Open Library of Humanities

A syntactic approach to gender assignment in Spanish–English bilingual speech

Abel Cruz, Santa Clara University, US, acruzflores@scu.edu

In this paper I argue for a syntactic analysis to gender assignment in codeswitched speech. To sustain this claim, I examine gender assignment in Spanish Det(erminer)–English noun switches (i.e., *el bishop* 'the.ᴍ') in 76 sociolinguistic interviews of approximately one hour each from a bilingual community in Southern Arizona, U.S. (The CESA Corpus, Carvalho 2012). Based on the findings from this dataset, I demonstrate that the distribution of gender assignment in codeswitched speech poses a serious challenge to current models of the bilingual architecture rooted in the distinct-lexicons perspective (MacSwan 2000 et seq.). Rather, I show that biological gender (interpretable gender) plays a crucial role in the assignment mechanism and the representation of gender features in the bilingual architecture. Taking gender assignment as a case study, I outline a single-lexicon approach to the bilingual grammar compatible with a *Late Insertion* view of the morphosyntactic model (Halle & Marantz 1993). In particular, I highlight the crucial relevance of a theme position at the morphological module to guide the bilingual speaker to the insertion of phonological matrices (language exponents) when codeswitching.

1 Introduction

It is well documented that Spanish–English bilingual speakers assign gender to English nouns occurring in asymmetrical switches whereby the determiner phrase (DP) consists of a Spanish Det(erminer) and an English noun, as in examples (1) and (2):

(1)   a.   era   su   primer clase […] porque ella era   **una junior**
          be-IMP.3SG her first   class […] because she be-IMP.3SG a.**F**   junior
          'it was her first class because she was a junior' (CESA024)

        b.   sí,   tengo   un- pues ahorita   tengo   **un stepson**
          yes, have-PRS.1SG um- well right now have-PRS.1SG a.**M** stepson.
          'yes, right now I have a stepson' (CESA024)

(2)   a.   él   hasta a   veces   lo   saco   sin **la**   **leash**
          'he [a   dog] sometimes CL.ACC take-PST.1SG out without the.**F** leash'
          'sometimes I even take him [the dog] out without the leash' (CESA028)

        b.   y   siempre tenía   **un wand** […] pero él cargaba   **el**   **wand**
          and always have-IMP.3SG a.**M** wand […] but he carry-IMP.3SG the.**M** wand
          'and he always had a wand […] but he would carry the wand' (CESA020)

In (1), the human-denoting nouns *junior* and *stepson* are assigned feminine (F) and masculine (M) gender, respectively. Similarly, the English inanimate nouns *leash* and *wand* in (2) are assigned F and M gender morphologically manifested on the singular Spanish Dets *el* 'the.M' and *la* 'the.F'. The morphological forms of the Dets in (1) and (2) suggest that bilingual speakers are not only uttering lexical entries from both languages, but crucially, they are also applying a binary-gender system when determining the gender feature of English nouns occurring in codeswitched speech. Without gender features involved, the morphological component of the grammar would not know whether to insert the Spanish masculine Det *el* or the feminine Det *la* in (2), and hence the crucial relevance of gender features in the syntax. In other words, the examples in (1) and (2) illustrate the morphosyntax of gender in the bilingual architecture, the subject of study in this paper.

     The ubiquitous manifestation of gender features in Spanish–English codeswitched speech across bilingual communities is of particular interest because Spanish, but not English, has a robust gender system whereby masculine and feminine nouns are distributed approximately equally in the Spanish lexicon, regardless of whether or not a given noun has a human referent: masculine 53%, feminine 47% (Teschner & Russell 1984). Moreover, Spanish exhibits a fine-grained correspondence between a noun's phonemic make up and its grammatical gender whereby nouns ending in the post-stem vowel *-o* are highly associated with masculine gender (i.e., *el libr-o* 'the.M book') (99.8%) and those ending in the post-stem vowel *-a* are strongly

associated with feminine gender (i.e., *la mes-a* 'the.F table') (96.30%) (Teschner & Russell 1984). English, on the other hand, has only a pronominal gender system whereby pronouns agree with the gender of some nouns, and it distinguishes animacy in 3rd person singular pronouns: *he, she, it* (Corbett 1991; Comrie 1999; Curzan 2003). In addition, English has a handful of human-denoting nouns morphologically marked for feminine gender (i.e., *actress, princess, duchess*, etc.) (McConnell-Ginet 2013). Importantly, if pronouns count as elements that agree, then both Spanish and English exhibit grammatical gender but, crucially, only Spanish exhibits a robust gender system whereby every noun must encode gender information for syntactic agreement.

On what basis do bilingual speakers assign gender to English nouns occurring in Spanish–English codeswitched speech? And what is the underlying representation gender features in the bilingual architecture? These are the broader questions that I focus on in this paper. In order to address these questions, I examine the distributional patterns of gender assignment in spontaneously elicited data from a bilingual community that habitually engages in codeswitching practices, namely the data from 76 sociolinguistic interviews documented in the Corpus del Español en el Sur de Arizona (The CESA Corpus, Carvalho 2012). I show that gender assignment in codeswitched speech poses a serious challenge to MacSwan's (2000) (and followers) distinct-lexicons view of the bilingual grammar. Gleaning from the empirical data of the CESA corpus on gender assignment in codeswitched speech, I develop a morphosyntactic analysis of gender in the bilingual grammar on the basis of the linguistic properties that correlate with gender assignment in codeswitched speech.[1] The analysis is framed within a *Late Insertion* view of the morphosyntactic model, viz. Distributed Morphology (henceforth DM) (Halle & Marantz 1993). Altogether, the present study aims to instantiate a syntactic analysis to gender assignment in codeswitched speech rooted in a single-lexicon approach to the bilingual grammar.

The paper is organized as follows. Section 2 provides a brief discussion of the fundamental machinery necessary to instantiate a syntactic approach to gender assignment in the bilingual grammar. Section 3 provides a summary of the previously reported data on gender assignment in Spanish–English codeswitched speech. Section 4 examines the distribution of gender assignment in Spanish–English codeswitched speech in the CESA corpus. Section 5 develops a syntactic analysis to gender assignment in Spanish–English codeswitched speech and provides further evidence for a *Late Insertion* approach to the analysis of bilingual data. Section 6 concludes this paper.

---

[1] I acknowledge that other approaches to language switching postulate different mechanisms to account for the type of data in (1) and (2). For instance, the Matrix Language Frame (MLF) model (Myers-Scotton 1993; Myers-Scotton & Jake 2017) has been widely cited in the literature concerned with language switching at the DP level. In this paper, I adopt the generative framework to language switching and take codeswitching to be an expression of a type of linguistic competence (i.e., González-Vilbazo & López 2011).

2 Theoretical background

The generative enterprise in linguistic theory has identified two essential components for language: (i) a repository of atomic elements (lexical entries) and (ii) rules for combining lexical entries into larger and more hierarchically complex structures. The nature of these components is captured in the so-called Y-model of grammar, according to which the narrow syntax feeds two interfaces: the Phonological Form (PF) and Logical Form (LF) (Chomsky 1995; also Bošković & Lasnik 2007; Hornstein 2018, among others). The analyses of gender assignment discussed in this section are rooted within the Y-model of the grammar.

2.1 Some features of Distributed Morphology

A note on lexical entries is in order here. Since the implementation of the Minimalist Program (MP) (Chomsky 1995), lexical entries are taken to be an array of *features* constructed in the generative lexicon and inserted "on-line" in the course of the derivation. In Chomsky's terms "[a] typical lexical entry consists of phonological matrix and other features, among them the categorial features N, V, and so on; and in the case of Ns, Case and agreement features (person, number, gender), henceforth *ϕ-features*" (Chomsky 1995: 33). In other words, in a lexicalist approach to word formation as envisioned in the MP, it is the inherent features that allow lexical entries to induce relationships with other lexical entries, and ϕ-features are taken to be part of these presyntactic lexical entries, which are in turn treated as the set of ϕ-features in the syntax.

In line with the standard Y-model of grammar, DM offers an alternative view to the nature of lexical entries. In particular, there are two core components that are crucial for the analysis of bilingual data compatible with the tenets of DM: (i) a lexical entry (or word) is decomposed into a category-neutral root (represented with '√' in this paper) devoid of any grammatical information and a categorizing head (i.e., *n, v*), that is, words are built in the syntax; (ii) *Late Insertion* is one of the core tenets of DM that distinguishes it from the lexicalist approach (Harley & Noyer 1999; Arad 2005; Embrick & Noyer 2001; Embick & Marantz 2008; Harley 2014, among many others). In a *Late Insertion* approach, the properties of traditional lexical entries are distributed across separate components of the grammar divided into three lists, each of which relevant to only a subset of the functions of the lexicon in a lexicalist approach to grammar. One list (the so-called *List 1*) contains a list of formatives which enter the syntactic computation including grammatical categories (i.e., *n, v*), functional features (i.e., ϕ-features, [+NEG], [+PAST], [DEF]), and category-neutral roots (√s). These feature bundles are the atomic elements of syntactic computation and are assembled onto hierarchical structures via successive applications of the operation *Merge*. The resulting syntactic representations, which are fully specified with syntactico-semantic features, are then sent off to the interfaces via the operation *Spell-Out*. Crucially, the emerging terminal nodes represent a 'position of exponence' and must receive some phonological content at PF and semantic information at LF (Harley 2014).

On the PF component of the Y-model, a second list (the so-called *List 2*) stores *Vocabulary Items* (phonological matrices). Vocabulary items (henceforth VIs) are correspondence rules for relating syntactic representations (terminal nodes) to phonological representation in context, that is, the pairing of a phonological exponent with grammatical context. VIs are accessed in an operation known as *Vocabulary Insertion*. Since VIs may be conditionalized to the input's local environment (i.e., grammatical context), they compete for insertion according to the Subset Principle (Halle 1997; Embick & Noyer 2007), which holds that a VI is inserted in the 'position of exponence' if all or a subset of the specified features on the terminal node match the VI's features.

On the LF side, a third list (the so-called *List 3*) is accessed. This is the Encyclopedia. Similar to VIs at PF, the Encyclopedia stores a set of correspondence rules for mapping syntactic feature bundles to semantic representations (i.e., information for "special" meanings) in context (Harley 2014). Correspondence rules in the Encyclopedia may be subject to the Subset Principle as with Vocabulary Insertion, but the nature of the Encyclopedia is still unsettled in the literature (see Harley 2014).

2.2 Gender assignment in linguistic theory

The prevailing assumption in linguistic theory is that gender assignment is a lexical representation in the sense that lexical entries (nouns) are assigned gender features via lexical rules in the lexicon (Roca 1989; Harris 1991; Carstens 2000). In particular, Harris developed an account of gender in Spanish and argued that "the formal representation of grammatical gender involves one privative (non-binary) gender mark, … *f*(eminine)" (Harris 1991: 29). Harris formulated lexical rules to enact the *f*-feature to nouns. First, he pointed out that Spanish has a declension class system and exhibits clear tendencies between its genders and its declension classes. However, Harris stressed that gender and declension class are two distinct phenomena. Although Harris's (1991) analysis of gender in Spanish is often cited in the literature on gender assignment in codeswitched speech, there are no formal analyses that have directly applied his lexical rules to the codeswitching facts. On the other hand, Harris's distinction of gender and declension class has been directly applied in formal analyses to gender in codeswitched speech, although in a misguided way as I explain in Section 5 of this paper.

The structural approach, which traces back to Picallo's (1991) analysis of gender in Catalan, offers an alternative to the lexicalist approach to gender assignment. In a structural approach, gender features are assigned to nouns on the basis of linguistic properties such as biological sex, animacy, humanness or phonological and morphological correlates (Kramer 2015; 2020). Kramer (2015; 2016) showed how Picallo's (1991) proposal that gender features project their own phrase in the syntax is unsustainable. As an alternative, in Kramer (2015), she pursued the structural approach to gender assignment and developed a comprehensive approach to gender within a DM

perspective. In particular, Kramer (2015) proposed that gender features are syntactically located on the nominalizer head *n*, which is involved in nominalization. It follows that, in a structural approach to gender assignment, gender information is constructed via syntactic merge of a head *n* with gender features and a category-neutral √ as illustrated in (3).

