KEYNOTE ARTICLE

Coactivation in bilingual grammars: A computational account of code mixing*

MATTHEW GOLDRICK
Northwestern University
MICHAEL PUTNAM
Pennsylvania State University
LARA SCHWARZ
Pennsylvania State University

A large body of research into bilingualism has revealed that language processing is fundamentally non-selective; there is simultaneous, graded co-activation of mental representations from both of the speakers' languages. An equally deep tradition of research into code switching/mixing has revealed the important role that grammatical principles play in determining the nature of bilingual speech. We propose to integrate these two traditions within the formalism of Gradient Symbolic Computation. This allows us to formalize the integration of grammatical principles with gradient mental representations. We apply this framework to code mixing constructions where an element of an intended utterance appears in both languages within a single utterance and discuss the directions it suggests for future research.

One of the more amazing feats of bilingual language production is the fluent integration of two languages within a single utterance. We refer to this phenomenon as CODE MIXING to emphasize the integration of two linguistic systems, using this synonymously with terms such as INTRA-SENTENTIAL CODE SWITCHING. An extensive body of research has identified important roles for the grammatical principles of the source languages in constraining code mixing (see, e.g., Deuchar, 2005; Muysken, 1995; Myers-Scotton & Jake, 1995; Poplack, 1980, for reviews). Parallel to this line of research, several decades of research has provided a wealth of evidence suggesting that bilinguals simultaneously co-activate elements from each language during production. For example, when intending to name a picture of a dog, a Spanish–English bilingual will SIMULTANEOUSLY activate, to varying degrees, representations corresponding to English (DOG) and Spanish (PERRO) forms (see e.g., Kroll & Gollan, 2014, for a review). This suggests that mental representations in bilingual speakers incorporate BLENDS of structures from each language (i.e., not DOG *or* PERRO, but a representation that is *both* DOG *and* PERRO).

Given the strength of the evidence for grammatical principles as well as blend representations, we argue below that an adequate theory of bilingual linguistic cognition must be able to incorporate both of these elements. Discrete grammatical principles must be integrated with gradient blend representations; currently, no existing framework does so. In this work, we propose such an integration using the Gradient Symbolic Computation framework (GSC; Smolensky, Goldrick & Mathis, 2014). This grammar-based formalism incorporates symbolic representations, whose elements are associated with continuous activation values. We show how a Gradient Symbolic approach to code mixing can allow us to account for grammatical constraints on blend representations that emerge in code mixing.

We begin by reviewing the evidence for blend representations in bilinguals across a variety of processing contexts. To highlight the interaction of blend representations and grammatical principles, we then examine in detail code mixing productions where blended elements are overtly produced – an element of the utterance is DOUBLED, appearing in both languages within a single utterance. With these empirical data in mind, we develop a GSC account of code mixing. We demonstrate how it accounts for empirically observed restrictions on doubling, and discuss the future research directions it suggests.

While our focus is on the interaction of grammatical principles and the gradient representational structure, it is important to note that many other factors contribute to code mixing. In particular, sociolinguistic factors play an important role in language choice and bilingual identity (for an overview, see Gardner-Chloros, 2009). While these are outside the scope of this current work, they define an important avenue for future development of our approach.

Address for correspondence:
Matthew Goldrick, Department of Linguistics, Northwestern University, 2016 Sheridan Rd., Evanston, IL 60208 USA
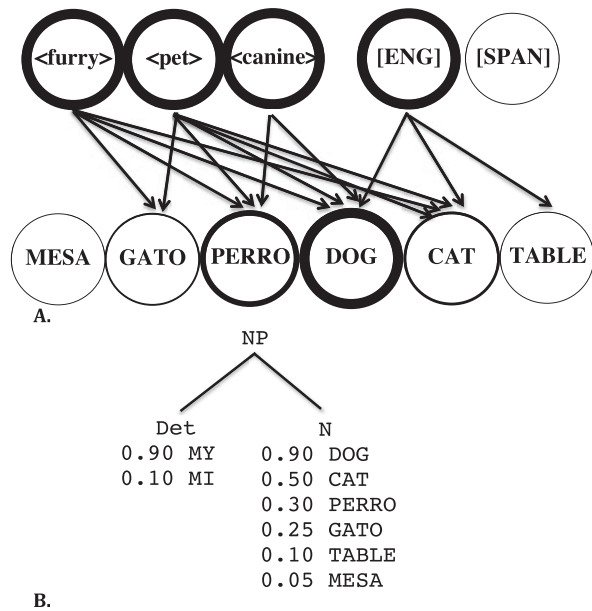*matt-goldrick@northwestern.edu*

Figure 1. **A.** Depiction of psycholinguistic processing model during production of DOG. Thickness of circle denotes relative activation of unit. **B.** Alternative depiction of the state of this system, focusing on gradient activation at the lexical level.

Blend Representations in Contexts without Code Mixing

Blend representations

Many psycholinguistic theories are framed within a spreading-activation or connectionist perspective (Rumelhart, Hinton & McClelland, 1986; see Goldrick, 2012, for a recent review). In such theories, mental representations are graded, distributed patterns of ACTIVATION, a numerical quantity associated with simple processing units. This allows for BLENDS: representational states in which multiple representational elements occupy (to varying degrees) a single position within a linguistic structure.

For example, suppose a native Spanish speaker is producing a sentence in English: "Yesterday I went to the park to walk my dog." While planning this utterance – in particular, while retrieving the appropriate final noun from memory – many psycholinguistic theories of bilingualism assume that the speaker's production system enters the state shown in Figure 1 (see, e.g., Kroll & Gollan, 2014, for a review of such proposals). In this network, there are three types of representational units. The input to the system consists of semantic features along with a representation of the intended language of response. Activation spreads from these units to a set of units corresponding to lexical items (e.g., 'lemmas'). Figure 1A shows the flow of activation through the connectionist network; Figure 1B provides an alternative view of the distribution of activation over the lexical items.

In such a representation, the intention to produce a single lexical item (a single noun in the phrase '*my* ___') results in the simultaneous co-activation of multiple mental representations. Lexical selection processes simultaneously consider the target (DOG), semantically related words within the same language (CAT), and non-target language words (PERRO). This can be seen in Figure 1A+B, where multiple representational elements have varying non-zero activation. The state of processing a single word is thus a blend of multiple linguistic representations. This representational hypothesis is often referred to as CO-ACTIVATION or PARALLEL ACTIVATION. We use the term BLEND to emphasize that the multiple elements are not simply simultaneously activated; they are co-present within a single position in the linguistic representation (e.g., head of a particular noun phrase; this is highlighted in the depiction in Figure 1B).

Empirical evidence for blend representations

Activation based representations do not require blends (one could assign 0 to all non-target representations in Figure 1, activating only DOG). It is also not immediately clear what functional motivation would require blend states. In a purely English utterance, why should one consider Spanish words? This makes it all the more striking that a substantial body of evidence supports such blend representations. Kroll and Gollan (2014) provide an extensive review of evidence from multilingual speakers (see Melinger, Branigan & Pickering, 2014, for a review of evidence from monolinguals). Here, we emphasize a few key recent studies that provide evidence of such representations during production of phrases and sentences.

A key prediction of blend representations is that the spread of activation will lead to the partial activation of non-target representations at other levels of processing. Following the example above, when producing target DOG, the partial activation of the lexical representation PERRO is predicted to lead to partial activation of representations of the /p/ sound. In contrast, when producing CAT, the PERRO representation should be less active, resulting in less activation of the /p/ sound. Consistent with this prediction, many studies have demonstrated that production is facilitated when there is a phonological relationship between the target utterance and non-target translation equivalents.

Spalek, Hoshino, Wu, Damian, and Thierry (2014) examined German–English bilinguals producing adjective-noun phrases. For both behavioral and electrophysiological measures, they found that second-language English production was facilitated when the

English adjective shared phonological structure with the noun's German translation equivalent (e.g., "*blue flower*": *blue* shares the initial sounds of the German translation equivalent *Blume*; contrast with "*green skirt*:" *green* shares no sounds with the German translation equivalent *Rock*). While significant, effects of non-dominant L2 English on production of dominant L1 German phrases were smaller, limited to electrophysiological measures. These results are consistent with the presence of blend representations, but suggest that the degree to which non-target representations are present in blends is modulated by the relative strength of each language (such that *Blume* is more active during processing of *flower* than vice versa).

Another set of studies arguing for co-activation of multiple representations has compared the production of targets that share phonological structure with their translation equivalent (e.g., English ANCHOR – Dutch ANKER) to those with no overlap (e.g., BOTTLE – FLES). The former are often referred to as 'cognates' in the psycholinguistic literature; however, no historical connection between the translation equivalents is required. The logic is that simultaneous activation of lexical representations in the two languages should facilitate processing of any shared phonological structure, producing a COGNATE FACILITATION EFFECT (Costa, Caramazza & Sebastián-Gallés, 2000). For example, simultaneously activating ANCHOR and ANKER will serve to facilitate retrieval/planning of shared segments /ŋ/, /k/, /ɚ/. Starreveld, De Groot, Rossmark, and Van Hell (2014) recently documented cognate facilitation during sentence planning. Dutch–English bilinguals read aloud sentences with an embedded picture (e.g., a picture of an anchor appeared in the position of the blank in the sentence "In the middle of the square was an ____ with a thick chain attached to it."). When producing picture names in L2 English, participants showed cognate facilitation. Following the study reviewed above (Spalek et al., 2014) as well as many other results, these effects were much stronger in L2 than L1 production. Furthermore, cognate effects were modulated by the sentence context. When the sentence placed greater constraints on the word that could fit in the space occupied by the picture (e.g., "Popeye the sailorman has a tattoo of an ____ on his arm."), cognate effects were diminished. As with the preceding study, these results suggest that while blends are part of language production, the degree to which non-target representations are present is modulated not only by the relative strength of each language, but also by the degree to which context supports the retrieval of a specific target word.

