Book Review

Torres Cacoullos, Rena and Catherine E. Travis. 2018. *Bilingualism in the community. Code-switching and grammars in contact*. Cambridge: Cambridge University Press. xiv, 240 pp. IBSN: 9781108235259, https://doi.org/10.1017/9781108235259.

Reviewed by **Francesca Di Garbo [fran'ʧeska di 'garbo]**, Department of Linguistics, Stockholm University, Stockholm 106 91, Sweden, E-mail: francesca@ling.su.se

https://doi.org/10.1515/lingty-2019-0020

1 Introduction

One of the most debated topics in research on linguistic diversity (historical, typological, and sociolinguistic alike) is how language and population contact affect (the evolution of) grammar. *Bilingualism in the community* by Torres Cacoullos and Travis addresses this question from the perspective of observed patterns of variation in bilingual language use, and through a principled method of analysis which rests upon (1) comparisons between usage patterns in distinct language varieties used by the same group of bilingual speakers, and (2) comparisons between these bilingual varieties and their monolingual benchmarks. The contact scenario investigated in the book is the Spanish and English-speaking community of northern New Mexico, in the southwest of the United States. The linguistic variable under investigation is subject pronoun expression.

While Spanish is a typical instance of a null-subject language, where subject pronouns often go unexpressed, English tends to overwhelmingly express them. This striking crosslinguistic difference in usage preferences has been taken as an ideal test case for the study of contact-induced change, under the assumption that Spanish varieties with English as a contact language would display a much stronger preference for subject pronoun expression, in alignment with the contact language. While this is a widely investigated issue in the field of Spanish contact studies (Silva-Corvalán 1994; Otheguy & Zentella 2012), what Torres Cacoullos and Travis take as the starting point for their own quest is the need for a better contextualization of the bilingual data. They argue that this can only be attained through a more sociolinguistically balanced population of target speakers than what has been used in previous studies of the phenomenon, better data on bilingual practices as experienced by these speakers, and a more rigorous baseline to test and assess influences between languages in contact. The proposed research design aims to address all three issues via quantitative analyses of over 10,000

instances of the variable of interest (1st and 3rd person singular subjects) as distributed over a community-based bilingual speech corpus of New Mexican Spanish and English, plus three corpora documenting an earlier variety of New Mexican Spanish, monolingual Spanish, and monolingual English, respectively.

The overarching questions addressed in the book are: *are grammars in contact different from grammars not in contact?* (p. 11) and, more specifically, *does code switching boost contact-induced change?* Answers to these questions are sought through the variationist comparative method, which focuses on the study of the probabilistic constraints that condition speakers' choices among competing variants in one and the same domain of linguistic encoding (*the variable context*). The variable context investigated in the book is the one pertaining to subject pronoun expression. The contents of the book are distributed over 11 chapters, which I summarize in Section 2.

2 Synopsis

Chapter 1 introduces the topics under study, research questions and methods.

Chapter 2 presents the community whose bilingual practices are investigated in the book. This is the Spanish-English bilingual community of northern New Mexico, where Spanish has been spoken for over 400 years and English for over 150, and thus the two languages have been in contact for at least 150 years. This qualifies New Mexican Spanish as an ideal object of study for testing hypotheses about contact-induced change, under the assumption that the longer the time of contact the higher the likelihood of convergence between the languages in contact. The Spanish-speaking community of the Santa Fe area was for centuries one of the most isolated areas within the whole Spanish speaking world (p. 14), and is the oldest truly bilingual community of the United States. Established as a Spanish settlement in 1693, first as a New Spain colony and later on as part of independent New Mexico (henceforth NM), the area became part of the US in 1848, but maintained its pronounced Hispanic identity. The Hispanic population of NM is still for the most part autochthonous (with only 16% of Spanish-speaking immigrants, mostly from Mexico). NM Spanish is generally described as retaining archaic features, which are otherwise lost in the standard, such as the archaic forms of the preterit of the verbs for 'see' and 'bring', *vide* and *truj-*, which coexist with the standard variants *vi* and *traj-* (p. 16). In addition, even though NM Spanish is historically related to Mexican Spanish, it has developed its own lexical, phonetic, and morphological features as, for instance, variable aspiration of syllable initial /s/, and the form *ha*, for the first person perfect of 'to have', corresponding to *he* in Mexican and standard Spanish. The acquisition of NM as a

