Complex Items and Units in Extra-Sentential Code Switching. Spanish and English in Gibraltar

Eugenio Goria

Assistant Professor of Linguistics, Department of Humanities, University of
Turin, Turin, Italy

eugenio.goria@unito.it

Abstract

As is well-known, code-mixing is particularly frequent at clause boundaries and with elements expressing pragmatic meaning. However, most of the literature has focussed on switching of simple elements such as conjunctions and discourse markers. This paper, in contrast, analyses clause peripheral switching involving two complex constructions: left dislocations and pseudo-clefts. The data are from English-Spanish bilingual conversations recorded in Gibraltar. A great majority of the bilingual constructions in the corpus belong to a few types occurring with a restricted set of lexical items. A vast amount of such highly recurrent strings in the data confirm the hypothesis that complex multi word strings that are switched together constitute units in code-mixing, i.e. they are processed as single lexical items.

Keywords

extra-clausal code-mixing – alternation – Gibraltar – left dislocation – pseudo-clefts – construction grammar

1 Introduction

In this paper, I investigate the process of unitisation (Backus, 2003) as an alternative explanation for some aspects of English-Spanish code-mixing in Gibraltar that in previous analyses were mostly attributed to functional-pragmatic features (Goria, 2016). Coherently with the other papers in the present special issue, I argue that the theoretical assumptions of Construction

Grammar (Fillmore et al., 1988; Goldberg, 1995, 2006; Croft, 2001; Hilpert, 2014 inter al.) and of usage-based linguistics (Bybee, 2006, 2010; Traugott and Trousdale, 2013 inter al.) have a great potential for explaining the emergence of patterns in code-mixing, both providing independent alternative evidence to previous findings, and providing explanations for questions that remain unsolved in functional analyses.

Based on the view of constructions as syntactic patterns stored and processed at different levels of abstraction (Diessel, 2016), I will consider two cases of clause-peripheral code-mixing, each involving a specific syntactic construction: left dislocations and pseudo-clefts. Relying on the notion of lexical specificity (Croft and Cruse, 2004 inter al.), I will show that code-mixing is found more frequently with lexically specific instances of these constructions, that is recurrent instances of the construction associated with particular lexical items. Conversely, the phenomenon occurs less frequently with instances attributable to abstracted constructional schemas. To give an example, the transitive construction is schematic in that it involves abstract relations between the predicate and the arguments, regardless of the individual lexical items occurring in the context. However, transitive clauses like *spill the beans* or *kick the bucket* have an idiosyncratic meaning that is not predictable from the more general pattern and that depends on the lexical entries occurring in the construction; cases like these are instances of lexically specific constructions and correspond to Fillmore et al.'s (1988) substantive idioms (see Croft, 2001: 15). Moreover, constructions are often specified lexically to an extent. For example, even a highly idiomatic expression such as *kick the bucket* has a minimum degree of schematicity in that the argument slot for the subject is left open. I will call constructions as these partially lexically filled. In this paper, I argue that the different behaviour of (partially) lexically filled instances of a construction and its lexically unspecific instances in bilingual speech can be explained with reference to Backus's (2003) Unit Hypothesis: lexically filled instances of constructions are favoured in code-mixing because they are already available to bilingual speakers as conventionalised multi-word lexical units.

The behaviour of multi-word strings in code-mixing has been explored in the type of code-mixing that Muysken (2000, 2013) calls insertion. This type of mixing involves the presence of a matrix language providing the grammatical frame for the whole sentence, and of insertions of variable sizes whose internal structure is determined by the principles of another language, often referred to as the embedded language (see Myers-Scotton, 1993, 2006 inter al.; Auer and Muhamedova, 2005). Backus (2003) argued that inserted multi word strings, or embedded language islands in Myers-Scotton's framework, are often switched together because they are lexical units with a fixed form and a

non-compositional meaning and emerge in the process of unitisation. In this paper, I argue that the same hypothesis applies to other types of code-mixing, namely to what Muysken (2000, 2013) refers to as clause-peripheral alternation.

I will adopt an operational definition of clause-peripheral alternational code-mixing, which applies to those cases where a switch at a clause boundary involves a clause expressed in language A and an extra-clausal constituent (henceforth ECC) expressed in language B. The definition of ECC that I will rely on is based on Dik's (1997) proposal, one of the first attempts to deal systematically with this type of items. Crucially, he points out that ECCs may have different degrees of complexity, ranging from single lexical items to more complex and/or more abstract patterns. The same perspective has recently been developed further in the works of Heine, Kaltenböck and associates (e.g. Kaltenböck et al., 2011; Kaltenböck et al., 2016), who demonstrate that at the extra-clausal level, which they call thetical grammar, the same functional values can be equally expressed by fixed formulaic expressions, constructions and even instantaneously produced elements. Therefore, it comes as no surprise that clause-peripheral alternation may involve elements with different degrees of syntactic complexity and lexical specificity. So far, research on ECCs in code-mixing has been based on their pragmatic functions (hence their definition as "utterance modifiers", Matras, 1998), or on their social-indexical value (e.g., Poplack, 1980). Instead, I will argue in favour of an approach that takes into account both form and function of ECCs, which in this respect are not different from any other construction. I will look in particular at the frequency with which specific lexical items are used in ECC constructions. This should help to provide a unified account of different types of alternational patterns, ranging from syntactically loose elements such as discourse markers or particles to more complex cases such as the left dislocation in (1), where the left-detached constituent, *les étrangers* occurs in French, whereas the rest of the clause, including the co-referential resumptive pronoun *ze*, is in Dutch.

(1) French-Dutch (Treffers-Daller 1994)¹

<i>Les étrangers</i>	<i>ze</i>	<i>habben</i>	<i>geen</i>	<i>geld</i>	<i>he?</i>
the foreigners	3PL	have:PRS.3PL	no	money	DM

'Foreigners, they have no money, isn't it?'

¹ To provide a better understanding of the non-English examples, I adopted a simplified version of the Leipzig Glossing Rules (<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>). A list of the abbreviations is provided at the end of the article. The gloss was not provided in case of minimal insertions into an English clause.

In Gorla (2016) I have reported the results of a qualitative analysis of clause-peripheral code-mixing, which was conducted along the lines of Functional Grammar and Discourse Grammar. I argued that the same regularities that are observed in morphosyntactically simple constituents also hold in elements of greater complexity. This result is partly expected from a functional-pragmatic perspective because the presence of extra-clausal pragmatic functions (Dik, 1997) in ECCs urges the speakers to deal with both complex and simple items in the same way. Similarly, the *principle of pragmatic detachability* proposed by Matras (1998, 2009) states that those items that are more “detachable” from the propositional content of an utterance, regardless of their complexity, will constitute an ideal *locus* for code-mixing. However, this is not where the story ends. I suggest that even if relying on the pragmatic properties of ECCs enables us to investigate alternational code-mixing *regardless* of syntactic complexity, we still need to provide a cognitively adequate explanation of how complex items behave in bilingual speech. Pragmatic motivations do not explain why some morphosyntactically complex items behave like simpler ones, and some others do not. Therefore, the research question underlying this paper is the following: given the general tendency of clause-peripheral elements to be expressed in a different language than the language of the rest of the utterance, what are the cognitive factors that facilitate the switching of syntactically complex items of this type?

The rest of this article is structured in the following way. In Section 2, I briefly describe the most relevant aspects of bilingual speech in Gibraltar, in order to provide a general framework for the interpretation of the more specific issues related to the constructions under examination. In Section 3, I discuss the theoretical assumptions that underlie this study; it is argued that, the different status of lexically filled and lexically open instantiations of the two constructions might be key to explaining the former type's higher usage frequency in code-mixing. Sections 4 and 5 describe the conducted case studies. In Section 6, a final discussion summarises the results of this work.

2 Bilingual Speech in Gibraltar

Gibraltar is a small peninsula in the region of Cadiz, Southern Andalusia, which has been part of the British Overseas Territories since 1713, when the presence of British military forces dating back to a few years earlier was ratified by the Treaty of Utrecht. Since then, Gibraltar has been, more or less directly, under British rule. Along with other social and political changes, English was introduced as the only official language of Gibraltar. However, the population

was extremely mixed in the 18th century, with a Ligurian immigrant community and Sephardic Jews constituting the majority (see the census data discussed in Levey, 2008b: 20). Although the distinctions between the various ethnic groups got blurred over the centuries, before the second half of the 20th century only a small part of the population had English as its first language. Instead, as Kramer (1986) argued, Andalusian Spanish was maintained as the major language of communication, mostly due to the intense commercial relationships with the neighbouring region of Spain, the presence of Spanish workers in Gibraltar, and the frequent mixed marriages.