Finally, some of the strongest evidence for blend representations has come from studies that have documented the literal co-production of multiple representations. The simultaneous co-presence of multiple linguistic representations during planning leads to the simultaneous production of actions associated with these representations. Pyers and Emmorey (2008) examined the oral and manual productions of BIMODAL BILINGUALS: native speakers of a spoken language (English) and a manual language (American Sign Language; ASL). During conversations with non-signers – where the bimodal bilinguals intend to speak a single (oral) language – they simultaneously produced ASL and English grammatical markers. At rates much higher than non-signers (but lower than in their ASL productions), the bimodal bilinguals furrowed their brows while producing *wh*-questions (e.g., "How many siblings does she have?"). This occurred in spite of the fact that spoken English explicitly marks *wh*-questions (making double-marking unnecessary to express the intended message). Note that this gesture is pragmatically dispreferred in spoken English, where it conveys negative affect. Pyers and Emmorey argued that this modulated the rate of co-productions, as co-productions were much higher for conditionals (e.g., "If it rains, class will be canceled"; associated with raised brows). This provides further evidence for constraints on the degree of activation of non-target representations in blends.

In the bimodal bilingual case, the two languages are not competing for expression on the same communication channel. More subtle co-productions can be found during production of two oral languages. While co-activation enhances retrieval of shared phonological structure, the heightened activation of non-target language representations should increase cross-language phonetic interference – the intrusion of non-target language phonetic properties into bilingual productions. For example, while Spanish and English share a common set of voicing contrasts in initial stops (e.g., /b/ vs. /p/), the phonetic realization of this contrast is distinct in each language (pre-voiced vs. short-lag voice onset time in Spanish; short vs. long-lag in English). This conflict leads non-native speakers to produce these sounds with phonetic properties intermediate between the two languages (Flege, 1991). Amengual (2012) showed that this cross-language phonetic interference is enhanced for cognates. When reading sentences aloud, Spanish–English bilinguals produced initial stops in Spanish with more English-like properties in cognates vs. non-cognates. No such difference was found in the productions of Spanish–Catalan bilinguals (where the two languages have similar phonetic realizations of this contrast). Other results suggest that these cognate effects are not simply word-specific phonetic patterns in bilingual speech, but rather reflect dynamic properties of bilingual production. Olson (2013) and Goldrick, Runnqvist, and Costa (2014) found that phonetic interference was increased when participants were required to unexpectedly code-switch during picture naming (vs. trials where participants did not switch languages). For voiceless stops, Goldrick et al. found that this context-specific phonetic interference

effect is enhanced during production of single cognate vs. non-cognate words – suggesting that the cognate effect reflects the context-specific activation of target and non-target language representations.

Summary: Blend representations in bilingual production

Even when intending to produce a single form in a single language, bilinguals simultaneously activate forms in both languages. The degree of co-presence in such blend representations, and our ability to observe the effects of this co-presence, is clearly constrained. To some degree this likely reflects physical constraints. It is impossible to place a single set of oral articulators in two contradictory positions. In these cases, the production system is limited to blended articulations, reflecting a partial compromise between contradictory actions. However, many of the other constraints on blends clearly reflect abstract, cognitive principles. Even when freed from physical constraints on co-production, the properties of bimodal bilinguals' blends are modulated by affective/pragmatic constraints. The properties of unimodal bilinguals' blends reflect the relative strength of the two languages and the context in which a target word is being produced.

Blend Representations in Code Mixing

Integration of grammatical principles in code mixing

Given the evidence for co-activation of the two languages in contexts where speakers intend to produce only one language, it is unsurprising that co-activation is a fundamental property of code mixing. Critically, speakers are not only uttering lexical items from both languages but are also integrating grammatical principles from each linguistic system.

The integration can be seen in cross-linguistic syntactic priming, where exposure to a structure in one language increases the probability that speakers will use a similar structure in another language (see Pickering & Ferreira, 2008, for a review). For example, Hartsuiker, Pickering, and Veltkamp (2004) found that when Spanish–English bilinguals heard a passive construction in Spanish, it increased the likelihood that they would produce a passive vs. active construction in English on a subsequent trial. Such priming does not only alter the probability of attested structures. In certain contexts, it can allow for the transfer of grammatical patterns from one language to another, reflecting the integration of knowledge of each language. For example, in many contexts Spanish does not allow for the word order adjective-noun, the typical word order pattern observed in English. Hsin, Legendre, and Omaki (2013) found that in Spanish–English bilingual children

priming could allow for transfer of this word order from English to Spanish.

Such integration also occurs in the context of intra-sentential code mixing. For example, Kootstra, van Hell, and Dijkstra (2010) elicited code mixed utterances from Dutch–English bilinguals. Participants described pictures by completing a Dutch sentence fragment that biased speakers to produce one of several word orders possible in Dutch (Subject-Verb-Object [SVO], SOV, or VSO). When cued to produce a mixed structure (i.e., using at least one English word to complete the Dutch fragment), participants preferred to use the word order common to both grammars (SVO). A similar preference for congruent grammatical patterns has been found in spontaneous mixing corpora (for reviews and discussion, see Deuchar, 2005; Muysken, 1995; Myers-Scotton & Jake, 1995; Poplack, 1980).

Blends and co-production in code mixing

When speakers intend to mix lexical items and grammatical principles from two languages, we also observe blends. Some of the most dramatic examples come from bimodal bilingual code mixing. For ASL–English bilinguals the predominant type of code mixing is code blending: co-production of oral and manual elements (Bishop, 2010; Emmorey, Borinstein, Thompson & Gollan, 2008). These cross-modal productions are typically semantically equivalent and synchronized in time. In examples (1) and (2), the English gloss of the sign production is shown in italics beneath the point in the sentence where the sign roughly occurred; underlining indicates the speech that co-occurred with that sign.

(1) And there's <u>the bird</u>. (Emmorey et al., 2008: 48)

   bird

(2) <u>Now I recently went back</u>. (Emmorey et al., 2008: 48)

   now I recently go-to

In unimodal bilinguals there is subtle evidence of co-activation in articulation. Analyzing a spontaneous code mixing corpus, Balukas and Koops (in press) found that phonetic interference effects in Spanish–English bilinguals increase at points closer to code switches. This suggests that co-productions are not unique to bimodal bilinguals.

Blends without co-production: Doubling constructions

Blending representations from two different languages can also yield non-simultaneous articulations. Languages differ in word order, mapping elements to different positions in the surface string. This raises the possibility that the grammatical principles from each language – both

active during code mixing – could both be satisfied without yielding simultaneous articulation. For example, a bilingual's L1 has the word order verb-object and L2 object-verb. The string verb(L1)-object-verb(L2) satisfies the word order constraints of both languages; the L1 verb precedes the object, while the L2 verb follows the object. Although such strings might violate structural constraints on linguistic representations, they would not suffer from articulatory incompatibility[1].

Patterns of attested doubling constructions: A review

Although such constructions are commonly discussed, only a few detailed references are devoted exclusively to them (Chan, 2009; Hicks, 2010, 2012; Muysken, 2000: 104–6). They manifest in a variety of constituents, although preference seems to be given to the doubling of functional elements (syntactic elements expressing grammatical relationships; e.g., complementizers, determiners, prepositions, and auxiliary verbs) over lexical item (e.g., nouns). The following examples[2] provide an overview of the range of doubled structures. The doubled elements are underlined in each example.

(3) Complementizers: English–Japanese (Azuma 1993: 199)

<u>if</u> it goes three rounds datta <u>ra</u> ne

if it goes three rounds was if TAG

'If it goes three rounds.'

Note that in (3), *if* is located in its canonical place in English (appearing at the start of the dependent clause), and the Japanese *ra* in its expected location (were the utterance fully Japanese, *ra* would appear at the end of the dependent clause). Examples (4)–(7) illustrate similar doubling for various other elements, respecting the contrasting word orders.

(4) Adpositions: English–Finnish (Poplack, Wheeler & Westwood, 1989: 405)

mutta se oli kidney-sta <u>to</u> aorta-<u>an</u>

but it was kidney-from to aorta-to

'But it was from the kidney to the aorta.'

(5) Adverbials: English–Tamil (Sankoff, Poplack & Vannianiarajan, 1990: 92)

<u>According to</u> the schedule

<u>paDi</u>   oNNutaan irukkaNum.

according to one only be must

'According to the schedule, there must be only one.'

(6) Coordinating: conjunctions Spanish–Aymara (Stolz, 1996: 146, citing Porterie-Guierrez, 1988: 355)

<u>pero</u> sorro -<u>sti</u> wali astuturi -tajna . . .

but fox -COO very keen -3.SG.PRT.EVI

'But the fox was very keen.'

(7) Verbs English–Tamil (Sankoff et al., 1990: 93)

they <u>gave</u> me a research grant <u>koḍutaa</u>

they gave me a research grant gave.3.PL.PAST

'They gave me a research grant.'

Multi-word chunks can be doubled, as shown in (8) (verb + adverb) and (9) (verb + complementizer).

(8) Verb + Adverb English–Japanese (Nishimura, 1986: 139)

We <u>bought about</u> two pounds <u>gurai kattekita</u> no

We bought about two pounds about bought TAG

'We bought about two pounds.'

(9) Verb + Complementizer English–Korean (Chan, 2008: 800)

everybody <u>think that</u> nay-ka yenge-lul   cal

everybody think C   I-NOM English-ACC well

hanta-<u>ko</u> <u>sayngkakhayyo</u>

do-C   think

'Everybody thinks that I'm a good English speaker.'

While these types of code mixing utterances have been consistently documented in corpora, they are clearly marked; in general, these structures are largely avoided. Poplack et al. (1989: 405) report that these blends are "exceedingly rare", citing that they only found 2 in their entire corpus; Furukawa (2008) found 7 examples in 5 hours of sociolinguistic interview data. However, Nishumura (1986) and Backus (1992) suggest that these blends occur in roughly 3–5% of their corpus materials. The rarity of such productions is unsurprising. As noted above, the integration of grammatical principles from both languages yields a preference for congruent grammatical patterns in code mixing. The examples above violate this principle; they involve doubling of elements that

---

[1] Consistent with an analysis where doubling can arise due to the co-presence of multiple elements in the input to the grammar, elements from multiple alternative formulations of an intended message are sometimes co-present in monolingual speech errors (Coppock, 2010; Menn & Duffield, 2013).

