

# A semantic analysis of bilingual compound verbs in two contact Spanish communities

Osmer Balam<sup>a/b</sup> – Lidia Pérez Leutza<sup>b</sup> – Ian Michalski<sup>c</sup> – María del Carmen Parafita Couto<sup>b/d</sup>

The College of Wooster<sup>a</sup> / United States

Leiden University<sup>b</sup> / The Netherlands

Roanoke College<sup>c</sup> / United States

Universidade de Vigo<sup>d</sup> / Spain

**Abstract** – Although previous work has contributed to our knowledge of bilingual compound verbs (BCVs) in different code-switching varieties, there is scant research on the semantic nature of these innovative constructions. To fill this gap, the present study examines semantic aspects of BCVs in Northern Belize and the Yucatan Peninsula in Mexico, two sociohistorically connected communities where Spanish *hacer* ‘do’ BCVs have been attested. Drawing on two datasets, we analyzed the semantic domains that are most open to other-language lexical verbs as well as the potential use of these structures as identity markers. The analysis of 1,140 BCVs (903 from Northern Belize and 237 from Yucatan) revealed that whereas ‘education’ particularly favored English lexical verbs in Northern Belize, ‘nourishment’ was the semantic sub-category most open to Yucatec Maya lexical verbs in the Yucatan Peninsula. Notably, only *hacer* BCVs from Yucatan evince the incorporation of cultural elements and linguistic practices such as *albur* ‘word play’ to index a Yucatec Maya ethnolinguistic identity. Our findings highlight the importance that the nature of bilingualism and community linguistic norms have on the semantic use of BCVs.

**Keywords** – bilingual compound verbs; semantic domains; Northern Belize; Yucatan Peninsula; ethnolinguistic identity; bilingual corpora

## 1. INTRODUCTION

Research on code-switching (henceforth, CS) has shown that nouns comprise the most frequently borrowed or switched element in bilingual discourse (Pfaff 1979: 305; Poplack and Meechan 1998: 127; Jake *et al.* 2002: 72; Gardner-Chloros 2009: 31; Clegg 2010: 223; Balam 2016a: 14).<sup>1</sup> As it pertains to the motivation behind this pattern in bilingual corpora, some scholars have examined the semantic domains that favor the incorporation

<sup>1</sup> We adopt Muysken’s (2000, 2013) broad conceptualization of CS, in which insertion, alternation, congruent lexicalization and backflagging are envisioned as different manifestations or ‘optimization strategies’ of CS. The use of these strategies depends on social, linguistic, and cognitive factors.



of other-language nouns.<sup>2</sup> In Spanish/English CS, it has been found that English-origin nouns are predominantly drawn from semantic categories such as years and numbers (Aaron 2015), technology (Clegg 2010; Aaron 2015; Balam 2016a), education (Clegg 2010; Balam 2016a) and work/money-related terms (Balam 2016a). There is scant research, however, on the semantic domains that favor other-language verbs in CS. To fill this gap, we analyze the openness of semantic domains to other-language lexical verbs in bi/multilingual speech.

Bilingual compound verbs (henceforth, BCVs) offer fertile ground for the study of this phenomenon given that while the light verb (*hacer* ‘do’ in this case) provides grammatical information such as tense and aspect, it is the other-language lexical verb that provides semantic content, as (1) illustrates.<sup>3</sup> The relevant question that arises is whether the semantic domains that have been previously found to be most open to other-language nouns can also be extended to other-language verbs, an issue we address here.

- (1) *No nos hacían* encourage  
 No CL.DO do.3PL.IMP encourage  
 ‘They did not encourage us’

Taken from New Mexico/Colorado, U.S. (Jenkins 2003: 197)

From a sociolinguistic perspective, an important finding in antecedent work is that whereas some semantic domains (e.g., technology) appear to favor other-language nouns irrespective of the sociolinguistic milieu, other domains are context-specific. In Spanish/English corpus data from New Mexico, Aaron (2015) found that kinship terms comprised the semantic domain most open to English nouns (e.g., *dad, daddy, grandma*). By contrast, Balam (2016a) found that this was the domain least open to English nouns in corpus data from Northern Belize; hence, revealing that in semantic patterns of CS there can be significant differences in terms of community linguistic norms.

Noteworthy is that BCVs can also markedly differ in terms of how they are used across communities (Gardner-Chloros and Finnis 2006; Balam *et al.* 2023). In more recent research on *hacer* BCVs in progressive and passive constructions, Balam and colleagues found significant differences across different groups of Spanish/English bilinguals (Balam *et al.* 2020; Balam *et al.* 2023). For instance, whereas bi/multilinguals

---

<sup>2</sup> For relevant discussion on what motivates borrowing or CS, see Muysken 2000 or Backus 2001, and references therein.

<sup>3</sup> In the linguistic examples, Spanish is italicized whereas Yucatec Maya is shown in bold and italicized. The English translation is provided between inverted commas.

in Belize give preferential ratings to *hacer* BCVs in present progressives, (e.g., ...*está haciendo* audit *el report* ‘...is auditing the report’), Puerto Rico and New Mexico bilinguals give the highest ratings of acceptability to *estar* ‘be’ +  $V_{\text{prog}}$  constructions (e.g., ...*está auditing el report* ‘...is auditing the report’). These findings indicate that the study of CS grammars requires careful consideration of both invariant and variable patterns, which are shaped not only by linguistic factors but historical and sociolinguistic conditions as well (Balam *et al.* 2023: 416).

As it relates to the study of BCVs, Demirçay’s (2017) work on Turkish/Dutch *yap* BCVs in the Netherlands is, to our knowledge, the only study that examines these hybrid constructions from a semantic lens. Demirçay analyzed 48 Turkish/Dutch BCVs extracted from self-recorded group conversations among 19 Turkish/Dutch bilinguals. Her results showed that the semantic domains most open to Dutch lexical verbs in *yap* constructions were 1) ‘school/education/learning-related’ (35.4%, e.g., *zakken* ‘fail’), 2) ‘life in the Dutch society – informal aspects’ (33.3%, e.g., *chillen* ‘chill’), and 3) ‘work-related’ (e.g., 14.6%, *verdiene* ‘earn’). Demirçay attributes her findings to the high degree of entrenchment<sup>4</sup> of Dutch lexical verbs, which are associated with things and activities that are typically taught, learnt, or experienced in Dutch rather than Turkish.

In Demirçay’s view, Dutch lexical elements and units from domains such as school, social life, and work get increasingly entrenched in the minds of second generation Turkish/Dutch bilinguals in the Netherlands due to their daily experiences. This subsequently “strengthens their storage and makes the further activation of such units easier” (Demirçay 1997: 112). Demirçay’s conclusion that language experience plays an important role on the semantic nature of BCVs echoes Jenkins’ (2003) previous observation. In a descriptive analysis of *hacer* BCVs, Jenkins (2003: 197) suggested that the use of these constructions among Spanish/English bilinguals in New Mexico/Colorado can be associated with “English-language domains” such as school and work. These are social contexts where monolingual English rather than Spanish is typically employed. Thus, BCVs are reflective of speakers’ experiences in environments where English is predominantly used.

---

<sup>4</sup> Following Croft (2000: 38), ‘entrenchment’ refers to a cognitive determinant of language use. The degree of entrenchment of a linguistic form depends on its frequency of use by speakers. The more speakers use a particular form, the more entrenched it becomes in speakers’ minds. This results in easier activation of a certain word or form in future speech events. Across time, a high degree of entrenchment may lead to conventionalization.

Although recent previous research has elucidated our knowledge of grammatical (see Balam *et al.* 2014; Balam 2015, 2016c, 2021; Balam and Prada Pérez 2017 —for Spanish/English data— and Pfeiler 2014; Michalski 2016, 2017, for Spanish/Maya data) and formal syntactic aspects of *hacer* BCVs ( see González-Vilbazo and López 2011, for Spanish/German data), no study has focused on the semantic nature of these verb constructions in contact Spanish. In the present paper, we shed light on *hacer* BCVs in two sociohistorically connected communities that have been previously noted for the use of *hacer* BCVs: namely, Northern Belize —where Spanish is in intense contact with English and Belizean Kriol— and the Yucatan Peninsula in Mexico, where Spanish is in contact with Yucatec Maya. More specifically, drawing on two datasets, we analyze two semantic aspects of *hacer* BCVs: 1) the semantic domains that are most open either to English or Yucatec Maya lexical verbs, and 2) the potential use of these hybrid constructions as identity markers in contact Spanish (see section 4.3).

The paper is divided as follows. In sections 2 and 3, we provide a brief overview of the two contexts under study and the differential use of BCVs in these communities. In section 4, we describe the methodology used in the analysis of semantic domains in *hacer* BCVs. In section 5, we present our results. Lastly, in section 6, we discuss the implications of our findings and offer concluding remarks.

