

Distributed Morphology and Bilingualism: Code-switching and Mixed Languages^{*}

Submitted to *The Cambridge Handbook of Distributed Morphology*, edited by Artemis Alexiadou, Ruth Kramer, Alec Marantz and Isabel Oltra-Massuet.

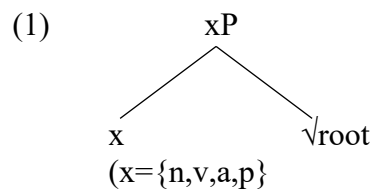
Luis López
April 2023

1. Introduction

It is said that a good scientific theory not only makes the predictions it was designed for, it also makes predictions that the original scientist did not know about. Since the early 2010s, there has been an explosion of work in the study of bilingual grammar and code-switching within the DM framework – or, more generally, exo-skeletal models of morphosyntactic structure (Alexiadou and Lohndal 2018, Alexiadou et al 2015, DenDikken 2011, Grimstad et al. 2014, Riksam 2017, López 2020, among others.) The reason for this interest is that the change in perspective on the theory of grammar provided by DM has given researchers an opportunity to gain fresh insights into bilinguals' I-languages as well as the tools to analyze long standing empirical problems. Most especially, DM has played a crucial role in the development of the *Integrationist Hypothesis*, according to which the linguistic competence of a bilingual speaker must be regarded as unitary, not as two separate systems. The purpose of this chapter is to explain what it is that DM brought to the study of bilingual grammar and what the study of bilingual grammar brought to DM.

At least three innovations of DM have been very helpful in bilingualism work:

(i) Marantz (1997) proposes that lexical categories like verb or noun are in fact syntactic structures made up of a categorizer and a root:



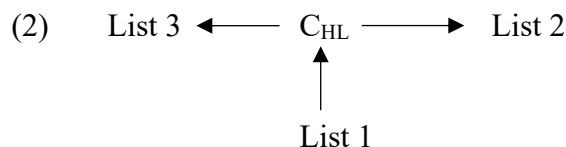
A direct consequence of this assumption is that so-called inherent properties of lexical items – such as the gender of nouns or the conjugation class of verbs – arise in particular instances of the syntactic configuration in (1).

(ii) DM rejects Lexicalism, understood as a specific proposal of grammar organization. Within Lexicalism, linguistic competence includes a lexicon, as a repository of words, collected with their semantic, morphosyntactic and semantic properties, and rules of word-formation; this lexicon feeds the syntax. Instead, DM proposes that word formation and syntax are one

^{*} I would like to thank Natascha Pomino for detailed and insightful comments on an earlier version of this chapter. The comments of an anonymous reviewer for CUP likewise forced me to sharpen my arguments and clarify my writing. I would also like to thank Ruth Kramer and Artemis Alexiadou for their kind invitation to contribute to this volume. All errors belong to me.

computational system (henceforth C_{HL}). Additionally, what feeds this computational system is a list of roots and bundles of morphosyntactic features, called List 1. Additionally, I incorporate Harley's (2014) proposal that roots are numerical indices whose only function is to map a semantic representation to a phonological interpretation.

The phonological shape of syntactic terminals is resolved in the PF section of the grammar by means of vocabulary insertion rules, referred to as List 2. Semantic interpretation of words and idioms is also resolved post-syntactically, by means of a List 3.



(iii) The vocabulary items (=exponents) of List 2 compete for insertion in syntactic terminals (Halle and Marantz 1993, Halle 1997). A syntactic terminal may be a root or a bundle of grammatical features, call it g_s . A vocabulary item includes a bundle of features, call it g_v , together with a rule that defines the context of insertion. In the ideal case, g_v matches g_s , but when that is not the case, the vocabulary item that has the biggest subset of g_s is inserted. This is referred to as the *Subset Principle*.¹

The rest of this chapter is organized as follows. Section 2 provides an overview of bilingualism and code-switching while I argue for the Integrationist Hypothesis. Section 3 shows how code-switching provides evidence for assumption (i). Section 4 discusses idioms and shows that switched idioms provide evidence for assumption (ii) because they show that the exponent of an idiom plays no role in idiom interpretation. Section 5 discusses two mixed languages and argues that they provide additional evidence for assumption (ii). Finally, Section 6 develops assumption (iii) by showing that the Subset Principle is active in the construction and spell-out of hybrid structures. Throughout, I will also argue that all the data presented here supports the Integrated Hypothesis.

Sections 2 and 6 are based on my own work (López 2020), while section 3 is also based on my own work as well as others who have converged on the same idea (Grimstad et al 2014 Alexiadou et al 2015, among others.) Section 4 is based on collaborative work (López et al. submitted). Section 5 is new and, consequently, it is longer and more detailed than the others.

2. Bilingual grammar. Code-switching.

We tend to understand “languages” as discrete, countable entities: Japanese is one language, Spanish is another, and so on. Consequently, bilingual persons are in possession of two distinct linguistic systems. Let's call this view *Separationism*. Separationism is the most common way of understanding bilingualism and has been argued for explicitly by acquisitionists (de Houwer 2005, Meisel 2019). It is a common sense notion: we understand languages as being independent

¹ “The phonological exponent of a Vocabulary Item is inserted into a position if the item matches all or a subset of the features specified in that position. Insertion does not take place if the Vocabulary Item contains features not present in the morpheme. Where several Vocabulary Items meet the conditions of insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen.” (Halle 1997)

entities: so somebody who can speak the local languages of Tokyo and Munich must be in possession of two languages.

(Post)Creole continua present a difficulty for Separationism: as pointed out many times by creolists, Creole speakers usually have a number of overlapping choices to express the same sentence. Bell (1976) lists the following forms in Guyanese Creole, among several others. All of them translate as ‘I gave him one’:

- (3)
- a. A giv im wan.
 - b. A did gi: i: wan.
 - c. Mɪ dɪ gi: i: wan.
 - d. Mɪ bɪn gi: i: wan.

How many languages does the Guayanese person speak? It is hard to see (1a) and (1d) as being sentences of the same language – the grammars that generate these sentences are very different. But what about all the intermediate forms? There is no answer to this question because the post-Creole situation constitutes a continuum. Creole continua provide stark evidence that regarding languages as discrete entities is misguided – probably a product of our educational and political systems that yield monoglossic ideologies. Monoglossic ideologies take monolingualism to be the norm and bilingualism is the marked phenomenon in need of study; it follows that a language is a homogenous, circumscribed, system.

The notion that a bilingual person is someone who has two languages in their head is so ingrained in our thinking, that the possibility that it should be challenged may sound absurd. However, empirical evidence has accumulated over the years that leads us no doubt that it should be rejected. Instead, we have to understand that every person on earth has a system of linguistic competence in their head, built on the bricks and mortar provided by the environmental input – and if the latter is complex, so will the resulting system be. Let’s call this theory of multilingualism *Integrationism*. Integrationism has been argued for in the context of the exo-skeletal approaches to code-switching mentioned in the introduction, as well as from constraint-based approaches to language (Hsin 2014, Goldrick et al 2016, Putnam et al 2018).

Integrationism makes some correct empirical predictions. For instance, many bilingual communities code-switch and when they do, they are likely to use an equivalent of a verb like ‘do’ or ‘make’ as a light verb to construct plain transitive predicates. Thus, a German-Spanish bilingual may regularly say something like (4) (see González-Vilbazo 2005, González-Vilbazo and López 2011, 2012):²

- (4)
- | |
|---------------------------|
| Spanish/ <i>German</i> |
| <i>hacer</i> <i>nähen</i> |
| do sew |
| ‘sew’ |

But interestingly, in order for the light verb meaning to come about, the lexical verb must be German: *Hacer coser* (‘make sew’) can only have a causative meaning. This means that the light verb *hacer* (as opposed to causative *hacer*) necessarily selects a “German” lexical verb. This should be completely impossible if *hacer* and *nähen* belonged to two different lexica. However,

² In the code-switched examples I use the general convention that one of the participating languages is written up in italics.

it is almost to be expected if they are members of the same lexicon (however this may be construed).

