 1999:29). Spanish has been spoken here longer than English (Bills & Vigil 2008:29-47). With U.S. statehood in 1912, however, Spanish was relentlessly displaced in the schools, being relegated to a foreign language subject. Today the local variety is

endangered, due to stigmatization and shift to English. Still, of U.S. states, New Mexico has the highest “Hispanic or Latino population” as a percent of the total, at 47%, followed by California and Texas, each at 38% (United States Census Bureau 2015). It also has the lowest proportion of first-generation immigrants (“foreign born”), at about one-fifth of the New Mexico Hispanic population, compared with approximately one third in Texas and nearly half in Florida (United States Census Bureau 2015). In northern counties, the percent of Hispanics is as high as 80% and the proportion of non-US born is as low as 5%. Since these demographics imply limited contact with monolingual Spanish, this bilingual community is a prime candidate for language mixing.

What kind of data offer insights on patterns of language use within the speech community? According to Labov (1989:2), “Word lists and formal elicitations are the primary sources of confusion in our descriptions of language.” The New Mexico Spanish-English Bilingual corpus (NMSEB) is a sociolinguistic-style corpus of spontaneously produced bilingual speech, recording 40 northern Hispano New Mexicans, minimally third-generation, who meet the criterion of regular use of both languages (Torres Cacoullous & Travis 2018:13-56). We do not ask participants to code-switch nor about their code-switching. Instead, the technique is to elicit narratives of personal experience, where speakers shift towards the vernacular, the mode of everyday speech, because they, not the university-affiliated interviewer, are the authority (Labov 1984:32).

Testimony to the bilingualism of the corpus is the distribution of [Main Clause + Complement Clause] instances according to language. In Figure 1 we see that the proportions of English-only, as in (2), and Spanish-only, as in (3), tokens are nearly equal, at 41% (467/1,133) and 43% (484/1,133), respectively. The even distribution of the two languages means that the notion of an overall matrix language or its converse, a salient language (e.g. Myslin & Levy 2015), is irrelevant for these data.

Figure 1: Language of [Main Clause + Complement Clause] (N=1,133)



Cases involving multiword code-switching (CS) constitute 9% (107/1,133) of [Main Clause + Complement Clause] tokens. In (4), code-switching is at the boundary between the clauses (@ represents one syllable of laughter). In (5) code-switching is within one of the clauses, here between the verb and an adverbial expression in the complement clause. Lone items add up to another 7% (75/1,133). These are single words with a language of provenance different from that of the surrounding syntax, as

in *me acuerdo que era una purple* ‘I remember that it was a purple one’ (06 El Túnico, 18:28-18:29). Because they universally tend to assume the linguistic features of borrowings (Poplack 2017), lone items are not counted with unambiguous (multiword) instances of code-switching.

(4) Code-switching at [MC + CC] boundary

Dolores: *so I told them*, [MC]
que van a salir en el <@ Sun @> @. [CC]
 ‘*so I told them*, [MC]
 that they’re going to be in the <@ Sun @> @. [CC]’
 (22 Farolitos, 17:05 -17:07)

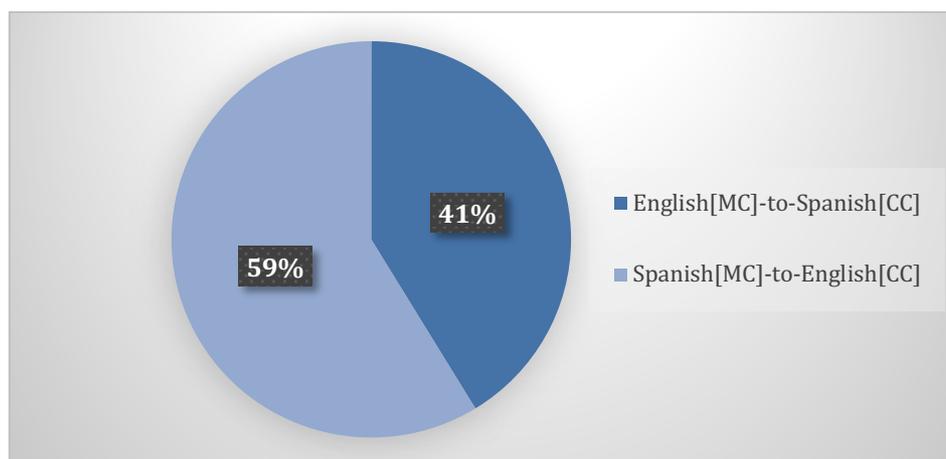
(5) Code-switching within MC or CC, but not at boundary