## 2. THE TWO COMMUNITIES UNDER STUDY

In line with previous cross-community research on BCVs (Gardner-Chloros and Finnis 2006; Balam *et al.* 2020; Balam *et al.* 2023), this study aims to contribute towards this emerging body of work by providing an analysis of *hacer* BCVs in two contact Spanish communities that have very close historical ties. The first community is Orange Walk, a district in Northern Belize located on the southeastern side of the peninsula. The second community is in the north and southwestern regions of the Yucatan Peninsula, a region encompassing the Mexican states of Campeche, Yucatan, and Quintana Roo.

In addition to being geographically adjacent, these two contact zones are connected sociohistorically and linguistically. Yucatec Maya and its related varieties were the predominant languages in both regions as early as the sixteenth century. It is worth noting, however, that the number of Maya speakers during the sixteenth century was significantly

lower in the area that today comprises Belize (Hagerty 1979; Jones 1998; Balam *et al.* In Press).

As a result of escalating tensions and political unrest in the region, the Maya revolution broke out in 1847. This marked the onset of the Caste War of Yucatan, which had a lasting impact on the sociolinguistic landscape of this region, in particular the area that is today Belize. Initially, the Mayans from the eastern regions defeated many of the Yucatecans of Spanish descent in the northern cities of Yucatan and Quintana Roo (Cal 1991). This prompted thousands of Yucatec Mayans and Mestizos<sup>5</sup> to seek refuge and establish hamlets in the present-day districts of Orange Walk, Corozal, and Northern Belize, as well as the district of Cayo in Western Belize (Balam 2014, 2015).

In 1871, the British Crown formally solidified its power by establishing British Honduras (now Belize) as one of its colonies (Balam *et al.* In Press). Following the British colonial takeover, the sociolinguistic situation in Belize gradually began to differentiate itself from the rest of the Yucatan Peninsula in the nineteenth and twentieth centuries (Balam 2014, 2016b). Language contact intensified between Spanish, English, and Belizean Kriol (Balam 2014, 2015, 2016b) due to different waves of migration of Spanish speakers, increased access to education in the 1950s, and access to American programming via satellite technology in the 1980s (Elliott 1995).

Despite its contact with other languages, Spanish has historically remained as the language of the majority in Northern Belize. Nowadays, the highest percentages of Spanish speakers are found in the northern districts of Orange Walk and Corozal. In Orange Walk, 86 percent of the population (especially the Mestizo population) speaks Spanish, 62 percent speaks English (the official language), 16.8 percent speaks Belizean Kriol (the lingua franca), and only 2.3 percent speaks Yucatec Maya (see Balam 2013, 2016b; Balam and Prada Pérez 2017, for further details). Thus, current speakers of Yucatec Maya constitute a very small minority in Northern Belize. This can be largely attributed to the stigmatization towards the culture and language and a rapid process of language shift to Spanish that took place in the 1930s and 1940s (Koenig 1975; Balam 2016b).

In contrast to Northern Belize, the Yucatec Maya-speaking population in the Yucatan Peninsula has constituted the demographic majority for centuries, which has

---

<sup>5</sup> In Belize, the term ‘Mestizo’ refers to any person of mixed indigenous Mayan and Spanish ancestry.

contributed to the use of Yucatec Maya as the *de facto* language of the peninsula. In the 1700s, for instance, Moseley (1980: 102–104) highlights that, while the Spanish population was 103,000, the number of Mayans was 254,000. The predominance of Yucatec Maya in the region, however, eventually came to an end in the twentieth century, when language shift towards Spanish monolingualism gained momentum from the 1970s onwards (Michalski 2017).

Census data show that the majority of Yucatec Maya speakers in Mexico live in the state of Yucatan (68%, 519,167), followed by Quintana Roo (23%, 174,817) and Campeche (9%, 70,603). At the moment, 85 percent of the population in the region are Yucatec Spanish monolinguals, while approximately 14.4 percent are Spanish/Maya bilinguals, and only 0.6 percent are Yucatec Maya monolinguals (Instituto Nacional de Estadística, Geografía e Información 2020; Sobrino 2010; Michalski 2017). Spanish/Maya bilingualism is particularly common in Yucatan, although census data indicate that the decrease in Yucatec Maya speakers has co-occurred with a concurrent decline in the number of Spanish/Maya bilinguals, over the last thirty years (Instituto Nacional de Estadística, Geografía e Información 2006; see Figure 8.2 in Pfeiler 2014: 210). Therefore, in contrast to Belize, where multilingual language practices have thrived in recent decades, Yucatan Spanish monolingualism is increasingly becoming the norm in the Yucatan Peninsula in Mexico today. It is important to underscore, however, that even though the Yucatec Maya language is not as predominant in the Yucatan Peninsula as in previous centuries, “[it still] enjoys a level of prestige uncommon among indigenous languages in Latin America” (Michnowicz 2015: 24).

### 3. THE USE OF BCVs IN ORAL PRODUCTION

In this section, we provide more detailed insights as regards the use of BCVs in oral production, which largely reflects the sociolinguistic nature of bi/multilingualism in each of the respective communities that we study.

#### 3.1. Northern Belize

The most remarkable characteristic of *hacer* BCV use in Northern Belize is its high degree of productivity. Previous research has shown that the light verb *hacer* overwhelmingly occurs with English lexical verbs, as illustrated in (2), taken from Balam

2021: 92). Although older bi/multilinguals from Northern Belize report having heard Spanish/Maya BCVs (e.g., *hacer chichís* ‘(lull a baby) to sleep’, *hacer hich* ‘tie a knot tightly’, etc.), they are not produced in spontaneous speech, and appear to have largely fallen into disuse (Balam 2014, 2015; Balam *et al.* 2014). Balam *et al.* (2021), nonetheless, assert that the existence of this Spanish/Maya template in earlier generations could have contributed to the diffusion and conventionalization of this structure during the community’s transition from Spanish/Maya to Spanish/English bilingualism.

- (2) *Cuando lo hago* do try, *no está nice*  
 When it do.1SG.PRES do try.INF, no be nice  
 ‘When I do try it, it [the food] is not nice’

A notable aspect of *hacer* BCVs in this context is that they have evolved across time. In a cross-generational analysis of 1,750 *hacer* BCVs, Balam (2015) found that speakers over 50 used *hacer* with transitive, intransitive, and ditransitive verbs only. Contrariwise, speakers ranging from 14 to 40 used *hacer* in a broader range of argument structures, including transitive, ditransitive, intransitive, copulative verbs, reverse psychological predicates, control structures, and passives. Results show diachronic development as there are novel BCV forms (e.g., *hacer* in control structures: *hacer choose hacer study* ‘choose to study’) that are only attested among younger generations with higher degrees of proficiency in English. Balam (2016c) subsequently found that higher levels of bilingual competence are associated with more innovative morphosyntactic uses of BCVs.

More recent research suggests that there are certain novel BCV forms that have emerged in Northern Belize but not in other Southwest U.S. communities, where similar *hacer* constructions are used. Drawing on intuitional data elicited via a two-alternative forced-choice (2AFC) task, Balam *et al.* (2023) comparatively analyzed the acceptability of stative and eventive passive BCVs among 149 Northern Belize and 36 Southwest U.S. bi/multilinguals. Results showed that both bilingual groups gave preferential ratings to stative passive BCVs without *hacer*, as in, for instance, *Jessica se molestó porque la batería no estaba* charged ‘Jessica got upset because the battery was not charged’). In the case of eventive passive BCVs, however, Southwest U.S. bilinguals rejected constructions with *hacer*, which are structures that have not been documented in Southwest U.S. By contrast, Northern Belize bi/multilinguals gave the highest ratings to eventive passive BCVs with *hacer* (e.g., *Hector se molestó porque la escuela no fue hecho*

recognized ‘Hector got angry because the school was not recognized’), despite the lack of gender agreement between the light verb and the feminine antecedent noun. Balam and colleagues advance that social conditions, such as positive attitudes towards CS and low levels of linguistic prescriptivism, have been instrumental in fostering a sociolinguistic environment that has allowed *hacer* BCVs to thrive and further grammaticalize in Northern Belize (see Balam 2015; Balam *et al.* 2020).

### 3.2. *The Yucatan peninsula*

Unlike the productivity that characterizes Spanish/English *hacer* BCVs in Northern Belize, the use of Spanish/Maya *hacer* BCVs appears to be rather infrequent in the Yucatan Peninsula in Mexico (Michalski 2017), as illustrated in example (3) taken from Kolmer (2006: 187). To date, documentation of *hacer* BCVs largely appears in descriptive analyses of Yucatan Spanish or language contact outcomes in this region (e.g., Suárez Molina 1996; Kolmer 2006; Sobrino 2010; Pfeiler 2014). Thus, there is still a notable gap in research concerning Spanish/Maya BCVs (Michalski 2016, 2017).