[2] Doubling of inflectional elements has also been reported, both when inflectional elements occur in distinct positions (e.g., prefixation vs. suffixation; Bokamba, 1988; Myers-Scotton, 1993) and when they occur in the same position (Backus, 1992). This latter type has not been reported with non-inflectional elements, which is the focus of the analysis here.

are subject to conflicting grammatical patterns (e.g., verb-object vs. object-verb word order). Furthermore, research on non-code mixed productions suggests there are strong limitations on blends; the degree of co-presence within a blend is highly limited. If doubling constructions reflect blend representations, we also expect them to be strongly dispreferred.

Surveying the reported instances of these constructions, Hicks (2010) identifies several cross-linguistic generalizations. First, as noted above, doubled elements locally respect the word order of the source grammars. Second, the doubled elements are typically HEADS; syntactic elements that define the syntactic properties of the phrase to which they belong. The doubled heads share a non-doubled COMPLEMENT; the other syntactic elements that belong to the phrase. For example, a verb phrase can be composed of a verb (the head) and an object (the complement). In the doubling construction $V_{L1}O_{L1}V_{L2}$, doubled verbs share a non-doubled object complement (see also Furukawa, 2008). Thus, strictly local doubling (e.g., $V_{L1}O_{L1}O_{L2}V_{L2}$) is typically not observed. Finally, while some languages exhibit doubling in monolingual contexts (discussed further below), Hicks notes that doubling of elements from the same language (e.g., analogous to (7), *gave grant gave, *gave gave grant) are not observed during code mixing.

In sum, while doubling is rare, it is consistently observed across various sources; a variety of elements participate in intrasentential code mixing blends. Doubling is not mere repetition of elements; its occurrence is constrained by grammatical principles; doubled heads share non-doubled complements; and equivalence for grammatical features. This suggests that doubling constructions are not an "ad hoc production strategy" (Sankoff et al., 1990: 92), but are rather coherent, syntactic objects that are governed by grammatical principles.

Blends in Grammatical Theories: Application to Doubling

Blend representations clearly play a role in bilingual language processing. In doubling constructions, we see that these blend representations interact with grammatical principles. How can this be formally specified? In this section, we develop a grammatical approach to code mixing that incorporates blended representations. We apply this to doubling constructions, showing how it accounts for the occurrence of doubling as well as the empirically attested constraints on this phenomenon.

Overview of the proposal

Our account is based around 3 general principles. We first provide an overview of these and then examine in some detail how they can be applied to the empirical patterns of doubling constructions.

Principle 1: Probabilistic grammars with weighted constraints

Language use – in mono- or multi-linguals – is defined in part by regular structural patterns (e.g., English requires SVO, while Dutch allows flexibility between SVO, SOV, and VSO). Grammars allow us to precisely specify the structure of the mapping between form and meaning that yields these patterns. The formalism we use specifies grammars through interaction of constraints on linguistic structure. For example, a constraint on word order might prefer that certain lexical categories appear at the left edge of a syntactic phrase. Constraints are associated with numerical weights that determine their relative importance; cross-linguistic variation (e.g., if a language categorically prefers SVO vs. SOV) is specified by changes in the relative weighting of constraints. Our grammatical formalism also allows us to specify not just categorical preferences, but also relative probabilities of different structures; this allows us to capture variation in the mapping between meaning and form within a speaker (e.g., variable word ordering in a Dutch speaker's productions, or variable structures observed in code switching).

Principle 2: Gradient blends of grammars

Bilinguals speakers have varying degrees of competence in multiple grammars, allowing them to produce distinct structures in each language. In our formalism, this is reflected by associating each language with a distinct weighting of constraints. These language-specific weightings contribute to the grammar, independently influencing the probability of different structures. However, as discussed above, the two linguistic systems of bilinguals interact. We model this by also incorporating into the grammar a weighting of constraints that blends the language-specific weightings. The degree to which each language contributes to this blend reflects the relative activation of that linguistic system.

Principle 3: Gradient blends in linguistic representations

Building on the connectionist formalisms that serve as the foundation of many psycholinguistic theories, we assume that there is simultaneous coactivation of representational elements in both the input and output of the grammar. This allows for representations that blend elements from multiple languages.

In the sections below, we elaborate the details of this grammatical proposal. It is important to note that grammars define cognitive processes at a high level of abstraction – in terms of mapping between inputs and outputs. This is key to developing a clear and rigorous

specification of what precisely a cognitive process does; what are the types of structures that are predicted to (probabilistically) emerge by our theory of language structure? We aim to develop such a framework for understanding the structure of code mixing. However, it is important to note that understanding the cognitive and ultimately neural processes that compute these input-output mappings is key to developing a complete theory of language processing (for discussion, see Goldrick, 2011; Smolensky, 2006). Grammar is the foundational component at the beginning of developing a complete theory, but is by no means the final step.

Relationship to other formal approaches to code mixing

Generative theories of code mixing – such as the one we further develop here – can be divided into two types. One set specifies grammars specific to code mixing. Rules or constraints refer specifically to code mixed structures, explicitly stating preferences for distinct types of code mixing (Belazi, Rubin & Toribio, 1994; Bhatt, 1997; Di Sciullo, Muysken & Singh, 1986; Joshi, 1985; Legendre & Schindler, 2010; Muysken, 2013; Myers-Scotton, 1993; Poplack, 1980; a.o.). A classic example is Poplack's (1980: 586) Equivalence Constraint: "Code-switches will tend to occur at points in discourse where juxtaposition of Ll and L2 elements does not violate a syntactic rule of either language." Here, the grammatical principle refers directly to code mixing, distinct from (but related to) syntactic patterns in non-code mixed contexts. Similarly, from an Optimality-Theoretic perspective, Muysken (2013: 715) makes use of constraints such as "*CSL = Don't switch between separate languages, either in their lexicon or in their grammar."

An alternative approach assumes the grammatical principles of two languages are integrated during code mixing, and that this integration yields the patterns (Chan, 2003, 2008, 2009; Lohndal, 2013; MacSwan, 1999, 2000; Mahootian, 1993; Woolford, 1983; a.o.). Since Mahootian (1993), this position is commonly referred to as the "null theory" of code mixing, according to which monolingual and bilingual grammars should be subject to identical representational/grammatical constraints and psychological principles.

Our approach incorporates elements of both perspectives. Following the null theory perspective, we assume that the features of code mixing reflect general principles of syntactic knowledge and sentence processing. Blend representations are a general feature of grammatical knowledge and processing; the emergence of doubling constructions in bilinguals is a consequence of the principles underlying these grammars. However, in contrast to the strongest version of the null theory (e.g., MacSwan, 1999), we assume that grammatical principles can refer to language membership (e.g.,
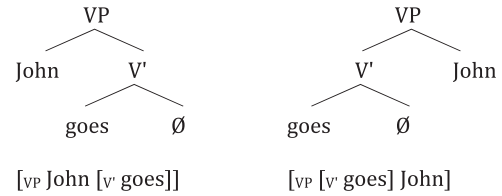

Figure 2. Two alternative surface syntactic structures corresponding to the input *goes* (*John*). The text below each provides a bracket notation corresponding to the tree, with the subscript on each open bracket denoting the category of the constituent.

distinguishing the well-formedness of lexical items in English vs. Tamil based not on syntactic features but on the language from which the item originates). This is critical to understanding attested doubling patterns. In such constructions, the doubled elements have (nearly) equivalent grammatical features (e.g., they match in agreement features (tense, aspect, case) and share argument structure requirements), yet surface in positions appropriate to the element's source language. If the grammar does not make reference to the source language, there is no means of capturing this restriction.

Weighted constraint interaction in stochastic generative grammars

Our theory utilizes the Gradient Symbolic Computation formalism (GSC; Smolensky et al., 2014). GSC is a constraint-based approached to generative grammar, building on work in Optimality Theory (Legendre, 2001; Legendre, Putnam, de Swart & Zaroukian, in press-a; Prince & Smolensky, 1993/2004) and Harmonic Grammar (Legendre, Miyata & Smolensky, 1990, 2006; Pater, 2009). Like other generative grammars, GSC defines a function that maps input structures (e.g., logical forms) to output structures (e.g., syntactic structures). In GSC (and Harmonic Grammar), the grammar is defined via a set of weighted violable constraints that assign a numerical well-formedness value (HARMONY) to each of the candidate outputs for a given input. GSC grammars are stochastic, generating a probability distribution over output forms (reflecting the relative harmony of the candidates).

To build up our theory, we begin by modeling monolingual grammars. Consider a simple input consisting of a subject and verb; as shown in Figure 2, this can be linearized using at least two surface syntactic structures.

These surface syntactic structures reflect the assumptions of X-bar theory (see Carnie 2010: Chapter 7, for a historical overview). Briefly, X-bar theory places restrictions on traditional phrase structure grammars. Abstracting away from more complex phenomena, basic X-bar theory assumes that the basic structure of an

Table 1. *Grammar fragment for English word order*

Input: *goes* (*John*)	SPECLEFT −20	HEADLEFT −1	*Harmony*	*Probability*
[VP John [V′ goes]]		−1	−1	≈ 1.0
[VP [V′ goes] John]	−20		−20	≈ 0

extended projection consists of two syntactic phrases: XP, consisting of a specifier and an X′ phrase; and X′, consisting of the head element X^0 and a complement. As illustrated here, the extended projection of the verb consists of a verb phrase (VP) with specifier *John*, and a V′ phrase consisting of the head *goes* (and a null complement). This simplified notation is sufficient for capturing the basic facts about syntactic constituency and allows us to specify the patterns in word order variation underlying the doubling examples we consider. We believe that the insights of this analysis would generalize to more recent representational frameworks (e.g., Bare Phrase Structure), which retained many of the insights of this basic system (see e.g., Chametzky, 2000).

To characterize the difference between languages that prefer the linear order subject-verb vs. verb-subject, we build on Grimshaw's (1997, 2001) analysis. Grimshaw proposed constraints on the alignment of specifiers, heads, and complements to edges of extended projections in X-bar theory[3]. The relative weighting of these constraints derives different word ordering preferences; i.e., determining whether a particular grammar prefers an SVO- vs. an SOV-ordering of arguments. We adapt these to develop a GSC analysis, using constraints on structural well-formedness (MARKEDNESS constraints). A subset of these is shown below:

(10) HEADLEFT: "Every X^0 is leftmost in X-max."