Dora *... se me hace* [MC] *que era four years ago*. [CC]
 ‘...I think [MC] that it was *four years ago*. [CC]’
 (14 Best of both worlds, 30:10-30:11)

When speakers switch from one language to the other at the boundary between main and complement clause, they do so in both directions at roughly the same rate. Speakers switch from Spanish main clauses to English complement clauses, as in (6), about 59% of the time, and the reverse, as in (4) above, 41% of the time in aggregate, as shown in Figure 2. The lack of a conspicuous directionality of code-switching is further testimony to the bilingualism of these speakers.

(6) Miguel *... (1.9) y dicen* [MC] *que he was .. real lazy*, [CC]
 ‘and they say [MC] that *he was real lazy*, [CC]’
 (04 Piedras y Gallinas, 19:07-19:11)

Figure 2: Direction of code-switching at [Main Clause + Complement Clause] boundary (N = 63)



The even distribution of the languages and lack of directionality of code-switching between them are consistent with the use of code-switching as a general discourse mode for the community. Rather than obeying local interactional or

discourse motivations, once the situation is seen as appropriate, code-switching cannot be predicted at the local level any more than we can predict, say, telling a joke (Poplack 1993:276; Sankoff 1998:39).

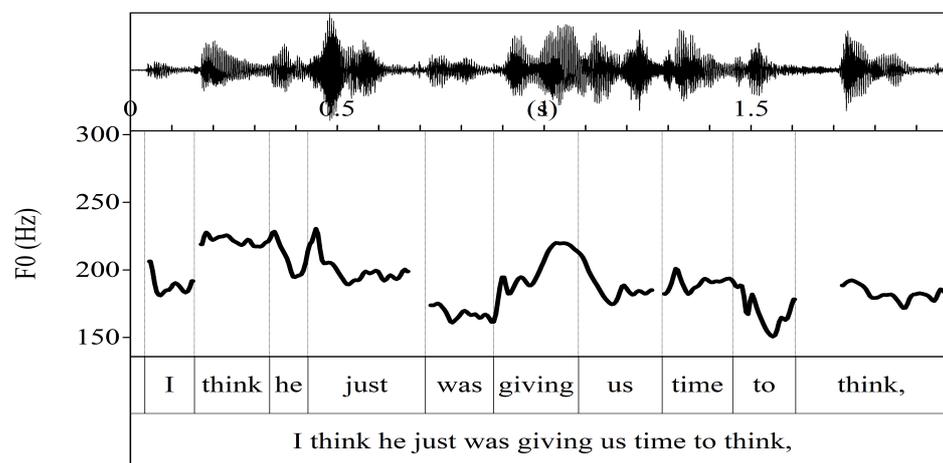
3. Prosodic and syntactic relationships

We do not usually speak in the well-formed sentences privileged by much syntactic analysis. Intonation Units are recognizable units of conversation (Chafe 1994:53-70). The Intonation Unit (IU) is “a stretch of speech uttered under a single, coherent intonation contour” (Du Bois et al. 1993:47). Each IU appears on a distinct line in the transcription followed by punctuation representing its transitional continuity (Du Bois et al. 1993:53). For example, the first IU in (7) has continuing intonation, indicated by the comma; the period following the second IU indicates final intonation contour.