- (3) *Hoy hago puts’ trabajo*  
 Today do.1SG.PRES skip.INF work  
 ‘Today I’m skipping work’

Pfeiler (2014: 218) observes that there is a distinction in the verb strategy that speakers in the Yucatan Peninsula employ when incorporating Maya verbs into Spanish. In contrast to Spanish monolinguals who borrow Maya verbal roots using the *do*-strategy (Amaro Gamboa 1987), as in (3), bilinguals typically integrate Maya verbs using the Spanish inflectional suffix *-ear* (e.g., *se ts’uk-ean* ‘they rot’). Pfeiler’s assertion is notable as it suggests that the use of *hacer* BCVs is not a distinguishing characteristic of bilingual language practices; thus, its employment in this region is limited. Michnowicz’s (2015) observation also points in this direction. According to Michnowicz, in Yucatan Spanish, *hacer* BCVs and other Mayan phrases are used by younger speakers of the middle or upper social classes. Importantly, these Mayan phrases are used to achieve a comic effect and to index local pride and identity (Kolmer 2006). This marked use of BCVs is somewhat parallel to what has been attested among young Cypriot Greeks in London. In a comparative analysis of Cypriot Greek/English *kano* BCVs from three different bilingual groups, Gardner-Chloros and Finnis (2006) found that only young Cypriot



Greeks from London employed BCVs in humorous or mocking contexts (in contrast to the other groups who used BCVs in serious contexts as well).

Michalski (2017) gives further insight into the linguistic features that distinguish the infrequent use of Spanish/Maya *hacer* BCVs in the Yucatan Peninsula from other language contact situations. Crucially, the inserted verbs in BCVs are restricted to a group of 12–15 Yucatec Maya monotransitive lexical verbs. Thus, in comparison to Northern Belize, the use of these constructions in Yucatan is much less productive. These verbs are borrowings (loanwords) that most likely originated from bilingual language practices when there was a high degree of Spanish/Maya bilingualism in the region. As an anonymous reviewer rightly suggests, these *hacer* BCVs may be fixed expressions or lexicalized bilingual phrases in the speech of predominantly Spanish monolinguals.

Now that *hacer* BCVs have been integrated into Yucatan Spanish, Michalski (2017) observes that they may be diachronically evolving into a grammatical construction. As shown in Table 1, Michalski (2016, 2017) provide a list of the most frequently used Yucatec Maya lexical verbs in BCVs, which are similar to those pointed out by Suárez Molina (1996: 110). For the present study, we adopt the translations given by Michalski (2016, 2017: 223–224), as they are more closely aligned with the contemporary use and meaning of these BCVs.

Suárez Molina (1996)		Michalski (2016–2017)	
Spanish/Maya BCV	English translation	Spanish/Maya BCV	English translation
<i>Hacer loch</i>	Hug	<i>Hacer loch</i>	Embrace/hug
<i>Hacer puch'</i>	Crush/squish and season vegetables	<i>Hacer puch</i>	Squish
<i>Hacer chuuk</i>	Soak	<i>Hacer chuuk</i>	Soak
<i>Hacer puts'</i>	Skip/not comply with an obligation	<i>Hacer putz</i>	Skip
<i>Hacer tamaychih</i>	Do evil eye, to prophesy	<i>Hacer tomochi</i>	Jinx
X	X	<i>Hacer chal</i>	Rinse/bathe
<i>Hacer hetsmek'</i>	Carry little children at the hip	<i>Hacer hetzmek</i>	Carry on hip
<i>Hacer ch'op</i>	Poke in the eye	<i>Hacer chop</i>	Poke in the eye
<i>Hacer hich</i>	Tie a knot tightly	<i>Hacer jich</i>	Tighten
X	X	<i>Hacer jach</i>	Scrub
X	X	<i>Hacer koy</i>	Pinch

Table 1: Most frequent Spanish/Maya BCVs [adapted from Suárez Molina (1996: 110) and Michalski (2016, 2017: 223–224)]

Suárez Molina (1996)		Michalski (2016–2017)	
Spanish/Maya BCV	English translation	Spanish/Maya BCV	English translation
X	X	<i>Hacer mek</i>	Embrace
<i>Hacer chuchú</i>	Suckle	X	X
<i>Hacer chichís</i>	Lull a baby to sleep (if intransitive)	X	X
<i>Hacer kuch</i>	Carry	X	X
<i>Hacer lit’í</i>	Tiptoe	X	X
<i>Hacer pats’</i>	Rub	X	X
<i>Hacer tirich</i>	Trick	X	X
<i>Hacer xix</i>	Crumble	X	X
<i>Hacer xuch</i>	Sip	X	X

Table 1: Continuation

Despite the limited number of monotransitive Yucatec Maya lexical verbs found in BCVs, the use of these constructions in the Yucatan Peninsula exhibits some similarities to those observed in Northern Belize. This includes their occurrence in different syntactic verb contexts, namely passive, active, reflexive, and with pronominalized objects (Michalski 2017).

The foregoing discussion has shown that even though contact Spanish varieties spoken in Northern Belize and the Yucatan Peninsula are sociohistorically and linguistically related, they are “sister dialects” that in the last century have followed divergent paths (Balam 2014: 91), especially in relation to the productivity of *hacer* BCVs. Whereas CS is unmarked in Northern Belize, Spanish monolingualism is the societal norm in the Yucatan Peninsula in Mexico. The differing nature of bi/multilingualism in these two communities, described in sections 3.1 and 3.2, is reflected in how *hacer* BCVs are used. In the case of Northern Belize, these constructions are productively used as a CS strategy to optimize the morphosyntactic and lexico-semantic resources available in Spanish and English (Balam 2015, 2016c, 2021). In contrast, *hacer* BCVs are employed in the Yucatan Peninsula mainly as a borrowing strategy used to integrate a limited set of Yucatec Maya lexical verbs. We acknowledge that the conceptualization of BCVs as illustrative of either CS or borrowing has been a topic of debate (see, for instance, Moinszadeh 1999; Balam 2015, 2021). This issue, however, goes beyond the purview of the present paper. Our main concern is to further contribute to the understanding of the semantic nature of *hacer* BCVs in contact Spanish more generally. In the ensuing section, we describe the methodology adopted in the study.

#### 4. METHODOLOGY

Recent research has shown that conducting comparative research among sociolinguistically related contexts allows us to unveil underlying social factors or grammatical outcomes that may not surface in the isolated study of different bi/multilingual communities (Balam *et al.* 2020; Balam *et al.* 2023). Following this line of research, we analyze two datasets that are representative of Northern Belize and the Yucatan Peninsula in Mexico.

##### 4.1. Data

The data for our study were drawn from Balam’s (2016b) corpus of oral production data from Northern Belize and from Michalski’s (2021) *Yucatan Spanish Twitter Corpus*. The main factor that led to the inclusion of *Twitter* data in our comparative analysis was the frequency of use of BCVs. As described in section 3.2, in the Yucatan Peninsula, *hacer* BCVs are infrequently used in oral production. However, one source that has proven fruitful for BCV data and linguistic data more generally is *Twitter* (see Bohmann 2020: 253–254, for discourse features of *Twitter*). Founded in 2006 and now rebranded as *X*, *Twitter* is a microblogging platform that is characterized by its informal nature, colloquial speech style, and use of creative grammatical structures (Rodríguez Riccelli 2018; Bohmann 2020). Within the last decade, there has been increasing interest in using *Twitter* as a rich source of linguistic data (e.g., Claes 2017; Michalski 2017; Hoff 2020, among others). It is noteworthy that while some scholars have questioned the suitability of *Twitter* data in the study of discourse markers (cf. De Smet and Enghels 2020), others envision *Twitter* speech as an intermediary register between oral and written discourse that is valuable in the study of infrequent morphosyntactic structures. As Rodríguez Riccelli (2018: 330) aptly underscores, *Twitter* data have “proven to be particularly useful for the analysis of forms that are relatively rare in the input and difficult to elicit in an interview or laboratory setting.” In light of this observation, we analyzed BCVs drawn from both oral production (Northern Belize) and *Twitter* (Yucatan in Mexico) to cast new insights into our knowledge of *hacer* BCVs in contact Spanish.

#### 4.1.1. Orange Walk, Northern Belize

A total of 903 Spanish/English *hacer* BCVs were manually extracted from sociolinguistic interviews with 18 speakers (ages 14–17:  $n = 6$ ; ages 18–20 = 2; ages 21–40 = 10) from Orange Walk, Northern Belize. These sociolinguistic interviews, which lasted between 20 and 30 minutes, were carried out by the first author in Orange Walk, Belize in 2014 (Balam 2016a). Participants in the interviews were all Spanish-dominant bi/multilinguals who were frequent Spanish/English code-switchers (see Balam 2015, 2016a, 2016b, for further details about the sample). The sample comprised speakers whose relative production of Spanish/English mixed nominal constructions was at least of 50 percent when compared to the overall production of nominal constructions (see Balam 2016a: 417). Thus, our findings particularly apply to Northern Belize bi/multilinguals, whose CS practices are frequent and unmarked.