(3)   *n*P
       ╱   ╲
    √ROOT   n *i*[+/−FEM]
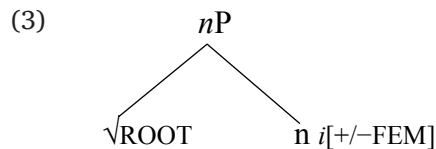
Moreover, Kramer (2015) delimited a clear distinction between biological gender (i.e., sex) and grammatical gender (arbitrary gender). In particular, she proposed that interpretable syntactic gender features [+ / − FEM] capture biological gender whereas an uninterpretable version of the same [+ FEM] feature represents grammatical gender in the syntax for a language like Spanish. In her analysis, only interpretable gender features "trigger some kind of effect at LF (e.g., insertion of a denotation) such that their presence/absence (or change in + / − value) changes the interpretation of a linguistic value" (2015: 58). In other words, a nominalized √ [$n_{[+/-\text{FEM}]}$, √] as in (3) is interpreted as male/female-denoting at LF, and semantic licensing conditions can be used to restrict interpretable gender (see Kramer 2015 for more details). The uninterpretable version of the [+ FEM] feature, on the other hand, is invisible to LF and, consequently, does not contribute denotations or licenses alternations at LF. In short, the property of (un)interpretability is used as the driving force to distinguish biological gender from grammatical gender in Kramer's (2015) comprehensive analysis of gender assignment. In this paper, I follow Kramer's proposal of gender features in the grammar and show how the codeswitching facts provide further support for a structural approach to gender assignment.

2.3 Formal approaches to codeswitching

In this section, I provide a brief summary of formal analyses of codeswitching with particular emphasis to MacSwan's (2000) model of the bilingual grammar. I conclude that a compartmentalized view of the bilingual grammar as impelled by MacSwan's model is empirical untenable and theoretically undesirable (see also López 2020 for similar criticism).

Formal approaches to codeswitching can be divided into two types. First, there are constraint-based models which invoke specific constraints or rules unique to codeswitching (i.e., Poplack 1980; Myers-Scotton 1993; Belazi, Rubin & Toribio 1994). On the other hand, MacSwan instantiated a research agenda of codeswitching in which "all grammatical relations and operations which are relevant to monolingual language are relevant to bilingual language, and only these" (MacSwan 2000: 43). That is, MacSwan's model dissented from previous codeswitching analyses which invoked rules and specific constraints. MacSwan's analysis is grounded in the tenets of

the MP (Chomsky 1995) and has inspired important minimalist analyses of gender assignment in codeswitched speech as I explain in Section 2.3.2.

Since the inception of the DM architecture of the grammar (Halle & Marantz 1993), scholars have also pursued a *Late Insertion* approach to codeswitching. This trend of research examines various bilingual corpora involving not only Spanish and English but also German, Basque, and Norwegian, among other languages (González-Vilbazo & López 2011; Alexiadou et al. 2015; Grimstad et al. 2015; Lillo-Martin et al. 2016; Vergara & López 2017; Grimstad et al. 2018; Burkholder 2018; Riksem 2018; López 2020). The analysis presented in this paper is in line with the DM trend of research. But first I scrutinize the theoretical assumptions of MacSwan's model of the bilingual architecture and discuss the predictions it makes in terms of gender assignment in codeswitched speech.

2.3.1 MacSwan's (2000) minimalist model of codeswitching

MacSwan's (2000) model of the bilingual grammar directly adopted minimalist machinery to explain codeswitching as a linguistic phenomenon. In particular, MacSwan took the following assumption as the point of departure of his minimalist model: "nothing constrains code switching apart from the requirements of the mixed grammars" (2000: 43). MacSwan further stressed that "*constrain* is used in a technical or theoretical sense, and as such implies that there are no statements, rules or principles of grammar which refer to CS" (2005: 5). Rather, MacSwan argued that any speaker, bilingual or not, has only one computational system that generates internal representations and maps these representations into the interfaces independently of the languages involved in the bilingual mind. In particular, MacSwan extolled the lexicalist approach of the MP and argued that "[t]his makes a rather different conception of bilingualism possible, since it is no longer necessary to regard grammars as compartmentalized in some way in the language faculty" (2000: 44). As for the lexicon, MacSwan suggested that "it appears reasonable to assume that bilinguals have distinct lexicons, each with their own internal morphological principles of word formation" (2000: 52).[2] In addition, since the computational system of the grammar is independent of the interfaces, MacSwan also proposed that "[a] bilingual speaker must therefore have separate and discrete phonological systems for each language" (2000: 52). The separate phonologies correspond to the two languages involved in the bilingual grammar.

With these assumptions in place, MacSwan (2000; 2005; 2014) argued that codeswitching falls out of the union of the interacting grammars ($\{G_X \cup G_Y\}$), where G_X a grammar of L_X and G_Y a grammar of L_Y and nothing more (L_X and L_Y represent the languages in the bilingual mind). In MacSwan's terms, "code switching within the syntactic component is possible because *Select* may

---

[2] It should be noted here that the distinct lexicons view of the bilingual grammar was explicitly laid out in Woolford (1983).

draw items from either lexicon, effectively forming the union of the two for the purpose of the derivation" (2000: 52). Given that the internal computation for language is language invariant (i.e., *Merge*), feature-checking constrains codeswitching because the requirements of the switched grammars must be satisfied throughout the derivation. These feature-checking mechanisms, namely *Agree*, operate as proposed in Chomsky's (1995) MP. The bilingual architecture consists of two phonologies, and hence two PFs whereby PF_X can only "process" phonological information from L_X. It follows that word-internal codeswitching at the determiner phrase (DP) is prohibited because heads are presumably the input to phonology. In short, the crux of MacSwan's model of the codeswitching facts is that the bilingual grammar operates on only one computational system, but it must consist of two lexicons and two phonologies (PF_X and PF_Y) as there are two languages (or more) involved in the bilingual mind.

As already pointed out, MacSwan's account of codeswitching predicts that Spanish Det–English noun switches ought to be impossible in Spanish–English codeswitched speech. Needless to say, language switching at the DP level is the most frequent type of codeswitching across Spanish–English bilingual communities in the U.S. (Poplack 1980). In fact, MacSwan's distinct-lexicons proposal makes clear predictions in terms of gender assignment in Spanish–English codeswitched speech.

One possibility is that gender features are assigned to English nouns, which are in turn stored in the English lexicon and retrieved when codeswitching. A second possibility, and more plausible, is that bilingual speakers retrieve gender features from the gendered lexicon when assigning gender to English nouns occurring in codeswitched speech. Although both possibilities are subject to empirical evidence and theoretical scrutiny as I explain in the upcoming sections, their implementation in the grammar poses serious confounds: (i) they allude to unnecessary mechanisms unique to codeswitching, contra MacSwan's own claim that "there are no statements, rules or principles of grammar which refer to CS" (2005: 5); and (ii) they require undesirable postulations about cognitive dimensions of the bilingual mind (i.e., Grosjean 1989; Ortega 2016; Putman et al. 2018).

Moreover, the claim that bilingual speakers have two separate phonologies is equally undesirable. For instance, MacSwan claimed that codeswitching is prohibited with "a lexical head (X^0) whose morphological composition has been determined internally within the lexicon" (2005: 11). In other words, MacSwan argued that codeswitching is prohibited in contexts of head movement. However, Alexiadou & Lohndal 2018 rightly pointed out that MacSwan's (2000; 2005) definition of "word" is based on phonological information, which is problematic because words are not phonetic units as there are phonological rules/constraints that go beyond the word.

Instead of accepting a minimalist model of the bilingual grammar that commits to a distinct-lexicons perspective, as MacSwan (2000 et seq.) did, a fruitful alternative is to adopt a *Late Insertion* approach. The *Late Insertion* approach to gender assignment in codeswitched speech can

be modeled in a single-lexicon view of the bilingual grammar. The validity of this theoretical alternative can be evaluated through a careful empirical analysis of cases of codeswitching that bear relevance to gender assignment in the bilingual grammar (see López 2020 for further empirical evidence). This is precisely what I will do in the remainder of this paper. But first, I briefly review the idea of analogical gender framed within MacSwan's model.

2.3.2 Gender assignment in the distinct-lexicons perspective of the bilingual grammar

There are several minimalist analyses of gender assignment in codeswitched speech (Radford et al. 2007; Cantone & Müller 2008; Liceras et al. 2008; Moro 2014). In particular, Liceras and colleagues have carried out much research on gender assignment in Spanish–English codeswitched speech in both bilingual children (Liceras et al. 2005; Liceras et al. 2016) and adult bilinguals (Liceras et al. 2008). Due to lack of space, this section is limited to Liceras et al. (2008), which developed a formal analysis of gender assignment in codeswitched speech.
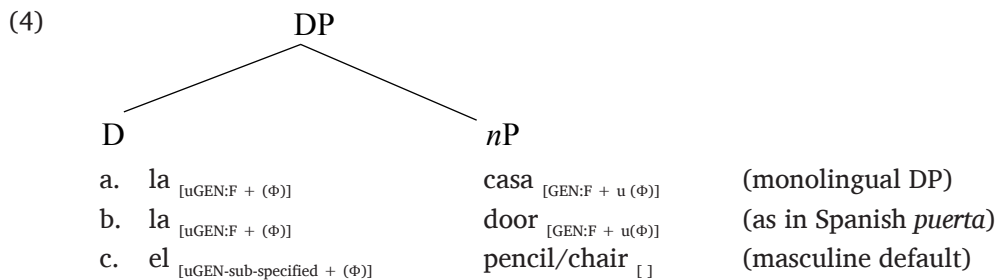
Liceras et al.'s (2008) study revealed that simultaneous bilingual children and adults favored Spanish Det–English noun switches (i.e., *la house*) over English Det–Spanish noun switches (i.e., *the casa* 'house'). In particular, nine Spanish–English bilingual children produced 104 instances of Spanish Det–English noun switches and only 5 instances of English Det–Spanish noun switches in spontaneous speech. In terms of gender assignment, and in agreement with the majority of existing research bearing on this point, it was reported that English nouns occurring in codeswitched speech were overwhelmingly masculine in both child and adult speech. Liceras et al. (2008) also reported experimental data of L2 learners of Spanish of L1 English background ($n = 61$) as well as L2 learners of English of L1 Spanish background ($n = 72$). Unlike simultaneous bilingual children and adults who favored Spanish Det–English noun switches in spontaneous speech, the experimental data revealed that L2 learners of Spanish of L1 English background shared a preference for English Det–Spanish noun switches followed by a strong preference for the masculine default gender. L2 learners of English of L1 Spanish background, on the other hand, retrieved the gender of the Spanish equivalent in an acceptability judgement task, seemingly favoring *analogical gender* whereby English nouns are assigned the gender of their Spanish equivalents.

In short, Liceras and colleagues' studies revealed two relevant findings: (i) in spontaneous speech, simultaneous bilingual children and adults favored masculine default gender in Spanish Det–English noun switches; (ii) in an experimental design, L2 learners of Spanish of L1 English background also adopted a masculine default gender, whereas L2 learners of English of L1 Spanish background retrieved the gender of the Spanish equivalent of the English noun occurring in codeswitched speech as the result of a gender-transfer strategy (analogical gender).

On the basis of these findings, Liceras et al. (2008) developed a formal analysis of gender assignment in codeswitched speech. First, they followed Liceras et al.'s (2005) proposal that the

preference for Spanish Det–English noun switches over English Det–Spanish noun switches can be explained in line with the *Grammatical Features Spell-out Hypothesis* (GFSH). The GFSH holds that "in the process of activating the features of the two grammars, the child, who will rely on the two lexicons, will make codeswitching choices which will favor the functional categories containing the largest array of uninterpretable features" (2005: 227). In particular, Liceras et al. (2005) postulated that the functional head D hosting a Spanish Det in switched DPs has an uninterpretable [GEN] feature. It naturally follows that the GFSH predicts that a Spanish Det ought to be favored over an English Det, which lacks an uninterpretable [GEN] feature.[3]

Liceras et al. (2008) proposed a Double-Feature Valuation mechanism which invokes two syntactic features that must be valued in order to establish agreement (concord) in Spanish DPs as well as in switched DPs: (i) a Gender feature ([GEN]) and (ii) a Gender Agreement feature ([Φ]).[4] More specifically, the [GEN] feature is an inherent functional feature of *n* on D and [Φ] is an inherent functional feature of D on *n*. In order for feature valuation to apply, the authors assumed that gender features are assigned in the lexicon, and hence the head *n* is lexically specified for either F or M gender features. The structure in (4) provides a simplified version of Liceras et al.'s (2008) account of gender in monolingual Spanish and Spanish–English codeswitched speech.