- (7) Sandra: *I think he just was giving us time to think, you know.*
(03 Dos Comadres, 18:49-18:51)

Transcription of the speech stream into IUs is done perceptually by trained linguists (Du Bois et al. 1993). For each hour of speech, at least 50 were required for transcription of NMSEB (Torres Cacoullous & Travis 2018:46-51). Nevertheless, the wave form and F0 trace in Figure 3 illustrate acoustic features transcribers attend to. For example, we can see the higher pitch at the beginning, gradually dropping over the course of the IU to the end, as well as the slower rate of speech at the end of the IU (compare the two instances of *think*).

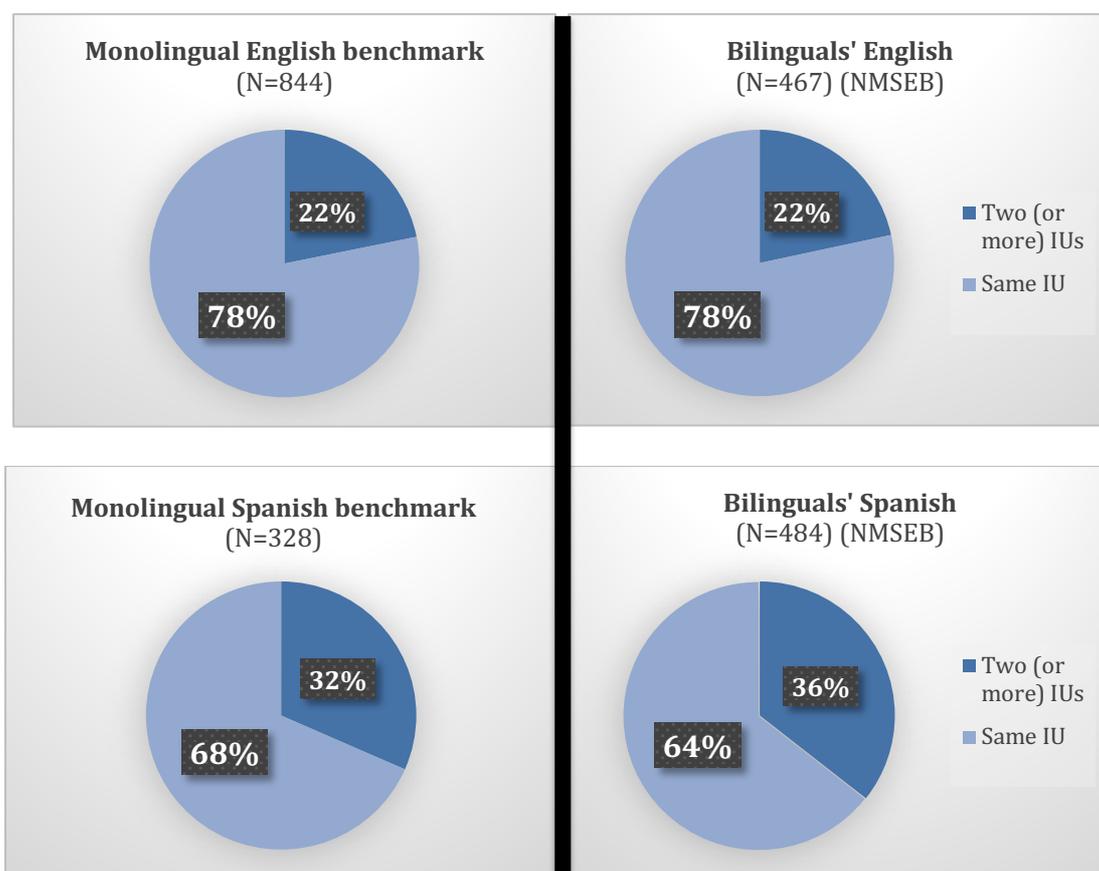
Figure 3 Acoustic properties of the Intonation Unit (IU)²



² Figure 3 was created in Praat (Boersma & Weenink 2018).