Some of the arguments for Integrationism rely on parsimony. Evidence that a polyglot has an integrated linguistic system can sometimes be recast in a Separationist framework as evidence that the two linguistic systems are “connected”, “co-activated”, or they “influence each other”; syntactic structures may “transfer” or be “calqued”, and so on. As these ad hoc labels accumulate, one should eventually wonder what they are worth. Here is an example: Vaughn-Evans et al (2014) tested the electrophysiological responses of a group of Welsh-English bilinguals as they processed mutated nouns *in English* (i.e.: ‘goncert’ for ‘concert’). Interestingly, the stimuli included mutations that obeyed the syntactic restrictions on mutation in Welsh, others that violated them and yet others were aberrant (‘doncert’ for ‘concert’). They found that the subjects processed the “good” mutations faster than the “bad” ones. Vaughn-Evans et al (2014) accounted for this in terms of co-activation: the Welsh mutation rule is active even while processing English. But why should an alien grammatical rule be active? If a speaker has two separate grammatical systems and is only using one, why should the other one be active? It seems more plausible to say that there is a mutation rule that is a component of the linguistic competence of Welsh speakers, not a component of their Welsh grammar. The performance systems make sure that it is inhibited when speaking English, as many other aspects of our linguistic competence are inhibited when we speak (taboos, inappropriate register and so on). From now on, although sometimes I will refer to the words of a bilingual as being e.g. “Spanish” or “English”, it should be understood that this is a shortcut for expository purposes

Within Integrationism, all lexical roots and functional bundles share the same real estate in List 1. For instance, a Spanish-English bilingual will have a feature that can be labeled as [past] and spells-out as *-ed* for regular roots, but she also has feature structures that bundle past with imperfect and perfect aspect, and these will spell-out in the forms of Spanish preterit and imperfect. Likewise, the vocabulary insertion rules that comprise List 2 are all integrated and – potentially – they may compete for the same position.

Let me now introduce code-switching. Code-switching, as a descriptive label, can be applied to different types of phenomena of bilingual speech in which heterogenous linguistic material is used in a single utterance. Here we are interested in the kinds of fluent intra-sentential switches that deep bilinguals can produce. The following is an example, produced by a Turkish-German bilingual consultant. As the reader can see, there are German words and Turkish words in one sentence:

(5) Turkish/German

Ama lütfen *vergessen* yapma. Biz Almancılar hepimiz *verschieden* konuşuyoruz. Bu *Sprache* çok flexibel. Yani hangi *Wort*’u daha önce *erinnern* yapıyorsan. Bu *Sprache*’de bir *Regel* yok.
‘But please don’t forget. We *Almancılar* speak differently. This language is very flexible. Namely, the first word you remember. There are no rules in this language.’

Code-switching is usually defined as the seamless interweaving of words from two languages in one utterance. However, this definition is based on a separationist view. If we adopt

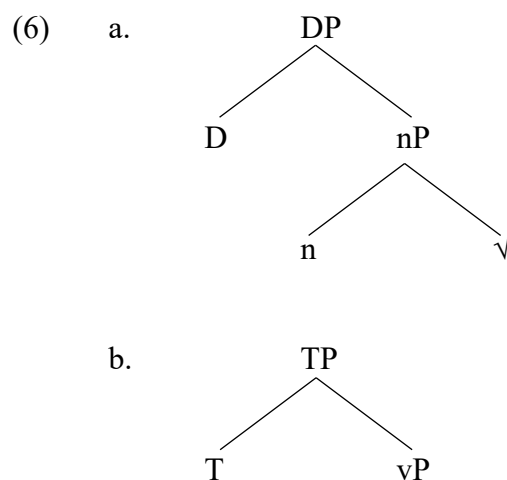
Integrationism, we can define code-switching as the usage of the speaker's entire set of linguistic resources in one utterance – without inhibiting a subset of them.³

The consultant expresses a very extended idea regarding code-switching: it consists of the combination of words from one language or the other without rhyme or reason. This idea, however, is mistaken. Scholars who study code-switching agree on only one thing: code-switching is rule-based, as rule-based as any other aspect of an individual's linguistic competence. For instance, notice that in the example, the German verbs (*vergessen* 'forget', *erinnern* 'remember') appear in their infinitive form to the left of a Turkish light verb, strictly following the rules of this language. The German word *Wort* appears with an accusative case mark *u*, as required by the transitive Turkish predicate. Harder to notice are the things the speaker does not do: for instance, code-switchers in many parts of the world do not code-switch between a pronoun and the verb (Koronkiewicz 2014). Since code-switching is rule governed, code-switchers can provide acceptability judgments on code-switched sentences as well as on unilingual sentences. Moreover, bilingual speakers from the same speech community have consistent judgments, thus confirming that code-switching is an expression of a form of linguistic competence.

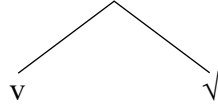
3. The features of nouns and verbs.

In this section I discuss assumption (i), the idea that syntactic categories are in fact syntactic structures consisting of a root and a categorizer. I introduce the original idea first and then I discuss how the code-switching data has enriched and provided additional evidence for it.

Famously, Marantz (1997) proposed that the input to the computational system is a list that consists of roots and bundles of grammatical features, which he called List 1. List 1 does not include nouns or verbs, these are syntactic products: a noun is a syntactic structure consisting of a root and a categorizing head, called *n*, a verb is a root selected by *v* and so forth. Broadly following Marantz' idea, assume the minimal structure of a noun phrase is (6a) and the clause is (6b):



³ For a typology of code-switching phenomena with many examples, see Muysken (2000). There are discussions of whether or how code-switching is distinct from lexical borrowing, see Grimstad (2017) for thoughtful discussion and numerous references.



The head D encapsulates the [\pm definiteness] feature. In some languages, D also bears unvalued gender and number features. The unvalued number feature of D is valued against a valued number feature that heads its own projection (not depicted). The unvalued gender is valued against the head *n* – following Kramer (2015) among others, I assume that gender is a feature of *n*.⁴ A language with gender will have different types of *n*: *n*[*f*], *n*[*-f*] in a two-gender system, with more complex systems for languages with three or more genders and noun classes. Some languages, like English, do not have any concord features in the noun phrase: for these languages, we posit a simple *n*.

Likewise, the head T of (4b) may have unvalued person, number and/or gender features, which will be valued against the closest c-commanded DP in the structure. Again, many languages do not have any agreement features in T.

Lexicalist theories take nouns and verbs to be listed in a lexicon with their inherent morpho-syntactic features. Most specially, gender/noun class features are seen as inherent to nouns. This view has some well-known empirical challenges. For instance, some Romanian nouns change gender concord in the plural from masculine to feminine (see Kramer 2015: 166-183 for discussion). In Bantu languages, nouns are not restricted to a particular class, but can appear within a range of possible noun classes, thus giving rise to diminutives, augmentatives and abstract nouns (López 2020). These and other phenomena have led linguists in the DM tradition to take gender to be a feature that emerges in a particular syntactic configuration (see Kramer 2015). Likewise, *v* comes in distinct types, referred to as flavors (Arad 2003), that relate to agreement and case assignment.