#### 4.1.2. Yucatan peninsula, Mexico

A controlled *Twitter* search of the ten most frequently used Spanish/Maya *hacer* BCVs was conducted by Michalski (2016, 2017, 2021), as shown in Table 2. The study yielded a total of 237 ( $n = 237$ ) tweets, which comprise the dataset analyzed here. Michalski employed *Twitter* as a data collection tool because this platform allows users to engage in free and unrestricted discussions, closely simulating informal and naturalistic language contexts. The participants in Michalski's (2021) *Yucatan Spanish Twitter Corpus* were 237 *Twitter* users. A subsequent analysis of the geotags and *Twitter* profiles conducted by the author confirmed that most *Twitter* users associated with the tweets were likely monolingual Spanish speakers from the Yucatan Peninsula in Mexico ( $n = 110$ ).

Spanish/Maya BCV	English translation
<i>Hacer</i> <b>loch</b>	Embrace/ hug
<i>Hacer</i> <b>puch</b>	Squish
<i>Hacer</i> <b>chuuk</b>	Soak
<i>Hacer</i> <b>putz</b>	Skip
<i>Hacer</i> <b>tomochi</b>	Jinx
<i>Hacer</i> <b>hetzmek</b>	Carry (on the hip)
<i>Hacer</i> <b>chop</b>	Poke in the eye
<i>Hacer</i> <b>jich</b>	Tighten
<i>Hacer</i> <b>jach</b>	Scrub

Table 2: Most frequently used *hacer* Spanish/Maya BCVs (adapted from Michalski 2017: 223–224)

It is important to note that we focused on bilingual constructions specifically as *hacer* BCVs do not have equivalent morphosyntactic structures in Spanish, English or Yucatec Maya (in contrast to other language pairs such as Persian/English: Moinzadeh 1999).

#### 4.2. Data analysis

All lexical verbs in *hacer* BCVs were individually coded for semantic domain. The semantic (sub-)categories used in this analysis were initially gleaned from previous studies that have examined the openness of semantic domains to other-language lexical items (Clegg 2010; Aaron 2015; Balam 2016a; Demirçay 2017). Modifications were necessary as none of the previous classification schemes were complex and meticulous enough to conduct a comparative analysis of two relatively large datasets from two different language contexts. Whereas Demirçay (2017) analyzed 48 BCVs, our study is based on 1,130 BCVs: 903 Spanish/English *hacer* BCVs and 237 Spanish/Maya *hacer* BCVs. Thus, during the coding process in our exploratory study, new sub-categories emerged.

Following Demirçay (2017), we first coded tokens according to three main categories which capture the overarching trends within the data across the two communities, namely 1) ‘life in the Belizean/Yucatecan society – informal aspects’, 2) ‘personality, personal life’, and 3) ‘life in the Belizean/Yucatecan society – formal aspects’. Nested under these main categories were 28 distinct sub-categories (A1, A2, A3, etc. See Appendix A) that provide more detailed insights into the semantic domain. It should be noted that while the three main categories were adopted from Demirçay (2017: 88), new sub-categories emerged during the coding process, which was guided by the data. Our analysis focuses primarily on the distribution of sub-categories which provide a more fine-grained analysis of the semantic use of BCVs (see section 5.2). As *hacer* BCVs on their own did not provide sufficient context for us to assign a sub-category, we expanded the scope to include the broader phrasal, sentential, or visual context.

As it relates to the Yucatan data, we analyzed the entire tweet, the thread of comments related to the tweet, and, in some cases, the *Twitter* biography of the user in order to determine sociolinguistic information. Although *Twitter* provides a rich source of language data, one disadvantage is that it is difficult to access sociolinguistic and

biographical information such as frequency of use of languages (Rodríguez Riccelli 2018). In the case of the Northern Belize data, we mainly focused on the sentential context. After a thorough examination and numerous revisions, the 28 distinct sub-categories shown in Table 3 were identified for the three main categories.

<b>Life in the Belizean/Yucatecan society – informal aspects (n = 18)</b>	<b>Personality, personal life (n = 6)</b>	<b>Life in the Belizean/Yucatecan society – formal aspects (n = 3)</b>
Entertainment, pop culture	Love, intimate relationships, affection	School, education, learning-related
Nourishment	Friendships, social communication	Work-related
Social activities	Sexuality	Government, police, law
Traffic, transportation	Personal thoughts, actions, feelings, dreams, advice	
Sports	Family, childhood, personal past	
Poetry	Future goals, personal development	
Local celebrations, traditions		
Household chores		
Daily routine, life		
Technology, social media		
Health, death		
Violence, drugs		
Travel, tourism		
Religion, superstition		
Language		
Society, country, history		
Nature, agriculture		
Money-related, economy		

Table 3: Categories and sub-categories

In addition to coding lexical verbs for semantic domain, we also analyzed references to identity in these constructions, as previous work has shown that BCVs may have an identity function. Makihara (2005: 747–749), for instance, describes a language shift situation in Easter Island where younger speakers, who predominantly speak Spanish positively, identify with their indigenous Rapa Nui heritage. Makihara reports that children commonly employ hybrid *hacer* BCVs (e.g., *Hizo hore el ñao* ‘The neck tube got cut’) and other lexical items from the Rapa Nui language in their Spanish variety to evoke a Rapa Nui voice and index their Rapa Nui identity. In the case of Yucatan, it has been suggested that these constructions may also be used to index group identity (Kolmer 2006; Michnowicz 2015). To examine whether this is the case, we coded tokens in both

datasets for references to identity (or not) in the phrasal, sentential, or visual context (i.e., images, punctuation, text effects, etc.). As shown in Table 4, we coded for cases in which there was no reference to identity, as well as for cases in which there was an explicit or implicit reference to ethnic or linguistic identity.<sup>6</sup> In the example for Category I shown in Table 4, for instance, there is no reference to identity, as the Spanish/Maya BCV is only used to express the idea of one’s heart being squished or broken. On the other hand, in the example for Category II, there is emphasis on what a Yucatecan should know regarding Yucatecan Maya cuisine and the way people eat food. In this example, *x’nipec* ‘dog snout’ refers to a spicy Yucatecan sauce that is typically made with habanero pepper, lime or bitter orange juice, purple onion, coriander, and salt.

Category	Reference to identity	Example
I	No reference to identity	<i>Me hacen <b>puch</b> el corazón.</i> ‘They squish my heart’.
II	Reference to ethnic identity – explicit	<i>#UnYucatecoSabeQue es hacer <b>puch</b> su comida, ponerle <b>x’nipec</b>, hacerle <b>chuc</b> y tomar en su pichel.</i> ‘#AYucatecanKnowsWhatitMeans to squish their food, put spicy pepper sauce (on their food) and drink from their pitcher’.
III	Reference to ethnic identity – implicit	<i>No s[é] ustedes pero yo voy por mi pan par[a] hacer chuuk en mi chocolate.</i> ‘I don’t know about you all, but I am going to get my bread to soak in my chocolate (drink)’.
IV	Reference to linguistic identity – explicit	<i>[N]o le hacen <b>loch</b> porque no saben qu[é] es.</i> ‘They don’t hug him because they don’t know what that means’.
V	Reference to linguistic identity – implicit	<i>Confieso que le hice “<b>Tomochi</b>” a un gran fanático del #RealMadrid!</i> ‘I confess that I jinxed a big Real Madrid fan’

Table 4: References to identity

Coding for semantic domain allowed us to shed light on the conceptual level of meaning that is found in the lexical verb in *hacer* BCVs. In contrast, our coding of potential references to identity enabled us to provide insight into the affective layer of meaning in BCVs, which relates more to the speaker’s or *Twitter* user’s personal feelings (see Leech 1981, for an overview of types of meaning).

To analyze the data, we carried out descriptive analyses. Furthermore, we conducted inferential statistical analyses using *R* (R Core Team 2023) to further analyze

<sup>6</sup> As references to identity were more characteristic of tokens in the Yucatan data, we provide here only examples from this dataset.

the relationship between the communities (categorical independent variables) and the semantic domains or references to identity (dependent variables).

#### 4.3. Research questions

The present study was guided by the following two research questions:

- 1) **RQ1:** What are the similarities and differences in the openness of semantic domains to other-language lexical verbs in *hacer* BCVs from Northern Belize and the Yucatan Peninsula in Mexico?

**Hypotheses:** Considering previous findings (Balam 2016a), we anticipated that semantic domains most open to English lexical verbs would be those related to ‘Life in the Belizean society – formal aspects’. Specifically, we expected that ‘school/education/learning-related’ and ‘work-related’ sub-categories would be most open to English lexical verbs. As it relates to the Yucatan data, we expected that semantic domains related to Life in the Yucatecan society – Informal aspects, would be most open to Yucatec Maya lexical verbs. We expected that there would not be a single category that markedly shows a greater degree of openness to Yucatec Maya lexical verbs. This would align with the fact that, among the most frequent Yucatec Maya lexical verbs in BCVs (see Table 2), there is no verb related to a particular context that is predominant.