territory of the US severely affected the use of NM Spanish as a primary language of communication, especially after the imposition of English at school. This paved the way to a dramatic process of stigmatization culminating today, when the population of NM Spanish speakers has diminished and there is evidence for ongoing shift to English, especially in the southwestern part of the state. NM Spanish is also under pressure from other non-local Spanish varieties (Mexican Spanish or standard Spanish), which speakers perceive as more prestigious. In this context of general language endangerment, those bilingual speakers of NM Spanish and English who still use both languages in daily life and freely alternate between them are the target population of the study pursued in the book. The sample consists of 40 speakers, who are all third generation New Mexican Hispanics, and whose demographics vary in terms of age, sex, occupation, educational level, and place of residence. A first investigation of rates of subject pronoun expression across demographic features shows that there is no effect of social factors on the distribution of the linguistic variable of interest, which is in line with findings from other Spanish varieties (Silva-Corvalán 2017). The linguistic practices and identities of the sampled population are studied through a unique data set of bilingual speech, the *New Mexican Spanish-English Bilingual* corpus (henceforth NMSEB), whose structure is the topic of Chapter 3.

The book follows a "community-based" (p. 35) method of data collection, which is presented in Chapter 3. The method aims at capturing community-specific patterns in bilingual language use through a large corpus of spontaneous speech, the NMSEB corpus. The NMSEB corpus is based on data stemming from sociolinguistic interviews conducted by experienced fieldworkers who are themselves speakers of the bilingual variety which is being documented.[1] The interviews comply to the "narrative of personal experience" format, which enhances the probability of the speakers' involvement in the topics under discussion and is likely to trigger more naturalistic patterns of language use, including code switching. The interviews mostly consist of conversations with individual speakers or between pairs of speakers; some group discussions are also included. As a result of this data collection procedure, the corpus captures an impressive amount of spontaneous code switching, which occurs both within and across

1 The fieldworkers were eight students of the University of New Mexico, all of them native members of the community. Their relationships with the interviewees ranged from family relationships to acquaintances. During data collection, the fieldworkers were instructed to interact with the interviewees as they would naturally do outside the recording setting and were not informed about the relevance of code switching to the study. Their speech is excluded from all analyses presented in the book.

conversational turns, and yields an overall even distribution of Spanish and English sentences. The transcriptions of the recorded data presuppose at least five rounds of revisions for each recording, and are orthographically and prosodically-informed so that intonation units can be identified. This is essential to the study of code-switching, whose occurrence in bilingual discourse tends to be delimited by prosodic boundaries. Prosodically-informed transcriptions are also remarkably useful for syntactic analysis and, more specifically, for the study of subject pronoun expression given that subject pronouns and verbs tend to occur in the same intonation unit. Data are thoroughly anonymized in order to ensure confidentiality.

Chapter 4 characterizes the sociolinguistic profile and language attitudes of the community under study by means of content analyses of the recorded conversations. These show that such notions as *language dominance* or *first* and *second language* are less meaningful to the linguistic identity of bilingual speakers than suggested by proficiency tests and/or sociolinguistic questionnaires, traditionally employed as a means to characterize the demographics and history of bilingual communities. The impact of three independent measures of contact as inferred through the corpus – self-reported preference of English vs. Spanish, self-rating of English proficiency, and proportion of English clauses produced by individual speakers – on the rate of subject pronoun expression is assessed through independent statistical analyses. These, however, do not bring support to the hypothesis of contact-induced change. The remaining chapters put this hypothesis to the test through the in depth analysis of the bilingual speech documented in the corpus and against the many facets of subject pronoun expression in both Spanish and English.