Interestingly, the investigation of bilingual grammars has uncovered that the role of *v* goes beyond case and agreement. González-Vilbazo and López's (2012) investigation of light verbs in German-Spanish code-switching shows that the light verb determines the VO or OV order of its complement. Later research has confirmed this claim in other language pairs: Japanese-Brazilian Portuguese, Telugu-English, Sranan-Hindustani, Japanese-English and Korean-English (see Veenstra and López (2016), López (2020), Shim (2016, 2021). Under the assumption that a light verb is a spell-out of *v*, we argued that *v* is responsible for the OV or VO order.⁵ Additionally, the light verb also decides whether O forms a prosodic phrase with V or they head distinct phonological phrases (González-Vilbazo and López 2012, López 2020).

Much code-switching data provides evidence for the view that nouns and verbs are syntactic structures composed of a root and a categorizer, as argued most forcefully in Alexiadou et al 2015, Grimstad 2017, Grimstad et al 2014, Riksem 2017. Consider the following examples:

- (7) Spanish/*German*
 La *Hose*

⁴ This analysis is not adopted universally. Picallo (1991) proposed that gender constitutes an independent functional category within the DP. This analysis is adapted to the study of code-switching in Grimstad et al (2014). For our purposes, either theoretical choice works.

⁵ There is now a wealth of work that splits the functions of *v* into a categorizer – *v* – and *Voice*, the latter involved in transitivity (see Alexiadou, Anagnostopoulou and Schäfer 2015 for an extensive argument). The reader can replace *v* with *Voice* in the following discussion regarding light verbs without affecting the argument.

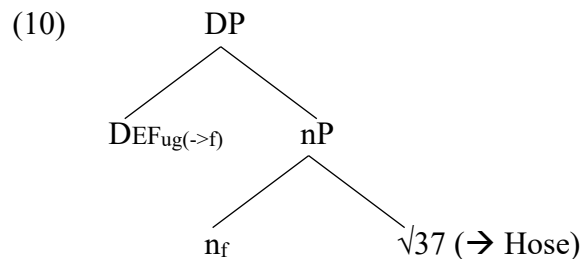
- (8) Spanish/English
 La house
 DEF.F

Delgado 2020:3

- (9) KiSwahili/English
 Ø-saa hi-yo i-na-depend na certificate 10-zako 10-za
 9-time 9-this 9-PRES-depend with 10-certificate 10-your 10-with
 Ø-shule
 10-school
 ‘At this time, it depends on the school certificates.’

Myers-Scotton and Jake 2009: 339

Let’s start with (7) and (8). I take it that *Hose* is the spell out of a root to which I arbitrarily assign the index 37. *Hose* in German is a feminine noun, a feature that has been kept in this code-switched structure. *Hose* is selected by a Spanish determiner that exhibits feminine concord. The Spanish definite determiner will get its feminine feature from n_f .



(8) is somewhat different. The English word *house* has no gender feature, but the determiner exhibits a feminine exponent.⁶ Therefore, (8) must include a n with a feminine feature, which must come from Spanish. From a lexicalist point of view, the feminization of ‘house’ is surprising: do bilingual speakers have two versions of ‘house’ in their repertoire, one of which is feminine and the other one has no gender? How does the gender of ‘house’ happen to be analogical to that of Spanish *casa*?

Within DM an analysis is easy. There is a root, say $\sqrt{67}$, which can be selected by n or by $n[f]$. The possibility that a root may be selected by two types of n is well documented in monolingual grammars, as in the Romanian and Bantu data mentioned above. Therefore, the idea that bilingual code-switchers also allow for a root to be selected by two flavors of n is not surprising. In List 2, $\sqrt{67}$ spells out as ‘house’ if selected by n . If $\sqrt{67}$ is selected by $n[f]$, it can spell out as ‘house’ or as *casa*.

The KiSwahili-English example (9) is included to show that the data can be complex. Take the noun *certificate*. If we were to analyze it as an English noun, we would say that it

⁶ This phenomenon is referred to as analogical gender, whereby a noun from a language that has no gender seems to inherit the gender of the equivalent noun in the gender language.

includes a root ‘certif-’ and the nominalizer ‘-ate’. As ‘certificate’ is used in a KiSwahili sentence, it is assigned to class 9/10. We can detect this because of the concord it triggers on the possessor and the preposition. Within DM, we can assume that the List 1 of a KiSwahili-English bilingual includes a root which can be selected by a less *n* or by a *n*[9/10]. Both types of *n* spell out as -ate.

Code-switching data provides rich evidence that gender cannot be inherent to a noun. In many language pairs, we find that a noun from a language that has no gender or noun class (like English) can appear as a constituent of a determiner phrase with gender/noun class exponents on determiners and adjectives (like KiSwahili, Spanish, German). This analysis builds on the fact that a root can be inserted in two different morphosyntactic frames – which is a feature of unilingual speech also - and supports the DM view of gender/noun class as being the result of a syntactic configuration and not something inherent of a lexical item. Simultaneously, it also provides evidence for the Integrationist Hypothesis: if the lexicons of the Spanish/English bilingual are separate, where does the feminine gender on ‘house’ come from? The lexicalist needs to posit some channels that connect the two lexicons; this connection is unmotivated beyond the empirical datum itself.

Let’s discuss verbs now. I use examples (11) and (12) for this purpose:

(11) German/*Spanish*

Wir	<i>utilisieren</i>	spanische	Wörter, die dann	<i>alemanisiert</i>	werden	y	<i>hacen</i>
We	use	Spanish	words	that then Germanized	are		and do
klingen		<i>un poco raro</i> .					
sound		a bit strange					

‘We use Spanish words, that are then Germanized and sound a bit strange.’⁷

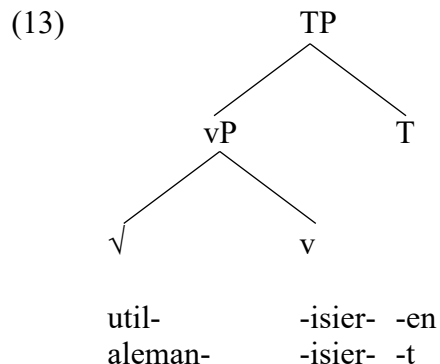
González-Vilbazo 2005: 1

(12) Norwegian/*English*

Jeg	<i>teach-a</i>	første	<i>grad-en</i>
I	teach-PAST	first	grade-DEF.M.SG

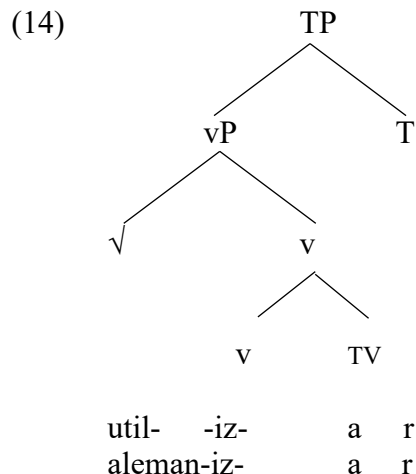
Grimstad et al 2014: 217

In example (9), the verbs *utilisieren* ‘use’ and *alemanisiert* ‘Germanized’ are constructed on a Spanish root and a German inflectional structure. I take it that the suffix *isier* spells out little *v* and *en* and *t* spell out T:



⁷ My gloss and translation.

The code-switched *utilisieren* has a fully Spanish version *utilizar*, and *alemanizar* is also an acceptable Spanish verb. The categorizer is *-iz-*. The Spanish verb belongs “inherently” to the first conjugation as revealed by the thematic vowel *a* (on thematic vowels in Spanish and *v* see Arregi and Oltra-Masuet 2005).



German has no conjugation classes. Following the logic sketched above for code-switched noun phrases, we could adopt the assumption that bilingual German-Spanish speakers have two types of little *v*, one with a conjugation class feature, the other without. The roots *util-* and *aleman-* could be selected by either of the two types. The existence of these two types of little *v* is a property that follows directly from the simple fact of bilingualism. On the other hand, within lexicalism, there should be two verbs that mean the same, *utilizar* and *utilisieren*, inhabitants of distinct lexica. We should ask why there is a reason to posit these duplicated lexica other than for describing the data.