(4)

$$
\begin{array}{c}
\text{DP} \\
\diagup \qquad \diagdown \\
\text{D} \qquad\qquad\qquad n\text{P}
\end{array}
$$

	D		nP	
a.	la $_{[uGEN:F + (Φ)]}$		casa $_{[GEN:F + u(Φ)]}$	(monolingual DP)
b.	la $_{[uGEN:F + (Φ)]}$		door $_{[GEN:F + u(Φ)]}$	(as in Spanish *puerta*)
c.	el $_{[uGEN\text{-sub-specified} + (Φ)]}$		pencil/chair $_{[\ ]}$	(masculine default)

In (4a), the [GEN_] feature is uninterpretable and unvalued, and it is consequently checked/valued against the interpretable [GEN:F] feature located on *n*. Similarly, the uninterpretable agreement feature [Φ] is checked/valued against its interpretable counterpart in (4a). Importantly, note that (4b) alludes to gender transfer in the sense that the bilingual speaker retrieves the gender of the Spanish equivalent of the English noun occurring in codeswitched speech. In other words, the double-valuation mechanism in (4b) not only imposes an interpretable [GEN] feature on *n* but also an Φ-feature on D as the result of a transfer mechanism. A sub-specified version of the

---

[3] It should be noted that Parafita Couto and Stradthagen-Gonzales's experimental data indicated that the preference of a Spanish Det in Spanish Det–English noun switches is an epiphenomenal consequence of "the co-occurrence relation between the source language of a determiner and of the verb in the same clause" (2017: 9). In other words, there is some empirical evidence suggesting that linguistic information other than the phi-features (ϕ) of the D head also influence the selection of the determiner (see also Herring et al. 2010 for a similar conclusion from spontaneous data).

[4] "Functional (syntactic)" in the sense that their interpretative value is only relevant at the level of narrow syntax rather than at the semantic interface (i.e., Liceras et al. 2016).

[GEN] feature on D explains masculine default gender, and thus any English noun can occur with this sub-specified D regardless of the gender of its Spanish equivalent; it is unclear whether the sub-specification of D in (4c) has any consequences for the morphophonological realization of the determiner system. Finally, note that D's [GEN] feature seems to remain unvalued in the syntax in (4c), which is prone to cause a crash at the interfaces. In sum, Liceras et al.'s analysis attempted to explain analogical gender and masculine default gender in codeswitched speech.

Liceras and colleagues' research is unusually rich in that it provides both spontaneous and experimental data from different types of bilinguals. That being said, Liceras et al. do not deliberately address biological sex as a potential variable for gender assignment. As we will see in the next section, biological sex is the only linguistic property that seems to categorically determine gender assignment across bilingual communities, suggesting that it might actually be the source of gender features in the bilingual grammar. Secondly, the implementation of a double-feature valuation to explain analogical gender directly invokes the distinct-lexicons view of the grammar as proposed in MacSwan (2000), namely because their analysis is based on the gender system of Spanish; rather, analogical gender seems to be subject to language dominance in the sense that Spanish dominant late bilinguals may invoke gender transfer. Finally, masculine default is treated as a 'sub-specified' functional D, which lures to a third functional head in the bilingual grammar. It is thus my goal to adjudicate this analysis with a novel analysis of gender assignment in codeswitched speech without invoking two lexicons.

3 Gender assignment in Spanish–English spontaneous codeswitched speech

Spanish Det–English noun switches are the most frequent type of codeswitching across Spanish–English bilingual communities. In this section, I briefly review the spontaneously elicited data on gender assignment in Spanish–English codeswitched speech and highlight the crucial role of biological gender in the assignment mechanism. In a crucial study of codeswitching, Pfaff (1979) reported 747 instances of Spanish Det–English noun switches. Pfaff pointed out that gender assignment to English nouns is conditioned on four factors: biological gender, analogical gender, phonological gender, and morphological gender. Biological gender is reported to be the strongest predictor for gender assignment to English nouns whereby female-denoting nouns are feminine (i.e., *la maid* 'the.F') and male-denoting nouns are masculine (i.e., *el trainer* 'the.M'). Analogical gender refers to gender transfer in the sense that English nouns are assigned the gender of their Spanish equivalents. In the phonological criterion, the phonological properties of the English noun correlate with its gender; for instance, it is indicated that "some nouns in -*er* are classed as feminine" (1979: 305). Finally, the morphological criterion refers to correlates between an English affix and their Spanish counterparts; for example, English nouns ending in -*ity* are assigned feminine gender because of the association with its Spanish counterpart -*idad*

(i.e., *la community* 'the.F'). While these four linguistic properties are claimed to influence gender assignment to English nouns in codeswitched speech, Pfaff (1979) emphasized that biological sex is the only determining factor for gender assignment in her data.

Another important study is reported in Poplack et al. (1982). Poplack and her colleagues were particularly interested in investigating how speakers of Puerto Rican Spanish "assign gender to English nouns when producing them in a host language [Spanish] context" (1982: 9). They analyzed over 300 hours of spontaneous speech recorded from 16 Puerto Rican working-class children and their parents who resided in New York, U.S. Children (age 5) produced a total of 172 Spanish Det–English noun switches, and their parents produced a total of 282 switches. Similar to Pfaff (1979), Poplack et al.'s study revealed that biological gender categorically determined gender assignment with human nouns. As for nouns that lack biological gender, their study showed that analogical gender had a strong effect in predicting gender assignment where 97% of analogically masculine nouns are masculine and 78% of the analogically feminine nouns are in fact feminine in codeswitched speech. Jake, Myers-Scotton, and Gross (2002) reported data from a bilingual community in California. Jake et al. reported 151 instances of Spanish Det–English noun switches. They noted that 78 (or 52%) of the English nouns marked for gender matched the gender of their Spanish equivalents, the resting 73 (or 48%) where defaulted to masculine gender. Of the 78 English nouns matched for gender, 68 (or 82%) were masculine, and thus only 13 (18%) were feminine.

There are other studies that do not corroborate previous findings for analogical gender. I review the studies here in some detail. In an analysis of 33 sociolinguistic interviews of Spanish–English bilingual speakers from New York, Otheguy & Lapidus (2003) found 234 English nouns in codeswitched speech marked for gender. Out of the 234 nouns, 87% are masculine and 13% are feminine. Otheguy and Lapidus also reported that biological gender (i.e., *una teenager* 'a.F') almost categorically determined gender assignment with human nouns. Clegg & Waltermire (2009) also reported 204 instances of Spanish Det–English noun switches in a corpus study of 15 sociolinguistic interviews from a bilingual community in New Mexico. Once again, the data indicated that biological gender categorically determined gender assignment with human nouns and that masculine is the default gender with inanimate nouns. Similar results were reported in Dubord (2004), who analyzed gender assignment in a Southern Arizona, U.S. corpus consisting of 18 sociolinguistic interviews.

More recently, Valdés Kroff (2016) reported 304 instances of Spanish Det–English noun switches in the Bangor Miami corpus (Deuchar et al. 2014). Out of these 304 instances, 296 (92%) are assigned masculine gender while only 8 (3%) are feminine. Importantly, Valdés Kroff reported that biological gender does not seem to be a strong predictor for gender assignment because 8 of the 11 nouns with a female referent were assigned masculine gender, but male nouns were categorically masculine. Perhaps even more striking is the distribution of gender

assignment in the Northern Belize bilingual community documented in Balam (2016), who reported a total of 3,587 Spanish Det–English noun switches marked for gender. Out of the 3,587 switches, 3,573 (or 99.6%) are masculine and only 14 (.4%) nouns are feminine. Importantly, 10 of the 14 feminine nouns reported in Balam's study could be assigned gender on the basis of the sex of their referents. In fact, these 10 nouns with female referents are feminine in codeswitched speech (i.e., *la principal, la nurse, la black woman*). Since only four non-human nouns are feminine in Balam's data (i.e., *la bell, la majority*, and *la half pound)*, this is robust evidence suggesting that masculine is a prevailing default gender.

To summarize so far, the studies reported in this section indicate that biological gender is a reliable predictor for gender assignment in codeswitched speech, although the number of human-denoting nouns is relatively small in spontaneous speech as illustrated in **Table 1**. Moreover, the analogical gender criterion is *not* sustained across bilingual communities, favoring a masculine default gender instead.

Importantly, some of the discordant findings may be related to the fact that the spontaneous bilingual data discussed in this section come from different bilingual communities with varying linguistic experience. For instance, the fact that Poplack et al.'s (1982) study included late bilinguals (i.e., the parents) could explain the preference for analogical gender, which is

Determiner	Male	Female	Total	Corpus & Study
Masculine **Feminine** **Total**	26 (96.30%) 1 (3.70%) 27 (77.15%)	2 (25%) 6 (75%) 8 (22.85%)	28 (80%) 7 (20%) 35 (100%)	Southern Arizona (Dubord 2004)
Masculine **Feminine** **Total**	17 (100%) 0 (0%) 17 (63.30%)	0 (0%) 11 (100%) 11 (36.70%)	17 (60.71%) 11 (39.29%) 28 (100%)	New Mexico (Clegg & Waltermire 2009)
Masculine **Feminine** **Total**	8 (100%) 0 (0%) 8 (42%)	8 (73%) 3 (27%) 11 (58%)	16 (84%) 3 (16%) 19 (100%)	The Bangor Miami Corpus (Valdés Kroff 2016)
Masculine **Feminine** **Total**	29 (100%) 0 (0%) 29 (64.45%)	1 (6.25%) 15 (93.73%) 16 (35.55%)	30 (66.70%) 15 (33.30%) 45 (100%)	New York (Poplack et al. 1982) *el cow* excluded
Masculine **Feminine** **Total**	60 (100%) 0 (0%) 60 (54.06%)	50 (98%) 1(2%) 51(45.94%)	110 (99.10%) 1 (0.90%) 111 (100%)	Northern Belize bilingual corpus (Balam 2016)

Table 1: A cross-tabulation of gender assignment with human-denoting nouns occurring in Spanish–English codeswitched speech.

the only study that provided support for analogical gender. For example, in a recent study Królikowska et al. (2019), cited in Beatty-Martínez & Dussias (2019), studied codeswitching patterns in four different Spanish–English bilingual communities (San Juan, Puerto Rico, El Paso, Texas, State College, Pennsylvania, and Granada, Spain) and reported that only bilinguals in two of these communities (Granada, Spain and El Paso, Texas) exhibited higher rates of feminine determiners in Spanish Det–English noun switches; nevertheless, masculine default was still overwhelmingly preferred across all four bilingual communities. Based on these findings, Beatty-Martínez & Dussias (2019) indicated that bilingual communities can exhibit different rates of codeswitching even when examining the same language pair and suggested that further research that carefully considers the context in which bilinguals use their two (or more) languages is needed. I hope to contribute to these efforts with an investigation of gender assignment in the CESA corpus.

4 Evidence of gender assignment from bilingual community in Southern Arizona

In the remainder of this paper, I examine the linguistic properties that determine gender assignment in the speech of a Spanish–English bilingual community in Southern Arizona, U.S., only 60 miles away from the U.S.–Mexico border. The corpus I analyze here is of particular interest because it comes with rich information about participants' bilingual experience. Moreover, all the participants of this bilingual corpus live and work in a bilingual community that is well documented to engage in codeswitching practices (DuBord 2004; Casillas 2013; Cruz 2016; 2018; Bessett 2017; Kern 2019).

4.1 Research questions

In line with the structural approach to gender assignment (Kramer 2015; 2020), the working hypothesis in the present study is that gender features are assigned on the basis of linguistic properties such as biological sex, analogical gender, and/or default genders, but not lexically specified in the bilingual lexicon. Building on the previous findings reported in Section 3, this study addresses the following research questions (RQs):

(5)   Research questions
    (a) RQ1: What role does biological gender play in assigning gender to human-denoting nouns occurring in Spanish Det–English noun switches in the CESA corpus?
    (b) RQ2: What support does the distribution of gender assignment in the CESA corpus provide for analogical gender versus a masculine default?
    (c) RQ3: To what extent do the results support a non-lexicalist (*Late Insertion*) approach to modeling codeswitching data within the realm of the generative enterprise?

4.2 Methods

4.2.1 Informants

The data for the present study were extracted from the Corpus del Español en el Sur de Arizona (The CESA corpus, Carvalho 2012). The CESA corpus is an on-going research project at the University of Arizona which aims at documenting and disseminating Spanish varieties spoken in Arizona, U.S. (https://cesa.arizona.edu). As of now, the corpus consists of 76 sociolinguistic interviews of approximately one hour each. The interviews were conducted in Spanish or codeswitching by Spanish–English bilingual speakers. Following a typical sociolinguistic interview protocol as discussed in Labov (1972), participants were informally asked about their families, childhood, current issues, memories, and specific questions about language use in their families and their community.

The interviews are transcribed and include a sound file. All 76 interviews were analyzed in the present investigation, but the data were extracted from 53 interviews only. The rest of the interviews were excluded from the analysis because the participants did not produce any instances of Spanish Det–English noun switches or their age of first exposure to English was after age 12. The study then includes 53 Spanish–English bilingual speakers (female = 35) whose age ranges from 22 to 63 years old with a mean age of 30 years old and a median age of 28. All participants completed a bilingual questionnaire, adopted from Birdsong, Gertken, & Amengual (2012). **Table 2** provides an overview of the participants' bilingual language profile.