- 2) **RQ2:** What role does the *hacer* BCV construction play in the indexing of a local or regional identity?

**Hypotheses:** In light of previous observations (Kolmer 2006; Michnowicz 2015), we anticipated that BCVs would be encoded with an affective meaning, and hence be used as identity markers only in the Yucatan dataset. Thus far, no descriptive or empirical study on BCVs in Northern Belize has suggested that these constructions may have an identity function in this context.

## 5. RESULTS

### 5.1. Distribution of main semantic categories

Contrary to our expectations, ‘life in the Belizean/Yucatecan society – informal aspects’ was the overarching semantic domain most open to other-language lexical verbs. This was especially attested in the Yucatan dataset, as semantic sub-categories open to Yucatec



Maya lexical verbs were predominantly related to informal aspects of Yucatecan society (57%, 135/237). In the case of Northern Belize, lexical verbs in *hacer* BCVs were more evenly distributed across the three main semantic domains: ‘life in the Belizean Society – informal aspects’ (37%), ‘personality, personal life’ (27%), and ‘life in the Belizean Society – formal aspects’ (36%).

The primary difference between the two datasets pertains to the main semantic category that is least open to other-language lexical verbs. Whereas in the Northern Belize data, ‘personality, personal life’ was least open to English lexical verbs (27%, 242/903), the main category ‘life in the Yucatecan society – formal aspects’ was least open to Maya verbs (9%, 21/237) in the Yucatan data. Considering that Yucatec Maya does not have official status in Mexico and is generally excluded from formal contexts (Sobrino 2010), it is not surprising that semantic sub-categories related to formal aspects of Yucatecan life (e.g., work-related) are least open to Yucatec Maya lexical verbs.

## 5.2. *Distribution of semantic sub-categories*

Table 5 provides a more detailed insight into the semantic sub-categories that predominantly favor (> 5%) English lexical verbs in the Northern Belize corpus. The remaining 23 domains, each with a token frequency equal to or lower than five percent, were combined under the semantic-subcategory ‘other’ (for full results across both corpora, see Appendix A). The data reveal that the sub-category ‘school, education, learning-related’ is most open to English lexical verbs, representing 20 percent (n = 184) of the data. This means that roughly one in every five *hacer* BCVs in the Northern Belize corpus incorporates an English lexical verb that relates to the educational context (e.g., *drop out, suspend, promote, attend, major, study, pass, fail, register, transfer, read, procrastinate, discipline, improve, graduate*, etc.). This is followed by the sub-category ‘work-related’ with 13 percent (n = 120) and the sub-category ‘B2 friendships, social communication’ with 10 percent (n = 90).

Semantic sub-category	Number of examples	Percentage
C1 School, education, learning-related	184	20
C2 Work-related	120	13
B2 Friendships, social communication	90	10
B4 Personal thoughts, actions, feelings, dreams, advice	61	7
A17 Society, country, history	52	6
Other	396	44

Table 5: Semantic sub-categories most open to English lexical verbs

Different patterns are attested in the Yucatan data. The data in Table 6 reveal that, contrary to our expectations, there were semantic sub-categories that favored Yucatec Maya lexical verbs. The sub-category ‘nourishment’ is most open to Yucatec Maya lexical verbs, representing 40 percent of *hacer* BCVs ( $n = 94$ ), followed by ‘love, intimate relationships, affection’ (19%,  $n = 45$ ) and ‘friendships, social communication’ (10%,  $n = 24$ ).

Semantic sub-category	Number of examples	Percentage
A3 Nourishment	94	40
B1 Love, Intimate Relationships, Affection	45	19
B2 Friendships, Social Communication	24	10
C1 School, Education, Learning-Related	16	7
A1 Entertainment, Pop Culture	14	6
Other	44	18

Table 6: Semantic domains most open to Yucatec Maya lexical verbs

As Tables 5 and 6 illustrate, only two semantic sub-categories were common in both corpora: ‘school, education, learning-related’ and ‘friendships, social communication’. With respect to the former, speakers from Northern Belize often discussed topics such as educational development, college majors, school-related events, or activities, etc. In the Yucatan data, *Twitter* users consistently expressed their desire to skip classes (e.g., ... *hacer putz escuela* ‘...skip school’).

As it relates to ‘friendships, social communication’, the Northern Belize data revealed that lexical verbs in this semantic sub-category were used to describe social situations or relationships between friends, among classmates or colleagues. By contrast, in Yucatan, lexical verbs in this sub-domain were often used in contexts of social banter among *Twitter* users, such as when vividly discussing sports events or describing

quotidian instances of daily social communication (e.g., shaking hands: *hacen puch' mi dedo* '[they] squish my finger').

With respect to differences, in the Northern Belize data, 'work-related' English lexical verbs (e.g., *resign, follow, offer, assess, work, sell, apply, retire, operate, deliver*, etc.) in BCVs were used to describe events or tasks related to the workplace. The sub-category 'personal thoughts, actions, feelings, dreams, advice' also favored English lexical verbs. This is a noteworthy finding, as it reveals that speakers are comfortable using English verbs in *hacer* BCVs when expressing deeply personal thoughts and feelings. Example (4) below is representative of many other BCVs from the Northern Belize corpus, where speakers produce a stream of consciousness, expressing their inner thoughts and opinions in a spontaneous and uninhibited manner.

(4) **Discourse**     **B4 Personal thoughts, actions, feelings, dreams, advice**

- a.                    if I do this *que es lo que va a pasa sino*, or you know, seeing things in different ways...then *ya vas a sabé cómo hace* tackle certain situations.  
                          'if I do this what is going to happen otherwise, or you know, seeing things in different ways...then [you] already know how to tackle certain situations'

In the case of Yucatan, the semantic sub-category 'nourishment' was most open to Yucatec Maya lexical verbs. Due to its large proportion, this semantic pattern can be considered specific to the Yucatan community. Notably, the majority of *hacer* BCVs (75 out of 94) included the Maya verb **chuk** 'soak'. In Yucatan, it is a very common habit to soak food in coffee or other hot beverages, especially during cold weather. As (5) below shows, the *Twitter* user employs the Spanish/Yucatec Maya BCV to visually illustrate how bread is soaked in the Yucatan region.

(5) **Tweet**             **A3 Nourishment**

- b.                    *Yo leo frente frío y enseguida!*  
                          *A: Todo el mundo, a esto se le llama Hacer **Chuk** el pan*  
                          'I read cold weather ahead and right away!  
                          Everybody, this is what is called to "Soak" the bread'



Food items that are commonly soaked or soakables, are typical of the Yucatecan cuisine in Mexico. While some tweets mentioned the soaking of regular bread or even a sandwich, the most popular soakable among Yucatecans are *globitos* and *pan dulce*. Whereas *globitos* are a type of small and round-shaped cookie from the Yucatecan brand *Dondé*, the term *pan dulce* ‘sweet bread’ refers to different types of local Yucatecan pastries. The use of the Yucatec Maya verb **chuk** is further illustrated in (6), where a disappointed *Twitter* user expresses their dissatisfaction at the sight of a plain cup of coffee without soakables.

(6) Tweet      A3 Nourishment

- a. *Gente de #Yucatán, quién tomó alguna vez así su café?*  
 ‘People from #Yucatán, who ever drank their coffee like this?’



*Faltan los globitos y el pan dulce para hacer chuk...*  
 ‘The globitos and sweet bread to soak...are missing’

Most of the remaining 19 BCVs included the use of the Maya verb **puch** ‘squish’. This was often used in the context of preparing food, in line with Suárez Molina’s (1996) translation of **puch** as ‘crush/squish vegetables and season them’. The items that were mentioned in the tweets usually included local Yucatecan foods or cooking ingredients such as rice, avocado, *frijoles con puerco* ‘beans with pork’, habanero and cilantro.

Finally, the semantic sub-category ‘love, intimate relationships, affection’ was also distinctive of the Yucatan data. This semantic sub-category accounts for 19 percent of the tweets analyzed. However, it was one of the least favorable sub-categories in the Northern Belize corpus data, with a proportion of only one percent. In the Yucatan data, this sub-category mostly occurred with the Maya verb **loch** ‘hug’ with a few exceptions of **puch** ‘squish’. These BCVs were used in the context of intimate and affectionate relationships between two individuals. In (7), for instance, the *Twitter* user expresses their desire for everybody to have someone to hug and be hugged during the cold weather. Most of the tweets referred to an intimate but platonic relationship between friends or family. Although not often explicitly mentioned, some tweets could also imply a romantic relationship but not in any sexual sense.