In the examples in (11), the entire functional structure is drawn from one language and only the root is switched. In fact, insertion of roots is a very easy and common form of code-switching, and all the sentences above instantiate it. On the other hand, code-switching between two functional categories is more problematic and often leads to unacceptable results. Forms like **utilisieraban* or **alemanisieron*, in which the German *v* is selected by a Spanish *T* (*aban*, *on*) are unacceptable. There are also restrictions on grammatical code-switching between *T* and *v* or between an auxiliary and a participle (González-Vilbazo and Struckmeier 2008); switching between *C* and *T* is unacceptable in several varieties of code-switching (Belazi et al 1994, González-Vilbazo 2005, among others). Code-switching between *n* and *D* is certainly possible, as shown in (8), but code-switching within the functional categories of the DP may also present difficulties (Bartlett and González-Vilbazo 2013). Why should this be the case? Notice that the distinction between root and categorized structure provided by DM already gives us an answer. Roots have no grammatical features that could clash with other grammatical features in the structure, so it follows that code-switching of roots should be freely available, as is the case. But, if switching involves functional categories, these functional categories will enter grammatical dependencies between them and other constituents in the clause, with the obvious potential of feature incompatibility and unacceptable results. The separation between categories and roots that DM argues for allows us to formulate this generalization about code-switching.

The English-Norwegian example (12) adds an interesting turn of the screw for DM practitioners. The English verb ‘teach’ appears as the complement of the Norwegian past morpheme *a*. However, as we know teach+past should yield ‘taught’. Thus, the challenge is to make sure that ‘teach’ yields an irregular past form only when surrounded by English functional categories. We can follow the system in Embick (2010) and assume that the form ‘taught’ is generated with the rules (15a,b) (see Embick 2010, 2015). The speakers of American Norwegian have an additional rule, (15c) according to which *-a* spells out past as a suffix on the verb:

- (15) a. $v + \sqrt{345} \leftrightarrow \text{taught} / \text{_____ past}$
 b. $\text{Past} \leftrightarrow \emptyset / \text{taught } \text{_____}$
 c. $\text{Past} \leftrightarrow a$
 d. $\text{Past} \leftrightarrow \text{ed}$

However, these rules will not produce the result we want if spell-out rules apply bottom-up (Bobaljik 2000, Embick 2010) because (15c) will never find an environment in which it can apply: $\sqrt{345}$, selected by past, will always yield ‘taught’, never *teacha*. One possible approach is to assume that speakers of Norwegian-English have two possible versions of *v*, call it *v1* and *v2*. Only *v1* has the property of triggering the spell out rule (15a).

4. Idioms

The study of code-switched idioms also provides empirical evidence for DM. The DM approach to idioms, as sketched in Marantz (1997) can be summarized thusly: (i) idioms are built in the computational system like any other phrase and (ii) the Encyclopedia assigns the structure a special meaning. Interestingly, this analysis entails that idiom interpretation is independent of whatever List 2 items spell out the constituents of the idiom. These assumptions are distinct from the other important current on the study of idioms, arising from Construction Grammar, in which an idiom is listed holistically in the lexicon with its morphosyntactic, semantic and phonological features. (Goldberg 1996, and many others)

Let’s further assume the Integrationist approach to bilingualism. If the lexica of the bilingual are integrated, one could expect that what we perceive as equivalent words in the two languages – say ‘puente’ and *bridge* – are in fact alternative exponents for the same List 1 item:

- (16) $\sqrt{456} \leftrightarrow \{[\text{puente}], [\text{bridge}]\}$

In fact, I suspect that large swaths of the bilingual nominal and adjectival vocabulary are structured as in (16), possibly even many verbs. Finally, allow me to remind the reader that some idioms are rigid and substituting one word for another breaks the idiomatic reading (17a) has an idiomatic meaning (=that is past) while (17b) and (17c) do not:

- (17) a. That is water under the bridge.
 b. That is a river under the bridge.
 c. That is water under the aqueduct.

Therefore, if something like (16) truly represents how parts of the lexicon of a bilingual speaker are organized, and if the DM analysis of idioms is correct, then we should be able to substitute one exponent of an idiom for another and maintain the idiomatic meaning:

- (18) English/ *Spanish*
 a. That is water under the *punte*. (punte=bridge)
 b. That is *agua* under the bridge. (agua=water)

López et al (submitted) decided to test this prediction in three bilingual communities: Papiamentu-Dutch speakers in the Netherlands, Spanish-Basque speakers in Gernika (Basque Country) and Spanish-English bilinguals in the Chicago area. We found that idiomatic meanings are indeed, by and large, maintained after the List 2 items are replaced. The following is an example:

- (19) Dutch/*Papiamentu*
 De aap komt uit de *manga*.
 The monkey comes out the sleeve
 ‘The truth is revealed’

In (19), we have a Dutch-Papiamentu code-switched sentence. This sentence is also an idiomatic expression in Dutch (there is no equivalent idiom in Papiamentu), from which we substituted the Papiamentu *manga* for the Dutch *mouw*. 15 out of 17 bilingual speakers accepted the code-switched result as natural in their speech community.

We take our results with code-switched idioms to be straightforward evidence for the DM model of grammar. First, code-switched idioms show that the idiom cannot be stored as a unit in the lexicon. As Mareike Keller, points out “One might assume that a string of words which appears as a unit on the level of conceptual representation should be barred from internal language mixing in order to preserve the exact meaning or pragmatic function of the unit.” (Keller 2020: 200).

Second, code-switched idioms show that the phonological form of a lexical item is separable from morphosyntax, and idiomatic meanings come about as a semantic property of a syntactic structure. This result is unexpected in a lexicalist framework where semantic, syntactic and phonological features are all inherent to the lexical item. In the next section we are going to see more examples of this separability.

5. Mixed languages

5.1 Introduction

Mixed languages arise in contact situations with pervasive bilingualism, and they are the result of combining resources from two languages into what appears to be a fully novel, stable, linguistic system. From the point of view of the organization of linguistic competence, the Integrationist Hypothesis leads us to conclude that there is no substantial difference between code-switching and mixed languages: both can be understood as the combination of ingredients that may appear heterogenous by the outside observer. There is a subtle difference, but it

involves usage, external perception, and social role: Code-switchers combine resources from their repertoire freely as the communicative needs arise; a mixed language is regarded by its speakers as a stable mode of expression and as a linguistic code independent from the participating languages. In fact, as shown by Lipski (2020b), a speaker might code-switch between a mixed language and one of the sources of the mixed language.

In the present context, the mixed languages I discuss here are of interest to us because they provide ample empirical evidence for the DM assumption (ii) of the introduction: the postulate that lexical items are distributed in three lists: the List 1 made up of roots and grammatical properties; the List 2 that consists of rules of exponence; and the List 3, a list of concepts, usually referred to as encyclopedia. The argument presented in this section builds on the property of manipulability: Mixed languages show that either List 2 or List 3 can be manipulated, leaving everything else unchanged. This property is unsurprising under DM because it follows from the design of the grammar. Within a Lexicalist framework, detaching inherent features should be surprising and require some ad hoc mechanism to account for it. In the following, I discuss Media Lengua and Sri Lankan Portuguese.

5.2 *Media Lengua*

The sentences in (20) exemplify Media Lengua (henceforth ML). Broadly, ML can be described as a language built on Spanish lexical roots and Quichua morphosyntax and phonology; although personal pronouns and question words are also built on Spanish material, and ML may also show some prepositions, which are absent from Quichua. Media Lengua has attracted the attention of some scholars and now we have a rich literature that I borrow from: Muysken 1979, 1986, 1997, Lipski 2020a,b, Gómez-Redón 2008, Deibel 2020, Stewart 2019, Shappeck 2011.