For each X^0 in candidate C, decrease C's harmony by 1 for each terminal node intervening between the X^0 and the left edge of its XP.

(11) SPECLEFT: "Every specifier is leftmost in X-max."

For each specifier in candidate C, decrease C's harmony by 1 for each terminal node intervening between the specifier and the left edge of its XP.

A pseudo-English weighting of these two constraints is shown below. The columns show the constraints. Cells in each column show the constraint's contribution to the

harmony of each candidate (scaled by the weight of the constraint). Here, since SPECLEFT has a stronger weighting than HEADLEFT, the subject-verb candidate has a higher harmony value. The final column gives the probability of each candidate. As in Maximum Entropy grammars (Goldwater & Johnson, 2003; Hayes & Wilson, 2008), the probability is an exponential function of its harmony relative to the other candidates[4]. In this example, the harmony of the subject-verb order is so much higher than verb-subject that its probability is extremely close to 1.0. While the second candidate technically has non-zero probability, it is extremely small; less than 1×10^{-8}. Thus, the grammar is essentially categorical.

If the ranking of the constraints shifts, the probability of different candidates will also shift. This can specify cross-linguistic variation; consider the grammar fragment in Table 2. Here, HEADLEFT has a much stronger weighting than SPECLEFT. This yields a language with post-verbal subjects.

In the above cases, the differences in harmony are quite large. However, as differences in harmony of candidates grow smaller, variation can result:

In these example fragments, the weighting of constraints has been arbitrarily decided. Our assumption is that such weightings are acquired by learners based on the probability distribution of forms in their linguistic experience (Goldwater & Johnson, 2003; Hayes & Wilson, 2008). For example, an English learner would acquire many examples with the word order subject-verb; all else being equal, this would lead her to favor constraint weightings similar to Table 1 over those in Table 2 or Table 3. As this discussion is focused on exploring the basic principles of the theory, we do not undertake a detailed study of this acquisition process. The conclusions we draw below will not be dependent on the particular weight values used to illustrate our analysis.

To provide concrete weight values for the purpose of illustration, we utilized Goldwater and Johnson's (2003) learning algorithm (as implemented in the MaxEnt Grammar Tool; Hayes, 2009; weights were rounded to

---

[3] This general approach is consistent with derivational/Minimalist approaches to grammar (see e.g., Broekhuis & Vogel, 2013; Legendre, Grimshaw & Vikner, 2001; Legendre, Putnam, de Swart & Zaroukian, in press-b) as well as constraint-based models such as Lexical Functional Grammar (see e.g., Bresnan, 2000; Kuhn, 2003; Sells 2001a, b).

[4] GSC assumes a stochastic optimization algorithm that converges to a distribution in which the probability of candidate c is: $\frac{e^{\frac{H(c)}{T}}}{\sum_x e^{\frac{H(x)}{T}}}$. Here, H(c) is the harmony of candidate c, x ranges over the set of all possible output candidates, and T is a parameter of the optimization algorithm. Here we assume that T has a lower limit of 1.

Table 2. *Grammar fragment for verb-subject word order*

Input: *goes* (*John*)	SPECLEFT −1	HEADLEFT −20	*Harmony*	*Probability*
[VP John [V′ goes]]		−20	−20	≈ 0
[VP [V′ goes] John]	−1		−1	≈ 1.0

Table 3. *Grammar fragment for variable word order*

Input: *goes* (*John*)	SPECLEFT −12	HEADLEFT −10	*Harmony*	*Probability*
[VP John [V′ goes]]		−10	−10	≈ 0.88
[VP [V′ goes] John]	−12		−12	≈ 0.12

yield integer values for ease of exposition). A weak uniform prior was used for each constraint ($\mu = 0$; $\sigma = 10^7$). The prior influences how constraint weights are updated during learning. This prior specifies a target value for each constraint weight (here, zero, so that constraint weights are as small as possible), along with a penalty for deviating from that target value (here, the very high variance implies an extremely small penalty). This reduces our examples to a single free arbitrary parameter: the variance on the prior. Using this training algorithm, this parameter (combined with the training data exemplifying a given word order) completely determines the constraint weights below.

Head-complement word order variation

For illustration purposes, the grammar fragments above are quite simple, considering only 2 possible candidate outputs and 2 constraints. In this section, we consider a somewhat more extended example, including two additional constraints and an explicitly defined space of possible output structures. This allows us to specify monolingual grammars that contrast in word order – specifically, subject-verb-object vs. subject-object-verb (again building on Grimshaw, 1997, 2001).

Modeling constructions including complements requires an additional markedness constraint, parallel to those proposed above:

(12) COMPLEFT: "Every complement is leftmost in X-max."

Foreach complement in candidate C, decrease C's harmony by 1 for each terminal node intervening between the complement and the left edge of its XP.

Extending the set of candidates, we consider not only those that vary in word order but also those that omit elements of lexical conceptual structure. These avoid violations of the constraints above by simply leaving out elements (a candidate with no complements cannot violate COMPLEFT). To insure that such candidates are dispreferred, we use a FAITHFULNESS constraint that assigns well-formedness based on the relationship between syntactic and semantic structure (after Legendre, Wilson, Smolensky, Homer & Raymond, 1995):

(13) PARSE: "Lexical conceptual structure is parsed."

Decrease candidate C's harmony by 1 for each element of lexical conceptual structure that does not have a corresponding element in C's surface syntactic structure.

For this discussion, we assume candidate outputs are limited to those including X-bar trees that parse all elements of lexical conceptual structure or any possible subset of elements. We further assume that elements of lexical conceptual structure are parsed into the appropriate syntactic positions (e.g., subject is parsed into Spec). For an input with a verb, subject, and object, this yields the candidate set shown in Table 4. Constraint weights were determined by training on data reflecting the context-neutral English word order for this particular input: 100% subject-verb-object (e.g., "They gave a grant"; Berk, 1999). Note that an English language learner might have somewhat different weightings for these constraints, as she would be exposed to different inputs (e.g., inputs with no object complement) and would have a much larger set of constraints.

This training procedure yields a strong weighting to our faithfulness constraint. Although deleting elements of lexical conceptual structure allows many candidate outputs in Table 4 to avoid violations of the markedness constraints, they incur one, two, or three penalties from the faithfulness constraint, substantially lowering their harmony. The relative ranking of the markedness constraints determines which of the first four fully faithful

Table 4. ***Grammar fragment: English subject-verb-object word order**. Blank cells indicate the candidate does not violate the constraint. Note that since* COMPLEFT *has a weighting of 0, violations of the constraint do not decrease harmony. Probabilities are rounded; those less than 1×10^{-4} are represented as 0.*

	SPECLEFT	HEADLEFT	COMPLEFT	PARSE		
Input: *gave (they, grant)*	−13	−12	0	−25	*H*	*Pr*
[VP they [V′ gave grant]]		−12	0		−12	1
[VP they [V′ grant gave]]		−24	0		−24	0
[VP [V′ gave grant] they]	−26		0		−26	0
[VP [V′ grant gave] they]	−26	−12			−38	0
[VP they [V′ gave]]		−12		−25	−37	0
[VP [V′ gave] they]	−13			−25	−38	0
[VP they [V′ grant]]			0	−25	−25	0
[VP [V′ grant] they]	−13			−25	−38	0
[V′ grant gave]		−12		−25	−37	0
[V′ gave grant]			0	−25	−25	0
[VP they]				−50	−50	0
[V′ gave]				−50	−50	0
[V′ grant]				−50	−50	0
ø				−75	−75	0

candidates are selected. The second most-highly weighted constraint prefers specifiers occur to the left of the V′ projection, ruling out the third and fourth candidates. The third constraint, preferring that heads be leftmost, then rules out the second, yielding the subject-verb-object word order.

Languages like Tamil exhibit a contrasting context-neutral word order pattern, subject-object-verb (Sarma, 1999; Schiffman, 1999). Training on these data for the same input yields a contrasting weighting.

The sole change to the weighting is the relative strength of HEADLEFT and COMPLEFT. Now that the latter has a higher weighting, there is a reversal of the relative harmony of the first two candidates; object complements, not verbal heads, are leftmost in V′. This yields the appropriate subject-object-verb word order.

Code mixing in constraint-based grammars

Having demonstrated that our formalism can represent cross-linguistic differences in word order, we consider the grammars utilized by bilinguals (e.g., an English–Tamil bilingual). As reviewed above, in intra-sentential code mixing bilinguals integrate grammatical principles from each linguistic system. In GSC, grammars are defined by the weighting of constraints. We therefore formalize this integration by having the weights of constraints in the grammar underlying code mixing reflect both linguistic systems.

We propose to associate each linguistic system present in a code mixed utterance (L1, L2) with an activation value (α_{L1}, α_{L2}; the sum of these values must be 1). This scales the amount each linguistic system contributes to the modulation of each constraint's violations. Specifically, violations of each constraint C are scaled by the sum of the C's ranking in each linguistic system, weighted by the activation of that system. This allows for interactions between the two linguistic systems (as each contributes to harmony for every element). This scaling value is additionally increased by the activation of a linguistic system if the constraint refers to an element in that system. This latter factor encodes a (violable) preference for linguistic elements to obey the properties of the source language.

For example, suppose HEADLEFT has weighting –10 in L1 and –5 in L2; the activation of L1 is 0.75 and L2 0.25. An L1 head that has 1 position intervening between it and the edge of XP will incur a harmony penalty of 1 * [–10*(0.75+0.75) + 1*–5*(0.25)] = –16.25. An L2 head that has 1 position intervening between it and the edge of XP will incur a harmony penalty of 1 * [–10*(0.75) + 1*–5*(0.25+.25)] = –10.0. The L1 head incurs a slightly greater penalty because of the stronger weighting of this constraint within the source grammar.