Under these conditions, the linguistic repertoire of the past can be described as a case of extended diglossia (Fishman, 1967), with English being almost restricted to official and written uses and Spanish being used in all other contexts. This changed in the second half of the 20th century and culminated in the closing of the border with Spain from 1969 to 1985. Under these circumstances, which had dramatic effects on the local population, a language shift towards English has been becoming more and more apparent, so that the linguistic repertoire at present can be best described as a case of diaglossia in Auer's (2005a) terms: English has developed a local colloquial variety (Kellerman, 1996, 2001; Levey 2008a, 2015) and is crucially expanding its domains, including informal conversation, whereas local Spanish is more and more confined to the family and the spoken domain.

With English and Spanish being equally present in conversation, code-switching is an expected outcome of language contact. Its conversational aspects have been studied by Moyer (1998), whereas Weston (2013) focused on differences in code-switching patterns between the generations. In Goria (2016, 2018) I carried out corpus-based analyses of bilingual speech in Gibraltar, focussing on the emergence of community-specific regular patterns. The results most relevant to the present study can be summarised as follows. The shape of code-mixing is ultimately determined by the relationship between English and Spanish in the community's repertoire and reflects the ongoing process of language shift from Spanish to English. We thus observe great differences between speakers of different ages: older speakers typically use discourse-related or participant-related code switching in the sense of Auer (1999) and code-mixing within the clause typically involves the insertion of English elements (mostly nouns) in a Spanish matrix structure. Furthermore, several English content words have entered as established loanwords in the local variety of Spanish and are listed in amateur dictionaries such as Vallejo (2001). Conversely, younger generations, and primarily speakers who were

younger than 20 years old at the time of the interviews, show a greater preference for English as the base language of the interaction, and code-switching serving conversational functions is less attested. Code-mixing of various types occurs, instead, with great frequency in all the recordings, with a general preference for alternational patterns occurring at the periphery of the clause. Furthermore, bilingual patterns involving extra-clausal constituents (ECCs) are unidirectional; consider the following examples:

- (2) *mira* i grew up in an environment where we had three television stations
 ‘look, I grew up in an environment where we had three television stations’
- (3) so you’ll need to converse in spanish. *y claro ya*. that becomes a part of you and your language your nationality
 ‘so you’ll need to converse in Spanish. *And of course*, that becomes a part of you, and your language, your nationality’
- (4) that was created *porque* the people of la_línea used to come to work in gibraltar
 ‘that was created *because* the people of La Línea used to come to work to Gibraltar’

In examples (2)-(4) there is alternation involving a Spanish peripheral element (conjunctions, subordinators, discourse markers, and so on) and an English clause, whereas opposite patterns involving an English ECC followed by a Spanish clause are restricted to very few individual cases. The emergence of this arbitrary and community-specific regularity was interpreted as a case of fusion in the sense of Auer (1999, 2014), that is, a situation where switching between the two languages loses its original pragmatic value and becomes only a residual behaviour, limited to a number of fixed patterns, where the speakers’ choice of language is heavily constrained. Single instances of (pragmatically motivated) code-switching become automatised routines and thus lose their conversational function, since switching is unavoidable at certain points. At the same time, these emerging fused lects (Auer, 1999) generally acquire social-indexical values as related to particular social groups and globally opposed to monolingual practices (see e.g., O’Shannessy, 2011 and to

appear, for an extensive study of fused lects as emergent in the speech of children and adolescents).

One of the aims of the present study is to carry out a finer-grained analysis of the bilingual patterns mentioned above, which systematically involve a Spanish left- or right-peripheral ECC and an English clause. Previous research has focussed on displaying and highlighting similarities between different types of peripheral elements, showing that:

- (i) regardless of their formal and functional differences, ECCs are easily switched even by speakers who do not code-switch a lot (Goria, 2015a: 138–153);
- (ii) all ECCs are subject to the same constraints, since switching seems to occur only in one direction.

Given these preliminary findings, the questions I raise in the present study are (a) what are the differences in terms of frequency between simple and complex ECCs, and, more specifically, (b) in which ways syntactic complexity can favour or disfavour the emergence of bilingual ECC-clause patterns.

The reported analyses draw on my unpublished Gibraltar corpus (Goria, 2015b; for details, see Goria, 2018). The data were collected in 2013 during field-work, yielding a total of over 20 hours of audio-taped semi-structured interviews involving speakers from three age-groups, namely “young” speakers, below 30 years old, “middle-age” speakers, between 30 and 60 years old, and “old” speakers, over 60 years old. While there are differences in code-mixing patterns across these three classes (see Kellerman, 2001; Levey, 2008b; Weston, 2013), clause-peripheral alternation involving ECCs is a ubiquitous phenomenon throughout the whole corpus and represents indeed its most salient feature. This makes it thus suitable for investigating the questions that were previously outlined in this Section.

3 Syntactic Complexity in Bilingual Speech

3.1 *The Unit Hypothesis and Alternational Code Mixing*

According to the view adopted by usage-based linguistics (Croft, 2001; Bybee, 2006, 2010, 2013), constructions represent complex form-meaning associations such that their meanings cannot be predicted by the semantics of the single elements involved. A distinction is made between lexically open (or schematic) constructions, which can be filled with any contextually appropriate lexical material (for example the English causative construction), and lexically filled (or substantive) constructions, in which all the elements forming the construction are fixed (for example: *to cut a long story short*). However,

“schematic [constructions]² vary considerably in their schematicity” (Croft and Cruse, 2004: 248), and in fact most constructions are partly schematic and partly fixed, in that they involve both lexically filled and lexically open slots. For example, the *way*-construction analysed in Goldberg (1995), as in *Frank kicked his way out of the bar*, has fixed slots represented by the lexical element *way*, the preceding possessive and the following prepositional phrase, but also open slots corresponding to the subject, the predicate and so on.

Research within the usage-based paradigm (e.g., Bybee, 2010) has provided a theory of how constructions, as elements of variable syntactic complexity, are dealt with in human cognition. Through repetition in use, linguistic representations of lexemes occurring in a given construction develop stable sequential relations, so that the whole string becomes stored in memory as a chunk. This holds true both for fully specific combinations, such as *break a leg!*, as well as for those that have a lexically unspecified slot, such as *to lend X a hand*. Chunking is thus the process that continuously shapes the lexicon/grammar, allowing for multi-word elements of any complexity to be autonomously stored. This has important consequences on how constructions with different degrees of specificity are treated by bilingual speakers. Research in this direction led to the formulation of the Unit Hypothesis (Backus, 2003), which states that “if two or more EL [embedded language] morphemes co-occur in CS [code-switching], and they form a conventional combination in the EL, then it would be too coincidental if the speaker had produced them as two or more independent switches” (ibid.: 91). In other words, multi-word strings that constitute units in monolingual speech, and which have thus lexical status regardless of their internal structure, are expected to maintain this status also in bilingual speech: this would make multi-word strings available for transfer from one language into the other under the same conditions as those that hold for simple lexical items. Now, the Unit Hypothesis has proven useful to describe different types of complex switches, involving elements such as nominal compounds (Backus, 2003), plural nouns (Backus, 2003; Hakimov, 2016a), prepositional phrases (Hakimov, 2016b). The aforementioned studies focus on cases of intra-clausal insertions *sensu* Muysken (2000). In this paper, instead, I show that the notion of unit, being applicable to speech production in general, is also relevant in the analysis of alternational code-mixing. Specifically, I will analyse a case of clause-peripheral alternation, which typically involves the juxtaposition of two structures: a clause and an extra-clausal constituent in the sense of Dik (1997). The extra-clausal constituent may be realised by elements

2 I have replaced the term *idiom*, used in the original quote, with *construction*, to avoid terminological redundancy.

of variable complexity. The most typical case is monomorphemic elements such as discourse markers (see Heine, 2013, for an overview), but complex elements are not infrequent and include fixed multi-word expressions (such as the *second level discourse markers* analysed by Siepmann, 2005) as well as fairly schematic patterns, such as the pseudo-cleft and the left dislocation constructions, in which the slots are distinguished by a high degree of lexical variability.