There generally is a tighter syntactic relationship between words in the same IU than between words positioned in different IUs (Croft 1995:849-864). The syntactic relationship between main and complement clause is realized prosodically in the tendency for complement clause verbs to occur in the same IU as their main clause verb. This is true for finite clausal complements in conversational English and Spanish data, at rates of 78% (660/844) and 68% (224/328) respectively (Steuck 2016:81), as shown in Figure 4 (left panel). The datasets consist of all declarative finite verbs with finite clausal complements. Conventionalized formulae such as *you know* or collocations such as *I think* that are prosodically independent or that occur as parentheticals were excluded (see Steuck [2016:77-80] for exclusion protocols).

Figure 4. Prosodic integration of [Main Clause + Complement Clause]: Proportion of occurrences in same IU versus two (or more) IUs in monolingual benchmarks and bilinguals' varieties



*Monolingual benchmarks from Steuck (2016): Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois et al. 2000-2005); Corpus of Conversational Colombian Spanish (CCCS) (cf. Travis 2005:9-25).

So, most of the time main and complement clauses are prosodically integrated in a single IU. When are they spread over two or more IUs? In Table 1, contexts hypothesized to influence the prosody of complementation are operationalized as factors in a multivariate analysis; their relative contribution to the probability of the *absence* of prosodic integration are given as factor weights between 0 and 1 (Sankoff et al. 2005). The higher the number, the greater the probability of the absence of

prosodic integration in the corresponding context. The second column gives the number of observations.

In both languages (monolingual benchmarks appear in the leftmost columns of Table 1), the absence of prosodic integration is favored in the context of intervening material, such as adverbials, other clauses, or pauses.³ In English, the presence of the complementizer, itself a kind of intervening material, also favors the absence of prosodic integration (in Spanish, the complementizer is invariably present). Grammatical person and form of the subject also contribute to speaker choices: the absence of prosodic integration is more likely with non-first person main clause subjects and with lexical rather than pronominal complement clause subjects. These factors indicate the role of syntactic distancing, as well as that of particular constructions (e.g., *I think, yo creo* ‘I believe’) and considerations of information status or heaviness of constituents (Croft 1995:850-864; Steuck 2016).

Table 1: Independent variable rule analyses of the contribution of linguistic factors to the absence of prosodic integration of [Main Clause + Complement Clause]

	English-Monolingual		Spanish-Monolingual		English-Bilingual		Spanish-Bilingual	
Average	22%, 184/844		32%, 104/328		22%, 101/467		36%, 172/484	
Corrected mean	.22		.27		.16		.32	
Intervening material								
Present	.92	126/183	.90	70/91	.94	62/80	.89	118/150
Absent	.34	58/661	.30	34/237	.37	39/387	.28	54/334
Complementizer								
Present	.81	60/107		328	.71	55/126		484
Absent	.45	124/737	NA	0	.42	46/341	NA	0
Person main clause subject								
3 rd or 2 nd person	.62	77/222	.61	61/150	[.57]	33/104	[.50]	106/310
1 st person	.46	107/622	.41	43/178	[.48]	68/363	[.50]	66/174
Form compl clause subject								
Lexical	[.59]	70/246	[.65]	32/69	[.58]	25/90	.65	123/377
Pronoun, Other	[.46]	114/598	[.46]	72/259	[.48]	76/376	.46	48/104

* Factor weights in variable rule analysis (Sankoff 1988) indicate the strength of constraints relative to the corrected mean. Monolingual benchmark analyses from Steuck (2016).

What, then, is the prosodic profile of bilinguals’ code-switched [Main Clause + Complement Clause] occurrences?

³ Pauses are represented by dots in the transcription; a number within parentheses indicates a timed pause of 0.7 seconds or longer, as in example (6). Though pauses often coincide with IU boundaries, they are not a required feature (Chafe 1994:58-60; Croft 1995:840). IU-internal pauses, as in (6), are notably less frequent; in a NMSEB sample, 33% (550/1662) of IUs have pauses marked initially, but only 5% (81/1662) have internal pauses (Steuck 2018:86-88).