Chapter 5 sets the stage for the NMSEB corpus analyses by reviewing and revisiting the conditioning factors for subject pronoun expression in (monolingual) Spanish. The chapter draws upon published work as well as the authors' own analyses of the Corpus of Conversational Colombian Spanish and, to a lesser extent, the NMSEB corpus. The analyses show that while contrast, emphasis, and ambiguity resolution are most commonly argued to be the main functional factors that motivate subject pronoun expression in Spanish, speakers' usage does not reflect these generalizations, which, in fact, account for a small portion of the dataset. The most robust effects documented by the authors are instead those engendered by accessibility/subject continuity and syntactic priming. The probability of pronominal subject expression is directly proportional to distance from previous mentions of the coreferential subject, and further increases if the intervening non-coreferential subjects are human, as in example (1), taken from chapter 5, p. 82 (example 66 in the reference text).

(1)	a. Omar: …**Ø** Tenía un=a pintura de= – –	'… **(She)** had a picture of – –
	b .. De un ángel.	..Of an angel.'
	c. Rocío: … Mhm	'… Mhm.'
	d. Omar: … en blanco y negro,	'…in black and white,
	e. e **yo** simplemente le eché color.	and **I** just added color to it.
	f: …Y **ella** quedó feliz	… And **she** was happy.'

In (1), the first 3rd person subject in (a) has zero pronominal expression, but the second in (f), which directly follows the intervening non-coreferential human subject *yo* 'I', is expressed pronominally. These effects pertain to reference accessibility. In addition, pronominal subject expression is favored by previous mention in pronominal form, which is an effect related to syntactic priming. This is illustrated in example (2), which is taken from chapter 5 p. 89 (example 75 in the reference text).

(2)	a. Ángela: **yo** ahorita no estoy trabajando.	'right now **I'**m not working.
	b. … Entonce=s,	…So
	c. Es de ahorros.	It's from savings.
	d. de unos de ahorritos que **yo** tengo.	from some savings that **I** have.'

In (2), previous pronominal expression favors subsequent pronominal expression. Subject continuity is enhanced by previous unexpressed subjects, which means that the two effects closely interact with each other.

The authors further show that clause linking disfavors subject pronoun expression, and that the strength of this effect is maximal in coordinated clauses, followed by subordinate clauses. In contrast, non-coordinated main clauses are more likely to favor pronominal subjects. Perfective aspect and dynamic verbs, that is, predicates encoding activities, are two additional factors that disfavor pronominal subjects, along with the quotative construction *decir*, 'say'. Conversely, the first person form of the present tense of the verb for 'believe' *(yo) creo* highly favors subject expression across monolingual Spanish varieties. Finally, while 1sg subjects are likely to occur at greater distance between each other, 3sg pronominal subjects and unexpressed subjects tend to cluster closer together, as full noun phrases are more likely to be featured at greater distances. The probabilistic constraints on subject pronoun expression examined in the chapter all stem from "general cognitive mechanisms and wider discourse patterns", which are not only specific to Spanish (p. 110). It is against this background, grounded in general and crosslinguistically common usage preferences, that comparisons across varieties, and in the light of divergent sociolinguistic scenarios, become meaningful.