This paper thus seeks to demonstrate the validity of the Unit Hypothesis beyond the now well-known cases of insertional code-mixing, by showing that the same hypothesis can actually explain the patterns of alternational mixing involving pseudo-clefts and left dislocations in my corpus. This assumption will be discussed in detail in Section 3.2.

3.2 *Syntactic Complexity in Extra-Clausal Constituents*

Previous approaches to extra-clausal constituents have identified the main rationale for their use in the pragmatic functions expressed by this class. However, if one only relies on a pragmatic function-based account of ECCs, there is a number of questions that cannot be fully answered. For one, Dik's (1997) account of ECCs downplays the importance of syntactic complexity, since he treats both simple and complex items on an equal basis if they have the same function. More recently, researchers working in the framework of Discourse Grammar such as Heine, Kaltenböck and associates have proposed a more accurate formalisation of the different types of phenomena occurring at the periphery of the clause, addressing syntactic complexity in a more specific way. In Kaltenböck et al. (2011), they argue for a major distinction between two domains of grammar, namely Sentence Grammar and Thetical Grammar. The latter is involved in the production of discourse-relevant entities, typically occurring at the periphery of the clause, which they call *theticals*. According to these authors, this sector of grammar is distinguished by three levels of complexity. The first is the level of *formulaic theticals*, which are generally monomorphemic unanalysable units. This level is illustrated by the use of the discourse marker *for example* in (5). The second is the level of *constructional theticals*, which are defined as "recurrent patterns or constructions of theticals, being compositional but having some schematic³ structure and function" (Kaltenböck et al., 2011: 875). These elements, such as the expression *as for X* in (6), correspond to constructions in usage-based linguistics. The third is the

3 Note that *schematic* is not used here in exactly the same way as in Construction Grammar works, and it points to the presence of morphosyntactic specifications on the structure of theticals.

level of *spontaneous theticals*, such as the parenthetical clause *I've forgotten what she was called* in example (7). These items are fully compositional and built with the principles of sentence grammar; they represent locally relevant interactional strategies and are used only occasionally.

- (5) (Huddleston and Pullum, 2002: 1355, quoted in Kaltenböck et al., 2011)
The therapist's level tone is bland and neutral – he has, for example, avoided stressing 'you'.
- (6) (Kaltenböck et al., 2011: 856)
(As for) Peter, he no longer is my friend.
- (7) (ICE-GB corpus: s1a-044-349; quoted in Kaltenböck et al., 2016b: 8)
Yes there was one woman particularly I've forgotten what she was called in this group

In Gorla (2016) I have argued for the application of this framework to bilingual data, showing that it leads to interesting generalisations about alternational code-mixing. One of them is that differently complex but functionally equivalent peripheral items behave in the same way in bilingual speech. Switches of varying sizes can thus be considered instances of the same alternation pattern, while in the code-switching literature they have sometimes been labelled differently, for instance as cases of extrasentential and intersentential code-switching. However, this approach is not devoid of shortcomings. For instance, it disregards the distributional properties of the various types of theticals in bilingual speech; therefore, relying on this analysis alone, it is impossible to predict the probability with which theticals of a specific level will occur in bilingual speech, nor to account for differences in the behaviour of elements belonging to the same class. If we focus on constructions, a finer grained analysis is needed in order to explain why there are instances of a construction that are more frequently switched than others and how this difference may be related to the organisation of constructions in the bilingual's lexicon/grammar.

The main claim of this paper is that lexically specific instances of a given construction are likely to be stored as lexical chunks and are thus more accessible in speech production. For this reason, they represent better candidates for code-mixing than instances of freely combined lexical items exploiting a

schematic template. A good example of this distinction is pseudo-cleft constructions (their behaviour in Spanish-English bilingual speech is analysed in Section 5). Pseudo-clefts constitute instantiations of the schematic pattern represented in (8), which has open slots that can be filled by any appropriate element, with very few constraints, as is illustrated in (9) (see also Collins, 1991).

(8) *wh- X is/was Y*

- (9) a. what I lost was my wallet
 b. what Sean wanted was a new life
 c. what the burglars took away was an old gift from my brother

However, as pointed out by Günthner and Hopper, (2010), lexically-specific instantiation of this general pattern have frozen into nearly-fixed lexical chunks such as the ones given in examples (10)-(12).

(10) (adapted from Günthner and Hopper, 2010: 12)

and so **what happens is**, your mic runs into it, and then it plugs into your receiver

(11) (Goria 2015b)

it's really central government really that (.) that is obviously all the time trying to make our life difficult really but (.) **the thing is that** we have to put up with it

(12) (2007 COCA: MAG quoted by Brinton, 2014)

what I'm saying is, don't get married in Vegas. Take your time.

According to Backus' (2003) Unit Hypothesis, it is possible to assume that lexically specific instances of this construction, like the ones given in (10)-(12), are stored as holistic units in the mental lexicon and are thus highly accessible during speech production. Thus, due to their unit status, such lexically specific instances are expected to occur with greater frequency in bilingual speech. This prediction is tested in Sections 4 and 5 by analysing the frequency distributions in bilingual speech of two types of ECCs involving schematic constructions: left dislocations and pseudo-cleft sentences.

4 Bilingual Left Dislocations

Bilingual left dislocations (LDs henceforth) are an interesting, and yet neglected, case of clause-peripheral alternation involving a schematic construction. LDs are constructions in that they represent complex syntactic structures expressing pragmatic and information-structure related functions in a non-compositional fashion. Moreover, LDs are interesting for the analyst because they allow two levels of generalisation corresponding to two different subtypes of the construction. The first one corresponds to what will be referred to as nominal LD and has a fully schematic structure; the second one will be referred to as pronominal LD and is partially lexically specific.⁴ In this Section I show that bilingual left dislocations occur more frequently as instances of code-mixing when the expression has a higher degree of specificity and corresponds to a lexical chunk, and less frequently when the construction is lexically open, and thus has a greater degree of schematicity. In fact, most bilingual LDs involve only a few highly frequent lexical items.

4.1 *Formal and Functional Characteristics of Left Dislocation*

The relevant constructions were identified in the Gibraltar corpus on the basis of both formal and functional features of LDs. Concerning formal parameters, Lambrecht (2001a) indicates four major features that cross-linguistically characterise LD. They are listed in (13) below:

(13) Defining features of the dislocation construction

- (i) Extra-clausal position of a constituent
- (ii) Possible alternative intra-clausal position
- (iii) Pronominal co-indexation
- (iv) Presence of a special prosodic contour

According to this definition, LD involves thus the presence of a constituent, typically an NP, occupying a clause-peripheral slot, and coindexed by an anaphoric element inside the clause. In (14), thus, *al partido carlista* ‘the Carlist party’ is a dislocated NP that occurs outside the clause, and it is anaphorically coindexed by the weak pronoun *lo* inside the clause.

4 An objection could be raised that pronominal LDs, being closed-class elements, cannot count as “lexically specific”. However, pronominal LDs only involve a few pronominal forms and, more importantly, they are associated with an overlay of pragmatic meaning; hence, greater specificity in form is mirrored here in a difference in meaning.

(14) (Rivero, 1980: 363)

Al partido carlista dicen que no lo
 DOM party carlist say:PRS.3PL COMP NEG 3SG.M.OB

legalizaron para las elecciones
 legalise:PST.3PL for the elections

'the Carlist party, they say they did not legalise it for the elections'

As for functional properties, LD has both informative and pragmatic-interactional values. Informatively, the construction marks the dislocated element as a Topic in the sense of Lambrecht (1994) and Dik (1997) and is often used to retrieve elements that are either textually or contextually accessible, but not topical at a certain point in the discourse (Hidalgo, 2002). Furthermore, Duranti and Ochs (1979), and more recently Geluykens (1992), Hidalgo (2002) and Hidalgo Downing and Hidalgo Downing (2007), have argued that, particularly in Romance languages, LDs also have more global interactional properties, both interpersonal and metatextual. These constructions allow the negotiating of “difficult” and potentially face-threatening operations such as keeping the floor, shifting the topic, and introducing the speaker’s perspective.