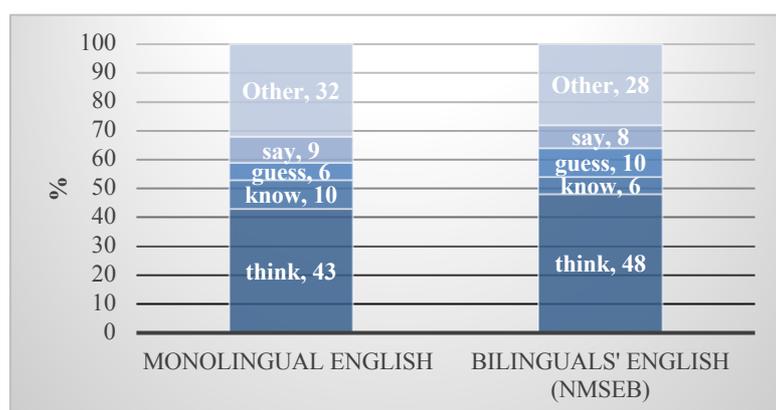
4. Variable equivalence and English-Spanish complementation

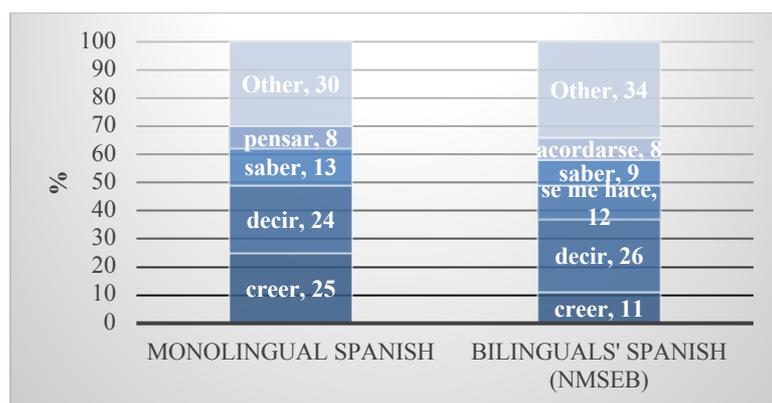
At junctures of *variable equivalence* the word strings of the two languages in contact are equivalent only sometimes. Such is the boundary of main and complement clauses: whereas in Spanish the complementizer *que* is always present, in English complementizer *that* is variably present. Bilinguals might circumvent the mismatch between the languages by mixing them, that is, through grammatical convergence (cf. Gumperz & Wilson 1967; Backus 2005). In other words, they might make the contact varieties more similar to each other than their non-contact (monolingual) counterparts are. They do not do so.

Evidence against grammatical mixing is that bilinguals' English-only and Spanish-only [Main Clause + Complement Clause] occurrences adhere to the same prosodic pattern as their respective monolingual benchmarks. The rates of integration of main and complement clauses in the same IU are 78% (N = 467) in bilinguals' English, but 64% (N = 484) in their Spanish (Figure 4, right panel). At the same time, the direction of effect of factors contributing to the prosody of complementation is consistent across monolingual *and* bilingual varieties of the two languages (Table 1, right-hand columns). (The lack of significance of main-clause subject person in bilinguals' New Mexican Spanish is because of the use of the impersonal expression *se me hace* 'it seems to me' [as in (5)], absent in the Colombian Spanish benchmark.)

A second piece of evidence countering convergence is that the distribution of main-clause verbs is strikingly similar in the bilingual and the monolingual varieties. In English, for both bilingual and monolingual varieties, *think*, *know*, *guess* and *say* account for two-thirds of the data (Figure 5, top panel). In Spanish, the most frequent complement-taking verbs are *creer* 'believe', *decir* 'tell', *saber* 'know', *pensar* 'think' and, in the New Mexican data, *se me hace* 'it seems to me' and *acordarse* 'remember' (Figure 5, bottom panel).