Chapter 6 studies similarities and differences between Spanish and English in the domain of subject pronoun expression. The two languages are usually classified as representing two opposite types, with Spanish being defined as a typical null subject language and English as a typical non-null subject language. While unexpressed subjects occur in English, they are much more rare than in Spanish, and it is generally assumed that the probabilistic constraints that condition their distribution in English categorically differ from those operating in Spanish. Chapter 6 puts this assumption to the test by studying similarities and differences between the two languages based on observed patterns of use. A dynamic method of crosslinguistic comparison, whereby languages are compared with respect to how they vary in relationship to the usage distribution of the variable of interest rather than in terms of discrete structural types, is put forward under the label *variationist typology*. The case study is based on the analysis of 1st and 3rd person singular subjects as occurring in comparable conversational corpus data of Spanish and English, which are prosodically transcribed based on the same criteria as the ones used for the NMSEB corpus. While it is generally assumed that the preference for unexpressed subjects in VP coordination sets English apart from other languages, the study conducted by the authors shows that (1) VP coordination is not a discrete trigger of unexpressed subjects, but rather a gradient type of phenomenon which operates at its highest in the occurrence of prosodic and syntactic linking with the preceding coreferential subject-clause, and (2) this pattern of variation is not specific to English but also attested in Spanish. What is more English-specific is the fact that unexpressed subjects tend to always occur at the beginning of an intonation unit. In addition, the analyses show that while probabilistic constraints related to accessibility, coreferential subject priming, and verb class operate in both languages, the strength of these effects varies across languages, with accessibility (measured through prosodic and syntactic linking) being the strongest constraint in English. Finally, lexically-determined differences between the two languages are also found. In English, constructions of the type [MOTION VERB$_i$ *and* Ø VERB$_i$] (*She goes and Ø gets the mother*) and [VERB$_i$ *and* Ø VERB-OF-SPEECH$_i$] (*He called me last week, and Ø said he wanted to be with me*) strongly favor unexpressed subjects. In Spanish, constructions of the type [(*yo*) + COGNITION VERB$_{1SF}$] (*yo no se* 'I don't know') strongly disfavor unexpressed subjects. By identifying differences in the probabilistic constraints that condition subject expression in English and Spanish, testable hypotheses about contact-induced change can be formulated. If subject pronoun expression in NM Spanish is changing under the influence of English, we should expect to observe: lower rates of expressed subject in prosodically initial positions (as in English), a stronger effect of accessibility (as in English), an overextension of unexpressed subjects in coordinated clauses linked by *y* 'and' (as in English) *vis à vis* a lower

effect of constructions involving cognition verbs (which are common triggers of pronominal subjects in Spanish). These hypotheses are tested in Chapter 8.

Chapter 7, in turn, compares probabilistic constraints on the expression of subject pronouns in NM Spanish with an earlier stage of the variety, allegedly less exposed to contact with English, and with monolingual Spanish varieties. Data on earlier stages of NM Spanish are drawn from a sample of the oldest, and less prone to code switching, speakers of the New-Mexico-Colorado Spanish Survey (Bills & Vigil 2008). Data on monolingual Spanish are taken from corpora of San Juan (Puerto Rico) and Madrid (Spain) use. Both case studies, which are based on a combination of multivariate analyses and cross-tabulations of the probabilistic constraints accounting for the distribution of subject pronoun expression, provide evidence "for continuity rather than change" (p. 159). With respect to the grammar of subject pronoun expression, present-day NM Spanish does not significantly differ from earlier stages of the variety nor from monolingual benchmark data.

Chapter 8 tests the hypothesis of convergence between bilingual Spanish and English, as formulated at the end of Chapter 6, through comparisons between four parallel datasets of prosodically annotated spontaneous speech: monolingual Spanish (through the Corpus of Conversational Colombian Spanish, CCCS), monolingual English (through the Santa Barbara Corpus of Spoken American English (SBCSAE), bilingual Spanish (through the NMSEB corpus), and bilingual English (also through the NMSEB corpus). The analyses presented in the chapter align with each other, showing no sign of convergence of bilingual Spanish and English along the three main diagnostic features established in chapter 6: prosodically initial position, accessibility, and language-specific verbal constructions. This pattern of negative evidence also holds for bilingual English in relationship with Spanish. The overarching finding of the chapter is thus that both bilingual varieties comply to their respective monolingual varieties in the distribution of expressed and unexpressed subjects. What remains to be tested is the hypothesis that code-switching boosts convergence effects in this domain. This is the topic of Chapter 9.

The NMSEB is ideal for testing synchronic hypotheses about the role of code switching in contact-induced change, as it abounds with instances of both single and multi-word incorporations from English. In Chapter 9, the authors test whether the presence of code switching in syntactic contexts which are maximally close to candidates for variable subject expression, and which are uttered by the same speakers, boosts the probability of occurrence of pronominal subjects. Once again, the results produce no evidence of convergence between the grammars in contact, in that they show no effect of proximity of English on the rate of Spanish pronominal subjects, and no changes in the probabilistic constraints that

condition subject expression in bilingual discourse as compared with monolingual discourse. This is a remarkable result, which goes against the expectation that at least reference accessibility (the strongest predictor of unexpressed subject in English) could have been impacted by code switching, either in the direction of weakening (as an outcome of simplification of pragmatic constraints under language contact) or strengthening (in alignment with English).