Table 2 shows that the average age for acquiring Spanish was markedly earlier (M = 1.86 years-old) than for acquiring English (M = 4.2), although onset for English was widely variable as illustrated in **Table 2**. While some of the participants might have understood "starting age" as the time of producing language, the first years of life must be interpreted as "from birth." In fact, 30 of the participants (or 56.60% of the 53 total) reported English and Spanish as their parents' languages, while 21 (or 39.63% of the 53 total) reported Spanish as their parents' language; only 2 (or 3.77%) said English was their parents' language. The participants' language background information indicates that these bilingual speakers were exposed to Spanish and English either simultaneously from birth or very early on (exposed to both Spanish and English inside the home) or sequentially but at a very young age (receiving outside the family exposure to English through daycare and school); and as such, these bilingual speakers can be classified as growing up as either simultaneous or sequential *child* bilinguals (i.e., Ortega 2020). For simplicity, I refer to these bilingual speakers as *early bilinguals*.

Overall, participants reported very good proficiency in both languages. They also use both languages on a weekly basis, although varyingly depending on the context. For instance, as illustrated in **Table 2**, participants reported using Spanish more often than English with their family members and English more than Spanish at work or school. In short, it is safe to say that the CESA corpus represents a bilingual community that uses both languages on a daily basis, that is, balanced Spanish–English bilinguals.

	Mean	SD	Min-Max
Age of acquisition[a]			
Spanish	1.86	1.46	1–6
English	4.2	2.7	1–12
Self-rated proficiency[b]			
Spanish	4.91	0.81	3.5–6
English	5.65	0.53	4.75–6
Self-reported language use[c]			
Spanish			
with friends	36.63	28.03	0–100
with family	64.37	30.15	0–100
at school/work	28.06	22.99	0–100
English			
with friends	68.06	24.96	5–100
with family	42.55	31.16	0–100
at school/work	73.62	21.45	0–100

Table 2: Bilingual language profile overview of the 53 CESA participants.
Note. [a] in years; [b] out of 6 = very good, averaged over speaking, comprehension, writing, and reading; [c] = percent per week.

4.2.2 Data coding

All instances of Spanish Det–English noun switches were extracted from the 53 interviews based on the transcriptions provided in the corpus and verified by the audio files. The coding system adopted here follows Valdés Kroff's (2016) study.[5] In particular, the extracted Spanish Det–English noun switches were coded for Gender of the DET (masculine or feminine), Codeswitch Type (single or multiple-word constituent), Biological Gender (male or female referent), and Analogical Gender (gender of the Spanish equivalent translation). Type of DET includes Spanish masculine (M) and feminine (F) determiners including articles and demonstratives. Codeswitch

---

[5] I am thankful to a reviewer for suggesting the coding system in Valdés Kroff (2016).

Type refers to whether Spanish Det–English noun switches were comprised of a single noun (i.e., *el curfew* 'the.M') or a multiple-word constituent (i.e., *el party dress* 'the.M'). As for biological gender (sex), context was used to determine whether a noun had a male or a female referent. Analogical gender refers to the gender of the Spanish equivalent translation so that *la leash* 'the.F' matches the gender of its Spanish equivalent *la cuerda*. **Table 3** illustrate the coding system adopted in this paper (see Appendix for the entire date set).

Importantly, there is an extra column in the coding system in **Table 3** compared to Valdés Kroff's (2016) study. This is the "Gender in switched DP" column. This column is necessary to analyze the analogical gender variable in which the gender of the Spanish Det–English noun switch must match the gender of its Spanish equivalent DP as in the case of *el curfew* 'the.M'. The DET column in **Table 3** cannot be used to analyze the analogical gender variable because switches like *el Kohl's* 'the.M' do not have a Spanish equivalent translation.

4.2.3 Results

This section provides frequencies and percentages for the variables that concern us here: biological gender, analogical gender, and default genders. But first, I present some examples extracted from the CESA corpus to illustrate the distribution of gender assignment in codeswitched speech in relation to the research questions in (5). The examples in (6) and (7) illustrate feminine and masculine gender assignment, respectively, with human-denoting nouns.

(6)   a.   porque yo era   **una tomboy** cuando era   chiquita
             because I be-IMP.1SG a.**F** tomboy when   be-IMP.1SG small-F
             'because I was a tomboy when I was little' (CESA013)

        b.   mi amiga   era   **una border patrol agent**
             my friend-F.SG be-IMP.3SG a.**F** border patrol agent
             'my friend was a border patrol agent' (CESA024)

(7)   a.   pero queríamos   **el**   **bishop**
             but   want-IMP.2PL the.**M** bishop
             'but we wanted the bishop' (CESA006)

        b.   Cristóbal Colón no es-no descubrió,   fue   **un intruder**
             C. C.   no is -no discover-PAST.3SG is.PAST.3SG a.**M** intruder
             'C. C. did not discover, he was an intruder' (CESA0045)

In (6a), the Spanish adjective *chiquita* 'little-**F**' identifies a female referent for the use of the English noun *tomboy*, and the Spanish noun *amiga* 'friend.F' denotes a female referent for the English noun *agent* in (6b). Interestingly, the nouns *tomboy* and *agent* in (6) are feminine in codeswitched speech as demonstrated through the use of the feminine singular article ***una***. Similarly, the

File	Sample	Token	Comments	Spanish translation	DET	Codeswitch type	Gender in switched DP	Gender in Spanish DP	Biological Gender
CESA001	…	los fields		los campos	M	I	M	M	N
CESA020	…	el wand		la varita mágica	M	I	M	F	N
CESA025	…	la leash		la cuerda	F	I	F	F	N
CESA042	…	el Kohl's	store name	NA	M	I	N	N	N
CESA076	…	the chancla		la chancla	E	I	N	N	N

Table 3: Sample of data coding in Excel Spreadsheet.

English nouns *bishop* and *intruder* in (7) have male referents and are assigned masculine gender in codeswitched speech. Together, the examples in (6) and (7) demonstrate that biological gender determines the grammatical gender of human-denoting nouns in the sense that female nouns are feminine and male nouns are masculine in codeswitched speech.

Although inanimate nouns lack biological gender, a binary gender system is also manifested with these nouns occurring in codeswitched speech as illustrated in (8); these examples have clear Spanish equivalents in the sense that *cash registrar* has a feminine equivalent translation *la caja* and *curfew* has a masculine translation *el toque de queda*.

(8) a. ehh, me pusieron   en **la   cash registrar** ahora
         ehh, CL put-PAST.3PL in the.**F** cash registrar now
         'ehh, they put me in the cash register' (CESA041)

     b. pero también **el   curfew** para mí se lo   quitaron   mis papás
         but also   the.**M** curfew for me CL CL.ACC take-PAST.3PL my parents
         'but the curfew my parents lift it out too' (CESA067)

There are also several examples of codeswitching that clearly demonstrate why a decompositional approach to word formation as posited in DM is more promising for the analysis of gender assignment in codeswitched speech. Consider the examples in (9):

(9) a. Y   él tenía   alto. Y   yo tenía   **el   go**
         and he have-IMP.3SG stop. And I   have-IMP.1SG the.**M** go
         'and he had a red light. And I had a green light' (CESA041)

     b. me   hicieron   **un drop** una   de mis clases
         to-me do-PAST.3PL a.**M** drop one-F of my classes
         'they dropped one of my classes' (CESA062)

The narrative in (9a) is about two drivers at an intersection. Interestingly, the bilingual speaker nominalizes *go* in a context where the action of *go* is more suitable. Similarly, *drop* is nominalized in (9b) in the context of a transitive construction with a light verb meaning 'do.' The verb-like nouns in (9) suggest that their roots syntactically merged with *n*, and not with *v*.

Let us now turn to a quantitative analysis of gender assignment in the CESA corpus. **Table 4** provides the distribution of the Det(erminer) in the CESA corpus.

As we can see in **Table 4**, there is a total of 708 instances of Spanish Det–English noun switches in the CESA corpus. Only two instances (or 0.29%) of English Det–Spanish noun switches (i.e., *the chancla*) were found in the corpus. Moreover, a total of 557 (or 78.45%) English nouns occur with a Det marked for masculine gender and 151 (or 21.26%) with a Det marked for feminine gender. Importantly, it should be noted that the 708 English nouns occurring with a masculine

or feminine Det include social media networks (i.e., *el Facebook*), store names (i.e., *el Kohl's*), and street names (i.e., *la Speedway*). Given that these nouns do not have a Spanish equivalent translation, they were not included for the analysis of analogical gender.

RQ1 asks whether biological gender is a reliable predictor for gender assignment in the CESA corpus. First, we note that there is a total of 50 human-denoting nouns in the CESA corpus, but only seven of these have a clear male referent and 15 a female referent as determined by context. Unfortunately, biological gender cannot be determined with the remaining 28 English nouns, and all of them occur with a masculine Det (see Appendix). In **Table 5**, I cross-tabulated the gender of the Det in codeswitched speech (masculine, feminine) with biological gender (male, female). As we can see, the seven nouns with male referents are categorically masculine in the CESA corpus. In addition, 11 of the nouns with female referents are categorically feminine, but four nouns with female referents are actually masculine. Importantly, the referent of these four nouns is context-dependent (i.e., *el navy seal*), but a female referent can be identified as a potential referent for these nouns given the context where they occur.

The analogical gender variable predicts that English nouns occurring in Spanish Det–English noun switches are assigned the gender of their Spanish equivalents. In order to investigate the analogical gender variable (RQ2) in this study, I cross-tabulated the gender of the Det in Spanish Det–English noun switches with the gender of their Spanish equivalents as illustrated in **Table 6**. There is a total of 466 English nouns that have a clear Spanish equivalent translation

	Determiner total	Percentage
English Det *the*	2	0.29
Masculine Det *el/un*	557	78.45
Feminine Det *la/una*	151	21.26
Total	710	100.00

Table 4: Distribution of Det(erminer) in the CESA corpus.

	Male Referent	Female Referent	Total
Masculine Det *el/un*	7 (100%)	4 (26.67%)	11 (50%)
Feminine Det *la/una*	0 (0%)	11 (73.33%)	11 (50%)
Total	7 (31.82%)	15 (68.18%)	22 (100%)

Table 5: Cross-tabulation of Det by biological gender in the CESA corpus.

	Spanish masculine	Spanish feminine	Total
Masculine Det in codeswitch	221 (99.55%)	145 (66.22%)	366 (83%)
Feminine Det in codeswitch	1 (0.45%)	74 (33.78%)	75 (17%)
Total	222 (50.35%)	219 (49.65%)	441 (100%)

Table 6: Cross-tabulation for analogical gender in the CESA corpus.

(see Appendix); only inanimate nouns are included in the analysis of analogical gender in **Table 6** since these nouns lack any sort of gender information in English.

We can draw several conclusions in terms of gender assignment in the CESA corpus in relation to the analogical gender variable, RQ2. First, there is a total of 221 English nouns marked for masculine gender in codeswitched speech whose Spanish equivalents are categorically masculine; there is only one English noun *las stereotypes* 'the.F.PL' which is feminine in codeswitched speech but has a masculine equivalent translation (*los estereotipos* 'the.M.PL'). As for feminine gender, there is a total of 74 (33.78%) English nouns marked for feminine gender in the CESA corpus. It should be highlighted, however, that 25 instances of these 74 English nouns marked for feminine gender represent the multiple-word constituent *high school*, and there are also six instances of the noun *scholarship* in the present data, all feminine. In short, the number of English nouns marked for feminine gender in the CESA corpus is relatively small (n = 29). With the exception of a few switches (i.e., *la Dream Act* 'the.F' or *la basis* 'the.F'), an interesting observation of these 29 feminine switches in the CESA corpus is that their Spanish equivalents have phonemic make up that strongly correlates with feminine gender in Spanish (i.e., *la leash* 'cuerd**a**', *una city* 'ciuda**d**', *una choice* 'decis**ión**', etc.).

Finally, there is a total of 145 (or 66.22%) English nouns that have feminine Spanish equivalents but are actually marked for masculine gender in the CESA corpus. This finding, combined with the small number of English nouns marked for feminine gender in **Table 6**, indicates that analogical gender is not a viable predictor for gender assignment in the CESA corpus. Put differently, English nouns that lack any sort of gender information are overwhelmingly masculine in the CESA corpus.