Figure 5. Frequency of complement-taking lexical types, in monolingual benchmarks and bilinguals' varieties





Monolingual benchmarks: English from Torres Cacoullós & Walker (2009:20); Spanish calculated from the Corpus of Conversational Colombian Spanish (CCCS) (cf. Travis 2005:9-25).

A third piece of evidence that bilinguals are adhering to separate grammars is that complementizer *that* is mostly absent in bilinguals' English (appearing at a rate of 27%, 126/467), as in the English monolingual benchmark, while complementizer *que* is never absent, as in the Spanish benchmark (Table 1).

In sum, at least as concerns finite complementation, when these bilinguals speak English, they speak as monolingual English speakers do. When they speak Spanish, they do so like other Spanish speakers. So, when switching at the main and complement clause boundary, bilinguals confront a word string mismatch between their languages, because of the language-internal structural variability in English. What strategy do they use, then, to deal with such a site of variable equivalence?

One hypothesis is that bilinguals employ prosodic distancing of code-switching boundaries (Torres Cacoullós & Poplack 2016). As we have seen, main and complement clauses tend to be prosodically integrated, that is, they tend to occur in the same prosodic unit (Figure 4, above). Therefore, the prediction is that, in the presence of code-switching at the boundary between main and complement clauses, absence of prosodic integration will be more frequent, as compared with Spanish-only and English-only baselines.

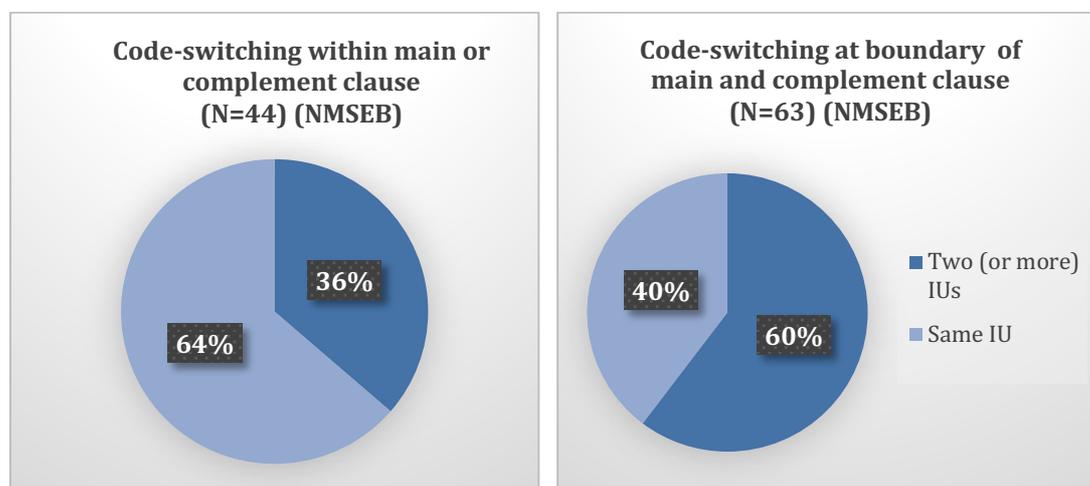
5. Code-switching through prosodic distancing of the boundary between main and complement clause

Code-switching in bilingual complementation may be at the boundary between main and complement clause—as in examples (4) and (6)—or within one of the clauses, as in (5). Are main and complement clauses prosodically integrated at the same rate when code-switching occurs at the clause boundary—a site of variable equivalence—as when it occurs within clauses at loci of equivalence? If code-switching entails a general processing cost, the absence of prosodic integration should be equally as high, compared with monolingual baselines. If, on the other hand, it is equivalence that matters and bilinguals use prosody to deal with variable equivalence, the absence of prosodic integration should be more frequent at the clause boundary.