"Speakers' choices are not independent of their prior choices" (p. 189). Priming is a driving force in the structuring of discourse, which has been argued to also impact bilingual discourse, as parallel language structures across languages in use may boost convergence. Chapter 10 tests this hypothesis by investigating how coreferential subject priming, whereby pronominal subjects are favored in the presence of a coreferential pronominal subject in previous mention, changes in the presence of code switching. The study shows that the effect of coreferential subject priming and its interaction with accessibility do not vary in the presence of code switching as compared with its absence. What instead varies under the pressure of code switching are the contextual distributions of the variable of interest. In the absence of code switching, only 27% of the data occurs with a pronominal prime (mostly a Spanish pronoun). Conversely, in the context of proximate code switching, pronominal subject primes go up to 61%, and the majority of the primes are English pronouns. This means that while the rate of subject pronoun expression stays flat, the probability of within-language (Spanish pronoun-to-pronoun) priming decreases, and the distribution of the conditioning factors for the variable of interest in turn changes (in this case, the proportion of pronominal primes, which favor subsequent subject pronouns, increases). Ultimately then, the study shows that, even under code switching, the grammars of the languages used in bilingual discourse are kept distinct.

Chapter 11 provides some concluding remarks and evaluates the main methodological and theoretical contributions of the study against the two initial research questions (cf. Section 1). The results of the investigation show no sign of convergence between NM Spanish and English in the domain of subject expression (question 1, *are grammars in contact different from grammars not in contact?*), and they also indicate that code-switching does not boost contact-induced change in subject expression, but only affects its contextual distribution (question 2, *does code switching boost contact-induced change?*).

3 Contributions to typology

Bilingualism in the community is a ground-breaking study on language variation and change and a must-read for anyone interested in the internal and external

dynamics of linguistic diversity. As I see it, this book makes two major contributions to linguistic typology, which I briefly summarize in the following and further discuss in Sections 3.1 and 3.2.

1. Defying the assumption that patterns of language structures can be readily captured into discrete types. The book promotes a methodology of crosslinguistic research, variationist typology, which instead aims to identify and assess the probabilistic contribution of a set of competing conditioning factors to the distribution of a given typological variable, and on the basis of the observation of language(s) in use.
2. Putting at stake common beliefs about language and population contact and, in particular, the idea that long-term bilingualism (always) leads to convergence between languages in contact. The book rejects this idea by demonstrating its lack of validity for the bilingual community under study. This is done through a series of related case studies, which rely on rigorous sampling of bilingual and monolingual speech data, and comparisons thereof, as well as on cutting-edge statistical methodologies.

While (1) has general implications for linguistic typology as a theory and method of crosslinguistic comparison, (2) specifically contributes to understanding the interplay between language structures and the sociolinguistic contexts in which languages are learned and used, which lies at the foundation of one of the most recent and fastest growing lines of research in our field, sociolinguistic typology.

3.1 Variationist typology

Typologists have long been concerned about how well the labels they use to classify and compare languages with each other in fact capture language-internal and external variation with respect to a given pattern of language structure. This is both a problem of *representativity* (do typological variables reflect how varied a given linguistic phenomenon is in language?) and *comparability* (can a given typological variable capture what is shared by otherwise potentially radically different languages?), and different approaches have been proposed to address these problems. Three such approaches are, for instance, multivariate typology (Bickel 2010, Bickel 2011), family- and/or areal-based typology (Greenberg 1978; Koptjevskaja-Tamm 2011), and corpus-based typology (cf. Cysouw & Wälchli 2007; Östling 2016), which are all characterized by a strong focus on micro- rather than macro- variation, as captured through fine-grained and maximally independent variable designs (multivariate typology), family and areal-specific

patterns of diversity (intragenealogical and areal typology), and context-specific usages as documented in text (corpus-based typology). *Bilingualism in the community* is an exemplary illustration of the latter type of appoach. It shows that crosslinguistic corpus studies of spontaneous speech data can shed new light on well-known and established typological generalizations by turning them into dynamic, multifactorial tendencies, which unfold in actual use through the conditioning effect of various contextual factors.