It is crucial for the present analysis to observe that, even if informative and interactional functions cannot be completely disentangled, differences in function tend to be reflected in differences in form. Therefore, as outlined above, I distinguish here between nominal and pronominal LDs (see also Goria, 2016). The “nominal LD” is characterised by clause-oriented informative functions, namely that of marking the sentence Topic, either introducing a new one or retrieving a previously activated referent. The dislocated constituent in this variant of the LD construction can be any NP; an example is given in (15).

(15) nominal left dislocation (Hidalgo Downing and Hidalgo Downing, 2007: 329)

oye y la revista esta ¿cada cuánto la sacáis?
 DM and the journal this how_often 3sg.F.OBJ publish:PRS.2PL
'Listen and this journal, how often do you publish it?'

The “pronominal LD” often expresses discourse-oriented interactive functions concerning the management of sentence topics in a broader discourse

setting and the avoidance of potentially face-threatening acts (see Duranti and Ochs, 1979; Downing, 1997; Hidalgo Downing and Hidalgo Downing, 2007). In Spanish, pronominal LD involves only a restricted set of elements, namely personal (16a) and demonstrative (16b) pronouns.

(16) pronominal left dislocation (Hidalgo Downing and Hidalgo Downing, 2007: 342)

a. *yo mucha gente me ha comentado*
 1SG.SBJ many people 1SG.OBJ have.AUX:3SG tell
'As for me, many people told me'

b. *eso yo creo que deberías*
 this:M.SG 1SG believe:PRS.1SG COMP must: COND.2SG
hacer = lo
 do:INF = 3SG
'This, I think you should do it'

According to Hidalgo Downing and Hidalgo Downing (2007), pronominal LD is particularly frequent in Spanish: they observe that of all instances of dislocation in the *Corpus oral de referencia de la lengua española contemporánea* (Marín 1992), 40% is represented by pronominal LDs. Moreover, both Downing (1997) and Hidalgo Downing and Hidalgo Downing (2007) indicate that among all possible pronouns, the first-person singular pronoun *yo* and the demonstratives *eso* and *esto* are the most frequent forms. In an exemplar model, these would therefore correspond to a prefab in Bybee's terms (see Bybee, 2010: 81 for a similar case). In contrast, for nominal LDs it is not possible to identify any prefabs, though the possibility cannot be ruled out that there are such frequent forms. This discrepancy between the two variants of the construction is of course biased by the different sizes of the class of nouns and pronouns, with possible effects also on their mental representations; nevertheless, this distribution suggests overall greater specificity of pronominal LD and greater schematicity of nominal LD.

Another issue that must be dealt with concerns the comparability of the LD construction across languages, and more specifically in the two contact languages considered here, English and Spanish. Differences between Romance languages and English were identified within the generative framework by Cinque (1990), Rizzi (1997) and also in functionalism by Hidalgo (2000). Authors from different theoretical background seem to agree that, while LDs share a number of core features across languages (also consider Lambrecht's,

2001a, typologically oriented definition), their functional values tend to be language-specific. Thus, in English, pronominal LD has in fact a contrastive value (Prince, 1981, 1984) that is not attested in Romance languages, as in example (17).

- (17) Carl Barks, Christmas on Bear Mountain. "Four colour comics". Dec. 1947
That silly season when everybody loves everybody else. Me – I'm different!
Everybody hates me and I hate everybody

To put it differently, LD has the same informative function in both English and Spanish, namely that of signalling a sentence Topic; the two languages differ, though, in the pragmatic uses of the construction. With respect to this issue, the analysis presented here is form-based in that the instances of the LD construction in the English-Spanish data have been considered in relation to Lambrecht's (2001a) interlinguistic definition and without consideration of their pragmatic function.

To sum up, two variants of the LD construction have been identified: nominal and pronominal LD. From a constructionist perspective, a nominal LD represents a fully schematic instance of the construction in that the extraclausal slot is lexically open and admits any informatively adequate NP. The pronominal LD on the contrary is more lexically specific in that it allows only a restricted set of items, namely personal and demonstrative pronouns, and has a few highly frequent instantiations. Moreover, the two different forms are also related to partially different functions: while nominal LD expresses a clause-oriented informative function, namely that of marking the Topic, pronominal LD has developed additional pragmatic and discursive values. For what concerns the behaviour of this construction in bilingual speech, I will show that lexically specific instances of the LD construction, i.e. lexicalised chunks, are more frequently switched from one language into the other.

4.2 *Data Analysis and Results*

The analysis of LDs in the corpus was preceded by manual annotation of the relevant constructions with the software ELAN⁵ (Sloetjes and Wittenburg, 2008). Based on the theoretical assumptions outlined above, two types of LD were identified: pronominal LDs, involving personal pronouns, demonstratives

5 ELAN is a software provided by Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands (<https://tla.mpi.nl/tools/tla-tools/elan/>).

and possessives, and nominal LDs, involving any nominal constituent. With regard to the language of the analysed construction and the adjoined clause, four possible patterns were considered, with a separate language value for the dislocated constituent and the following clause. The bilingual speaker may produce both the dislocated constituent and the following clause in the same language, English or Spanish, and in this case the LD is labelled as monolingual, or she may switch the language at the juncture between the dislocated constituent and the clause. The monolingual patterns were labelled “e-e”, or “s-s”; the bilingual patterns were labelled “s-e” when a Spanish dislocand precedes an English clause and “e-s” when an English dislocand precedes a Spanish clause. Therefore, I distinguished between eight possible combinations of the dislocation type, the language of the dislocand and the language of the clause. The patterns are exemplified below:

(18) English monolingual (nominal) LD

you go there even if we're british [the cultural shock]_i you have [it]_i

(19) Spanish monolingual (nominal) LD

[y la pequeña]_i que va vení dentro una semana
and the youngest REL go:3SG come:INF within a week

[ella]_i tiene tres hijos
3SG.SBJ have:3SG three sons

'And the youngest one, who's going to come within a week, she has three sons'

(20) “s-e” bilingual nominal LD

sin embargo ahora [mucho' niño' jovene']_i
however now many children young

[they]_i 're writing with Americanisms *tambien*

'No doubt now many young children are writing also with americanisms'

(21) “s-e” bilingual pronominal LD

you 'd be really really struck *po'que* [yo]_i [i]_i 'm very innocent with those things

'You'd be really really struck, because me, I'm very innocent with those things'

- (22) “e-s” bilingual nominal LD
and [the next child he caught]_i
[eso]_i era él que se quedaba
this.M be:PST.3SG 3SG:SBJ REL 3SG.REFL stop:PST.3SG
‘And the next child he caught, *that was the one who would stop*’
- (23) “e-s” bilingual pronominal LD
cosa’ como son [us]_i economically
things as be:3PL.PRES

tamo demasiado bien
be:PRS.3PL too good
‘As things stand us economically we are way too good’

The total number of LDs found in the corpus is 81. The analysis of the distribution of the different constructions, which is given in Table 1, yielded the following results.

As for monolingual utterances, Spanish LDs appear to be more frequent than English ones, confirming the view expressed in Hidalgo (2000) about a broader diffusion of the construction in Spanish. Also, English shows a strong preference for nominal LDs (14 out of 16 tokens), whereas Spanish clauses more frequently contain pronominal LDs (23 out of 28 tokens). This too is partly expected in line with the literature discussed in Section 4.1, according to which Spanish pronominal LD was found, in some cases, to be predominant in number over nominal LD and associated with several interactive functions, whereas English LD is much more marginal. As for bilingual LDs, the “s-e” type, occurring 33 times, outnumbers the “e-s” type. This tendency is also attested in all cases of clause-peripheral alternation in the Gibraltar corpus: with very few exceptions, bilingual patterns involve a Spanish extra-clausal constituent and an English clause. In Goría (2016, 2018) I have argued that the same

TABLE 1 Distribution of nominal and pronominal left dislocations.

	e-e	s-s	e-s	s-e	Total
Nominal LD	14	5	2	14	35
Pronominal LD	2	23	2	19	46
Total	16	28	4	33	81

unidirectionality applies also to discourse markers, conjunctions and subordinators.⁶ We also observe that pronominal LDs with a Spanish dislocand are more frequent than nominal LDs both in monolingual and bilingual sentences.