Blends in grammatical representations

Weighted constraints are not a novel claim of GSC; these overlap with existing formalisms including Harmonic

Grammar and Maximum Entropy models. A novel feature of GSC is the incorporation of blends. Specifically, GSC proposes that elements of symbolic grammatical representations are associated with activation values. This includes all elements of syntactic representations: the nodes of the tree (both terminal and non-terminal elements) as well as the links between nodes. This allows for the specification of blends; multiple representational elements that would occupy a single position or role in a discrete symbolic representation can be co-present, to varying degrees. In previous work, we have examined the role that blends play in monolingual language processing. For example, in phonological speech errors (mispronouncing *bat* as *pat*), there is evidence that target and error sound representations are co-activated (the onset of the error syllable *pat* is a blend simultaneously containing elements of both /b/ and /p/; Goldrick & Chu, 2014; Smolensky et al., 2014). Here, we extend this very general representational principle to the domain of bilingualism, focusing on blends involving elements from two distinct source languages.

For our initial discussion, we focus on cases where multiple elements are co-present to the same degree; this suffices to illustrate the general analysis. We illustrate this with example (7), repeated below for convenience.
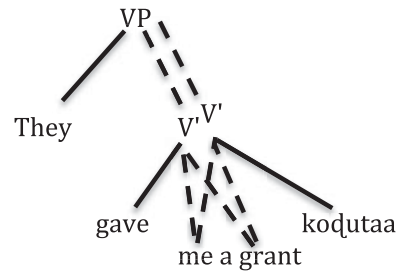
(14) Doubling: Verbs English–Tamil (Sankoff et al., 1990: 93)

they <u>gave</u> me a research grant <u>koḍutaa</u>

they gave me a research grant gave.3.PL.PAST

'They gave me a research grant.'

We analyze the input to the grammar in code mixing contexts as consisting of blends of semantic elements. For the example above, we analyze the input as the co-presence of two verbal elements – drawn from two distinct languages – which share multiple arguments (shown in (15) below). This representation instantiates a core claim of the blend analysis of doubling constructions: the simultaneous presence, in the input to the grammar, of the semantic representation underlying the doubled elements.

(15) $\genfrac{}{}{0pt}{}{\text{gave}}{\text{koḍutaa}}$ (3rd Plural, grant)

Blending in syntactic representations is a key part of our analysis of doubling constructions. In general, we analyze such constructions involving two V′ phrases, with distinct heads, that occupy the same position in the tree. These V′-projections share two complements: an indirect (me) and a direct object (grant), with these elements simultaneously associated to both V′. Figure 3 illustrates the representation hypothesized for (15). Note

[VP They [V′ gave [V′ me a grant] koḍutaa]]

Figure 3. Hypothesized blend structure for English-Tamil doubling construction *they gave me a (research) grant koḍutaa*. The text below each provides a bracket notation corresponding to the tree. Dashed lines highlight blended components of the representation.

that we adopt a ternary-branching structure for the double object construction. Our analysis does not hinge on this assumption; a binary branching VP-shell structure (Larson, 1988) could also include blends, and would yield similar results here.

In this example, there are two V′ (headed by *gave* vs. *koḍutaa*), sharing the complements *me* and *grant*. These two V′ simultaneously serve as the head of VP (a blend of nodes in the same position in the tree). This is a key part of our analysis, as it places each of the doubled elements in the same role within the syntactic structure. This sets up the structural relationships that ensure the doubled elements exhibit parallel tense, aspect, agreement, and case features.

Thus far, our discussion of blending has focused on the blending of elements within a structural position (e.g., two V′ projecting from the head position; two verbs occupying the same role in semantic structure). However, as each element of a representation can be associated with an activation value, blending is predicted to extend to the roles themselves. This is depicted in Figure 3, where the same arguments are shared across the two phrases. This represents a blending of positions or syntactic relationships; one node in the tree (e.g., the indirect object) has the same type of link to two distinct nodes (e.g., V′ headed by an English verb and the V′ headed by a Tamil verb). In our analysis, this sharing of the complements is critical, as it allows the blended structure to be linearized. If two lexical items simultaneously occupy the head of VP, there is no way to determine which one of them should occur first in the surface string; simultaneity means there is no precedence relationship between the phrases. However, because the precedence relationships WITHIN each phrase share a common element (*gave* precedes *me* and *grant*, *me* and *grant* precedes *koḍutaa*), a complete ordering of the terminals can be determined (by transitivity, *gave* must precede *koḍutaa*).

Table 5. ***Grammar fragment: Tamil subject–object-verb word order***. *Blank cells indicate the candidate does not violate the constraint. Note that since* SPECLEFT *has a weighting of 0, violations of the constraint do not decrease harmony. Probabilities are rounded; those less than 1×10^{-4} are represented as 0.*

	SPECLEFT	HEADLEFT	COMPLEFT	PARSE		
Input: *gave (they, grant)*	−13	0	−12	−25	*H*	*Pr*
[VP they [V' gave grant]]		0	−24		−24	0
[VP they [V' grant gave]]		0	−12		−12	1
[VP [V' gave grant] they]	−26		−12		−38	0
[VP [V' grant gave] they]	−26				−26	0
[VP they [V' gave]]		0		−25	−25	0
[VP [V' gave] they]	−13			−25	−38	0
[VP they [V' grant]]			−12	−25	−37	0
[VP [V' grant] they]	−13			−25	−38	0
[V' grant gave]		0		−25	−25	0
[V' gave grant]			−12	−25	−37	0
[VP they]				−50	−50	0
[V' gave]				−50	−50	0
[V' grant]				−50	−50	0
ø				−75	−75	0

As reviewed in preceding sections, there is ample empirical evidence that blends are subject to both physical constraints (e.g., unimodal bilinguals cannot place articulators in contradictory positions) as well as cognitive constraints (e.g., affective/pragmatic constraints, relative strength of the two languages, grammatical context). GSC suggests a clear theoretical motivation for some of these cognitive constraints (Smolensky et al., 2014). In many cases, purely grammatical constraints will prefer blends that do not reflect the structural principles of the source grammars. For example, if constraints prefer that all elements be at the left edge (i.e., specifier) of XP, why not place all elements in that position simultaneously? This candidate would satisfy all the constraints above; in the absence of other principles, it would have the highest harmony. By allowing the grammar to avoid making choices between structures exhibiting different word orders, this blend representation would prevent GSC from capturing key properties of cross-linguistic variation. GSC, unlike many previous connectionist proposals, therefore incorporates an explicit dispreference for blend representations (Smolensky et al., 2014). For the purposes of this discussion (focused on simultaneous presence of equally active elements), we represent this as a constraint that simply refers to the presence vs. absence of blended elements[5]:

(16) QUANTIZATION: "Candidates must be discrete symbolic structures."

For each blended structural element in candidate C, decrease C's harmony by 1.

It is important to emphasize that QUANTIZATION plays a key role in MONOLINGUAL grammars. Although we have emphasized their role in bilingual language processing, blend representations are a ubiquitous feature of monolingual processing as well (Melinger et al., 2014). ALL grammatical computations – not only those involved in code mixing – must therefore evaluate representations where multiple elements occupy the same structural role (Smolensky et al., 2014).

With respect to the current discussion, it is important to note that QUANTIZATION is VIOLABLE; other constraints can compel the presence of blends. The next section examines a situation in which this can occur.

Analysis of doubling constructions

Combining the results of the previous sections, we examine conditions under which doubling constructions can be produced. We limit ourselves to candidates that parse all elements of lexical conceptual structure or any possible subset of elements (in this code mixing case, this includes elements from multiple languages). Parsed elements are associated with appropriate syntactic

---

[5] See Smolensky et al., 2014, for details on the stochastic optimization processes that generalize this idea to varying levels of activation of

elements (which results in non-linear changes to relative harmony of different representational states).

Table 6. ***Grammar fragment: Doubling construction, Tamil–English code mixing*** (*see appendix for full set of candidates*). *Blank cells indicate the candidate does not violate the constraint. Probabilities are rounded; those less than 1×10^{-4} are represented as 0.*

Input: gave koḍutaa (3rd Plural, grant)	SpecLeft	HeadLeft	CompLeft	Parse	Quant	H	Pr
English: 0.5 activation	−6.5	−6	0	−12.5			
Tamil: 0.5 activation	−6.5	0	−6	−12.5			
Combined weighting	−13	−6	−6	−25	−8	*H*	*Pr*
a. [$_{XP}$ they [$_{X'}$ gave [$_{X'}$ grant koḍutaa]]]		−30	−12		−16	−58	0.039
b. [$_{XP}$ they [$_{X'}$ koḍutaa [$_{X'}$ grant] gave]]		−42	−12		−16	−70	0
c. [$_{XP}$ they [$_{X'}$ grant koḍutaa]]		−12	−6	−37.5		−55.5	0.48

positions. We extend this candidate set to include the blend structure depicted in Figure 3, simplifying the example by omitting the indirect object. Following the section on code mixing, the constraint rankings in Table 4 [English] and Table 5 [Tamil] are combined to determine the grammar used to evaluate code-mixed constructions. We include the QUANTIZATION constraint as a language-independent constraint, examining how its relative weighting affects the probability of blend constructions.

The tableau in Table 6 illustrates one ranking that qualitatively approximates the empirical distribution of doubling constructions – non-zero, but relatively small probability.

The observed doubling construction (candidate a) violates several markedness constraints:

- With respect to HEADLEFT, it receives a total harmony penalty of 30 due to:
  ○ Language general constraints: –24 = –6 * 4 violations (3 for *koḍutaa* and 1 for *gave*)
  ○ English-specific constraints: –6 = –6 * 1 violation (for *gave*)
  ○ Tamil-specific constraints: 0 = 0 * 3 violations (for *koḍutaa*)
- For COMPLEFT the penalty is –12:
  ○ Language general constraints: –12 = –6 * 2 violations (for *grant*)
  ○ English-specific constraints: 0 = 0 * 2 violations (for *grant*)
- Two violations of the QUANTIZATION constraint yield a penalty of –16 (for the two V′ simultaneously projecting from the head of VP, as well as *grant* occurring as a complement in both V′).

However, unlike candidates that delete English or Tamil verbs (e.g., candidate c), the observed doubling constructions avoid violations of the faithfulness constraint PARSE. The probability of the doubling construction relative to non-doubled candidates like (c) is

therefore related to the weighting of faithfulness relative to the markedness constraints above.