4.3 Discussion

The findings of the present study can be summarized as follows. With respect to RQ1, the CESA corpus provides support for biological gender as a reliable predictor for gender assignment in codeswitched speech; this is consistent with all previously reported evidence (Poplack et al. 1982; Otheguy & Lapidus 2003; Clegg & Waltermire 2009) but not in line with the few studies that have found negative evidence for biological gender as a reliable predictor in other bilingual

communities (Balam 2016; Valdés Kroff 2016). As for RQ2, we found no strong evidence for analogical gender, contrary to the few studies that found some evidence for analogical gender in other bilingual communities (Poplack et al. 1982; Jake et al. 2002). Only a handful of English inanimate nouns that have feminine Spanish equivalents whose phonemic make up strongly correlates with feminine gender are feminine in the CESA corpus; this finding corroborates Delgado's (2018) experimental study, which demonstrated that Spanish heritage speakers preferred the singular feminine article *la* as opposed to the singular masculine article *el* with some highly frequent English nouns whose Spanish equivalents are feminine. Instead, masculine gender is a prevailing default gender in the CESA corpus; and this is in line with previously reported data across bilingual communities (Dubord 2004; Aaron 2015; Balam 2016; Valdés Kroff 2016; Królikowska et al. 2019, cited in Beatty-Martínez & Dussias 2019). As for RQ3, the distribution of gender assignment in the CESA corpus can best be accounted for by a *Late Insertion* approach to the grammar. In the next section, I elaborate on this answer and present a theoretical account that best fit the CESA evidence.

5 A syntactic approach to gender assignment in the bilingual grammar

This section delves into the nature of gender features in the Spanish–English bilingual lexicon and develops an analysis to gender assignment in codeswitched speech from a Distributed Morphology perspective. The analysis takes a single-lexicon of the bilingual architecture as cynosure and shows that a *Late Insertion* model of the grammar provides a compelling approach for modeling the codeswitching facts under a single-lexicon view of the bilingual architecture. I share López's position that the linguistic competence of bilingual speakers is "a single cognitive module" whereby the bilingual grammar has "one List 1 and one List 2 of VIRs [Vocabulary Insertion Rules], which may therefore compete for insertion" (2020: 46).

5.1. The nature of roots ($\sqrt{}$) in the bilingual architecture

The nature of $\sqrt{}$s in the bilingual lexicon requires a careful analysis.[6] Here, and in line with López (2020), I follow Harley's (2014) proposal that $\sqrt{}$s have no phonological or semantic content in *List 1* but must be individuated from one another in the syntax. In other words, syntactic $\sqrt{}$s are pure units of structural computation free of semantic and phonological content in the syntax. To distinguish one $\sqrt{}$ from another, I follow Acquaviva (2008), Pfau (2009), and Harley (2014) in assuming that $\sqrt{}$s are identified by non-phonological "index notations" (i.e., $\sqrt{}_{77}$) that serve "as the linkage between a set of instructions for phonological realization in context and a set of instructions for semantic interpretation in context" (Harley 2014: 226).

---

[6] When represented by the root symbol "$\sqrt{}$", the term "root" names a particular theoretical construct which plays an important role in the DM framework (Harley 2014).

In particular, I propose that √s have no language specific information in *List 1*, and each terminal node emerging from the syntax represents a position of exponence that must receive phonological and semantic content post-syntactically at PF and LF, respectively. More specifically, I suggest that formatives stored in *List 1* which enter the syntactic computation—that is, feature bundles including grammatical categories (i.e., *n, v*), functional features (i.e., [+FEM], [+NEG], [+past], [DEF]), and category-neutral √s—are not duplicated in the bilingual lexicon. Instead, a single √ like [√$_{77}$] from *List 1* has two Vocabulary Items for insertion at PF in the bilingual architecture: [√$_{77}$] ↔ <mes-a, table>. Consequently, VIs stored in *List 2* can compete to realize particular terminal nodes emerging from the syntactic derivation as explained Section 5.3. If √s are devoid of phonological content, it follows that the bilingual architecture does not consist of distinct phonologies as proposed in MacSwan (2000). Rather, there is a single PF module in the bilingual architecture (see also López 2020), and the bilingual speaker has more VIs in her *List 2* repertoire to realize a particular emerging terminal from *List 1*.

5.2 Gender features in the bilingual lexicon

With the exception of several nouns (i.e., *persona* 'person.F', *víctima* 'victim.F', etc.), human nouns are sexed in Spanish, and Harris (1991) demonstrated that biological gender categorically determines gender assignment whereby nouns denoting male referents are masculine and those denoting female referents are feminine. Moreover, as discussed in Section 2.2, Kramer (2015) postulated that interpretable [+/−FEM] gender features encode biological gender in the grammar. Here, I suggest that biological gender (interpreatble gender) is the driving force of gender assignment in Spanish–English codeswitched speech. I explain how.

In Sections 3 of this paper, I showed that there is ample evidence indicating that biological gender almost categorically determines gender assignment with human-denoting nouns across bilingual communities, including the CESA corpus examined in Section 4.[7] In addition, there is some evidence indicating that bilingual speakers also associate Spanish adjectives with English human nouns, and the grammatical gender of the Spanish adjective is determined on the basis of the biological gender of the noun's referent as illustrated in (10).

(10) a. The old man está enojad**o.**   (Woolford 1983: 527)
          The old man is   mad-M

      b. I'm not terc**a.**   (Pfaff 1979: 305)
          I'm not stubborn-F

---

[7] I believe studies on gender assignment/agreement in heritage grammars would also benefit from considering the biological gender variable as a potential factor for maintaining a gender system in the heritage grammar (see, for example, Laleko 2018 for Heritage Russian).

In (10b), the Spanish adjective *terca* 'stubborn-F' refers to a female referent identified in the discourse in Pfaff's (1979) study, and the English noun *man* triggers masculine gender in (10a). Furthermore, McConnell-Ginet (2013) rightly pointed out that English has a handful of human nouns which are morphologically marked for feminine gender: *actress, countess, princess, duchess, empress, abbess*, and *poetess*.

In the case of the Spanish–English bilingual lexicon, I propose that there is only one gender feature, namely [FEM], whose (un)interpretability becomes crucial for interpretation at the semantics system and the assignment mechanism invoked in codeswitched speech. In particular, I follow Kramer's proposal that only interpretable gender features "trigger some kind of effect at LF [Logical Form, semantics system] (e.g., insertion of a denotation) such that their presence/absence (or change in $+/-$ value) changes the interpretation of a linguistic value" (2015: 58). Similarly, I suggest that an uninterpretable version of the same [+FEM] feature represents feminine grammatical gender in codeswitched speech for agreement purposes only because, unlike interpretable gender, uninterpretable gender is irrelevant for interpretation at the semantics. Importantly, (un)interpretable gender features form part of *List 1* in the bilingual architecture, meaning that they are not language specific and/or transferred from one lexicon to the other as it is often proposed in the literature on codeswitching. Although [FEM] differs in interpretability, interpretable [+FEM] and uninterpretable [+FEM] are not independent of each other since they are the same kind of feature in the grammar (Kramer 2015; 2020), but only interpretable gender is relevant for meaning and a reliable predictor for gender assignment.

With these assumptions in place, I propose that the affix *-ess* of a noun like *actress* is the exponent of a $n_{[+FEM]}$ bundle in the context of certain English √s as shown in (11).

(11)   *Vocabulary Item for feminine affixes of human nouns in English*
       $[n_{[+FEM]}] \leftrightarrow$ */ess/*   //√ ___ List of √: {poet-, princ-, actr-…}

In (11), a *n* combined with interpretable [+FEM] represents the femaleness of certain English √s for purposes of interpretation at LF. According to the Vocabulary Insertion rule in (11), the $n_{[+FEM]}$ bundle is morphologically exponed as */ess/* at PF with the English √s listed in (11). The logical question at this point is whether interpretable gender represented via the syntactic [+/−FEM] feature is visible to the head D in an English DP, (i.e., the girl). I sustain that it is for the following reasons. First, the syntactic [+/−FEM] feature triggers agreement with pronouns in English, which in turn suggests that English has grammatical gender if pronouns count as elements that agree. Secondly, the syntactic [+/−FEM] feature is invoked in the spell-out form of the Det in Spanish Det–English noun switches with human-denoting nouns as demonstrated throughout this paper. Crucially, note that interpretable gender does not pose a threat to the interfaces because it is legible at LF.

Gender assignment with English nouns that lack biological gender occurring in codeswitched speech raises intriguing questions about the nature of gender features and the assignment mechanism. Analogical gender as discussed in Section 3 is a possible explanation for gender assignment with inanimate nouns in codeswitched speech. However, the results of the CESA corpus combined with previous studies indicate that analogical gender is not a viable predictor for gender assignment in codeswitched speech since feminine gender is limited to a handful of English inanimate nouns, and masculine gender seems to be the default gender rather than the result of analogical gender. Here, I propose that the uninterpretable version of the [+FEM] feature represents feminine analogical gender in codeswitched speech for purpose of syntactic agreement. Importantly, note that the presence of uninterpretable [+FEM] in codeswitched speech does not imply "transferring" feminine gender from the gendered lexicon as proposed in Liceras et al. (2008). In my view, it is an epistemic matter whether a speaker invokes the gender of the Spanish equivalent. If so, I suggest that the concept of analogical gender can be attributed to the Encyclopedia as I explain in the next section—where I also address masculine default gender in codeswitched speech.

5.3 The analysis

Table 7 illustrates the distribution of gender with human-denoting nouns in the Spanish–English bilingual lexicon. First, note that both Spanish and English have different-root nouns with lexico-semantic gender and epicene nouns whose gender is context-dependent. However, only Spanish has post-stem vowels that alternate according to the sex of the noun's referent: *-a* for female referents and *-o* for male referents. The analysis of gender assignment in codeswitched speech proposed in this paper is primarily based on the distribution of gender in **Table 7**.

Let us first see how we can combine human-denoting nouns that have lexico-semantic gender with those that exhibit vowel alternation in **Table 7**. Following the premises of the single-lexicon

	English	**Spanish**
1. Lexico-semantic gender (different-root)	boy, girl, son, daughter, husband, wife, sister, brother, etc.	yerno, nuera, marido, mujer, etc.
2. Context-dependent gender (invariant same-root)	child, teacher, coach, nurse, clerk, engineer, doctor, etc.	estudiante, periodista, artista, rehén, etc.
3. *-o/-a* vowel change (same-root)	NA	chico/a, amigo/a, tío/a, hijo/a, sobrino/a, etc.

**Table 7:** Gender with human-denoting nouns in the Spanish–English bilingual lexicon.

approach outlined in Sections 5.1 and 5.2, I propose that one and the same √ (i.e., $[\sqrt{}_{75}]$) has two phonological matrices stored in *List 2* of the bilingual architecture, whereby √s are paired with a specific type of *n* in the syntax. Consider the examples in (12).

(12)   *Vocabulary Insertion rules for different- and same-root nouns*
      a. $[\sqrt{}_{75}]$ $\leftrightarrow$ */girl, niñ-/*   //__$n_{[+FEM]}$ = female referent
      b. $[\sqrt{}_{76}]$ $\leftrightarrow$ */boy, niñ-/*   //__$n_{[-FEM]}$ = male referent

The examples in (12) illustrate that a √ merged with $n_{[+FEM]}$ is interpreted as female- and one merged with $n_{[-FEM]}$ is interpreted as male-denoting, respectively, at LF. Consequently, the emerging √ combined with a type of *n* represents a position of exponence and must receive phonological content at PF via contextual restrictions.

According to the Vocabulary Insertion rules in (12), phonological matrices are inserted in the position of a syntactic terminal in the context of a bundled *n*. Given the VIs available for exponence in (12), I must say something about how language exponents from *List 2* can compete for insertion at PF when codeswitching. I suggest that morphophonological information guides the bilingual speaker to the phonological realization of a target terminal emerging from the syntax. This information must be part of the linguistic competence of the bilingual speaker since Spanish and English exhibit different morphophonological properties. For instance, Oltra-Massuet & Arregi (2005) showed that Spanish has a morphological well-formedness condition which holds that all syntactic functional heads require a theme node (TH) position at the morphological module in the derivation. In the case of nouns, a TH node is post-syntactically adjoined to the nominalizer *n* and corresponds to nominal declension classes (I, II, III) whose realization depends on three theme vowels (*-o, -a, -e/Ø*).

When codeswitching, I propose that a Spanish phonological matrix like /niñ-/ in (12) is inserted in the context of a *n* hosting a TH position as illustrated in (13).