The picture drawn so far suggests a slight preference for pronominal LDs when the dislocand is Spanish, and the absence of a similar pattern with English dislocands. This tendency can be better understood if one considers the single lexical items involved in LD, as represented in Table 2.

The table lists the different instantiations of the first slot in the construction and their frequencies in monolingual and in bilingual combinations. The non-recurring lexical items were coalesced into two categories “other Spanish” and “other English”. It can be observed that English elements are underrepresented in LD constructions, with only the semantically general noun *people* occurring more than once. In contrast, the Spanish types that recur in LDs, according to Table 2, belong to a few well identifiable classes: personal pronouns (*yo*, *ello*, *nosotro*), demonstratives (*eso*), semantically general nouns (*gente*, *personas* ‘people’) and proper nouns. It is also important to note that all these elements represent prototypical sentence Topics. Furthermore, the two most frequent lexical types that enter the construction are the Spanish pronouns *yo* and *eso*, which together represent 42% of all LDs and 46% of all bilingual LDs. This distribution seems thus to mirror the distribution found in monolingual Spanish: Hidalgo Downing and Hidalgo Downing (2007: 333) observe that: “[i]n Marcos Marín’s *Corpus oral de referencia de español contemporáneo* (1992), pronouns amount to the 40% of detachments. Of these pronominal detachments, 68% corresponds to personal pronouns, whereas the remaining 31.8% is composed of demonstrative pronouns”.

This parallelism between monolingual and bilingual data supports the hypothesis that lexical specificity of particular instances of the LD construction is one of the major factors influencing the production of bilingual LDs. The items that occur more frequently in these constructions in bilingual speech, namely the first-person singular pronoun *yo* and the proximal deictic *eso* are highly frequent items also in monolingual speech.

To conclude, the analysis of bilingual left dislocation in my corpus supports the view that code-mixing, and in particular the emergence and distribution of patterns such as “*yo* + English clause” or “*eso* + English clause” are heavily affected by the frequency with which lexical items appear in the construction. Given a schematic pattern such as the LD construction, those subtypes of the

6 Unidirectionality in code-mixing patterns is also a diagnostic feature of incipient fusion in the sense of Auer (1999, 2014), in that it shows that code-mixing is restricted to a number of fixed patterns.

TABLE 2 Monolingual and bilingual LDs involving single lexical items occurring more than once in the corpus.

	monolingual	bilingual	total
<i>yo</i>	6	12	18
<i>eso</i>	11	5	16
<i>people</i>	4	0	4
<i>ello'</i>	2	1	3
<i>nosotros</i>	2	1	3
proper nouns	3	0	3
<i>la gente</i>	0	2	2
<i>la' persona'</i>	0	2	2
other spa	7	10	17
other eng	9	4	13
total	44	37	81

construction that are attested in monolingual Spanish as lexicalised routines are also the ones more frequently used in code-mixing. An assumption behind this effect is that lexicalised instantiations are more accessible to the speakers than purely schematic patterns. To conclude, a constructionist approach to bilingual LDs is complementary to the observations made on the same topic (among others, Goría, 2016). Besides being more frequent in monolingual speech and more lexically specific, the subtypes of LD that are switched more frequently are also the ones that typically express interpersonal and discourse-oriented functions, and that are thus more pragmatically detachable (Matras, 1998, 2009) from the propositional content of the utterance.

5 Bilingual Pseudo-Clefts

Pseudo-cleft sentences are another example of a partially schematic construction that may have lexically filled and lexically open slots. These constructions are used to express pragmatic and informative values of an utterance in a non-compositional way, i.e. through the realisation of a simple propositional content in a biclausal construction involving a relative clause and a copular clause. As occurs with other elements with similar functions, pseudo-clefts in my corpus often appear in bilingual utterances, with one element of the construction being realised in a different language than the other, which results in language alternation. In this section I analyse this type of code-mixing in the

light of the Unit Hypothesis, and I argue that clause-peripheral alternation is more frequent with those instances of pseudo-clefts that have a higher degree of lexical specificity, and which therefore constitute lexical chunks, whereas more schematic instantiations of the construction tend to be monolingual.

5.1 *Form and Function of Pseudo-Clefts*

Pseudo-clefts, as part of a broader family of cleft constructions (e.g., Collins, 1991; Lambrecht, 2001b; Traugott, 2008; Traugott and Trousdale, 2013; Patten, 2012), are characterised by a number of formal and functional properties. Among them are some that pseudo-clefts share with related constructions, while others are unique to this particular type. Here, I will concentrate on those aspects of pseudo-clefts that are relevant for their identification in the bilingual corpus.

Pseudo-clefts are biclausal constructions that consist of a headless relative clause, often introduced by a *wh*- element, and a subsequent copular clause. The complement of the copula is coreferent with the relative clause and often expresses the pragmatic function of Focus, whereas the relative clause contains presupposed content, for example:

- (24) a. what we are talking about is your future
b. we are talking about your future

In (24), the two sentences have the same propositional content, but different informative values. The pseudo-cleft in (24a) has the function of marking the constituent *your future*, which occurs as the complement of the copula, as the Focus of the sentence. At the same time it signals that the content of the relative clause is presupposed or retrievable from the context (see Lambrecht, 1994). From this definition, pseudo-clefts are thus schematic form-function pairings involving two fixed structural positions, one for the subordinator introducing the relative clause and one for the copula, and two open slots, one for the relative clause and one for the argument of the copula, as represented below:

- (25) *wh*- X *be* Y

The element indicated as *wh*- in (25) corresponds to a relative-like subordinator. As recently argued by Patten (2012), this slot has three main realisations, corresponding to three subtypes of the pseudo-cleft construction, which are referred to as *WH*-clefts, *TH*-clefts and *ALL*-clefts, as in (26).

- (26) (Patten, 2012: 62)
- a. What I like best is grape soda [WH-cleft]
 - b. The thing that I like best is grape soda [TH-cleft]
 - c. All he drinks is grape soda [ALL-cleft]

However, since ALL-clefts have to be regarded as a more specific and in a way a less prototypical instance of the construction (see Traugott, 2008), my analysis will address only the first two types. The pseudo-cleft constructions in English and Spanish exhibit obvious similarities, both in form and in function; the structure is thus highly comparable between the two languages in contact. As for *wh*-clefts, the most significant difference between English and Spanish is the form of the *wh*- element. In English, a personal pronoun and a relative marker are fused into a single portmanteau form, which is commonly referred to as fused relative (Agar Marco, 2014). In Spanish, like in most Romance languages, the *wh*-element is expressed analytically by a personal pronoun followed by a relative pronoun, with the neuter form *lo que* being the default case, for instance:

- (27) (Fernández Soriano, 2009: 88; my glosses)
<i>lo</i>	<i>que</i>	<i>quiero</i>	<i>es</i>	<i>un</i>	<i>coche</i>
3SG	REL	want:PRS.1SG	be:PRS.3SG	a	car
- ‘What I want is a car’*

Th-clefts have the same structure in English and Spanish. According to Collins (1991), the *wh*- element is formed by a semantically general noun such as *thing*, *time*, *place*, *one*, *way* and so on, followed by a relative clause. Examples (26b) and (28) show that TH-clefts are structured similarly in English and Spanish.

- (28) (Agar Marco, 2014: 180)
<i>la</i>	<i>persona</i>	<i>a la que</i>	<i>más</i>	<i>echa de menos</i>
the	person	to_which	most	miss:3SG.PRES
<i>es</i>	<i>a</i>	<i>su</i>	<i>madre</i>	
be:3SG.PRES	to	POSS.3SG	mother	
- ‘The person that misses him most is his mother’*

Both in English and Spanish, the second argument of the copular clause indicated as Y in (25), can be a nominal element, as in (26a-c), as well as a clausal one, as in (29) below.

- (29) (Goria, 2015b)
all the companies are here. and what happens is that they have (.) a
lot of (.) workers that do a lot of work for customer services o:r (.) or
IT work.