Torres Cacoullos and Travis take the typological generalization whereby languages can be classified into two opposite types depending on whether they show a bias towards pronominal or null subject expression and demonstrate how this is in fact a probabilistic, distributional bias, grounded in a number of functionally motivated conditioning factors, namely: accessibility, syntactic priming, (verbal) aspect, temporal sequencing, person, and language-specific lexical constructions. This radically empirical heuristic enterprise is made possible by the availability of extraordinary corpus data on the bilingual variety under study (NM Spanish) and on the respective monolingual benchmarks, English and Spanish, which are, in fact, two of the major (in terms of population and geographical spread) and better resourced languages of the world. At first sight, the issue of comparably adequate and accessible corpus materials may call into question the feasibility of this approach with lesser-resourced languages, for which only limited corpora exists, and often only restricted to elicited discourse. However, especially in recent years, significant investment has been put into enhancing text-based (narrative, spontaneous, task-driven) language description, and on refining procedures for data archiving and accessibility.[2] This, paired with recent advancements in computational methods for the study of multilingual corpus data, makes it plausible, now probably more than ever before, that the type of in-depth investigations conducted by Torres Cacoullos and Travis for NM Spanish will become more of a norm than an exception in research practices in linguistic typology. *Bilingualism in the community* thus offers an outstanding theoretical and methodological yardstick for future typological variationist studies on other languages and domains of grammar.

2 Some of the currently available digital repositories of corpus data on minority and endangered languages of the world are ELAR (https://elar.soas.ac.uk/), AILLA (for native languages of the Americas, https://ailla.utexas.org/), and DOBES (http://dobes.mpi.nl/). SCOPIC (https://scopicproject.wordpress.com/.) is a collection of crosslinguistically matched naturalistic corpus data, featuring grammatical categories related to social cognition in a variety of languages. An older and well-known effort to elicit comparable naturalistic discourse data from a variety of languages comes from the Pear Stories project (http://pearstories.org/). I am grateful to Karolina Grzech for this list of relevant links.

3.2 Sociolinguistic typology

Over the past couple of decades, typologists have become increasingly interested in disentangling the relationship between the distribution of linguistic diversity, and the social and natural environment in which languages are learned and used. Ever since Trudgill's (2011) book, this approach has been known under the label *sociolinguistic typology*. Many publications have appeared over the last couple of decades, which provide both negative and positive evidence for adaptive responses of language structures to the sociolinguistic environment (for a recent overview see Nettle 2012). Particularly debated is the question of how to identify appropriate variables, both at the linguistic and extra-linguistic level, for testing hypotheses on linguistic adaptation. *Bilingualism in the community* makes a significant contribution to this ongoing discussion, and it does so in at least two ways.

First, the book demonstrates that demographic data based on speakers' self-reports of proficiency levels in their languages of use, and most often inferred through sociolinguistic questionnaires, may fail to represent the multifaceted identities of bilingual and multilingual speakers. Better insights into speakers' demographics can instead be obtained via sociolinguistic interviews, which facilitate more spontaneous interactions with the speakers of the community under study, and in turn increase the likelihood of obtaining more authentic data on their linguistic biographies and language attitudes. In sociolinguistic typology, common sources of demographic data have been population size figures (both L1 and L2), mostly obtained via censuses, which correspond to the kind of data type that Torres Cacoullos and Travis demonstrate to be likely to misrepresent bilingual practices in the community. While the problem of using population figures as a proxy for language contact is discussed in the sociolinguistic-typological literature (Sinnemäki & Di Garbo 2018), these approximative measures are still in use in large-scale sociolinguistic-typological investigations, essentially due to lack of alternatives. The insightful discussion of demographic data provided by Torres Cacoullos and Travis throughout their study invaluably contribute to raising further awareness on this topic. This, in combination with the fact that currently ongoing sociolinguistic-typological projects are explicitly targeting the necessity of developing principled ways to refine sociolinguistic variables for large-scale comparative studies[3] makes the perspective of further advancement in the field not just desirable, but also gradually more achievable.