(13)
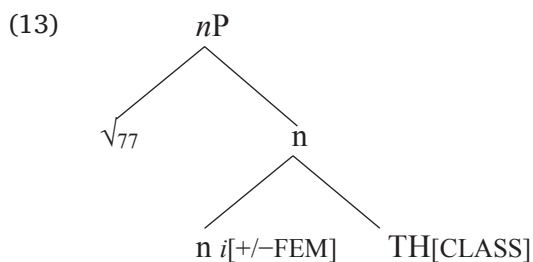
In (13), the TH node is purely morphological and hosts class features (I, II, III) which are in turn morphologically realized as theme vowels via Vocabulary Insertion (see Oltra-Massuet & Arregi 2005; Bermúdez-Otero 2013; Kramer 2015). But how can the TH position guide the competition of phonological matrices from *List 2* when codeswitching?

The TH position is of crucial relevance for the computation of stress in Spanish: it determines stress placement with verbs and non-verbal environments (Oltra-Massuet & Arregi 2005). In other words, the TH position adjoined to *n* in (13) requires that the phonological matrix inserted in such context obeys the phonotactics of Spanish in order for a Spanish word to be constructed. The phonological matrices /girl/ and /boy/ in (12) cannot be inserted in (13) because they do not obey the requirements imposed by the TH position adjoined to *n*. Again, the reference to the internal syntactic structure of nouns as illustrated in (13) is part of the bilingual's linguistic competence. Later, in Section 5.4, I will provide further support on the crucial relevance of a TH position in the bilingual architecture.

Adopting a *probe-goal* relation for agreement (Chomsky 2000; 2001), I assume that the head D enters the derivation with an unvalued gender feature ([GEN:_]) in codeswitched speech. The unvalued feature *probes* (dotted line) in (14) and finds a valued (and interpretable) gender feature on the head *n* in both instances in (14); gender and number are independent probes in (14).

(14)   a.

    b.
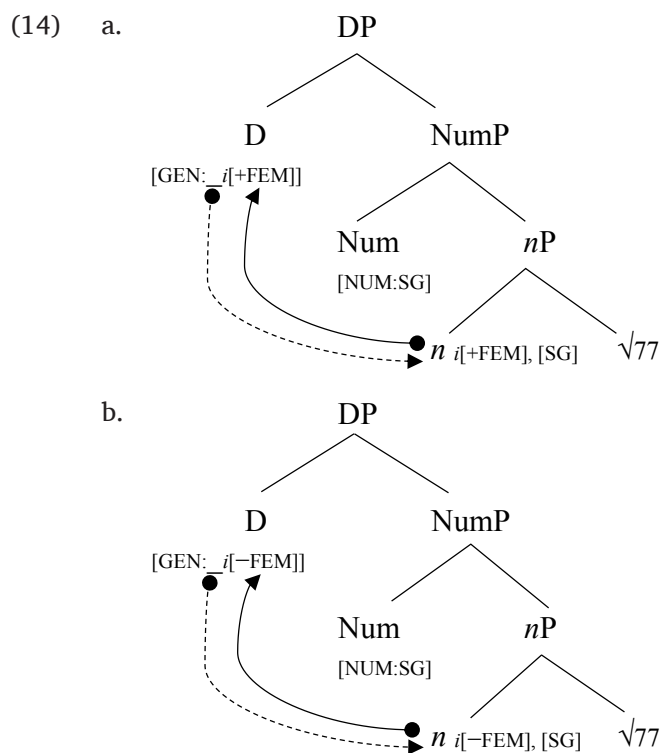
Valuation occurs via Agree: the *probe* acquires the [+FEM] feature in (14a) and the [−FEM] in (14b) (solid line). It follows that a valued [+/−FEM] feature is acquired by the bundle head D for morphological realization at PF. An interpretable [+/−FEM] gender feature does not pose a threat at the interfaces, and I speculate that the unvalued [GEN:_] feature of the head D is invisible when gender is irrelevant for syntactic agreement (i.e., in the context of English mode).

Phonology	Feature specification
el	D, [DEF], [−PL]
los	D, [DEF], [PL]
la	D, [DEF], [+FEM], [−PL]
las	D, [DEF], [+FEM], [PL]
the	D, [DEF]
a	D,
un	D, [−PL]
unos	D, [PL]
una	D, [+FEM]
unas	D, [+FEM], [PL]

Table 8: Vocabulary Items for the Determiner System in Spanish–English bilingual lexicon.

Vocabulary Items for the Det system in the Spanish–English bilingual lexicon are presented in **Table 8**. The Spanish Det system has a contrastive distinction for definiteness, number, and gender. I assume that the bilingual speaker has access to these contrastive features.

In the analysis in (14), an interpretable [+/−FEM] feature is acquired by the head D as the result of a *probe-goal* relation. Now consider the VIs of the Det system in (15).

(15) *Vocabulary items for definite Dets in Spanish–English codeswitched speech*
      a. [D], [DEF], [+FEM], [−PL] ↔ **la** *junior*
      b. [D], [DEF], [−PL] ↔ **el** *stepson*

Vocabulary Insertion in (15) obeys the Subset Principle (Halle 1997; also Embick & Noyer 2007), which holds that given the set of VIs competing for insertion into a given terminal with a set of morphosyntactic features, only the one with more features in common with the target node without having any more will win the competition and be inserted. Note that the VI for the feminine Det *la* in (15a) contains a [+FEM] morphosyntactic feature, and the head D emerging from the syntax in (14a) has the same feature inducing the VI in (15a). The head D with a [−FEM] feature is realized with *el* in (15b) because *la* in (15a) has a non-matching feature; in other words, the article *el* is the underspecified form (see also Riksem 2018 for the application of the Subset Principle to language mixing).

Although inanimate nouns lack biological gender, English inanimate nouns are also assigned grammatical gender in codeswitched speech as illustrated in (16).

(16) a. él   hasta a   veces lo   saco sin   **la   leash**
          'he [a dog] sometimes CL   take-PST.1SG out   without the.**F** leash'
          'sometimes I even take him out without the leash' (CESA028)

  b. pues yo voy   a entrar en **el   army**
      well I   go-PST.1SG to join-INF in the.**M** army
      'well, I will join the army' (CESA007)

The analysis of the CESA corpus reported in Section 4 revealed that English inanimate nouns are overwhelmingly masculine (16b) in codeswitched speech. We also saw that only a handful (n = 29) of inanimate nouns whose Spanish equivalents are feminine (16a) and exhibit phonemic make up that strongly correlates with feminine gender are feminine in the CESA corpus.

As pointed out in Section 5.2, I propose that one and the same feature, namely [FEM], accounts for gender assignment with both human nouns and unsexed entities in the Spanish–English bilingual lexicon. In particular, I suggest that an uninterpretable version of the [+ FEM] feature captures feminine grammatical gender with inanimate nouns. An important question at this point is how do we prevent interpretable gender from merging with non-human roots. In DM, Encyclopedia entries are pairings of √s and categorizing heads (Arad 2005). This opens the possibility that some licensing condition invoked at the Encyclopedia ensures that human-denoting √s are only interpretable in the context of a *n* with interpretable gender. Again, this is possible because √s combine with different flavors of *n*s, so that √s denoting unsexed entities are prohibited with interpretable gender at the Encyclopedia because interpretable [+ / − FEM] gender features are incapable of finding a referent in the world for interpretation. Similarly, the Encyclopedia also ensures that the √ for a human noun like *mother* is incompatible with a *n* with interpretable [− FEM]. This analysis captures the fact that grammatical feminine gender is restricted to a handful of English nouns in codeswitched speech, presumably because gender agreement is easier with frequent English nouns that have feminine equivalents and there is no need to adopt a default gender. Future work is needed in this direction, but attributing the (un)interpretability of [+ FEM] to a licensing condition applied at the Encyclopedia is suitable with the single lexicon approach of the bilingual architecture.

A a *probe-goal* relation as in (14) also applies in the context of *n* with uninterpretable [+ FEM], and the bundle D hosting the English noun in (16) spells out as *la* with inanimate nouns like *la leash* according to the rules in (15a). In other words, the same morphological realization applies for the (un)interpretable [+ FEM] feature.

English inanimate nouns are overwhelmingly masculine in the CESA corpus. In fact, masculine gender is a prevailing default gender across Spanish–English bilingual communities as

reported in Section 3. Masculine default gender must be amenable in the bilingual architecture. In particular, I suggest that these inanimate nouns lack gender features altogether. In other words, their √s combine with a *n* lacking gender features, a plain *n.* In this case, the unvalued [GEN:_] of the head D in (14a) does not find its valued counterpart because the potential goal (plain *n*) lacks gender features. I assume that Preminger's (2009) *failed Agree* applies in this case: the agreement operation applies as normal, but D's unvalued feature is 'reset' to a default value as part of sending the syntactic output to the interfaces. It follows that the bundle D arrives to PF with an unvalued feature (failed probe), which is in turn incompatible with *la* in (15a), which has a valued [+FEM] feature. Therefore, the underspecified form *el* of the Det system in (15b) is inserted with a failed probe.

A note on previous accounts to gender assignment in codeswitched speech framed within *Late Insertion* is in order here. At first glance, the analysis in (13) seems to resemble Burkholder's (2018) proposal of gender assignment in Spanish Det–English noun switches, but it is fundamentally different. Burkholder (2018) argued that a Spanish–English bilingual speaker has two *List 1*s (and two *List 2*s) in the (duplicated) bilingual lexicon, and that a binary uninterpretable [±f] feature represents gender in codeswitched speech, namely because her analysis is based on grammatical gender in Spanish, that is, biological gender is not considered. Burkholder further proposed that a bilingual speaker can merge a √ like '√house' in *la house* with a Spanish *n* hosting uninterpretable [+f], or uninterpretable [−f] for '√book' in *el book.* It follows that a Spanish $n_{[+f]}$ taken from the Spanish lexicon combined with '√house' can trigger a theme node (TH) at the morphological module, but such TH position must be realized with a null element /Ø/ at PF with a √ like '√house' because English lacks theme vowels. In other words, Burkholder collapsed gender and declension class in the sense that the presence of the [±f] feature (the result of gender transfer) triggers a TH position at the morphological module, imposing gender and declension class information onto English nouns in codeswitched speech.

López (2020) analyzed gender assignment as a key feature of his integrated model of the bilingual architecture. He postulated a [±f] gender feature in the bilingual lexicon. Assuming that √s are devoid of any grammatical information and phonological content as proposed in this paper, López (2020) suggested that a single √ represents different phonological matrices in *List 2*. As for gender assignment in codeswitched speech as in *la key,* López suggested that the English noun *key* can spell out the complement of $n_{[+f]}$ in analogy to its Spanish feminine counterpart *llave.* In other words, the Spanish–English bilingual speaker has two phonological matrixes for spelling out of the concept *key,* {/key/, /llav-/}, as the complement of a single $n_{[+F]}$. López also proposed that word markers (i.e., /a/, /e/, /Ø/, etc.) spell out $n_{[+F]}$, so that /e/ spells out $n_{[+F]}$ in *la llave* and the null element /Ø/ spells out $n_{[+F]}$ for the codeswitch *la key.*

Although they take different positions on the nature of a theme (TH) position, Burkholder (2018) and López (2020) proposed a binary [±f] feature to account for gender assignment

in codeswitched speech.[8] However, their analyses do not capture the distribution of gender assignment in codeswitched speech; that is, the empirical fact that biological gender almost categorically determines gender assignment, whereas feminine grammatical gender is restricted in codeswitched speech as demonstrated in this paper, and first pointed out in Delgado (2018). Moreover, if the bilingual speaker has two phonological matrixes for a single √ as proposed in López (2020), it is unclear how the insertion of language exponents is accomplished at PF when codeswitching, and Burkholder's proposal that an English noun occurring in codeswitched speech triggers a TH position is unwarranted.

In this paper, I outlined an alternative analysis to gender assignment in bilingual speech. The analysis can be summarized as follows: (i) gender information is constructed from combining a category-neutral √ and the nominalizer *n* in the syntax, whereby gender features are located on *n*; (ii) the bilingual architecture consists of a single *List 1*, a single *List 2*, and a single *List 3* and the Encyclopedia can handle issues of feature interpretability; (iii) the single (un)interpretable feature [FEM] captures gender assignment with human and non-human referent nouns, and biological gender (interpretable gender) plays a crucial role in the assignment mechanism, in line with typological data demonstrating that biological gender and grammatical gender are not independent of each across the world languages (Kramer 2020); (iv) the competition of phonological matrixes from *List 2* when codeswitching makes crucial reference to the internal syntactic structure of words, namely the TH position; (v) gender and declension class are distinct phenomena, but they are in close interaction as first noted in Harris (1991). In the next section, I say more about this last point.