In example (20a), the lexical items *kumi* ‘eat’ and *kiri* ‘want’ are built on the Spanish *comer* and *querer* respectively. Notice that the phonology of the roots is adapted to Quichua phonology, in particular the Spanish mid vowels are raised following the Quichua pattern (although there appear to be some words where this adaptation does not take place, see Gómez Redón 2008, Stewart 2014). The suffixes are all drawn from Quichua. The OV order also comes from Quichua, Spanish is a VO language. In (20b) *yu* ‘I’, *bus* ‘you’, *abla* ‘speak’ and *kiri* ‘want’ are drawn from Spanish, respectively: *yo*, *vos*, *hablar* and *querer*. In (20c) *unu*, *fabur*, *pidi*, and *bini* are also drawn from Spanish.

- (20) a. *kumi-na kiri-ni*
 eat-INF want-1
 ‘I want to eat.’

Muysken 1979: 45

- b. *Yu-ka bus-kuna-man midialingua-pi abla-ngapa-mi kiri-ni*
 I-TOP you-PL-DAT MediaLengua-LOC speak-DES-AFFIRM want-1
 ‘I want to speak to y’all in MediaLengua.’

Lipski 2020: 334

- c. *Unu fabur-ta pidi-nga-bu bini-xu-ni*
 One favor-ACC ask-NOM-BEN come-PROG-1
 ‘I am coming to ask a favour.’

ML is spoken in some valleys in Ecuador with considerable differences between one location and another. The population are typically bilingual or trilingual, using ML in the village, Quichua with surrounding villages and Spanish in town or with outsiders, although recent fieldwork suggests that there is shift to Spanish (Muysken 1997, Shappeck 2011). As originally described in Muysken (1979), Media Lengua arose as the outcome of a process of relexification, involving the replacement of the phonological form of a lexical item in Quichua with the phonological form of an equivalent lexical item in Spanish, while retaining the semantics and morphosyntax of the replaced Quichua item. Muysken's original analysis has been questioned in light of the complex contact situation in Ecuador, where multiple linguistic processes take place simultaneously, such as code-switching, calquing and borrowing, which suggests that relexification is not all there is (Shappeck 2011). However, this criticism modulates but does not affect, in my view, Muysken's original idea that Media Lengua consists, at least to a large extent, in relexification.

The first question is to find out what relexification really involves. In the examples in (20) we can see that the morphosyntactic structure is fully Quichua, unaffected by the change of roots. Following, Muysken's description, it looks like the semantics is also Quichua. Muysken (1979) argues that the semantics of the roots is Quichua by using the root [sinta]. The root [sinta] is built on the Spanish verb *sentar* 'to seat', but carries additional meanings absent from the Spanish verb, such as 'to live (in a place)' or 'to be (locative)', which map precisely with Quichua *tiya* (Muysken 1979: 45). Since both the morphosyntactic and semantic structures come from Quichua, we can see the relexification in ML involves the replacement of the phonological form of the roots only, leaving the semantic and morphosyntactic features untouched. It is not surprising that Muysken (1979) calls ML a Quichua dialect.

A DM analysis of ML is straightforward: the only change from Quichua to ML is that the exponents of some roots have been replaced with exponents from Spanish: an addition or alteration to the List 2 of these speakers, while maintaining the List 1 and List 3 items untouched. This is exactly what could be expected from a model like Marantz' (1997). Once again, Lexicalism would force us to posit some operation that allows the grammar to detach what is otherwise regarded as an inherent property, such as the phonological representation, of a lexical item and replace it with another inherent property.

We turn our lens now on the examples in (21). The first question is: what is really in the structure? Is the Spanish component just the root or does it embrace some functional structure? At first blush, the "Spanish" component would seem to consist of the root followed by a categorizer. Consider the following examples:

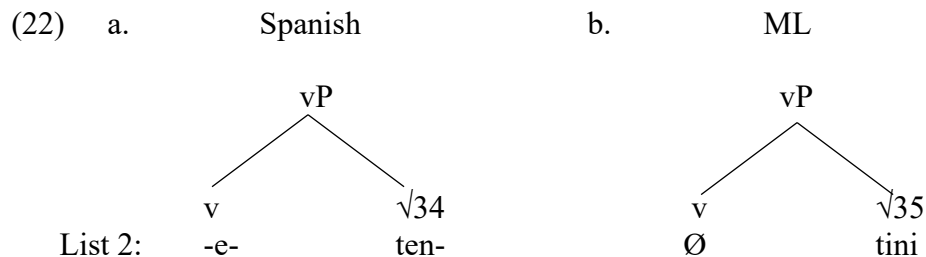
- | | | | |
|------|----|-------|------------------------------------|
| (21) | | ML | SP |
| | a. | tini | ten-e-r
ten-TV-INF
'to have' |
| | b. | esta | est-a-r
be-TV-INF
'to be' |
| | c. | sinta | sent-a-r
seat-TV-INF |

The ML verbs end in *i* or *a*, replicating the thematic vowels of the corresponding Spanish verbs. Spanish thematic vowels can be regarded as spell-outs of the verbalizer *v* or morphemes associated with *v* (see Arregi and Oltra-Massuet 2005 and example 12 above). This could lead to the (mistaken) conclusion that the structure of a ML verb consists of a Spanish vP on top of which a Quichua structure has been built. This conclusion however would raise a question: It has been shown that *v* regulates a number of morphosyntactic properties of its complement (Arad 2005, Embick 2010, González-Vilbazo and López 2012, Delgado 2020, López 2020) and even phonetic properties (Delgado 2020). As mentioned, González-Vilbazo and López (2012) show that the choice of *v* determines the VO or OV order. This body of work would lead us to expect that the complement of a Spanish *v* should have Spanish properties, but actually every constituent of the clause is fully adapted to Quichua phonology and morphosyntax and the word order is OV, as in Quichua. If *v* were Spanish, it would be a fully inert *v*.

Additionally, the thematic vowels regulate other properties of the verbal inflection. For instance, the imperfect tense takes the suffix *-ba* with the *a* thematic vowel (yielding *sentaba*) while the *e, i* thematic vowels take the suffix *-(i)a* (yielding *tenía*). None of this is visible in ML because the inflectional structure is entirely drawn from Quichua. Thus, there is no evidence that ML has thematic vowels as normally understood.

Finally, one more observation. Spanish verbal roots can be highly irregular. For instance, the root [ten-] is diphthongized to [tien-] when it carries word stress, it becomes [tuv-] in the past tense and it is turned into [teng-] in the first person singular of the present indicative and all the forms of the present subjunctive. All of these irregularities disappear in ML, which shows only the form [tini] (cf. the American Norwegian datum in (12)). This suggests that the roots of ML are not the Spanish verbs within a Quichua structure: If the ML form included a Spanish little *v* and thematic vowel, it would be subject to at least the diphthongization rule. The fact that it isn't suggests that the ML roots are not Spanish roots but fully adapted forms, built upon the Spanish stem root+TV, but Spanish functions only as a source language.