Second, the book provides a rigorous methodological baseline for studies of contact-induced change. According to the procedure proposed and implemented

3 https://www.helsinki.fi/en/researchgroups/linguistic-adaptation

by the authors, convergence between two languages in contact, X and Y, should be ideally assessed through a research design consisting of four steps: (1) comparisons between a contact variety of language X and corresponding non-contact varieties; (2) comparisons with earlier documented versions of the contact variety of X; (3) comparisons with monolingual benchmarks of the contact language Y; and (4) comparisons between the contact varieties of X and Y, as bilingual varieties of the same speakers' population (with systematic control for code switching). The issue of corpora availability, which I discussed in Section 3.1 with respect to crosslinguistic comparisons of domains of language-internal variation, also applies here. When working with varieties and contact scenarios that are less well described than those of Spanish and English, several phases of the multistep procedure laid out by Torres Cacoullos and Travis may, in fact, be missing. For instance, earlier stages of the contact variety under study may be undocumented and a monolingual benchmark may not be easy to identify. Similarly, the contact language as bilingual variety of the same community of speakers, and its monolingual benchmarks, may also lack appropriate documentation. This lack of adequate data probably reflects an inescapable bias in comparative linguistics, at any breadth and scale of analysis, but it should not, as such, hinder the feasibility of the approach. In view of this bias and of the groundbreaking results which, as shown by Torres Cacoullos and Travis, can be achieved when relying on a full-fledged set of linguistic and sociolinguistic data, generalizations about contact-induced change in situations of long-term bilingualism should be perhaps scaled down. However, while it is a fact that for many bilingual communities around the globe, we simply cannot have access to the whole spectrum of data that would be needed in order to fully apply the four-step design developed by Torres Cacoullos and Travis, we can still strive to make the most out of what is available.

In sum, sociolinguistic typology is a growing field in the language sciences, and more work is still needed in refining its theoretical assumptions and methodologies. By zooming in to the specifics of one bilingual community and studying it through a wealth of data and impressive analytical tools, *Bilingualism in the community* pushes the debate on the comparability between language structures and sociolinguistic profiles one step further, setting a gold standard for assessing general and language-specific claims on grammatical convergence and contact-induced change.

References

Bickel, Balthasar. 2010. Capturing particulars and universals in clause linkage: Multivariate analysis. In Isabelle Brill (ed.), *Clause-hierarchy and clause-linking: The syntax and pragmatics interface*, 51–101. Amsterdam: John Benjamins.

Bickel, Balthasar. 2011. Multivariate typology and field linguistics: A case study on detransitivization in Kiranti (Sino-Tibetan). In *Proceedings of Conference on Language Documentation and Linguistic Theory*, vol. 3, 3–13.
Bills, Garland & Neddy Vigil. 2008. *The Spanish language of New Mexico and southern Colorado: A linguistic atlas*. Albuquerque: University of New Mexico.
Cysouw, Michael & Bernhard Wälchli. 2007. Parallel texts: Using translational equivalents in linguistic typology. *STUF-Sprachtypologie und Universalienforschung* 60(2). 95–99.
Greenberg, Joseph. 1978. Diachrony, synchrony and language universals. In Joseph Greenberg, Charles Ferguson & Edith Moravcsik (eds.), *Universals of human language*, vol. 1: Method and theory, 62–92. Stanford: Stanford University Press.
Koptjevskaja-Tamm, Maria. 2011. Language contact. In Jae-Jung Song (ed.), *The Oxford handbook of linguistic typology*, 568–590. Oxford: Oxford University Press.
Nettle, Daniel. 2012. Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B* 367. 1829–1836.
Östling, Robert. 2016. Studying colexification through massively parallel corpora. In Päivi Juvonen & Maria Koptjevskaja-Tamm (eds.), *The lexical typology of semantic shifts*, 157–176. Berlin: Mouton de Gruyter.
Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. New York: Oxford University Press.
Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.
Silva-Corvalán, Carmen. 2017. *Sociolingüística y pragmática del español*, 2nd edn. Washington, DC: Georgetown University Press.
Sinnemäki, Kaius & Francesca Di Garbo. 2018. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology* 9. 10.3389/fpsyg.2018.01141. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01141.
Trudgill, Peter. 2011. *Sociolinguistic typology: Social determinants of linguistic complexity*. New York: Oxford University Press.