5.4 Further support for the crucial role of the theme node position at PF

I believe there is further evidence indicating that the analysis to gender assignment in codeswitched speech developed in this paper is on the right track. More specifically, there is a handful of novel nouns that are fully integrated in the bilingual lexicon obeying the morphosyntax of Spanish. They are illustrated in **Table 9** (see Barkin 1980; Poplack et al. 1982; Smead 2000). The novel nouns *troca* and *breca* in **Table 9**, whose etymology traces to English *truck* and *break*,

---

[8] A reviewer points out that López (2020) found instances of English Det–Spanish noun switches (i.e., *the harina* 'flour') in natural speech and highlights that the presence of such codeswitches is one of the reasons for the proposal of a binary [±f] feature in López's (2020) account of gender assignment in codeswitched speech. The same reviewer asks how my analysis can explain codeswitches like *the harina*. I do not believe a binary [±f] feature is necessary to explain such codeswitches. In fact, López emphasized that "[t]he acceptability of *the mesa* suggests that a $n_{[+f]}$ can be selected by *a featureless English determiner* [my emphasis]" (2020: 94). It follows that gender information is irrelevant for competition of the Det system when "a featureless English determiner" is invoked in codeswitching. In my view, the uninterpretable [+FEM] can be invoked for the presence of the TH position spelled out as /a/ in *the harina*, /o/ is the default word marker. This is in contrast to López (2020) who suggested that the categorizer *n* determines externalization of the root √*mes* in *the mesa*. Again, the TH position is needed for computing stress in my analysis, whereas López attributes any Spanish phonotactic properties to $n_{[+f]}$, or a 'Spanish-like n.'

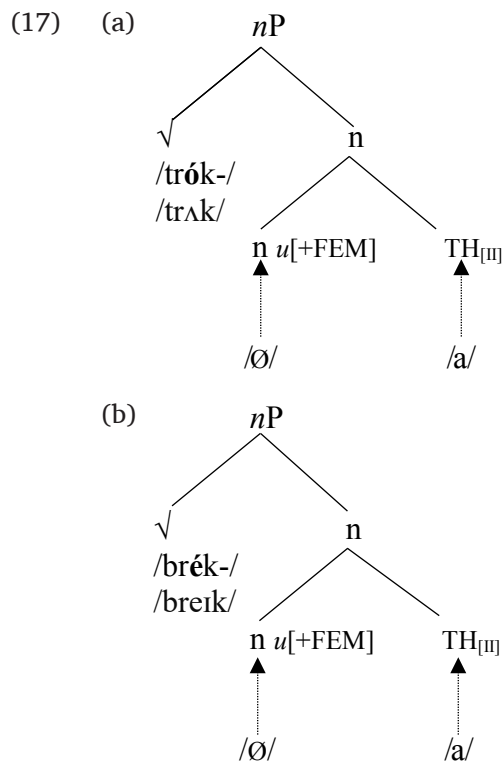|	*Closest Spanish equivalent*	*English Etymology*	*Meaning*
la troca	camión.M/camioneta.F	truck	a loading pick up
la traila	caravana.F	trailer	home-like trailer
la norsa	enfermera.F	nurse	female nurse
la baica	bici.F	bike	bicycle
la suera	suéter.M	sweater	piece of clothing
la breca	freno.M	break	car break
la ploga	enchufe.M	plug	material used to stop an aperture
la estofa	cosa.F	stuff	a thing that is not specified
la tena	inquilino/a.F/M	tenant	a person who rents a property
la lonchera	NA	lunch	lunch pail
la jira	Radiador.M	heater	heater used to warm up a home

Table 9: Word class and gender of loanwords incorporated in the bilingual lexicon.

respectively, are well-cited examples that serve as a way to illustrate the role of the TH position in the incorporation of such novel nouns in the bilingual lexicon.

Note that the nouns in **Table 9** are feminine and end in the post-stem vowel *-a*, the marked form of the *inner core* in Harris's (1991) analysis of word classes in Spanish. In the analysis developed in Section 5.3, I proposed that a √ representing an English noun like *table* and its Spanish equivalent *mesa* can be merged with an uninterpretable [+FEM] feature in the syntax for syntactic agreement purposes in codeswitched speech. I also suggested that a TH node is adjoined to the nominalizer *n* at the morphological module when Spanish phonotactics are required at PF, namely for computing stress in Spanish. The data in **Table 9** aim to demonstrate the crucial role the TH node, and the morphological module more generally, plays in the bilingual architecture.

In particular, I speculate that the nouns in **Table 9** underwent a phonological change in order to be incorporated into the bilingual lexicon so that they can function as independent phonological matrices in *List 2* of the bilingual architecture. This is evident by the fact that the presence of the TH node with these nouns induced stress alignment so that these nouns can obey the phonotactics of Spanish as schematized in (17).

(17)   (a)

```
              nP
            /    \
          √        n
       /trók-/    /  \
       /trʌk/    /    \
              n u[+FEM]   TH[II]
                 ↑          ↑
                /Ø/        /a/
```

(b)

```
              nP
            /    \
          √        n
       /brék-/    /  \
       /breɪk/   /    \
              n u[+FEM]   TH[II]
                 ↑          ↑
                /Ø/        /a/
```

In (17), we can see that *truck* and *break* underwent a vocalic change, and the presence of the TH node, which is in turn spelled-out with the word marker /a/, shifted the stress pattern to the vowel preceding the TH node: /trók-a/ and /brék-a/. Again, the TH position plays a crucial role for computing stress in Spanish and for the competition of language exponents when codeswitching. In this way, the phonological matrix <trók-> competes for insertion with its counterpart <trʌk>, the former being inserted in the context of a *n* bearing a TH node.[9]

Note again that the novel nouns in **Table 9** exhibit the marked form of Harris's (1991) *inner core* of word markers in Spanish, namely -*a*. Moreover, the Spanish equivalents of *break* and *truck* are actually masculine, and so analogical gender does not apply in the case *breca* 'break' and *troca* 'truck'. The data in **Table 9** then suggest a tight relation between the marked form -*a* of the *inner core* and the gender feature [+FEM] in the bilingual grammar in the sense that all the feminine nouns in **Table 9** exhibit the word marker -*a*. As far as I am aware of, there is no evidence indicating that English nouns incorporated in the bilingual lexicon trigger the default word marker -*o* even when these nouns are overwhelmingly defaulted to masculine gender (i.e.,

---

[9] López (2020) also refers to the novel noun *troca* as evidence for his analysis of gender, but he collapses gender and declension class, with the word marker -***a*** being the spell out of *n* in his analysis.

el club, el bar, el swing, etc.).[10] In short, the analysis in (17) points to the tight relation between the syntax and the morphological module of the bilingual grammar as predicted by the syntactic analysis to gender assignment outlined in this paper. I hope future studies will explore such relation of the modules in connection to the codeswitching facts (see, for example, Grimstad 2017 for a *Late Insertion* approach to English-origin verbs in American Norwegian). In this paper, I attempted to provide an analysis to gender assignment in codeswitched speech within a DM perspective, but gender is only one of many linguistic phenomena that are currently being debated in the literature (see López 2020 for a relevant discussion).

6 Conclusion

In this paper, I investigated the distribution of gender assignment in Spanish–English codeswitched speech and argued for a syntactic analysis to gender assignment in line with Kramer's (2015) proposal of gender features in the grammar. New evidence on gender assignment stemming from a bilingual community in Southern Arizona, U.S. documented in the CESA corpus (Carvalho 2012) was provided. In particular, I showed that the distribution of gender assignment in the CESA corpus poses a serious challenge to MacSwan's (2000) distinct-lexicons perspective of the bilingual grammar. While the distinct-lexicons perspective has been the standard assumption in formal approaches to bilingualism, the empirical evidence emerging from carefully designed psycholinguistic studies on codeswitching, and bilingualism more generally, renders such a view increasingly implausible (Kroll et al. 2015). Instead, in this paper I developed a syntactic analysis to gender assignment in the bilingual grammar on the basis of the linguistic properties that correlate with gender assignment in codeswitched speech. Contra the lexicalist view of gender assignment, in the analysis proposed here gender information is constructed from combining a category-neutral $\sqrt{} + n$ and interpretable gender is the primary source of gender information in codeswitched speech. Finally, I outlined a single-lexicon approach to the bilingual grammar compatible with a *Late Insertion* view of the morphosyntactic model.

---

[10] In fact, the feminine version of *el bar* was also reanalyzed for stress in Spanish, and it triggered a TH node which is spelled out with the marked word class *-a*: *la barra* 'the.**F** bar.'

Abbreviations

Abbreviations follow the Leipzig glossing rules unless otherwise noted.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** Gender assignment in the CESA Corpus. DOI: https://doi.org/10.16995/glossa.5878.s1

Acknowledgements

Competing Interests

The author has no competing interests to declare.

References

Acquaviva, Paolo. 2008. Roots and lexicality in Distributed Morphology. *York Papers in Linguistics Series* 10. 1 − 21.

Alexiadou, Artemis & Lohndal, Terje. 2018. Units of language mixing: A cross-linguistic perspective. *Frontiers in Psychology* 9. 1–15. DOI: https://doi.org/10.3389/fpsyg.2018.01719

Alexiadou, Artemis & Lohndal, Terje & Åfarli, Tor & Grimstad, Maren B. 2015. Language mixing: A Distributed Morphology approach. In Bui, Thuy & Özyildiz, Deniz (eds.), *Proceedings of NELS* 45. 25 − 38. Amherst, MA: University of Massachusetts, Graduate Linguistic Student Association.

Arad, Maya. 2005. *Roots and Patterns: Hebrew Morpho-syntax*. Dordrecht: Springer. DOI: https://doi.org/10.1007/1-4020-3244-7

Balam, Osmer. 2016. Semantic categories and gender assignment in contact Spanish: Type of code-switching and its relevance to linguistic outcomes. *Journal of Language Contact* 9. 405–435. DOI: https://doi.org/10.1163/19552629-00903001

Barkin, Florence. 1980. The role of loanwords assimilation in gender assignment. *Bilingual Review/Revista bilingüe* 2. 105 − 112. http://search.proquest.com/docview/1297886305/.

Beatty-Martínez, Anne & Dussias, Paola. 2019. Revisiting Masculine and Feminine Grammatical Gender in Spanish: Linguistic, Psycholinguistic, and Neurolinguistic Evidence. *Frontiers in Psychology* 10(751). 1 − 11. DOI: https://doi.org/10.3389/fpsyg.2019.00751

Belazi, Hedi & Rubin, Edward & Toribio, Almeida. 1994. Code-switching and X-bar theory: the functional head constraint. *Linguistic Inquiry* 25(2). 221–237.

Bermudez-Otero, Ricardo. 2013. The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features. *Probus* 25(1). 3–103. DOI: https://doi.org/10.1515/probus-2013-0009

Bessett, Ryan. 2017. Exploring the phonological integration of lone other-language nouns in the Spanish of Southern Arizona. *U. Penn Working Paper in Linguistics* 23(2). 31−39.

Birdsong, David & Gertken, Libby M. & Amengual, Mark. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL; University of Texas at Austin.

Bošković, Zeljko & Lasnik, Howard. 2007. *Minimalist syntax: the essential readings*. Malden, MA: Blackwell Pub.

Burkholder, Michèle. 2018. Language Mixing in the Nominal Phrase: Implications of a Distributed Morphology perspective. *Languages* 3. 1−10. DOI: https://doi.org/10.3390/languages3020010

Carstens, Vicki. 2000. Concord in minimalist theory. *Linguistic Inquiry* 31. 319−355. DOI: https://doi.org/10.1162/002438900554370

Carvalho, Ana. 2012. Corpus del Español en el Sur de Arizona (CESA); University of Arizona. https://cesa.arizona.edu.

Casillas, Joseph. 2013. La fricativización del africado /tʃ/: actitudes lingüísticas cerca de la frontera. In Carvalho, Ana M. & Beaudrie, Sarah (eds.), *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*, 177–188. Summerville, MA: Cascadilla Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries. In Martin, Roger & Michaels, David & Uriagereka, Juan & Keyser, Samuel Jay (eds.), *Step by step: Studies in honor of Howard Lasnik*, 89−155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Kenstowicz, Michael (ed.), *Ken Hale: A Life in Language*, 1–52. Cambridge, MA: MIT Press.

Comrie, Bernard. 1999. Grammatical gender systems: a linguist's assessment. *Journal of Psycholinguist Research* 28. 457–466. DOI: https://doi.org/10.1023/A:1023212225540

Corbett, Greville. 1991. *Gender*. Cambridge, UK: Cambridge University Press.

Clegg, Jens & Waltermire, Mark. 2009. Gender assignment to English-origin nouns in the Spanish of the Southern United States. *Southwest Journal of Linguistics* 28(1). 1–17. http://search.proquest.com/docview/85708379/.