- (7) **Tweet**      **B1 Love, intimate relationships, affection**
- a.      *Excelente fin de semana!*  
*Que todos tengamos alguien para hacer loch en este frío :)*  
 ‘Excellent weekend!  
 May we all have someone to hug in this cold weather :)’

To further examine the association between the semantic domains and the two communities (Northern Belize and Yucatan, Mexico) we conducted a chi-square test, which revealed that there is a probable relation between the semantic sub-categories and the two communities ( $\chi^2 = 528.89$ ,  $df = 26$ ,  $p < 0.001$ ). Likewise, the calculation of the Cramér’s V test, which measures the strength of the association between the two variables, confirms that the strength of the relation is relatively strong (Cramér’s V = 0.482). Our results, therefore, indicate that there are context-specific patterns in the semantic sub-categories that are most open to other-language lexical verbs in *hacer* BCVs across Northern Belize and Yucatan in Mexico.

### 5.3. References to identity

In this section, we shift our focus to results that shed light on whether *hacer* BCVs are used as identity markers. Table 7 shows the distribution of references to identity across the two datasets. In Northern Belize, only six percent of BCVs ( $n = 52$ ) have an implicit or explicit reference to identity. Contrariwise, in the Yucatan data, there is a relatively frequent occurrence of references to ethnic or linguistic identity, accounting for 47 percent ( $n = 111$ ) of the data analyzed.

		Northern Belize		Yucatan	
Reference to Identity		Number	Percentage	Number	percentage
I	No reference to identity	851	94%	126	53%
II	Reference to ethnic identity – explicit	0	0%	23	10%
III	Reference to ethnic identity – implicit	43	5%	60	25%
IV	Reference to linguistic identity – explicit	0	0%	21	9%
V	Reference to linguistic identity– implicit	9	1%	7	3%
<b>Total</b>		<b>903</b>	<b>0%</b>	<b>237</b>	<b>100%</b>

Table 7: Distribution of BCVs across references to identity

In the Northern Belize data, there were only a few examples that refer to identity. For example, (8) highlights the bi/multilingual linguistic identity of Spanish speakers in

Northern Belize. In other cases, speakers made specific reference to their mixed Creole/Mestizo identity or to Yucatec Maya folklore and superstitious beliefs.

(8) **Discourse**    **A16 Language: V Reference to linguistic identity – implicit**

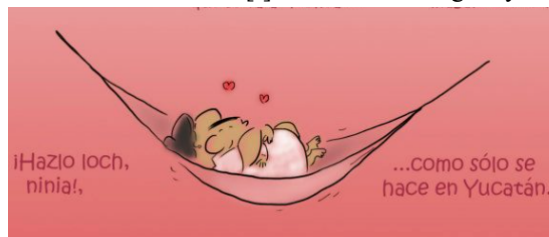
- a.                    *los hago* text in English *y ellos me contestan en Spanish*  
                          ‘[I] text them in English and they reply to me in Spanish’

In the Yucatan data, BCVs with ‘nourishment’ lexical verbs frequently occurred in contexts where references to the Yucatec Maya identity were made. Nourishment BCVs accounted for 45 percent (50/111) of the tweets that primarily made implicit or explicit reference to ethnic identity (76%, 38/50). The sub-category ‘love, intimate relationships, affection’ is also another domain where references to identity occur, accounting for 14 percent (15/111). Thus, whereas *hacer* BCVs cast light on the bilingual or bicultural identity of speakers from Northern Belize, the Yucatan data show that these constructions are more frequently used as markers of a Yucatec Maya ethnolinguistic identity by *Twitter* users. Crucially, the data reveal that the indexing of this group identity was expressed in three primary ways:

- 1) By mentioning or visually including objects or lexical items that are inherently linked to the Yucatec Maya culture, the Yucatecan cuisine, or the Yucatec Maya language, such as (a) food and local ingredients (e.g., *globitos* ‘little globe cookies’, *frijoles con Puerco* ‘beans with pork’, *x’nipec* ‘spicy Yucatecan sauce’, etc.), (b) everyday objects or items which are essential in the Yucatecan culture (e.g., hammocks, as illustrated in (9), Yucatecan sombrero, etc.), and (c), Yucatec Maya loanwords or Yucatan Spanish expressions (e.g., *xic* ‘armpit’, *xixito* ‘remainder, rest, residue typically of a meal, ¿vera hija? ‘right child?’, etc.)

(9) **Tweet**                    **B1 Love, intimate relationships, affection; III Reference to ethnic identity, implicit**

- a.                    *#NoEsPorPresumir pero sé hacer loch bien bonito*  
                          ‘#NotToShowOff but [I] know how to hug very nicely’



‘Hug him, child! ... as it is only done in Yucatan’

- 2) By emphasizing their Yucatec Maya identity with (a) hashtags (e.g., *#OrgullosamenteYucateco* ‘#ProudlyYucatec’, *#UnYucatecoSabeQue* ‘#AYucatecanKnowsThat’, *#esunatradicionyucatecanuestra* ‘#It sourYucatecanTradition’, *#UnYucatecoSiempre* ‘#AYucatecanAlways’, or *#tuitandoenmaya* ‘#tweetinginyucatecanmaya’), (b) texts in Twitter bios (e.g., *Nacido en Yucatán, tierra de maravías, ija* ‘Born in Yucatan, wonderland, child’, *100% Yucateca* ‘100% Yucatecan’, or *Yucateca amante de las siestas en hamaca después del #maldelpuerco* ‘Yucatecan lover of hammock naps after a #foodcoma’), and (c) the use of different fonts, italics, bold letters, or quotation marks to highlight Yucatec Maya lexical items.
- 3) By employing *albur*, which is ‘a game of words’ in which the double meaning is typically sexual or eschatological in nature (Anaya and Cózar Angulo 2014: 144). *Albur* is very typical of Mexican popular culture and is generally used among friends and colleagues (Beristáin 2000; Anaya and Cózar Angulo 2014). In (10), for example, the use of the *hacer* BCV has a double entendre: while it conveys the literal meaning of ‘soaking sponge cake’, it can also imply a sexual act. In Mexico, *albur* is a sociolinguistic practice that is generally viewed as a skill that requires mental dexterity, quick wit, and linguistic creativity. In Yucatan, *albur* has taken its own form known as *bomba*. These are picaresque short rhymes with a sexual connotation that are typically accompanied by traditional Yucatecan music and dancing (see Anaya and Cózar Angulo 2014: 148). Example (11) is linguistically reminiscent of *bomba*.

- |      |       |  |
|------|-------|--|
| (10) | Tweet | <b>A3 Nourishment: III Reference to ethnic identity – implicit</b><br>a. @SoyYuca <i>La tarde esta para hacer <b>chuk</b> el bizcocho...vera hija?</i><br>‘[It] is the type of afternoon to soak the sponge cake, right child?’                                  |
| (11) | Tweet | <b>A7 Poetry: III Reference to linguistic identity – implicit</b><br>a. <i>Frío tu <b>tuch</b>, cuando me haces <b>loch</b>, pues ya no aprieta tu <b>pirix</b>.</i><br>‘Your belly button [is] cold, when [you] hug me, well now it does not squeeze your butt’ |

Our results show that while there are some similarities between communities, there are also significant differences in the semantic nature of *hacer* BCVs in Northern Belize and

the Yucatan Peninsula in Mexico. In the following section, we discuss the implications of our findings.

## 6. DISCUSSION AND CONCLUSION

Previous research has shown that even though Northern Belize and the Yucatan Peninsula in Mexico share sociohistorical ties, they have followed very different sociolinguistic paths. Whereas the former is characterized by frequent Spanish/English CS and the productive use of *hacer* BCVs, the latter is distinguished by Yucatan Spanish monolingualism and infrequent use of *hacer* BCVs. Our study has shed light on the similarities and differences in the semantic use of these constructions.

In relation to our first research question, we found that there was a relatively even distribution across the three main semantic domains in the Northern Belize data. There was also a broader range of sub-categories that were open to other-language lexical verbs in comparison to the Yucatan data (see Appendix A). This is consonant with the frequent use and acceptability of bilingual language practices in Northern Belize, where CS is prolific (Balam 2016b; Balam and Prada Pérez 2017). In the case of Yucatan, bilingual language practices are more marked and fewer semantic domains are open to Yucatec Maya lexical verbs. Sub-categories most open to English lexical verbs in the Northern Belize dataset were ‘school, education, learning-related’ (20%) and ‘work-related’ (13%), which are semantic patterns that remarkably mirror previous findings for the nominal domain (Balam 2016a). In the case of Yucatan, even though *hacer* co-occurs with a very limited set of Yucatec Maya lexical verbs, the semantic categories, ‘nourishment’ (40%) and ‘love, Intimate relationships, affection’ (19%) particularly favored Maya lexical verbs.