These considerations lead me to conclude that the final vowel of ML roots is not a thematic vowel or in any way related to categorizer *v*. The ML [tini], [esta] and the rest are in fact pure roots without category specification, the category specification is introduced by a fully Quichua *v*. This is exemplified in (22) with the roots *ten* and *tini*, meaning 'have', to which I have given arbitrary numbers. The Spanish verbal root combines with *v* building the vP structure that ends up being spelled as *estar* or *tener*, as shown in (22a). In ML, this vP structure is reanalyzed as a root, the categorizer being a Quechua one that constitutes the foundation of the clause's extended projection, as shown in (22b):



I conjecture that this is the origin of ML: ML consists of creating a new battery of exponents for already present List 1-List 3 combinations by fusing the Spanish Root + *v/n*

structure into a new Root. This suggestion can be pictured as in (23). It shows a fragment of a List 2 of a (hypothetical, possibly non-existent) trilingual Spanish-Quichua speaker. Recall that I showed above (following Muysken's 1979 analysis) that Quechua *tiya* and ML *sinta* occupy the same semantic field – in our terms, a unique root expones as *tiya* or *sinta*. The Spanish vocabulary item *sent*, which forms the basis for ML *sinta*, is associated with a more reduced semantic field, which suggests that it expones a root distinct from the root that expones *tiya* and *sinta*. Additionally, *sinta* and *tiya* are selected by functional categories with properties very distinct from Spanish *sent*. Thus, I take it that the ML and Quechua vocabulary items are alternative spell-outs of the same root while the Spanish one is a separate root:

- (23) √234 ↔ { [tiya], [sinta] }
 √235 ↔ [sent]

5.3 *Sri Lankan Portuguese*

Sri Lankan Portuguese (henceforth SLP) originated as a Portuguese creole that developed in Ceylon during the Portuguese occupation in the XVIth century and became isolated from the metropolitan language when the Dutch – and later, the British – colonizers took over. Since speakers of SLP are also typically bilingual in Tamil, the resulting morphosyntactic structure is largely Tamil while the exponents of both the lexical roots and the affixes are recognizably Portuguese. (For a description of SLP, see Smith 1977, 1979, for a summary of its external history, see Smith 1979, Canelas Cardoso 2009). Here are some examples from Smith's discussion:

- (24) a. kəzmæ:ntu-pə dispo:s
 marriage-DAT after
 'after the marriage' (depois do casamento)
 b. iskə:lə ənda:-se:, iskə:lə-ntu tudus məlwa:r
 school go-COND school-LOC all Tamil
 'If (they) go to school, everyone in school is Tamil.'
 c. E:w eli-pə diñe:ru ja:-dá:
 I him-DAT money PAST-give
 'I gave him the money.'

Smith 1979: 202, 198, 201.

Almost every morpheme in these examples can be traced to a Portuguese origin. Consider (24a). The two lexical items *kazmentu* and *dispo:s* are drawn from Portuguese. Even the dative marker is probably adapted from the Portuguese preposition *para*. But the structure is head final, as in Tamil and unlike Portuguese. In (24b) the morpheme [-se] is drawn from Portuguese, where it has the same function of introducing a conditional clause but, whereas in this language it is a free morpheme located in first position, in SLP it has become a verbal suffix, again as in Tamil. (24b) also exemplifies the head final property of SLP. Finally, (24c) shows a ditransitive sentence which is also head final, as in Tamil, except for the past tense marker, which is head initial. This past tense marker is drawn from a Portuguese adverb that indicates perfect aspect (as English

‘already’), but its grammar is not Tamil or Portuguese - rather it has all the markings of a Creole origin (Clements (2009: 49) points out that using *ja* as a past-perfect marker is a feature of several Portuguese creoles).

As Smith (1979) argues, the interpretation of morphemes is also Portuguese. This can be seen in the distribution of Differential Object Marking (DOM) in SLP (Smith 1979: 197-8). In SLP, human and animate beings bear accusative case, which spells out as the suffix [-pə]. This morphosyntactic pattern reflects an aspect of the grammar of Portuguese as it was spoken in the XVIth Century, and which has a very limited distribution in contemporary Portuguese. In this “old” Portuguese the DOM marker was the preposition [a] (see Pires 2021 for a full description). Notice also that both the Portuguese [a] and the SLP [-pə] are also used to expone dative case. There has been a transposition of the semantic properties of [a] onto [-pə], while the latter is a suffix in the Tamil manner. In Tamil there is also DOM, and it is a suffix, but the distribution follows the [±definiteness] feature. The Tamil DOM affix is [-a(y)], which is distinct from dative [(u)kku]. Thus, the morphosyntactic structure of SLP is very much like Tamil’s, while the meaning and phonology of the words remain Portuguese.

If we translate the previous description into a DM model, in SLP all List 1 features are Tamil while the List 2 and List 3 units remain Portuguese. SLP then complements what we saw with ML: while the latter shows that new (Spanish) List 2 items can attach to old (Quichua) List 1-List 3 pairs, SLP shows that new or modified List 1 (Tamil) items can attach to old List 2-List 3 (Portuguese) pairs. I take this to constitute empirical evidence that the DM conception of words is correct. The mixed languages phenomena suggest that words should not be regarded as units of morphosyntactic, semantic and phonological features but rather, each set of features forms its own list and the unit we call word is the output of the computational system.

(25) shows the two stages in the formation of SLP from Iberian Portuguese - although I am not positing a diachronic change from (25a) to (25b) for which I have no evidence; (25a) is simply an abstract representation of what a speaker in Ceylon may have heard as input and (25b) what the current patterns reflect. (25a) represents a description of an Iberian Portuguese conditional. The Portuguese conditional is defined as a complementizer that selects for a finite Tense and spells out as /se/. (25b) shows the resulting structure in SLP, as described in Smith (1979): the only apparent change is that /se/ has become a suffix:

(25) a. <i>Iberian Portuguese – Ceylon Portuguese circa XVIthC</i>			
	List3	List1	List2
	[cond]	C, selects T ^{fin}	/se/
b. <i>SLP</i>			
	List3	List1	List2
	[cond]	C, selects T ^{fin}	/se/
		Verbal clitic	

That is, the List 1 item has simply acquired the property of being a verbal clitic, while the List 3 and List 1 units are (as far as we can tell) untouched.

We can do the same work with DOM. The result is a little more complicated because the Portuguese [a] has been replaced with a new item, [-pə], which is itself derived from the Portuguese preposition *para*, which has been reanalyzed to take on some of the functions that [a] has in Portuguese, i.e., as a dative marker and [+animate] accusative marker. We adopt the

assumption that DOM is grammatically an instance of K that selects for a particular type of complement. From the morphosyntactic point of view, the only change from Old Portuguese to SLP is that the DOM morpheme has become a suffix. Since I am uncertain as to what the meaning of a case marker might be, I leave List 3 slots blank:

(26)	a.	List3	List1	List2
		[?]	K, selects +animate	/a/
	b.	List3	List1	List2
		[?]	K, selects +animate NP suffix	/pə/

5.4 *Conclusions*

To sum up, mixed languages confirm that the different components that supposedly conform a word or a morpheme are in fact independent of each other and detachable. Separability is a property that one should expect within a DM understanding of grammar architecture.

A linguistic model that takes as fundamental unit a lexical item in possession of a bundle of inherent semantic, phonological and morphosyntactic features cannot account for mixed languages without some awkward acrobatics. The only way to account for mixed languages data within such a model would be by positing “fission” operations that allow one to break the set of lexical items of a language into its constituent parts. If a lexical item is a bundle of inherent semantic, morphosyntactic and phonological features, the development of ML and SLP can only be the result of creating a new lexicon from bits and pieces of the existing mental lexicons of speakers. This raises the already familiar question of motivation.

Finally, I would argue that the existence of mixed languages provides a parsimony argument for the Integrated Hypothesis: in fact, I would say that their appearance is almost expected if bilingual individuals carry one single system of competence in their heads and not two autonomous systems. Within a separationist architecture, some ad-hoc mechanism has to be built that allows for contact, influence and even amalgamation between the two systems. The issue of explanation comes about again: if a bilingual speaker is in possession of two separate languages, why would these languages happen to be so inextricably mixed? What purpose could such permeability accomplish? Or to turn this onto its head: what is the use of positing a theory (separationism) that needs to be weakened with numerous ad hoc clauses to account for the data?