It must be noted that pseudo-clefts that introduce a nominal constituent typically express a clause-internal function of focus-marking, while the ones introducing a clausal complement do not have this function. As Hopper and Thompson (2008) and Traugott and Trousdale (2013) have argued, this variant of the construction expresses a discourse-oriented cataphoric function, in that it is used to project (Auer, 2005b) an upcoming unit of talk. This variant of the construction has been also identified for Spanish pseudo-clefts, for example by Curnow and Travis (2004) and Fernández Soriano (2009). Interestingly, studies of pseudo-clefts in both languages (see Fernández Soriano, 2009; Reig Alamillo, 2011; Pannunzi, 2016, for Spanish, and Hopper and Thompson, 2008, for English) point out that constructions expressing a discourse-oriented function are strongly associated to a few highly frequent lexical types, such as *what happens is* in English and *lo que pasa es (que)* in Spanish. Therefore, a difference in function is reflected by greater specification of one of the constructional slots, namely, the first part of the construction is often represented by a lexicalised chunk that constitutes a unit in Backus' (2003) sense.

Moreover, an even greater level of fixedness is found in connectives such as *es que* in Spanish and *(the) thing is* in English, as in examples below:

- (30) (Goria, 2015b)
it's really central government really that that is obviously all the time
trying to make our life difficult really but *the thing is that* we have to
put up with it.

- (31) (Goria, 2015b)
that's what i wanted to say
<i>é</i>	<i>que</i>	<i>no</i>	<i>me</i>	<i>acordaba</i>
be:3SG.PRES	that	NEG	REFL.ISG	remember:PST.ISG
'that's what I wanted to say, <i>it's that I didn't remember</i> '				

As pointed out by Travis (2005), in some varieties of Spanish, these items arise as a diachronic development of pseudo-clefts and have a discourse-oriented function, in that they are used to project an upcoming new content. From a formal perspective, they represent an outcome of chunking, in that they are lexical

units that no longer admit internal lexical variation. To sum up, researchers distinguish between two variants of the pseudo-cleft construction. On the one hand, there are schematic instantiations resulting from the pairing of a headless relative clause with a copular clause, with few constraints on the elements that can enter the construction. These constructions have a discourse function of focus marking and typically take a nominal argument. On the other hand, lexically specific instances of the pseudo-cleft construction involve the presence of a few highly frequent verbs in the relative clause, meaning ‘do’, ‘say’ or ‘happen’. Such constructions typically take a clausal argument and express discourse-oriented functions, in that they serve to project an upcoming propositional content. The same holds for the two fixed expressions *es que* and *the thing is*. Based on these observations, I hypothesise that, in a pseudo-cleft construction, alternational code-mixing between the relative clause and the copular clause is more likely to occur if the first part of the construction is a lexical chunk corresponding to a highly frequent verb. Empirical evidence in favour of this hypothesis is provided in Section 5.2.

5.2 *Data Analysis and Results*

Relying on the description of the pseudo-cleft constructions given in Section 5.1, I have conducted an analysis of switch placement between the two parts of the construction, i.e. between the relative clause and the copular clause, as in (32).

- (32)

<i>lo</i>	<i>que</i>	<i>pasó</i>	<i>era</i>	<i>que</i>
3SG	REL	happen:PST.3SG	be:PST.3SG	COMP

they started the youth as a normal thing
'What happened was that they started the youth (club) as a normal thing'

The analysis followed the same methodology as was used for left dislocations. The instances of the pseudo-cleft construction were labelled according to the language in which each component of the construction was realised. The patterns where both the relative and the copular clause were in the same language were labelled “e-e” (English monolingual) or “s-s” (Spanish monolingual), as in (33a) and (33b). Instances where the relative clause occurs in Spanish and the copular clause in English were labelled “s-e”, as in (34a), and instances where the relative clause occurs in English and the copular clause in Spanish, as in (34b), “e-s”.

(33) Spanish monolingual pseudo-cleft

a. *tu lo que tiene' decí a la gente*
 2SG:SBJ 3SG REL have:PRS.2SG say:INF to the people

es mira e'to é lo que tu
 be:PRS.3SG DM this be:PRS.3SG it REL 2SG.SBJ
puede' hace
 can:PRS.2SG do:INF

'You, what you have to tell the people is 'look, this is what you can do''

b. English monolingual pseudo-cleft

what they sell to their nation is 'no no, gibraltar is a tax heaven'

(34) a. s-e bilingual pseudo-cleft

la gracia y la suerte que tenemo' é que el
 the flair and the luck REL have:PRS.1PL is that the
llanito is funny in itself

'And the flair and the luck that we have is that the llanito is funny in itself'

b. e-s bilingual pseudo-cleft

so basically what you do is

te queda' en Casemates and
 2SG:REFL stay:PRS.2SG in Casemates

te toma' una cerveza
 2SG:REFL take:PRS.2SG a beer

'So basically what you do is, you stay in Casemates and you have yourself a beer'

In the following step, the instances were examined as to whether they introduced a nominal or a clausal argument. The frequency distribution of each type is given in Table 3.

As can be observed, constructions taking clausal arguments are clearly predominant in the data, and bilingual constructions represent 25% of the observed instances (36 out of 143 tokens). The bilingual patterns reveal two clear tendencies. The first tendency is the general preference for "s-e" over "e-s"

TABLE 3 Distribution of pseudo-clefts taking clausal and nominal arguments. In the two cases labelled as “?”, it was not possible to label the languages involved in the pattern, because the argument of the construction was the adverb *no*, which may be considered a homophone in English and Spanish; they were therefore excluded from the analysis.

	?	s-s	e-e	e-s	s-e	total
clausal	2	51	25	3	32	113
nominal	0	26	3	0	1	30
total	2	77	28	3	33	143

(33 *vs.* 3 tokens): this pattern was found to be dominant for all types of bilingual extra-clausal constituents (see Goría 2016, 2018), but what makes pseudo-clefts particularly interesting is that they are not fully-fledged extra-clausal constituents like discourse markers or conjunctions, but constructions exhibiting a complex syntactic structure and varying degrees of schematicity. The second finding is that code-mixing almost only involves constructions taking a clausal argument (35 out of 36 tokens), which, as said, express discourse-oriented function. Thus, the analysis so far supports the functionalist claim that elements expressing pragmatic functions are more available for switching than those expressing clause-internal functions (see, for example, Matras, 1998). However, a functional account alone leaves a number of open questions. The main issue is that, relying on functional distinctions such as “clause-oriented” *vs.* “discourse-oriented”, we are forced to treat all the instances of the pseudo-cleft at the same level, regardless of potential differences in frequency. Therefore, this does not give us any insight on how the pseudo-cleft may be represented in the mental lexicon and accessed by code-switchers. On the contrary, a distribution analysis of the construction’s lexically specific instantiations enables us to uncover bilingual patterns that are potentially overlooked using a function-based approach.

Table 4 reports the frequency distribution of all the lexical types⁷ occurring more than once as the predicate of the relative clause. The data show that both monolingual and bilingual instances of the construction are strongly associated either with the connectives deriving from reduced pseudoclefts, such as *es que* (‘it is that’) and *the thing is*, or with a small set of recurrent verbs, which correspond to lexically specific instantiations of the construction. Bilingual constructions occur most frequently with the connective *es que*

7 I chose to consider lexical types instead of surface forms and examined the predicates occurring in the present tense and those in the past tense as instances of the same type.

TABLE 4 Frequency distribution of single lexical items occurring more than once in monolingual and bilingual pseudo-clefts. See table 3 for the interpretation of the “?” column.

	?	monolingual	bilingual	total
<i>es que</i>	2	23	16	41
<i>pasar</i>	0	16	8	24
<i>tener</i>	0	1	3	4
<i>thing is</i>	0	17	2	19
<i>hacer</i>	0	6	2	8
<i>do</i>	0	1	1	2
<i>decir</i>	0	6	0	6
<i>afectar</i>	0	2	0	2
<i>happen</i>	0	2	0	2
<i>querer</i>	0	2	0	2
<i>ser</i>	0	2	0	2
other	0	16	1	17
total	2	102	33	137

(16 over 33 tokens), and with the lexical type *pasar* ‘happen’ (8 over 33 tokens), all of which correspond to the lexical chunk *lo que pasa es* ‘what happens is’ This result is in line with the examined hypothesis that those instances of the construction that are more directly associated with specific lexical elements, or that have already developed into formulaic expressions in monolingual speech, are more likely to function as units in code-mixing, as opposed to their schematic counterparts that require the productive application of a lexically open template.