In terms of similarities, the sub-categories ‘school, education, learning-related’ and ‘friendships, social communication’ evinced openness ( $> 7\%$ ) across the two datasets. Similar patterns were reported by Demirçay (2017). In the case of Northern Belize, the high degree of openness of ‘school, education, learning-related’ and ‘work-related’ can be attributed to the more frequent use of English in educational and professional settings. Our findings coincide with Demirçay’s (2017) and reveal that ‘education’ is the sub-domain most open to other-language verbs, especially in contact situations where CS is common and other-language verbs are drawn from the official or dominant language of



instruction in schools. In our dataset, this finding reflects the experiences of speakers in Northern Belize who are linguistically Spanish-dominant, but whose educational experiences at school (i.e., content courses, textbooks, etc.) are almost exclusively in English. Importantly, in our sample, ‘education’ is the semantic sub-domain most open to English lexical verbs not only among students, but also among participants who held full-time jobs and were not enrolled in any academic program.

In terms of main cross-community differences, while ‘work-related’ and ‘personal thoughts, actions, feelings, dreams, advice’ favored English lexical verbs in Northern Belize, ‘nourishment’ and ‘love, intimate relationships, affection’ showed a higher degree of openness to Maya lexical verbs in Yucatan. In the Yucatan dataset, these hybrid constructions highlight specific cultural aspects of the Yucatecan society, mainly food. Thus, this particular use of *hacer* BCVs constitutes a community linguistic norm that is distinctive of the Yucatan Peninsula in Mexico.

As regard our second research question, we found that *hacer* BCVs are not used as markers of identity in Northern Belize. In the Yucatan data, however, these linguistic structures function as identity markers sometimes, as we had hypothesized. We found that 47 percent of tweets in the Yucatan dataset had references either to ethnic or linguistic identity. As shown in section 5.3, BCVs are used “to [express] a sense of local pride and identity” (Micnowicz 2015: 34). The use of *albur* (including the Yucatecan variant *bomba*) in BCVs is especially noteworthy as it is “a way of expressing, appropriating and manipulating language, as well as a means for establishing social ties and identity of social groups” (Anaya and Cózar Angulo 2014: 159). Given that BCVs in this language contact situation comprise Yucatec Maya lexical verbs and linguistic practices that are context-specific, we analyze this morphosyntactic frame as an emblematic grammatical structure that can be deliberately used by Yucatan Spanish speakers to index their indigenous Yucatec Maya ethnolinguistic identity.

The main difference between Northern Belize and Yucatan is that whereas in the former context bilingual CS has emerged as a distinctive marker of bi/multilingual speakers’ national and ethnic identities (Balam 2016b: 33–43; Balam and Prada Pérez 2017), in the latter it is a morphosyntactic structure that sometimes has an ethnolinguistic identity function (see Bucholtz and Hall 2005, for relevant discussion on the use of linguistic systems or structures to index identity). Thus, in Northern Belize, bi/multilinguals productively use Spanish/English CS to project their Belizean and

Maya/Mestizo identities. Contrariwise, in the case of Yucatan, Spanish speakers employ Spanish/Maya BCVs (albeit infrequently) to index a Yucatec Maya ethnolinguistic identity.

Collectively, our findings show that semantic patterns in the use of BCVs are intricately tied not only to the nature of bilingualism (i.e., stable bilingualism vs. language shift) and CS practices (i.e., frequent CS vs. less CS/more borrowing of lexical items), but also to community-specific linguistic norms, which may be related to the indexing of a social or group identity. In relation to the study of bilingual speech patterns more generally, a notable finding is that ‘technology’ was not a sub-category with a high degree of openness to other-language verbs in both communities. Instead, our results suggest that the semantic sub-categories most open to other-language verbs in BCVs are closely related to the status and social functions of languages. Given that English is the official language of instruction in Belize and that the classroom is the main social context where Belizeans develop their lexical repertoire from early childhood, it is not surprising that ‘education’ is the semantic domain most open to English lexical verbs in *hacer* BCVs. In the case of Yucatan in Mexico, however, the language of instruction in schools is predominantly Spanish. On the other hand, Yucatec Maya is largely restricted to the home. This accounts for the openness of ‘nourishment’ to Yucatec Maya lexical verbs in BCVs. Based on our results, we posit that classroom discourse (student-to-student, student-to-teacher, and teacher-to-student) likely plays a deterministic role in how unilingual and other-language verbs are learnt and used in conversations outside of schools, a topic that merits further investigation.

Although our study contributes to our understanding of the semantic nature of BCVs, there were limitations. The two datasets markedly differed in terms of the total number of tokens analyzed. The small size of the Yucatan dataset limited our ability to use inferential statistical methods to further analyze the data from a comparative lens. Our results, therefore, should be taken with caution. More research on a larger number of Spanish/Maya BCVs is needed to further examine semantic patterns in the use of these constructions in Yucatan, both in terms of their openness to semantic domains and their use as identity markers. Another primary limitation of our study is that although tweets share similarities with oral production, the two datasets are not fully comparable. The anonymous nature of *Twitter* allows users to express their thoughts and feelings in ways that may differ from everyday conversations. Tweets also allow users to employ different

strategies (e.g., images, text effects, etc.) to highlight semantic aspects of their discourse. Future work, therefore, could examine the semantic nature of BCVs from datasets that are more comparable in terms of modality. Comparative research of data from two communities where BCVs are frequently used in oral production may reveal additional or more granular insights. Finally, studies can investigate whether in other CS communities, education is the semantic sub-category most open to other-language lexical items in both the nominal (Balam 2016d) and verbal domains, as the Northern Belize data have shown.

#### REFERENCES

- Aaron, Jessi Elana. 2015. Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism* 19/4: 459–480.
- Amaro Gamboa, Jesús. 1987. *Vocabulario del Uayeísmo en el Habla de Yucatán*. Mérida, México: Universidad Autónoma de Yucatán
- Anaya, Yosi and Xavier Cózar Angulo. 2014. The albur and refrán as tropes for identity construction in Mexico. *Bulletin of the Transilvania University of Braşov Series IV: Philology and Cultural Studies* 7/2: 141.
- Backus, Albert. 2001. The role of semantic specificity in insertional codeswitching: Evidence from Dutch-Turkish. In Rodolfo Jacobson ed. *Codeswitching Worldwide*. Berlin: Mouton de Gruyter, 125–154.
- Balam, Osmer. 2013. Overt language attitudes and linguistic identities among multilingual speakers in Northern Belize. *Studies in Hispanic and Lusophone Linguistics* 6/2: 247–278.
- Balam, Osmer. 2014. Notes on the history and morphosyntactic characteristics of Spanish in Northern Belize. *Kansas Working Papers in Linguistics* 35: 79–94.
- Balam, Osmer. 2015. Code-switching and linguistic evolution: The case of ‘hacer + V’ in Orange Walk, Northern Belize. *Lengua y Migración* 7/1: 83–109.
- Balam, Osmer. 2016a. Semantic categories and gender assignment in contact Spanish: Type of code-switching and its relevance to linguistic outcomes. *Journal of Language Contact* 9/3: 405–435.
- Balam, Osmer. 2016b. *Language Use, Language Change and Innovation in Northern Belize Contact Spanish*. Gainesville, FL: University of Florida dissertation.
- Balam, Osmer. 2016c. Mixed verbs in contact Spanish: Patterns of use among emergent and dynamic bi/multilinguals. *Languages* 1/1: 1–21.
- Balam, Osmer. 2021. Beyond differences and similarities in codeswitching and translanguaging research. *Belgian Journal of Linguistics* 35/1: 76–103.
- Balam, Osmer and Ana de Prada Pérez. 2017. Attitudes toward Spanish and code-switching in Belize: Stigmatization and innovation in the Spanish classroom. *Journal of Language, Identity and Education* 16/1: 17–31.
- Balam, Osmer, Ana de Prada Pérez and Dámaris Mayans. 2014. A congruence approach to the study of bilingual compound verbs in Northern Belize contact Spanish. *Spanish in Context* 11/2: 243–265.