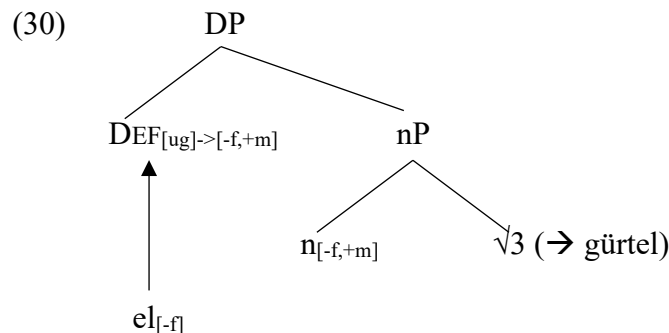
6. **Vocabulary insertion**

Code-switching also provides evidence for assumption (iii), the idea that the Vocabulary Items in List 2 compete for insertion in syntactic terminals following the Subset Principle. As it turns out, the Subset Principle can be used to account for some puzzling restrictions on code switching. Integrationism leads to a clear prediction regarding the Vocabulary Insertion Rules (VIRs): The VIRs should be blind to the “language” they apply to, they should be sensitive only to the feature structures they are meant to match.

González-Vilbazo (2005) discovered that code switching between a determiner and a noun in German-Spanish code-switching reveals some puzzling judgments. If the determiner is Spanish and the noun is German, it is generally acceptable. German is a three-gender language while Spanish is a two-gender language. González-Vilbazo found that when the original German noun was feminine, the Spanish determiner was also feminine; and when the German noun was neuter or masculine, the Spanish determiner was masculine:

- (27) *l-a* Hose
DEF-F pant(f)
- (28) *el* gürtel
DEF.M belt(m)
- (29) *el* brötchen
DEF.M bun(n)

We can assume that in Spanish the feature [feminine] suffices to account for gender in Spanish: so-called masculine nouns are [-feminine]. In German, on the other hand, the feature [-feminine] branches into two categories: [-feminine, +masculine] and [-feminine, - masculine] (the neuter gender). Determiners in both languages have unvalued features that may agree with *n*. This bare description is enough for us to account for the effects of the subset principle: although the Spanish *el* has no value for [\pm masculine], it is the best one available for German neuter and masculine nouns.



Why does *l-* not compete with (and lose out against) German *der*(masculine) or *das*(neuter)? To answer the last question let me note a morphological difference between Spanish and German determiners. In Spanish there are separate terminals for gender and number, and both can be spelled out, obtaining forms like *las* ‘the’ with exponents for gender *a* and number *s*. In German, determiners express either gender or plural but not both. I assume that the German syntactic terminal D has only one terminal for the unvalued Φ -features. If there is a Num head in the structure, the German D agrees with Num and the determiner surfaces as plural. If there is no Num, D agrees with the gender on *n*.

- (31) a. Spanish determiner *l-*: [D D, [ug], [u#]]
b. German determiner *d-*: [D D, [u Φ]]

This suffices to account for lack of competition. The VIR for the Spanish determiner *l-* specifies an environment with separate terminals for gender and number. The German determiner *d-* can

only be inserted in a context in which there is only one Φ -terminal. Since *d-* and *l-* are specified for different contexts, they never compete.

The more puzzling data arises when we have a German determiner and a Spanish noun. González-Vilbazo reports judgments such as the following:

(32)		d-ie	<i>torre</i>
		DEF-F.NOM/ACC	tower(f)
(33)	??	d-er	<i>tenedor</i>
		DEF-M.NOM	fork(-f)
(34)	??	d-as	<i>tenedor</i>
		DEF-N.NOM/ACC	fork(-f)
(35)	??	d-en	<i>tenedor</i>
		DEF-M.ACC	fork(-f)
(36)		d-em	<i>interruptor</i>
		DEF-M/N.DAT	switch(-f)

We take it that the structure now includes a determiner of the form (31b); for simplicity, we take it that *n* and root are Spanish and therefore *n* is either [+feminine] or [-feminine]. (32) is grammatical because D has copied the [+feminine] feature from *n* and there is a *d-* determiner that matches this feature perfectly: *die*.

In (33), (34) and (35) the terminal D only has the feature [-feminine], copied from Spanish *n*[-f], but the vocabulary items *der* and *das* have a feature composition that includes [\pm masculine]: this is the reason why they can't be inserted, and the result is ungrammatical. What about (36)? In the dative case, the distinction between masculine and neuter nouns disappears. We can take the dative exponent to be simply [-feminine] and therefore it matches the [-feminine] feature in D.

The indefinite determiners confirm the analysis:

(37)		ein	<i>interruptor</i>
		INDEF.M/N.NOM	switch(-f)
(38)		ein	<i>interruptor</i>
		INDEF.M/N.ACC	switch(-f)
(39)	??	ein-en	<i>interruptor</i>
		INDEF-M.ACC	switch(-f)

As above, *n* only has the feature [-feminine] and this is what is copied on D. As shown in (37), (38), the exponent *ein* neutralizes the distinction between masculine and neutral determiners: its feature structure only includes [-feminine]. Consequently, it can be inserted in a terminal that has only this gender feature. As for (39), *einen* is specified as [+masculine] but the D terminal only bears the [-feminine] feature, which leads to unacceptability. Let me tie up one loose ending: The German *ein* is unspecified for [\pm masculine] but in the accusative case it is never used with masculine nouns; this is because the form *einen*, which is specified for [+masculine], always beats it to it by the Subset Principle.

To conclude, the code-switching data involving Spanish-German code-switching discovered by González-Vilbazo provides direct evidence of competition among vocabulary items for insertion in syntactic terminals and of the Subset Principle as the acting referee for the

competition. The basic datum is that Spanish *n* has only one gender feature while German exponents of *D* may have one or two; as a consequence, code-switching creates potential clashes that bring to the surface the properties of the VIRs that DM has argued for.

General conclusion

I have shown that the study of bilingual code-switching and of mixed languages provides empirical evidence in favor of three DM assumptions and against lexicalist alternatives. First, code-switching shows that what we normally call a noun or a verb consists of a syntactic structure that includes a root and a categorizer, as initially envisioned in Marantz (1997). Second, the organization of the grammar in three independent lists is supported by the fact that the items from one list can be detached from the rest, as most clearly seen in idiom interpretation and in the two mixed languages discussed here. Finally, the acceptability patterns in German-Spanish code-switching that were carefully documented and analyzed in González-Vilbazo (2005) provide evidence that Vocabulary Insertion and the Subset Principle are alive and well. Along the way, I hope I have convinced the reader that the Integrationist Hypothesis is on the right track and – closer to the goal of this volume - that the Integrated Hypothesis can take off the ground as a genuine theory of bilingualism because of the theoretical framework furnished by DM.

Distributed morphology and bilingualism Bibliography.

Alexiadou, Artemis and Terje Lohndal. 2018. Units of language mixing: A cross-linguistic perspective. *Frontiers in Psychology* Volume 19, Article 1719.

Alexiadou, Artemis, Elena Anagnostopoulou and Florian Schäfer. 2015. *External Arguments in Transitive Alternations*. Oxford: Oxford University Press.

Alexiadou, Artemis, Terje Lohndal, Tor Åfarli and Maren Berg Grimstad. 2015. Language mixing: A Distributed Morphology approach. In Thuy Bui and Denis Ozildiz (eds) *Proceedings of NELS*. Amherst, Mass: GLSA Publications, pp 25-38.

Arad, Maya. 2003. Locality constraints on the interpretation of roots: the case of Hebrew denominal verbs. *Natural Language and Linguistic Theory* 21: 737-778.

Arregi, Karlos and Isabel Oltra-Masuet. 2005. Stress by structure in Spanish. *Linguistic Inquiry* 36: 43-84.