6 Discussion

In this paper, I proposed a novel approach to explaining regularities emerging in bilingual speech. As outlined above, extra-sentential switching has been frequently explained on a pragmatic basis, under the assumption that linguistic items expressing pragmatic and discourse-related meanings are more frequently switched from one language into the other *regardless* of their syntactic complexity and lexical specificity. I have adopted a different perspective, arguing that alternational switching of syntactically complex items is heavily constrained by the nature of the constructions’ individual

realisations, as determined by their mental representations. Specifically, those items that are available as lexical units in monolingual speech represent the best candidates for switching. By adopting as a major criterion the distinction between schematic and lexically specific instances of a construction, we do not merely observe that code-mixing involves simple and complex units to the same extent, but we can even predict which instances of a construction will be more likely to be selected in bilingual speech. In fact, even in typically schematic constructions, like dislocations and pseudo-clefts, switching was found to involve only a few lexical types, which behave as lexical units also in monolingual corpora.

The analysis of the two case studies presented in Sections 4 and 5 provides further evidence in support of the Unit Hypothesis (Backus, 2003), showing that the hypothesis applies not only to lexical insertions but also to some types of alternational code-mixing. In fact, bilingual patterns involving left dislocations and pseudo-clefts in the Gibraltar English-Spanish corpus provides positive evidence towards this view. In both cases it was demonstrated that the instances of the examined bilingual constructions vary as to the language in which their parts are realised, and the frequencies with which these parts occur in the sample. The most frequent pattern involves for both constructions the presence of a Spanish left-peripheral element, followed by an English clause, but not all their instantiations equally contribute to this tendency: this pattern is favoured when the constructions include highly frequent elements like the first singular pronoun *yo* and the deictic *eso* in left dislocations, or lexicalised chunks like *es que* 'it is that' and *lo que pasa es que* 'what happens is' in pseudo-clefts. Interestingly, these results may also be interpreted in terms of functional-pragmatic distinctions between clause-internal and clause-external functions as well. In both case studies, the elements that were recognised as lexical chunks also convey discursive and pragmatic meaning, whereas schematic instances of the constructions were more typically associated with local clause-oriented functions.

To conclude, the reported case studies suggest that lexically-filled slots of particular constructions occurring at a high rate in bilingual speech may have the status of lexical chunks in the speakers' mental lexicon/grammar. According to the usage-based approach, these items are more frequent in code-mixing than schematic instantiations of the same constructions because they develop stronger mental representations. This enables us to assume that code-mixing is cognitively easier to perform than interference, or pattern-replication in the sense of Matras and Sakel (2007). While the former involves the use of surface forms whose mental representations are readily available in the lexicon/grammar, the latter involves more complex

operations than activation and retrieval. Namely, the speaker requires to select an empty (schematic) template in language A as well as appropriate surface material from language B (or both). Further evidence in favour of this view could be obtained by adopting the same perspective in future experimental research involving similar constructions, characterised by some degree of internal complexity and by the possibility to have both lexically-filled and lexically-open instantiations.

ACC	Accusative
SG	Singular
PL	Plural
AUX	Auxiliary
COMP	Complementiser
COND	Conditional
DM	Discourse marker
DOM	Differential object marking
F	Feminine
INF	Infinite
M	Masculine
N	Neuter
NEG	Negation
NOM	Nominative
PRS	Present
PST	Past
REFL	Reflexive
REL	Relative
1	1st person
2	2nd person
3	3rd person

Acknowledgements

The research that led to this publication was funded by the Ph.D. programme in Linguistics at the University of Pavia (xxviii ciclo). I wish to express my deepest gratitude to the editors of this special issue for their attentive and insightful observations on my paper. Thanks in particular to Nikolay Hakimov for encouraging me in this enquiry and for sharing his views on code-mixing. All the errors in this paper are obviously my own.

References

- Agar Marco, Rocío. 2014. Pseudo-cleft sentences. Italian-Spanish in contrast. In Anna-Maria De Cesare (ed.) *Frequency, forms and functions of cleft constructions in Romance and Germanic*, 177–216. Berlin: De Gruyter.
- Auer, Peter. 1999. From *code-switching via language mixing to fused lects*: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3 (4): 309–332.
- Auer, Peter. 2005a. Europe's sociolinguistic unity, or: A typology of European dialect / standard constellations. In Nicole Delbecque, Johan van der Auwera, Dirk Geeraerts (eds.), *Perspectives on variation* (Trends in Linguistics; 163), 7–42. Berlin: de Gruyter.
- Auer, Peter. 2005b. Projection in interaction and projection in grammar. *Text* 25: 7–36.
- Auer, Peter. 2014. Language mixing and language fusion: When bilingual talk becomes monolingual. In Juliane Besters-Dilger, Cynthia Dermarck, Stefan Pfänder, Achim Rabus (eds.), *Congruence in contact-induced language change*, 294–336. Berlin: de Gruyter.
- Auer, Peter, Raihan Muhamedova. 2005. 'Embedded language' and 'matrix language' in insertional language mixing: some problematic cases. *Italian Journal of Linguistics / Rivista di linguistica* 17(1) (special issue "Commutazione di codice e teoria linguistica", edited by Gaetano Berruto): 35–54.
- Backus, Ad. 2003. Units in codeswitching: Evidence for multimorphemic elements in the lexicon. *Linguistics*, 41: 83–132.
- Brinton, Laurel. 2014. The recent rise of "saying" ECCs: (*I'm*) *just saying*, *All I'm saying* (*is*), and *What I'm saying* (*is*). Paper presented at the *Outside the Clause Workshop*. University of Vienna, 4–5 July 2014.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82: 711–733.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan. 2013. Usage-based theory and exemplar representation. In Thomas Hoffmann, Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 46–49. Oxford: Oxford University Press.
- Cinque, Guglielmo. 1990. *Types of A-dependencies* Cambridge (MA), MIT Press.
- Collins, Peter C. 1991. *Cleft and pseudo-cleft constructions in English*. London: Routledge.
- Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William and D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Curnow, Timothy Jowan, Catherine E. Travis. 2004. The emphatic *es* construction of Colombian Spanish. In Christo Moskovsky (ed.), *Proceedings of the 2003 Conference*

- of the Australian Linguistic Society*. Downloadable at <http://www.als.asn.au/proceedings/als2003.html>. Last access: 1st July 2016.
- Diessel, Holger. 2016. *Frequency and lexical specificity. A critical review*. In Heike Behrens and Stefan Pfänder (eds.) *Experience Counts: Frequency Effects in Language*. Linguae & Litterae, Freiburg Institute for Advanced Studies, 209–238. Berlin: de Gruyter.
- Dik, Simon. 1997. *The theory of functional grammar*. Vol 2, *Complex and derived constructions*. Dordrecht: Foris.
- Downing, Angela. 1997. Discourse and pragmatic function of the Theme constituent in spoken European Spanish. In John Connolly, Roel Vismans, Chris Butler, Richard Gatward (eds.) *Discourse and pragmatics in functional grammar*, 137–161. Berlin: de Gruyter.
- Duranti, Alessandro and Elinor Ochs. 1979. Left-dislocation in Italian conversation. In Talmy Givón (ed.) *Syntax and semantics. Discourse and syntax*, vol. 12, 377–416. New York: Academic Press.
- Fernández Soriano, Olga. 2009. On the nature of covert operations. The case of focus in Spanish pseudoclefts. In Enoch O. Aboh, Elisabeth van der Linden, Josep Quer and Petra Sleeman (eds.) *Romance languages and linguistic theory. Selected papers from 'Going Romance' Amsterdam 2007*, 87–108. Amsterdam: Benjamins.
- Fillmore, Charles, Paul Kay and Catherine O'Connor (1988). Regularity and *idiomaticity in grammatical constructions*: The case of *let alone*. *Language* 64: 501–38.
- Fishman, Joshua. 1967. Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of Social Issues* 23 (2): 29–38.
- Geluykens, Ronald. 1992. *From discourse process to grammatical construction: On left dislocation in English*. Amsterdam: Benjamins.
- Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goria, Eugenio. 2015a. *Il discorso bilingue a Gibilterra. L'emergere di schemi regolari nella commutazione di codice*. PhD Thesis, Pavia University.
- Goria, Eugenio. 2015b. *English and Spanish in Gibraltar. A corpus of bilingual speech*. Unpublished data.
- Goria, Eugenio. 2016. The role of extra-clausal constituents in bilingual speech. The emerging of regular patterns in a bilingual corpus. In Gunther Kaltenböck, Evelien Keizer, Arne Lohmann (eds.) *Outside the clause: Form and function of extra-clausal constituents*. Studies in Language Companion Series 178, 273–302. Amsterdam: Benjamins.
- Goria, Eugenio. 2018. *Inglese e spagnolo a Gibilterra. Le dinamiche del discorso bilingue*. Cesena: Caissa.
- Günthner, Susanne and Paul Hopper. 2010. Zeitlichkeit & sprachliche Strukturen: Pseudoclefts im Englischen und Deutschen. *Gesprächsforschung. Online Zeitschrift*