Figure 6 (left panel) shows that in the presence of within-clause code-switching, the main and complement clause tend to occur in the same IU, at a rate of 64% (28/44), nearly the same as in the monolingual benchmarks and—notably—as in bilinguals' own non-code-switched speech (Figure 4). But with code-switching at the

clause boundary, we see the opposite tendency, namely *not* to prosodically integrate the clauses, at a rate of 60% (38/63) (Figure 6, right panel).

Figure 6. Prosodic integration of [Main Clause + Complement Clause]: Proportion of occurrences in same IU versus two (or more) IUs, in the presence of code-switching



* $p < .05$, by Fisher's exact test.

In summary, if cross-language equivalence matters, bilingual strategies for dealing with *variable* equivalence should be observable in code-switching preferences (Torres Cacoullous & Poplack 2016). The absence of prosodic integration of main and complement clauses is indeed more frequent when code-switching occurs at the boundary between them. The rate is higher as compared with both Spanish and English monolingual benchmarks, the very same bilinguals' Spanish-only and English-only speech, and, finally, when code-switching occurs elsewhere—not at clause boundary. These prosodic patterns of complementation are evidence that prosodic distancing is due not so much to a problem of general processing cost of code-switching per se, but rather due to a problem of cross-language mismatches at particular sites. The prosody of complementation also supplies evidence that grammatical mixing is not a necessary accompaniment of code-switching since, where there are differences between English and Spanish (in the rate of prosodic integration and, especially, in the presence of the complementizer), these are maintained in bilinguals' varieties.

The conclusion is that, just like monolingual speech, bilingual speech is syntactically and prosodically structured. When confronted with quantitative cross-language mismatches—sites of variable equivalence—bilinguals may mitigate them through prosodic distancing of code-switching boundaries.

References

- Backus, Ad. 2005. Codeswitching and Language Change: One Thing Leads to Another? *International Journal of Bilingualism* 9 (3-4): 307-340.
- Bills, Garland. D., and Neddy A. Vigil. 2008. *The Spanish Language of New Mexico and Southern Colorado: A Linguistic Atlas*. University of New Mexico Press.
- Boersma, Paul, and David Weenink. 2018. Praat: doing phonetics by computer. Version 6.0.39, retrieved from <http://www.praat.org/>
- Chafe, Wallace. 1994. *Discourse, Consciousness and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Croft, William. 1995. Intonation Units and Grammatical Structure. *Linguistics* 33 (5): 839-882. doi: 10.1515/ling.1995.33.5.839
- Davies, Mark. 2008-. *The Corpus of Contemporary American English (COCA): 520 million words, 1990-present*. Available online at <https://corpus.byu.edu/coca/>.
- Delbecque, Nicole. 1990. Word Order as a Reflection of Alternate Conceptual Construals in French and Spanish: Similarities and Divergences in Adjective Position. *Cognitive Linguistics* 1 (4): 349-416.
- Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming, and Danae Paolino. 1993. Outline of Discourse Transcription. In *Talking data: Transcription and coding in discourse*, ed. by Jane Edwards, and Martin Lampert, 45-89. Hillsdale, NJ: Lawrence Erlbaum.
- Du Bois, John. W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson, and Nii Martey. 2000-2005. *Santa Barbara corpus of spoken American English* (Files 1-30). Philadelphia: Linguistic Data Consortium.
- Gonzales, María Dolores. 1999. Crossing social and cultural borders: The road to language hybridity. In *Speaking Chicana: Voice, power and identity*, ed. by D. Leticia Galindo, and María Dolores Gonzales, 13–38. Tucson: University of Arizona Press.
- Gries, Stefan Th, and Gerrit Jan Kootstra. 2017. Structural Priming within and across Languages: A Corpus-Based Perspective. *Bilingualism: Language and Cognition* 20 (2): 235-250.
- Gumperz, John J., and Robert Wilson. 1971. Convergence and creolization: A case from the Indo-Aryan / Dravidian border in India. In *Pidginization and creolization of languages*, ed. by: Dell Hymes, 151-167. Cambridge: Cambridge University Press.
- Labov, William. 1984. Field Methods of the Project on Linguistic Change and Variation. In *Use: Readings in Sociolinguistics*, ed. by John Baugh, and Joel Sherzer, 28–54. Englewood Cliffs, NJ: Prentice-Hall.
- Labov, William. 1989. Exact Description of the Speech Community: Short A in Philadelphia. In *Language Change and Variation*, ed. by Ralph Fasold, and Deborah Schiffrin, 1-57. Philadelphia: John Benjamins.
- Lausberg, Hedda, and Han Sloetjes. 2009. Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, and Computers* 41(3):841–849 (Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. <http://tla.mpi.nl/tools/tla-tools/elan/>).
- Myslín, Mark, and Roger Levy. 2015. Code-Switching and Predictability of Meaning in Discourse. *Language* 91 (4): 871-905.