Assuming faithfulness has a strong enough weight to compel the presence of doubling, the attested candidate (a) will be preferred to unattested candidate (b) due to the influence of language-specific constraints. The two candidates incur equal violations of the language general constraints, but (b) incurs extra violations of language-specific constraints (2 additional English-specific violations for *gave*). So long as these language-specific constraints have a non-zero weighting, the grammar will assign higher probability to the attested form.

Predicted limitations on doubling constructions

Our analysis above focused on an example that, following the empirical patterns of doubling, consisted of two heads with a shared complement. We assume that doubled complements – specifically, arguments of verbs – are unattested because such structures would violate Chomsky's Theta Criterion, which states that "each argument bears one and only one theta-role, and each theta-role is assigned to one and only one argument" (1981: 35; see also Chan, 2003, 2008, 2009). For example, contrast the attested *they gave$_{English}$ grant gave$_{Tamil}$* with the unattested *they grant$_{Tamil}$ gave grant$_{English}$*. In the latter, the Theta Criterion is violated as both *grant$_{Tamil}$* and *grant$_{English}$* share the same role (theme). In contrast, in the attested example the theme is occupied by a single entity (*grant$_{English}$*, shared across the two verbs) so there is no violation. We assume that violation of this constraint either causes the grammar to categorically rule out such structures or, alternatively, greatly reduces their probability (if the Theta Criterion is realized via a strongly weighted constraint).

A novel prediction of our account is that different distributions of doubling should be observed for expletive vs. non-expletive elements. Expletive elements are those

that appear solely for structural considerations and are semantically vacuous (e.g., in *It's raining*; the pronoun *it* does not actually refer to a specific agent). Our analysis attributes the presence of doubling to the co-presence of multiple elements in the input to the grammar. This is reflected by the crucial role of faithfulness; PARSE provides an advantage for doubling constructions, in spite of their increased violations of alignment constraints and QUANTIZATION. This makes a novel prediction: we should not observe doubling of expletive elements (e.g., English *do*) alone. Following Grimshaw (1997, 2001, 2013, a.o.), the occurrence of such elements can be attributed to structural (i.e., markedness) constraints rather than to the presence of expletives in the input to the grammar. As they only appear to satisfy the structural requirements of other elements, our account predicts that expletives should not be doubled in isolation.

To make this concrete, consider a case where doubling could be predicted. English and Korean both utilize *do*-support in negatives (Grimshaw, 2013), but exhibit contrasting word order. Like the verb, Korean negatives (and *do*) appear following the object, the opposite of English:

(17) Chelswu-ka   ppang-ul   mek-ci ani

Chelswu-NOM book-ACC read-CI NEG

ha-ess-ta (Hagstrom, 1996: 169)
do-PAST-DECL

'Chelswu did not read the book.'

An account that attributed the presence of doubling to contrasting surface word orders would predict that doubling of either the verb, negative morpheme, or *do* alone could occur in Korean–English code mixing. In contrast, our analysis predicts that the doubling of *do* alone should not occur. The presence of *do* here reflects structural well-formedness constraints, triggered by the presence of negation (Grimshaw, 1997, 2013). There is no independent motivation to include this expletive aside from this. Thus, doubling of *do* alone would violate constraints such as QUANTIZATION while providing no benefit with respect to constraints such as PARSE. (In fact, insertion of expletive elements not present in the input may violate faithfulness constraints such as Grimshaw's FULLINT.)

Bilingual doubling cannot be analyzed as movement

Having outlined our proposal, we briefly consider whether existing analyses of doubling could provide an alternative to our analysis. Doubling of elements in monolingual grammars has been a focus of recent generative research (see, e.g., the contributions in Barbiers, Koeneman, Lekakou & van der Ham, 2008). This is typically analyzed as resulting from phonological realization of multiple links in a representational structure linking a syntactic element from its location in the surface syntactic structure to other distal locations in the syntactic tree (e.g., derivational chains; Jónsson, 2008; Nunes, 2004). Such an analysis does not appear to be tenable for the attested examples of bilingual doubling. Consider the Tamil–English example analyzed above, where there is doubling of the verb *gave*. There is no clear motivation for such movement in the grammar of either English or Tamil. Even if we were to entertain such an analysis, it would violate a basic principle of the locality of head movement – the head of a projection (here, the verb) cannot undergo movement within that projection (Abels, 2003).

Gradient co-activation and directions for future research

Gradient activation in blends – the key to accounting for the range of psycholinguistic data reviewed in the introduction to this paper – is clearly outside the scope of any traditional grammatical theory. In the GSC framework, such representations are possible inputs and outputs to the grammar, and are assigned Harmony values by constraints. Specifically, the violation of each constraint reflects the ACTIVATION of the constituents referred to by the constraint. For example (c.f. (13), emphasis added to show contrast in definitions):

(18) PARSE: "Lexical conceptual structure is parsed."

Decrease candidate C's harmony by **the activation of each element of lexical conceptual structure** that does not have a corresponding element in C's surface syntactic structure.

The incorporation of activation stems from principles of connectionist computation (Legendre et al., 1990; Smolensky & Legendre, 2006). GSC-representations are realized by real-valued activation vectors over simple processing units. Over the course of computation, activation spreads among these units via weights that implement grammatical constraints (Smolensky et al., 2014). Critically, the activation values are continuously updated; the network does not simply 'jump' from one gradient symbolic representational state to another. In order to insure that this continuous update respects the well-formedness conditions specified by the grammar, constraints must assign well-formedness values to the full range of intermediate, gradient representational states. Harmony therefore varies with the activation of each representational constituent.

In the context of code mixing, gradient activation of elements in the input will alter the relative probability of these elements appearing in the output. This is because violations of faithfulness constraints like PARSE will be

Table 7. ***Effects of variation in input activation: Strong bias towards English vs. Tamil***.
Note that competitors involving additional deletion of input elements have been omitted (due to violations of PARSE*, they have very low Harmony and thus output probability near 0).*

Input : gave.6 kodutaa.4 (3rd Plural)	SPECLEFT	HEADLEFT	PARSE		
English: 0.5 activation	−6.5	−6	−12.5		
Tamil: 0.5 activation	−6.5	0	−12.5		
Combined weighting	−13	−6	−25	*H*	*Pr*
[XP they [X′ gave]]		−12	−15	−27	0.82
[XP they [X′ kodutaa]]		−6	−22.5	−28.5	0.18

Table 8. ***Effects of variation in input activation: Weaker bias towards English vs. Tamil***.
Note that competitors involving additional deletion of input elements have been omitted (due to violations of PARSE*, they have very low Harmony and thus output probability near 0).*

Input : gave.56 kodutaa.44 (3rd Plural)	SPECLEFT	HEADLEFT	PARSE		
English: 0.5 activation	−6.5	−6	−12.5		
Tamil: 0.5 activation	−6.5	0	−12.5		
Combined weighting	−13	−6	−25	*H*	*Pr*
[XP they [X′ gave]]		−12	−16.5	−28.5	0.18
[XP they [X′ kodutaa]]		−6	−21	−27	0.82

scaled by activation. For a given weighting of PARSE, the harmony penalty incurred by deleting an element will be less if the element has a lower vs. higher activation value. Less active elements will therefore be more likely to be deleted. The tableaux in Table 7 and Table 8 illustrate this for a simple subject-verb sentence (here, we assume QUANTIZATION is strongly weighted, blocking the appearance of output blends).

In these tableaux, violations of PARSE are scaled by the activation of the input element. For example, in Table 7, the first candidate deletes *kodutaa*. Language general constraints assign a violation of −10 = 0.4 * −25 and Tamil specific constraints assign a violation of −5 = 0.4 * −12.5. Compare this to the first candidate in Table 8. Here, the violation of language general constraints increases to −11 = 0.44 * −25 and Tamil-specific constraints to −5.5 = 0.44 * −12.5. As the activation of an input element increases, the cost of deleting it also increases – it becomes more critical to preserve the element (reflected in the shift in output probabilities across the two tableaux).

Gradient activation of elements in the OUTPUT provides a mechanism for modeling the data reviewed in the first sections of this paper; gradient blends observed in phonological and articulatory processing in spoken and signed languages (e.g., the co-activation of <DOG> and <PERRO> during Spanish–English bilingual production, depicted in Figure 1). Clearly, gradient symbol structures have the expressive capability to represent such structures; throughout this discussion, we have assumed graded activation of elements of the input to the grammar. Our claim is that the degree of blending present in the output reflects GRAMMATICAL computations.

Outside of numerical simulations, the final blend states of our first implementation of the QUANTIZATION constraint (Smolensky et al., 2014) cannot be determined. In more recent work (Tupper and Smolensky, in progress) we have therefore developed new realizations of this constraint that are more amenable to analysis. Using these methods, we can calculate the optimal blend state predicted by the grammar.

To illustrate this approach, we considered the scenario shown in Figure 1 – the coactivation of two nouns in the head of an NP consisting of a determiner and noun. For this computation, we simplified our grammar, focusing only on QUANTIZATION and FAITHFULNESS (as Spanish and English agree on word order for nouns in these phrases). Following the scenario depicted in Figure 1, we assumed that English has greater activation than Spanish. As shown in Figure 4, this set of constraint weightings[6]
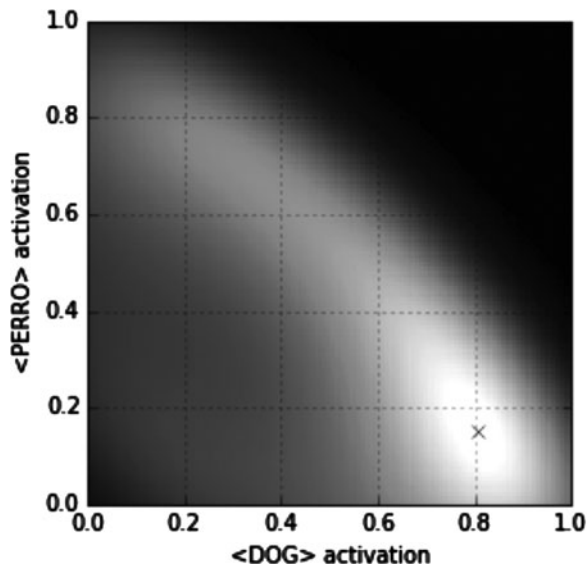
Figure 4. Relative Harmony (e^H) of various blends of
<DOG> (activation shown on X axis) and <PERRO>
(activation shown on Y axis); lighter color indicates higher
Harmony. The optimal blend (0.81 <DOG>, 0.15
<PERRO>) is marked with an *x*.

assigned highest Harmony to a blend state that is closest
to <DOG> (reflecting the higher activation of English
vs. Spanish), yet contains some partial activation of the
translation equivalent <PERRO> (reflecting the relative
weighting of QUANTIZATION). Critically, this degree of
blending is not assumed, but is rather DERIVED from the
constraints of the grammar.