- Balam, Osmer, María del Carmen Parafita Couto and Mia Amanda Chen. 2021. Being in bilingual speech: An analysis of estar 'be' constructions in Spanish/English code-switching. *Journal of Monolingual and Bilingual Speech* 3/2: 238–264.
- Balam, Osmer, María del Carmen Parafita Couto and Jacob Shelton. In Press. Belizean Spanish: Past, present and future. In Leonardo Cerno, Hans-Jörg Döhla, Miguel Gutiérrez Maté, Robert Hesselbach and Joachim Steffen eds. *Contact Varieties of Spanish and Spanish-lexified Contact Varieties*. Berlin: Mouton de Gruyter.
- Balam, Osmer, María del Carmen Parafita Couto and Hans Stadthagen-González. 2020. Bilingual verbs in three Spanish/English code-switching communities. *International Journal of Bilingualism* 24/5–6: 952–967.
- Balam, Osmer, Hans Stadthagen-Gonzalez, Eva Rodríguez-González and María del Carmen Parafita Couto. 2023. On the grammaticality of passivization in bilingual compound verbs. *International Journal of Bilingualism* 27/4: 415–431.
- Beristáin, Helena. 2000. El albur. *Acta Poética* 21: 399–422.
- Bohmann, Axel. 2020. Situating Twitter discourse in relation to spoken and written texts: A lectometric analysis. *Zeitschrift für Dialektologie und Linguistik* 87/2: 250–284.
- Bucholtz, Mary and Kira Hall. 2005. Identity and interaction: A sociocultural linguistic approach. *Discourse Studies* 7/4–5: 584–614.
- Cal, Angel. 1991. *Rural Society and Economic Development: British Mercantile Capital in Nineteenth-Century Belize*. Tucson, AZ: University of Arizona dissertation.
- Claes, Jeroen. 2017. La pluralización de haber presentacional en el español peninsular: Datos de Twitter. *Sociolinguistic Studies* 11/1: 41–64.
- Clegg, Jens. 2010. An analysis of the motivations for borrowing in the Spanish of new Mexico. In Susana Riviera-Mills and Daniel Villa eds. *Spanish of the U.S. Southwest: A Language in Transition*. Madrid: Iberoamericana Vervuert, 223–238.
- Croft, William. 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Pearson Education.
- Demirçay, Derya. 2017. *Connected Languages: Effects of Intensifying Contact between Turkish and Dutch*. Tilburg: Tilburg University dissertation.
- De Smet, Emma and Renata Enghels. 2020. Los datos en Twitter como fuente del discurso oral coloquial: Estudio de caso del marcador discursivo *en plan*. *ORALIA* 23/2: 199–218.
- Elliott, Larry S. 1995. *National Identity and Media System Dependency in Belize*. Gainesville, FL: University of Florida dissertation.
- Gardner-Chloros, Penelope. 2009. *Code-switching*. Cambridge: Cambridge University Press.
- Gardner-Chloros, Penelope and Katerina Finnis. 2006. *Using LIDES to correlate compound verbs with other factors*. Communication at *Sociolinguistics Symposium 16*, University of Limerick, 6–8 July.
- González-Vilbazo, Kay and Luis López. 2011. Some properties of light verbs in code-switching. *Lingua* 121/5: 832–850.
- Hagerty, Timothy W. 1979. *A Phonological Analysis of the Spanish of Belize*. Los Angeles, CA: University of California dissertation.
- Hoff, Mark. 2020. Cerca mí/a or cerca de mí? A variationist analysis of Spanish locative + possessive on Twitter. *Studies in Hispanic and Lusophone Linguistics* 13/1: 51–78.
- Instituto Nacional de Estadística, Geografía e Información (INEGI). 2006. *Segundo Censo de Población y Vivienda 2005*. Aguascalientes. <https://www.inegi.org.mx/programas/ccpv/2005/>

- Instituto Nacional de Estadística, Geografía e Información (INEGI). 2020. *Censo de Población y Vivienda. [Census of Population and Housing]*. Aguascalientes, Mexico: INEGI. <https://www.inegi.org.mx/programas/ccpv/2020/>
- Jake, Janice L., Carol Myers-Scotton and Steven Gross. 2002. Making a minimalist approach to code-switching work: Adding the matrix language. *Bilingualism: Language and Cognition* 5/1: 69–91.
- Jenkins, Devin L. 2003. Bilingual verb constructions in southwestern Spanish. *Bilingual Review* 27/3: 195–204.
- Jones, Grant. 1998. *The Conquest of the Last Maya Kingdom*. Stanford: Stanford University Press.
- Kolmer, Katrin. 2006. ¡Chuch, Qué Bueno! Vom wiederaufblühen der maya-kultur und ihrer präsenz im spanischen von Mérida (Yucatán, Mexiko). *Romanistik in Geschichte und Gegenwart* 12/2: 179–192.
- Koenig, Edna L. 1975. *Ethnicity and Language in Corozal District, Belize: An Analysis of Code-switching*. Austin, TX: University of Texas dissertation.
- Leech, Geoffrey. 1981. *Semantics: The Study of Meaning*. London: Penguin Books.
- Makihara, Miki. 2005. Rapa Nui ways of speaking Spanish: Language shift and socialization on Easter Island. *Language in Society* 34/5: 727–762.
- Michalski, Ian. 2016. *Yucatecan Verbs that ‘do’: A syntactic analysis of Hacer + MayaV in Yucatan Spanish*. Communication at the *Hispanic Linguistic Symposium*, October 5–8, 2016, Washington D.C.
- Michalski, Ian. 2017. Morphological case and argument structure variation with hybrid Spanish-Yucatec Maya verbs: ‘hacer + V’ in monolingual Yucatán Spanish. In Julia Nee, Margaret Cychosz, Dmetri Hayes, Tyler Lau and Emily Ramirez eds. *Proceedings of the Forty Third Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society, 235–260.
- Michalski, Ian. 2021. *Yucatan Spanish Twitter Corpus of Spanish/Maya Bilingual Compound Verbs*. Unpublished corpus.
- Michnowicz, Jim. 2015. Maya-Spanish contact in Yucatan, Mexico: Context and sociolinguistic implications. In Sandro Sessarego and Melvin González-Rivera eds. *New Perspectives on Hispanic Contact Linguistics in the Americas*. Madrid: Iberoamericana Editorial Vervuert, 21–42.
- Moinzadeh, Ahmad. 1999. Bilingual phenomena: Towards the fateful triangle of language mixture. *Cahier Linguistiques d’Ottawa* 27: 31–63.
- Moseley, Edward H. 1980. From conquest to independence: Yucatan under Spanish rule, 1521–1821. In Edward Mosely and Edward Terry eds. *Yucatan: A World Apart*. Alabama: University of Alabama Press, 83–121.
- Muysken, Pieter. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
- Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16/4: 709–730.
- Pfaff, Carol W. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55/2: 291–318.
- Pfeiler, Barbara. 2014. Maya and Spanish in Yucatan: An example of continuity and change. In Salikoko Mufwene ed. *Iberian Imperialism and Language Evolution in Latin America*. Chicago: University of Chicago Press, 205–224.
- Poplack, Shana and Marjory Meechan. 1998. Introduction: How languages fit together in codemixing. *International Journal of Bilingualism* 2/2: 127–138.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Rodríguez Riccelli, Adrian. 2018. Espero estén todos: The distribution of the null subordinating complementizer in two varieties of Spanish. In Jeremy King and Sandro Sessarego eds. *Language Variation and Contact-Induced Change: Spanish Across Space and Time*. Amsterdam: John Benjamins, 299–333.
- Sobrino, Carlos M. 2010. Contacto lingüístico maya-español: Transferencias en la morfosintaxis del español yucateco. *Temas antropológicos: Revista científica de investigaciones regionales* 32/1: 79–94.
- Suárez Molina, Victor M. 1996. *El Español que se Habla en Yucatán: Apuntamientos Filológicos*. Yucatán: Universidad Autónoma de Yucatán.

*Corresponding author*

Osmer Balam  
 The College of Wooster  
 Department of Spanish  
 1189 Beall Ave  
 Wooster, OH.  
 44691  
 United States  
 Email: [obalam@wooster.edu](mailto:obalam@wooster.edu)

received: January 2024  
 accepted: April 2024

APPENDIX A: DISTRIBUTION OF THE SEMANTIC SUB-CATEGORIES OPEN TO OTHER-  
LANGUAGE LEXICAL VERBS IN TWO CONTACT SPANISH COMMUNITIES

Semantic sub-category	Northern Belize		Yucatan Peninsula	
	N	%	N	%
A1 Entertainment, pop culture	48	5	14	6
A3 Nourishment	30	3	94	40
A4 Social activities	13	1	0	0
A5 Traffic, transportation	8	1	10	4
A6 Sports	27	3	3	1
A7 Poetry	3	0	5	2
A8 Local celebrations, traditions	5	1	3	1
A9 Household chores	0	0	1	0
A10 Daily routine, life	2	0	2	1
A11 Technology, social media	17	2	1	0
A12 Health, death	9	1	1	0
A13 Violence, drugs	10	1	0	0
A14 Travel, tourism	11	1	1	0
A15 Religion, superstition	22	2	0	0
A16 Language	16	2	0	0
A17 Society, country, history	52	6	0	0
A18 Nature, agriculture	15	2	0	0
A19 Money-related, economy	48	5	0	0
B1 Love, intimate relationships, refection	8	1	45	19
B2 Friendships, social communication	90	10	24	10
B3 Sexuality	1	0	5	2
B4 Personal thoughts, actions, feelings, dreams, advice	61	7	7	3
B5 Family, childhood, personal past	49	5	0	0
B6 Future goals, personal development	33	4	0	0
C1 School, education, learning-related	184	20	16	7
C2 Work-related	120	13	4	2
C3 Government, police, law	21	2	1	0
<b>Total</b>	<b>903</b>	<b>100</b>	<b>237</b>	<b>100</b>