Cruz, Abel. 2016. The Spanish discourse marker *o sea* in the speech of bilinguals from Southern Arizona. *Revista de Estudios Lingüísticos y Literarios* 14(1). 70–81.

Cruz, Abel. 2018. The past persists into the present: a multivariate analysis of Present Perfect in Southern Arizona Spanish. In MacDonald, Jonathan (ed.), *Contemporary Trends in Hispanic and Lusophone Linguistics: Selected Papers from the Hispanic Linguistics Symposium*, 169–189. John Benjamins. DOI: https://doi.org/10.1075/ihll.15.09cru

Curzan, Anne. 2003. *Gender shifts in the history of English*. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486913

Delgado, Rodrigo. 2018. The familiar and the strange: Gender assignment in Spanish/English mixed DPs. In López, Luis (ed.), *Code-switching–Experimental Answers to Theoretical Questions: In honor of Kay González-Vilbazo*, 39–61. Amsterdam: John Benjamins Pub. DOI: https://doi.org/10.1075/ihll.19.03del

Deuchar, Margaret & Davies, Peredur & Herring, Jon & Parafita Couto, M. Carmen & Carter, Diana. 2014. Building bilingual corpora. In E. Thomas & I. Mennen (eds.), *Advances in the study of bilingualism*, 93–110. Bristol: Multilingualism Matters. DOI: https://doi.org/10.21832/9781783091713-008

DuBord, Elise. 2004. Gender assignment to English words in the Spanish of Southern Arizona. *Divergencias. Revista de estudios lingüísticos y literarios* 2(2). 27–40.

Embrick, David & Marantz, Alec. 2008. Architecture and blocking. *Linguistic Inquiry* 39. 1–53. DOI: https://doi.org/10.1162/ling.2008.39.1.1

Embick, David & Noyer, Rolf. 2001. Movement operations after syntax. *Linguistic Inquiry* 32. 555–595. DOI: https://doi.org/10.1162/002438901753373005

Embick, David & Noyer, Rolf. 2007. Distributed Morphology and the syntax–morphology interface. In Ramchand, Gillian & Reiss, Charles (eds.), *The Oxford Handbook of Linguistic Interfaces*, 289–324. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199247455.013.0010

González-Vilbazo, Kay & López, Luis. 2011. Some properties of light verbs in code-switching. *Lingua* 121. 832–850. DOI: https://doi.org/10.1016/j.lingua.2010.11.011

Grimstad, Maren B. 2017. The code-switching/borrowing debate: evidence from English-origin verbs in American Norwegian. *Lingue e linguaggio, Rivista semestrale* 1. 3 − 34.

Grimstad, Maren B. & Riksem, Brita & Lohndal, Terje & Åfarli, Tor. 2018. Lexicalist vs. exoskeletal approaches to language mixing. *The Linguistic Review* 35. 1 − 32. DOI: https://doi.org/10.1515/tlr-2017-0022

Grimstad, Maren B. & Lohndal, Terje & Åfarli, Tor. 2015. Language mixing and exoskeletal theory: A case study of word-internal mixing in American Norwegian. *Nordlyd* 41. 213 − 237. DOI: https://doi.org/10.7557/12.3413

Grosjean, François. 1989. Neurolinguists, Beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36. 3–15. DOI: https://doi.org/10.1016/0093-934X(89)90048-5

Halle, Morris. 1997. Distributed Morphology: Impoverishment and fission. *MIT Working Papers in Linguistics* 30. 425–449.

Halle, Morris & Marantz, Alec. 1993. Distributed Morphology and the pieces of inflection. In Hale, Kenneth & Keyser, Samuel Jay (eds.), *The view from Building 20*, 111 − 176. Cambridge, MA: MIT Press.

Harley, Heidi. 2014. On the identity of roots. *Theoretical Linguistics* 40. 225 − 276. DOI: https://doi.org/10.1515/tl-2014-0010

Harley, Heidi & Noyer, Rolf. 1999. Distributed morphology. *Glot International* 4. 3–9.

Harris, James. 1991. The exponence of gender in Spanish. *Linguistic Inquiry* 22. 27–62.

Hornstein, Norbert. 2018. The Minimalist Program after 25 years. *Annual Review of Linguistics* 4(1). 49–65. DOI: https://doi.org/10.1146/annurev-linguistics-011817-045452

Jake, Janice & Myers-Scotton, Carol & Gross, Steven. 2002. Making a minimalist approach to code-switching work: Adding the Matrix language. *Bilingualism: Language and Cognition* 5. 69–91. DOI: https://doi.org/10.1017/S1366728902000147

Kern, Joseph. 2019. Like in English and como, como que, and like in Spanish in the speech of Southern Arizona bilinguals. *International Journal of Bilingualism* 24(2). 184–207. DOI: https://doi.org/10.1177/1367006919826329

Kramer, Ruth. 2015. *The Morphosyntax of Gender*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199679935.001.0001

Kramer, Ruth. 2016. The location of gender features in the syntax. *Language and Linguistics Compass* 10(11). 661–677. DOI: https://doi.org/10.1111/lnc3.12226

Kramer, Ruth. 2020. Grammatical gender: A close look at gender assignment across languages. *Annual Review of linguistics* 6(1). 45–66. DOI: https://doi.org/10.1146/annurev-linguistics-011718-012450

Kroll, Judith & Dussias, Paola & Bice, Kinsey & Perrotti, Lauren. 2015. Bilingualism, Mind, and Brain. *Annual Review of Linguistics*, 377–394. DOI: https://doi.org/10.1146/annurev-linguist-030514-124937

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Laleko, Oksana. 2018. What is Difficult about Grammatical Gender? Evidence from Heritage Russian. *Journal of Language Contact* 11(2). 233–267. DOI: https://doi.org/10.1163/19552629-01102004

López, Luis. 2020. *Bilingual Grammar: toward an integrated model*. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/9781108756181

Liceras, Juana M. & Spradlin, Kenton T. & Fuertes, Raquel Fernández. 2005. Bilingual early functional-lexical mixing and the activation of formal features. *International Journal of Bilingualism* 9(2). 227 − 252. DOI: https://doi.org/10.1177/13670069050090020601

Liceras, Juana M. & Fuertes, Raquel Fernández & Klassen, Rachel. 2016. Language dominance and language nativeness. The view from English-Spanish codeswitching. In Guzzardo, Rosa & Mazak, Catherine & Parafita Couto, M. Carmen (eds.), *Spanish-English Codeswitching in the Caribbean and the US*, 107–138. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/ihll.11.05lic

Liceras, Juana M. & Fuertes, Raquel Fernández & Perales, Susana & Pérez-Tattam, Rocío & Spradlin, Kenton T. 2008. Gender and gender agreement in bilingual native and non-native grammars: A view from child and adult functional–lexical mixings. *Lingua* 118. 827–851. DOI: https://doi.org/10.1016/j.lingua.2007.05.006

Lillo-Martin, Diane & Müller de Quadros, Ronice & Pichler, Deborah. 2016. The development of bimodal bilingualism Implications for linguistic theory. *Linguistic Approaches to Bilingualism* 6(6). 719–755. DOI: https://doi.org/10.1075/lab.6.6.01lil

MacSwan, Jeff. 2000. The architecture of the bilingual language faculty: Evidence from code-switching. *Bilingualism: Language and Cognition* 3. 37–54. DOI: https://doi.org/10.1017/S1366728900000122

MacSwan, Jeff. 2005. Codeswitching and generative grammar. A critique of the MLF model and some remarks on "modified minimalism." *Bilingualism: Language and Cognition* 8. 1−22. DOI: https://doi.org/10.1017/S1366728904002068

McConnell-Ginet, Sally. 2013. Gender and its relation to sex: The myth of 'natural' gender. In Corbett, Greville (ed.), *The Expression of Gender*, 3−38. Berlin: De Gruyter Mouton.

Moro, Monica. 2014. The semantic interpretation and syntactic distribution of determiner phrases in Spanish-English codeswitching. In MacSwan, Jeff (ed.), *Grammatical Theory and Bilingual Codeswitching*, 213−226. Cambridge, MA: MIT Press.

Myers-Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Code-Switching*. Oxford: Oxford University Press.

Myers-Scotton, Carol M. & Jake, Janice L. 2017. Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism* 21. 340−366. DOI: https://doi.org/10.1177/1367006915626588

Oltra-Massuet, Isabel & Arregi, Karlos. 2005. Stress by-structure in Spanish. *Linguistic Inquiry* 36. 43–84. DOI: https://doi.org/10.1162/0024389052993637

Ortega, Lourdes. 2020. The Study of Heritage Language Development From a Bilingualism and Social Justice Perspective. *Language Learning* 70. 15–53. DOI: https://doi.org/10.1111/lang.12347

Ortega, Lourdes. 2016. Multi-competence in second language acquisition: inroads into the mainstream? In Cook, Vivian & Wei, Li (eds.), *The Cambridge Handbook of Linguistic Multi-Competence*, 50–76. Cambridge, UK: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781107425965.003

Ortheguy, Ricardo & Shin, Naomi Lapidus. 2003. An adaptive approach to noun gender in New York contact Spanish. In Núñez-Cedeño, Rafael & López, Luis & Cameron, Richard (eds.), *A Romance perspective on language knowledge and use*, 209−29. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/cilt.238.17oth

Parafita Couto, Maria & Stadthagen-Gonzalez, Hans. 2017. El book or the libro? Insights from acceptability judgments into determiner/noun code-switches. *The International Journal of Bilingualism* 23. 349–360. DOI: https://doi.org/10.1177/1367006917728392

Pfaff, Carol W. 1979. Functional and structural constraints on syntactic variation in code-switching. *Language* 55. 297−318. DOI: https://doi.org/10.2307/412586

Pfau, Roland. 2009. *Grammar as processor a distributed morphology of spontaneous speech errors*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/la.137

Picallo, M. Carme. 1991. Nominals and nominalization in Catalan. *Probus* 3. 279–316. DOI: https://doi.org/10.1515/prbs.1991.3.3.279

Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: toward a typology of code-switching. *Linguistics* 18. 581–618. DOI: https://doi.org/10.1515/ling.1980.18.7-8.581

Poplack, Shana & Pousada, Alicia & Sankoff, David. 1982. Competing influences on gender assignment: variable process, stable outcome. *Lingua* 57. 1 − 28. DOI: https://doi.org/10.1016/0024-3841(82)90068-7

Putnam, Michael & Carlson, Matthew & Reitter, David. 2018. Integrated, Not Isolated: Defining Typological Proximity in an Integrated Multilingual Architecture. *Frontiers in Psychology* 8. 2212–2212. DOI: https://doi.org/10.3389/fpsyg.2017.02212

Preminger, Omer. 2009. Breaking agreements: distinguishing agreement and clitic-doubling by their failures. *Linguistic Inquiry* 40(4). 619–666. DOI: https://doi.org/10.1162/ling.2009.40.4.619

Radford, Andrew & Kupisch, Tanja & Köppe, Regina & Azzaro, Gabriele. 2007. Concord, convergence and accommodation in bilingual children. *Bilingualism: Language and Cognition* 10(3). 239–256. DOI: https://doi.org/10.1017/S1366728907003057

Riksem, Brita R. 2018. *Language Mixing in American Norwegian Noun Phrases: An Exoskeletal Analysis of Synchronic and Diachronic Patterns*. Ph.D. Dissertation, Trondheim; NTNU q Norwegian University of Science and Technology.

Roca, I. M. 1989. The organization of grammatical gender. *Transactions of the Philological Society* 87. 1 − 32. DOI: https://doi.org/10.1111/j.1467-968X.1989.tb00617.x

Smead, Robert N. 2000. On the Assignment of Gender to Chicano Anglicisms: Processes and Results. *The Bilingual Review* 25(3). 277–297.

Teschner, Richard & Russell, William. 1984. The gender patterns of Spanish nouns: An inverse dictionary-based analysis. *Hispanic Linguistics* 1. 115–132.

Valdés Kroff, Jorge. 2016. Mixed NPs in Spanish-English bilingual speech: Using a corpus-based approach to inform models of sentence processing. In Guzzardo Tamargo, Rosa E. & Mazak, Catherine M. & Parafita Couto, M. Carmen (eds.), *Spanish-English Codeswitching in the Caribbean and the US*, 281 − 300. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/ihll.11.12val

Vergara, Daniel & López, Luis. 2017. Obliteration after Vocabulary Insertion. In Bellamy, Kate & Child, Michael W. & González, Paz & Muntendam, Antje & Parafita Couto, M. Carmen (eds.), *Multidisciplinary Approaches to Bilingualism in the Hispanic and Lusophone World*, 261 − 282. Amsterdam: John Benjamins.

Woolford, Ellen. 1983. Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14. 520–536. http://search.proquest.com/docview/1297364297/.