Given input activations and relative constraint
weightings, a GSC theory will make predictions about
multiple facets of code mixed productions: discrete (e.g.,
output probabilities of various structures) as well as
gradient (e.g., coactivation of translation equivalents).
To develop this account, it is important that we gain a
more precise understanding of the factors that facilitate
(and inhibit) the activation of representations within each
of a bilingual's languages during sentence processing as
well as how bilinguals learn the relative weightings of
grammatical constraints. Critically, GSC provides us with
a framework that can integrate these various influences on
code mixing – allowing us to develop a unified account
of discrete and gradient properties of bilingual linguistic
knowledge and processing.

---

[6] QUANTIZATION's contribution to harmony is based on the
activation of each representational element e in a given structural
position: $\sum_e a_e^2 (1 - a_e)^2 + ([\sum_e a_e^2] - 1)^2$; this is weighted by –10.
The PARSE constraint, weighted at +1, is defined as $\sum_e a_e^i a_e$, where
a_e^i is the activation of each element in the input (<DOG> = +2,
<PERRO> = +1). Finally, following other Harmony networks
(Smolensky, 2006a), there is a contribution from unit Harmony (a term
ensuring the harmony maximum is a finite value): $\sum_e \frac{1}{2}(a_e - \frac{1}{2})^2$,
weighted at –11.

Conclusions

We have sought to bring together two traditions in
bilingual research. Studies of on-line behavior have
established that blend representations – where multiple
elements are co-present within a single structural position
– play a key role in bilingual language processing at all
levels of linguistic structure. Studies of code mixing have
emphasized the role that grammatical knowledge plays
in constraining bilingual sentence production. We used
the phenomenon of doubling to highlight the connection
between these two lines of research: the integration of
blend representations and grammar. To formally link
these two aspects of bilingual cognition, we introduced an
account of code mixing based in the Gradient Symbolic
Computation (GSC) formalism. Using violable, ranked
constraints, we characterized the probabilistic grammars
underlying code mixing. The ranking of such constraints
reflects the weighted sum of rankings in each language
involved in a code mixed utterance along with a contribu-
tion from the source language of each element. Crucially,
blend representations are part of the input and output of the
grammar. This provides a predictive account of doubling
constructions; specifically, we predict restrictions on the
insertion of expletive elements in blended structures. Fi-
nally, our approach can be extended to account for graded
blend representations in bilingual language processing.

The principles of our account of code mixing – blend
representations; probabilistic grammars with weighted
constraints – come from general principles of GSC.
They are not postulated to account for bilingual language
processing specifically, but rather reflect principles of the
cognitive system that hold for all speakers. Similarly,
the grammatical principles we use to account for code
mixing are the same principles that underlie non-code
mixed utterances. Our account therefore does not assume
that bilingualism in general or code mixing specifically
represents atypical, exceptional circumstances. That said,
these two aspects of linguistic cognition provide a key
test case for discovering the principles that underlie
the cognitive architecture of language processing. Code
mixing is an 'experiment' in the natural 'laboratory'
of bilingualism, revealing the interaction of blend
representations and grammar that is at the heart of
Gradient Symbolic Computation.

Appendix

Note that this grammar fragment predicts that in the
absence of doubling the most probable code mixed
productions are ones that insert the Tamil verb, either
in the Tamil or English word order. Note that both
types of constructions are empirically attested in code
mixing (Bhatt, 1997). Why does this occur in this specific
analysis? The HEADLEFT constraint has a weighting of

Table A1. *Grammar fragment: Doubling construction, Tamil–English code mixing, showing full set of candidates. Blank cells indicate the candidate does not violate the constraint. Probabilities are rounded; those less than 1×10^{-4} are represented as 0.*

Input :
gave
koḍutaa (3rd Plural, grant)

	SPECLEFT	HEADLEFT	COMPLEFT	PARSE	QUANT	H	Pr
English: 0.5 activation	−6.5	−6	0	−12.5			
Tamil: 0.5 activation	−6.5	0	−6	−12.5			
Combined weighting	−13	−6	−6	−25	−8	*H*	*Pr*
[VP they [V' gave [V' grant] koḍutaa]]		−30	−12		−16	−58	0.039
[VP they [V' koḍutaa [V' grant] gave]]		−42	−12		−16	−70	0
[VP they [V' gave grant]]		−12	−12	−37.5		−61.5	0.001
[VP they [V' koḍutaa grant]]		−6	−12	−37.5		−55.5	0.48
[VP they [V' grant gave]]		−24	−6	−37.5		−67.5	0
[VP they [V' grant koḍutaa]]		−12	−6	−37.5		−55.5	0.48
[VP [V' grant gave] they]	−39	−12		−37.5		−88.5	0
[VP [V' grant koḍutaa] they]	−39	−6		−37.5		−82.5	0
[VP [V' gave grant] they]	−39		−6	−37.5		−82.5	0
[VP [V' koḍutaa grant] they]	−39		−6	−37.5		−82.5	0
[VP they [V' gave]]		−12		−75		−87	0
[VP they [V' koḍutaa]]		−6		−75		−81	0
[VP [V' gave] they]	−19.5			−75		−94.5	0
[VP [V' koḍutaa] they]	−19.5			−75		−94.5	0
[V' gave grant]			−6	−75		−81	0
[V' koḍutaa grant]			−6	−75		−81	0
[V' grant gave]		−12		−75		−87	0
[V' koḍutaa grant]		−6		−75		−81	0
[VP they [V' grant]]			−6	−75		−81	0
[VP [V' grant] they]	−19.5			−75		−94.5	0
[VP they]				−112.5		−112.5	0
[V' gave]				−112.5		−112.5	0
[V' koḍutaa]				−112.5		−112.5	0
[V' grant]		0	0	−112.5		−112.5	0
Ø		0	0	−150		−150	0

0 in the Tamil linguistic system vs. −6 in the English system. If only one verb is retained (resulting in a violation of faithfulness), it is therefore more harmonic to retain the Tamil verb – it incurs fewer violations of alignment constraints. In this fragment we have also included harmonically bounded candidates to illustrate all possible representational output forms.

References

Abels, K. (2003) *Successive cyclicity, anti-locality, and adposition stranding*. Doctoral dissertation, University of Connecticut-Storrs.

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition, 15,* 517–530.

Azuma, S. (1993). Word order vs. word class: portmanteau sentences in bilinguals. In P.M. Clancy (Ed.), *Japanese/Korean linguistics 2* (pp. 193–204). Stanford: CSLI.

Backus, A. (1992). *Patterns of language mixing: a study of Turkish-Dutch bilingualism*. Wiesbaden: Harrassowitz.

Balukas, C., & Koops, C. (in press). Spanish–English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*.

Barbiers, S., Koeneman, O., Lekakou, M., & van der Ham, M. (eds.) (2008). *Syntax and semantics 36: Microvariation in syntactic doubling*. Bingley: Emerald.

Belazi, H. M., Rubin, E. J., & Toribio, A. J. (1994). Code switching and X-Bar theory: the Functional head constraint. *Linguistic Inquiry, 25,* 221–237.

Berk, L. M. (1999) *English syntax: From word to discourse.* Oxford: Oxford University Press.

Bhatt, R. (1997). Code-switching, constraints, and optimal grammars. *Lingua, 102,* 223–251.

Bishop, M. (2010). Happen can't hear: An analysis of code-blends in hearing, native signers of American Sign Language. *Sign Language Studies, 11,* 205–240.

Bokamba, E. (1988). Code-mixing, language variation, and linguistic theory: Evidence from Bantu languages. *Lingua, 76,* 21–62.

Bresnan, J. (2000). Optimal syntax. In J. Dekkers, F. van der Leeuw, & J. van de Weijer (eds.) *Optimality theory: phonology, syntax and acquisition* (pp. 334–385). Oxford: Oxford University Press.

Broekhuis, H., & Vogel, R. (eds.) (2013). *Linguistic derivations and filtering: Minimalism and optimality theory.* London: Equinox.

Carnie, A. (2010). *Constituent structure.* Oxford: Oxford University Press.

Chametzky, R. (2000). *Phrase structure: From GB to minimalism.* Malden, MA: Blackwell-Wiley.

Chan, B. H.-S. (2003). *Aspects of the syntax, the pragmatics, and the production of code-switching, with special reference to Cantonese-English.* New York: Peter Lang.

Chan, B. H.-S. (2008). Code-switching, word order and the lexical/functional category distinction. *Lingua, 118,* 777–809.

Chan, B. H.-S. (2009). Code-switching between typologically distinct languages. In: B. Bullock & A.J. Toribio (eds.), *The Cambridge handbook of linguistic code-switching* (pp. 182–198). Cambridge: Cambridge University Press.

Chomsky, N. (1981). *Lectures of government and binding.* Dordrecht: Foris.

Coppock, E. (2010). Parallel grammatical encoding in sentence production: evidence from syntactic blends. *Language and Cognitive Processes, 25,* 38–49.

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1283–1296.

Deuchar, M. (2005). Congruence and Welsh-English codeswitching. *Bilingualism: Language and Cognition, 8,* 255–269.

Di Sciullo, A.-M., Muysken, P., & Singh, R. (1986). Government and code-mixing. *Journal of Linguistics, 22,* 1–24.

Emmorey, K., Borinstein, H. B., Thompson, R., & Gollan, T. H. (2008). Bimodal bilingualism. *Bilingualism: Language and Cognition, 11,* 43–61.