Steuck, Jonathan and Torres Cacoullós, Rena. May 2018. Complementing in another language: Prosody and code-switching. Under consideration for *Language variation. European Perspectives, VII*.

- Poplack, Shana. 1980. Sometimes I'll Start a Sentence in Spanish y Termino en Español: Toward a Typology of Code-switching. *Linguistics* 18 (7-8): 581-618.
- Poplack, Shana. 1993. Variation Theory and Language Contact: Concepts, Methods and Data. In *American Dialect Research*, ed. by Dennis. R. Preston, 251–286. Amsterdam: John Benjamins.
- Poplack, Shana. 2017. *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.
- Sankoff, David. 1988. Variable Rules. In *Sociolinguistics: An international handbook of the science of language and society*, vol. 2., ed. by Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier, 984-997. Berlin: Walter de Gruyter.
- Sankoff, David. 1998. A Formal Production-Based Explanation of the Facts of Code-Switching. *Bilingualism: Language and Cognition* 1 (1): 39–50.
- Sankoff, David, and Shana Poplack. 1981. A formal grammar for code-switching. *Papers in Linguistics: International Journal of Human Communication* 14(1): 3–45.
- Sankoff, David, Sali A. Tagliamonte, and Eric Smith. 2005. Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.
- Shank, Christopher, Koen Plevoets, and July Van Bogaert. 2016. A multifactorial analysis of *that*/zero alternation: The diachronic development of the zero complementizer with *think*, *guess* and *understand*. In *Corpus-based approaches to construction grammar*, ed. by Joyoung Yoon, and Stefan Th. Gries, 201-240. Amsterdam: John Benjamins.
- Silva-Corvalán, Carmen. 1994. *Language Contact and Change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Steuck, Jonathan. 2016. Exploring the Syntax-Semantics-Prosody Interface: Complement Clauses in Conversation. In *Inquiries in Hispanic Linguistics: From theory to empirical evidence*, ed. Alejandro Cuza, Lori Czerwionka, and Daniel Olson, 73-94. John Benjamins.
- Steuck, Jonathan. 2018. The prosodic-syntactic structure of intra-sentential multi-word code-switching in the New Mexico Spanish-English bilingual community. Ph.D. dissertation, Penn State University.
- Thompson, Sandra. A., and Anthony Mulac. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* (15) 3: 237–251.
- Torres Cacoullós, Rena, and Shana Poplack. 2016. Code-switching in spontaneous bilingual speech, National Science Foundation 1624966.
- Torres Cacoullós, Rena, and Catherine E. Travis. 2018. *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Torres Cacoullós, Rena, and James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics* 47 (1): 1-43. doi: 10.1515/LING.2009.001
- Travis, Catherine E. 2005. *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin, Germany / New York, New York: Mouton de Gruyter.
- United States Census Bureau. 2015. 2011-2015 5-Year American Community Survey, <http://factfinder2.census.gov>, Last accessed: 13 July 2017.