Bartlett, Laura and Kay González-Vilbazo. 2013. The Structure of the Taiwanese DP in Taiwanese-Spanish Bilinguals: Evidence from Code-Switching. *Journal of East Asian Linguistics* 22: 65-99.

Bell, R.T. 1976. *Sociolinguistics: Goals, Approaches and Problems*. Batsford.

Bobaljik, Jonathan. 2000. The ins and outs of contextual allomorphy. In Kleanthes K. Grohmann and Caro Struijke (eds.) *Proceedings of the Maryland Mayfest on Morphology 1999*. Maryland Working Papers in Linguistics 10. College Park, MD: University of Maryland, Department of Linguistics. (pp 35-71).

Canelas Cardoso, Hugo. 2009. *The Indo-Portuguese language of Diu*. Utrecht: LOT.

Clements, Clancy. 2009. *The linguistic legacy of Spanish and Portuguese. Colonial expansion and language change*. Cambridge University Press.

Deibel, Isabel. 2020. *Language representations in the presence of a lexical-functional split: An experimental approach targeting the Quichua-Media Lengua-Spanish interface*. PSU, doctoral dissertation

Delgado, Rodrigo. 2020. *Gender and Phonetics in the Spanish/English code-switched Determiner Phrase*. UIC dissertation

DenDikken, Marcel. 2011. *The distributed morphology of code-switching*. Paper presented at UIC Bilingualism Forum.

Embick, David. 2010. *Localism vs globalism in morphology and phonology*. Cambridge, Mass, MIT Press.

Embick, David. 2015. The morpheme. Berlin: Mouton De Gruyter.

Goldberg, Adele. 1996. Constructions: A construction grammar approach to argument structure. Chicago University Press.

Gómez-Redón, Jorge. 2008. Mestizaje lingüístico en los Andes: genesis y estructura de una lengua mixta. Quito, Ecuador: Ediciones Abya-Yala.

González-Vilbazo, Kay. 2005. Die Syntax des Code-Switching. PhD dissertation, Universität zu Köln.

González-Vilbazo, Kay and Luis López. 2012. Little v and parametric variation. *Natural Language & Linguistic Theory*, 30:1, 33-77.

González-Vilbazo, Kay and Volker Struckmeier. 2008. Asymmetrien im Code-Switching: Eine DM-Lösung zur Partizipselektion. In Guido Mensching and Eva Remberger (eds) *Romanistische Sprachwissenschaft – minimalistisch*. Tübingen: G. Narr Verlag. Pp 83-102.

Grimstad, Maren Berg. 2017. The Code-Switching-Borrowing Debate: Evidence from English-Origin Verbs in American Norwegian. *Lingue e Linguaggio* 16: 3-34.

Grimstad, Maren Berg, Terje Lohndal and Tor A. Åfarli. 2014. Language mixing and exoskeletal theory: A case study of word-internal mixing in American Norwegian. *Nordlyd* 41.2: 213-237, special issue on Features edited by Martin Krämer, Sandra Ronai and Peter Svenonius. University of Tromsø – The Arctic University of Norway.
<http://septentrio.uit.no/index.php/nordlyd>

Halle, Morris. 1997. Distributed Morphology: Impoverishment and Fission. In Benjamin Bruening, Young Kand and Martha McGinnis (eds.), *MITWPL 30: Papers at the Interface*. Cambridge: MITWPL. (pp 425-449).

Halle, Morris and Alec Marantz. 1993. Distributed morphology and the Pieces of inflection. In Ken Hale and S. Jay Keyser (eds.) *The View from Building 20*. Cambridge, MA: MIT Press, pp 111-176.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40:3/4, 225-276.

Hsin, Lisa. 2014. *Integrated Grammatical Architecture: Insights from Syntactic Development*. PhD dissertation, Johns Hopkins University.

Keller, Mareike. 2020. Compositionality: Evidence from Code-Switching. In Elisabeth Piirainen, Natalia Filatkina, Sören Stumpf and Christian Pfeiffer (eds) *Formulaic Language and New Data*. Berlin: De Gruyter. Pp 197-224.

- Koronkiewicz, Bryan. 2014. Pronoun categorization: Evidence from Spanish/English code-switching. UIC dissertation.
- Kramer, Ruth. 2015. The morphosyntax of gender. OUP.
- Lipski, John. 2017. Language Switching Constraints: More than Syntax? Data from Media Lengua. *Bilingualism: Language and Cognition* 20(4): 722-746.
- Lipski, John. 2020a. Pronouns, interrogatives, and (Quichua-Media Lengua) code-switching: The eues have it. *Languages* 5,11.
- Lipski. 2020b. Can a bilingual lexicon be sustained by phonotactics alone? Evidence from Ecuadoran Quichua and Media Lengua. *The Mental Lexicon* 15(2): 330-365.
- López, Luis. 2020. *Bilingual Grammar. Toward and Integrated Model*. CUP.
- López, Luis, Irati De Nicolás, Rodi Laanen, Charlotte Pouw, José Sequeros, María Carmen Parafita Couto. Submitted. *Make my mouth agua: Idioms and the Integrated Lexicon*.
- Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In Alexis Dimitradis et al. (eds.) *Proceedings of the 21st Penn Linguistics Colloquium*. Philadelphia: University of Pennsylvania. Upenn Working Papers in Linguistics, pp 221-225.
- Muysken, Pieter. 1979. La mezcla de Quichua y castellano: el caso de la “media lengua” en el Ecuador. *Lexis* 2: 41-56.
- Muysken, Pieter. 1996. Media Lengua en Ecuador. In S.A. Wurm, P. Mühlhäusler, D.T. Tyron (eds.) *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*. Berlin: Mouton de Gruyter.(pp1335-1337).
- Muysken, Pieter. 1997. Media Lengua. In S.G. Thomason (ed.) *Contact Languages: A wider perspective*. Amsterdam: John Benjamins. (pp101-124).
- Muysken, Peter. 2000. *Bilingual Speech. A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
- Picallo, Carme. 1991. Nominals and Nominalizations in Catalan. *Probus* 3: 279-316.
- Pires, Aline Jessica. 2021. The adequacy of the animacy scale to describe differential object marking in Portuguese. *Estudos linguístico e literários* 72: 366-389.
- Putnam, Michael, Mathew Carlson and David Reitter. 2018. Integrated, not Isolated: Defining Typological Proximity in an Integrated Multilingual Architecture. *Frontiers in Psychology* Volume 8 Article 2212.

- Riksem, Brita. 2017. Language mixing in American Norwegian noun phrases. Dissertation, Norwegian University for Science and Technology, Trondheim, Norway.
- Shappeck, Marco. 2011. Quichua-Spanish language contact in Salcedo, Ecuador: Revisiting Media Lengua syncretic language practices. UIUC, doctoral dissertation.
- Shim, Ji Young. 2016. Mixed Verbs in Code-Switching: The Syntax of Light Verbs. *Languages* 1, 8.
- Shim, Ji Young. 2021. *OV and VO Variation in Code-Switching*. Current Issues in Bilingualism 1. Language Science Press.
- Smith, Ian R. 1977. Sri Lanka Creole Portuguese phonology. Cornell University, doctoral dissertation.
- Smith, Ian R. 1979. Convergence in South Asia: A Creole example. *Lingua* 48: 193-222.
- Stewart. 2019.
- Vaughan-Evans, Awel, Jan R. Kuipers, Guillaume Thierry and Manon W. Jones. 2014. Anomalous transfer of syntax between languages. *The Journal of Cognitive Neuroscience* 34(24): 8333-8335.
- Veenstra, Tonjes and Luis López. 2016. Little v and cross-linguistic variation: Evidence from Code-Switching and Suriname Creoles.” In Bidese, Ermenegildo, Federica Cognola and Manuela Caterina Moroni (eds.) *Theoretical Approaches to Linguistic Variation*, [LA 234]. Pp 317- 336.