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America, 89,* 395–411.

Furukawa, T. (2008). Head/complement relations: Portmanteau code-switching between Japanese and English. *Language and Information Science 6* (pp. 283–292). Tokyo: University of Tokyo, Institute for Integrated Cultural Studies, Division of Language Information Science.

Gardner-Chloros, P. (2009). Sociolinguistic factors in code-switching. In B.E. Bullock & A.J. Toribio (eds.) *The Cambridge handbook of linguistic code-switching* (pp. 97–113). Cambridge, UK: Cambridge University Press.

Goldrick, M. (2011). Utilizing psychological realism to advance phonological theory. In J. Goldsmith, J. Riggle, & A. Yu (eds.) *The handbook of phonological theory, 2nd edition* (pp. 631–660). Oxford: Wiley-Blackwell.

Goldrick, M. (2012). Neural network models of speech production. In M. Faust (Ed.) *Handbook of the neuropsychology of language* (vol. 1, Language processing in the brain: Basic science, pp. 125–145). Chichester, UK: Wiley-Blackwell.

Goldrick, M., & Chu, K. (2014). Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010). *Language, Cognition and Neuroscience, 29,* 452–458.

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less native-like. *Psychological Science, 25,* 1031–1036.

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In J. Spenader, A. Eriksson, Ö. Dahl (eds.), *Proceedings of the workshop on variation within Optimality Theory* (pp. 111–120). Stockholm, England: Stockholm University.

Grimshaw, J. (1997). Projection, heads and optimality. *Linguistic Inquiry, 28,* 373–422.

Grimshaw, J. (2001). *Economy of structure in OT.* ROA-434, Rutgers Optimality Archive, http://roa.rutgers.edu.

Grimshaw, J. (2013). Last resorts: A typology of *do*-support. In H. Broekhuis & R. Vogel (eds.) *Linguistic derivations and filtering: Minimalism and Optimality Theory* (pp. 267–295). United Kingdom: Equinox Publishing.

Hagstrom, P. (1996). Do-support in Korean: Evidence for an interpretive morphology. In H.-D. Ahn, M.-Y. Kang, Y.-S. Kim, & S. Lee (eds.) *Morphosyntax in generative grammar: Proceedings of the 1996 Seoul international conference on generative grammar* (pp. 169–180). Seoul, South Korea: Hankuk Publishing Co.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science, 15,* 409–414.

Hayes, B. (2009). MaxEnt Grammar Tool [Software]. Retrieved from http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry, 39,* 379–440.

Hicks, C. (2010). *Morphosyntactic doubling in code switching.* MA thesis, University of North Carolina-Chapel Hill.

Hicks, C. (2012). A dual-structure analysis of morphosyntactic doubling in code switching. In D. Ross (Ed.) *Studies in the linguistic sciences: Illinois working papers 2012* (pp. 44–57). http://hdl.handle.net/2142/35295.

Hsin, L., Legendre, G., & Omaki, A. (2013). Priming cross-linguistic interference in Spanish–English bilingual

children. In Baiz, S., Goldman, N., & Hawkes, R. (eds.) *Proceedings of the 37th annual Boston university conference on language development* (pp. 165–77). Somerville, MA: Cascadilla Press.

Jónsson, J. G. (2008). Preposition reduplication in Icelandic. In Barbiers, S., Koeneman, O., Lekakou, M., & van der Ham, M. (eds.) *Syntax and semantics 36: Microvariation in syntactic doubling* (pp. 403–417). Bingley: Emerald.

Joshi, A. (1985). Processing of sentences with intrasentential code switching. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (eds.) *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 190–205). Cambridge: Cambridge University Press.

Kootstra, G., van Hell, J., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language, 63,* 210–231.

Kroll, J. F., & Gollan, H. (2014). Speech planning in two languages: What bilinguals tell us about language production. In M. Goldrick, V. Ferreira, & M. Miozzo (eds.) *The Oxford handbook of language production* (pp. 165–181). Oxford: Oxford University Press.

Kuhn, J. (2003). *Optimality-theoretic syntax - A declarative approach*. Stanford: CSLI.

Larson, R. (1988). On the double object construction. *Linguistic Inquiry, 19,* 335–391.

Legendre, G. (2001). An introduction to optimality theory in syntax. In G. Legendre, J. Grimshaw, & S. Vikner (eds.). *Optimality-theoretic syntax* (pp. 1–27). Cambridge, MA: MIT Press.

Legendre, G., Grimshaw, J., & Vikner, S. (eds.) (2001). *Optimality-theoretic syntax.* Cambridge, MA: MIT Press.

Legendre, G., Miyata, Y., & Smolensky, P. (1990). Harmonic Grammar – A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the cognitive science society* (pp. 388–395). Hillsdale, NJ: Lawrence Erlbaum.

Legendre, G., Putnam, M., de Swart, H., & Zaroukian, E. (in press-a). Introduction. In G. Legendre, M. Putnam, H. De Swart, & E. Zaroukian (eds.), *Optimality-theoretic syntax, semantics, and pragmatics: From uni- to bidirectional optimization*. Oxford: Oxford University Press.

Legendre, G., Putnam, M., de Swart, H., & Zaroukian, E. (eds.) (in press-b). *Optimality-theoretic syntax, semantics, and pragmatics: From uni- to bidirectional optimization*. Oxford: Oxford University Press.

Legendre, G., & Schindler, M. (2010). Code switching in Urban Wolof: A case for violable constraints in syntax. *Revista Virtual de Estudos da Linguagem-ReVEL, 8,* 47–75.

Legendre, G., Wilson, C., Smolensky, P., Homer, K., & Raymond, W. (1995). Optimality and wh–Extraction. In J. Beckman, S. Urbanczyck, & L. Walsh (eds.) *Papers in Optimality Theory* (University of Massachusetts Occasional Papers 18, pp. 607–636). Amherst, MA: Graduate Linguistics Student Association.

Lohndal, T. (2013). Generative grammar and language mixing. *Theoretical Linguistics, 39,* 215–224.

MacSwan, J. (1999). *A minimalist approach to intrasentential code switching: Spanish-Nahuatl bilingualism in Central Mexico*. London/New York: Routledge.

MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language and Cognition, 3,* 37–54.

Mahootian, S. (1993). *A null theory of code-switching*. PhD dissertation, Northwestern University.

Melinger, A., Branigan, H. P., & Pickering, M. J. (2014). Parallel processing in language production. *Language, Cognition and Neuroscience, 29,* 663–683.

Menn, L., & Duffield, C. J. (2013). Aphasias and theories of linguistic representation: representing frequency, hierarchy, constructions, and sequential structure. *Wiley Interdisciplinary Reviews: Cognitive Science, 4,* 651–663.

Muysken, P. (1995). Code-switching and grammatical theory. In L. Milroy & P. Muysken (eds.) *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 177–198). Cambridge: Cambridge University Press.

Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.

Muysken, P. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition, 16,* 709–730.

Myers-Scotton, C. (1993). *Dueling languages: Grammatical structure in code-switching*. Oxford: Clarendon Press.

Myers-Scotton, C., & Jake, J. J. (1995). Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism: Language and Cognition, 17,* 511–525.

Nishimura, M. (1986). Intrasentential code-switching: The case of language assignment. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives* (pp. 123–143). Hillsdale: Lawrence Erlbaum Associates.

Nunes, J. (2004). *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics, 41,* 407–420.

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science, 33,* 999–1035.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134,* 427–459.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics, 18,* 581–618.

Poplack, S., Wheeler, S., & Westwood, A. (1989). Distinguishing language-contact phenomena: evidence from Finnish–English bilingualism. *World Englishes, 8,* 389–406.

Porterie-Gutierrez, L. (1988). *Étude linguistique de l'aymara septentrional (Pérou-Bolivie) (= Thèse Amerindia)*. Paris: A.E.A.

Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Technical report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ. Technical report CU-CS-696–93, Department of Computer Science, University

of Colorado, Boulder. Revised version, 2002: ROA-537, Rutgers Optimality Archive, http://roa.rutgers.edu. Published 2004, Oxford: Blackwell.

Pyers, J. E., & Emmorey, K. (2008). The face of bimodal bilingualism grammatical markers in American Sign Language are produced when bilinguals speak to English monolinguals. *Psychological Science, 19,* 531–535.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 110–146). Cambridge, MA: MIT Press.

Sankoff, D., Poplack, S., & Vannianiarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change, 2,* 71–101.

Sarma, V. M. (1999). *Case, agreement and word order: Issues in the syntax and acquisition of Tamil*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Schiffman, H. F. (1999). *A reference grammar of spoken Tamil*. Cambridge: Cambridge University Press.

Sells, P. (2001a). *Structure, alignment and optimality in Swedish*. Stanford: CSLI.

Sells, P. (Ed.) (2001b). *Formal and empirical issues in optimality theoretic syntax*. Stanford: CSLI.

Smolensky, P. (2006). Computational levels and integrated connectionist/symbolic explanation. In P. Smolensky & G. Legendre (eds.), *The harmonic mind: From neural computation to Optimality-Theoretic grammar. Vol. 2:, Linguistic and philosophical implications* (pp. 503–592). Cambridge, MA: MIT Press.

Smolensky, P., & Legendre, G. (2006). Formalizing the principles II: Optimization and grammar. In P. Smolensky & G. Legendre (eds.), *The harmonic mind: From neural computation to Optimality-Theoretic grammar. Vol. 1: Cognitive architecture* (pp. 207–234). Cambridge, MA: MIT Press.

Smolensky, P., Goldrick, M., & Mathis, D. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science, 38,* 1102–1138.

Spalek, K., Hoshino, N., Wu, Y. J., Damian, M., & Thierry, G. (2014). Speaking two languages at once: Unconscious native word form access in second language production. *Cognition, 133,* 226–231.

Starreveld, P. A., De Groot, A. M. B., Rossmark, B. M. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition, 17,* 258–276.

Stolz, T. (1996). Grammatical Hispanisms in Amerindian and Austronesian languages: The other kind of transpacific isoglosses. *Amerindia, 21,* 137–160.

Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry, 14,* 520–536.