- zur verbale Interaktion: 1–28*. Downloadable at <http://www.gespraechsforschung-online.de/heft2010/heft2010.html>. Last access: 1st July 2016.
- Hakimov, Nikolay. 2016a. Explaining variation in plural marking of single German noun insertions in Russian sentences. In Heike Beherens and Stefan Pfänder (eds.) *Experience counts: Frequency effects in language*, 21–60. *Linguae et litterae* 54. Berlin: de Gruyter.
- Hakimov, Nikolay. 2016b. Effects of frequency and word repetition on switch-placement: Evidence from Russian-German code-mixing. In Justyna Robinson and Monika Reif (eds.) *Cognitive perspectives on bilingualism*, 91–125. Trends in Applied Linguistics 17. Berlin: de Gruyter.
- Heine, Bernd. 2013. On discourse markers. Grammaticalization, pragmaticalization, or something else? *Linguistics* 51: 1205–1247.
- Hidalgo, Raquel. 2002. Establishing Topic in conversation: A contrastive study of left-dislocation in English and Spanish. In Angela Downing, Jesus Moya, José I. Albentosa (eds.), *Talk and text. Studies on spoken and written discourse*, 31–50. Cuenca: Ediciones de la Universidad de Castilla-La Mancha.
- Hidalgo Downing, Raquel and Laura Hidalgo Downing. 2007. Metadiscursive and interpersonal values of pronominal topics in spoken Spanish. In: Christopher S. Butler, Raquel Hidalgo Downing and Julia Lavid (eds.) *Functional perspectives on grammar and discourse. In honour of Angela Downing*, 327–348. Amsterdam/Philadelphia: Benjamins.
- Hilpert, Martin. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
- Hopper, Paul J. and Sandra A. Thompson. 2008. Projectability and clause combining in interaction. In Ritva Laury (ed.), *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions*, 99–123. Amsterdam: Benjamins.
- Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Kaltenböck, Gunther, Bernd Heine, Tanja Kuteva. 2011. On thetical grammar. *Studies in language* 35: 852–897.
- Kaltenböck, Gunther, Evelien Keizer and Arne Lohmann. 2016. *Outside the clause: Form and function of extra-clausal constituents*. Amsterdam: Benjamins.
- Kellerman, Anja. 1996. When Gibraltarians' speak we're quite unique! Constructing Gibraltar identity with the help of English, Spanish and their respective local varieties. In: Sandi Michele De Olivera (ed.) *The linguistic construction of social and personal identity. First international conference of sociolinguistics in Portugal*, 73–78. Evora: Cidehus.
- Kellerman, Anja. 2001. *A new new English. Language, politics and identity in Gibraltar*. Heidelberg: Heidelberg Schriften zur Sprache und Kultur.
- Kramer, Johannes. 1986. *English and Spanish in Gibraltar*. Hamburg: Buske Verlag.

- Lambrecht, Knud. 1994. *Information structure and sentence form. Topic, focus and the mental representation of discourse referents*. Cambridge: Cambridge University Press.
- Lambrecht, Knud. 2001a. Dislocation. In: Martin Haspelmath (ed.) *Typologie des langues et les universaux linguistiques: Manuel international*. Cap. 80, 1056–1078. Berlin: De Gruyter.
- Lambrecht, Knud. 2001b. A framework for the analysis of cleft constructions. *Linguistics*, 39(3):463–516.
- Levey, David. 2008a. The changing face of Gibraltar English. TH-fronting on the Rock. In: Ewa Waniek-Klimczak (ed.), *Issues in accents of English*, 91–102. Newcastle: Cambridge Scholar Press.
- Levey, David. 2008b. *Language change and variation in Gibraltar*. Amsterdam: Benjamins.
- Levey, David. 2015. Gibraltar English. In Jeffrey P. Williams, Edgar Schneider, Peter Trudgill and Daniel Schreier (eds.), *Further studies in lesser-known varieties of English*, 51–69. Cambridge: Cambridge University Press.
- Marcos Marín, Francisco. 1992. Corpus oral de referencia de español contemporáneo. Madrid.
- Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2): 281–331.
- Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
- Matras, Yaron and Jeanette Sakel. 2007. *Grammatical borrowing in cross-linguistic perspective*. Berlin: De Gruyter.
- Moyer, Melissa. 1998. Bilingual conversation strategies in Gibraltar. In Peter Auer (ed.) *Codeswitching in conversation. Language, interaction and identity*, 215–234. New York: Routledge.
- 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. In *Bilingualism: Language and Cognition* 16(4), 709–730
- Myers-Scotton, Carol. 1993. *Dueling languages: grammatical structure in codeswitching*. Oxford: Clarendon.
- Myers-Scotton, Carol. 2002. *Contact linguistics: bilingual encounters and grammatical outcomes*. Oxford: OUP.
- O'Shannessy, Carmel. 2011. Young children's social meaning-making in a new mixed language. In Ute Eickelkamp (ed.) *Growing up in Central Australia – indigenous experiences of childhood, youth and transformations*, 131–155. New York: Berghahn Books.
- Pannunzi, Alessandro. 2016. Pseudo-cleft sentences in spoken Italian. Syntax, semantics, information structure and a comparison with Spanish. In Anna-Maria

- De Cesare and Davide Garassino (eds.) *Current issues in Italian, Romance and Germanic non canonical word-orders*, 227–250. Frankfurt am Main: Peter Lang.
- Patten, Amanda. 2012. *The English it-cleft: A constructional account and a diachronic investigation*. Berlin: De Gruyter.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish *y termino en español*. *Linguistics* 18: 581–618.
- Prince, Ellen. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.) *Radical pragmatics*, 223–255. Amsterdam: Benjamins.
- Prince, Ellen. 1984. Topicalization and left-dislocation: A functional analysis. In Sheila White, Virginia Teller (eds.), *Discourses in reading and linguistics. Annals of the New York Academy of Sciences* 433, 213–225.
- Reig Alamillo, Asela. 2011. The pragmatic meaning of the Spanish construction *lo que pasa es que*. *Journal of Pragmatics* 43(5), 1435–1450.
- Rivero, María Luisa. 1980. On Left-Dislocation and Topicalization in Spanish. *Linguistic Inquiry* 11(2), 363–393.
- Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In Liliane Haegeman (ed.), *Elements of grammar*, 281–337. Kluwer: Dordrecht.
- Sloetjes, Han and Peter Wittenburg. 2008. Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation* (LREC 2008).
- Traugott, Elizabeth. 2008. “All that he endeavoured to prove was ...”: On the emergence of grammatical constructions in dialogical and dialogic contexts. In Robin Cooper and Ruth Kempson (eds.), *Language in flux: Dialogue coordination, language variation, change and evolution*, 143–177. London: Kings College Publications.
- Traugott, Elizabeth Closs and Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.
- Travis, Catherine E. 2005. *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin: De Gruyter.
- Treffers-Daller, Jeanine. 1994. *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin: De Gruyter.
- Vallejo, Tito. 2001. *The Yanito dictionary*. Gibraltar.
- Weston, Daniel. 2013. Code-switching variation in Gibraltar. *International Journal of Bilingualism* 17(1): 3–22.