



Integrated, Not Isolated: Defining Typological Proximity in an Integrated Multilingual Architecture

Michael T. Putnam*, Matthew Carlson and David Reitter

Center for Language Science, The Pennsylvania State University, Pennsylvania, PA, United States

OPEN ACCESS

Edited by:

Kleanthes K. Grohmann,
University of Cyprus, Cyprus

Reviewed by:

Lisa Hsin,
University of Alabama, United States
Terje Lohndal,
Norwegian University of Science and
Technology, Norway

*Correspondence:

Michael T. Putnam
mtp12@psu.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 30 June 2017

Accepted: 05 December 2017

Published: 04 January 2018

Citation:

Putnam MT, Carlson M and Reitter D
(2018) Integrated, Not Isolated:
Defining Typological Proximity in an
Integrated Multilingual Architecture.
Front. Psychol. 8:2212.
doi: 10.3389/fpsyg.2017.02212

On the surface, bi- and multilingualism would seem to be an ideal context for exploring questions of typological proximity. The obvious intuition is that the more closely related two languages are, the easier it should be to implement the two languages in one mind. This is the starting point adopted here, but we immediately run into the difficulty that the overwhelming majority of cognitive, computational, and linguistic research on bi- and multilingualism exhibits a monolingual bias (i.e., where monolingual grammars are used as the standard of comparison for outputs from bilingual grammars). The primary questions so far have focused on how bilinguals balance and switch between their two languages, but our perspective on typology leads us to consider the nature of bi- and multi-lingual systems as a whole. Following an initial proposal from Hsin (2014), we conjecture that bilingual grammars are neither isolated, nor (completely) conjoined with one another in the bilingual mind, but rather exist as integrated source grammars that are further mitigated by a common, combined grammar (Cook, 2016; Goldrick et al., 2016a,b; Putnam and Klosinski, 2017). Here we conceive such a combined grammar in a parallel, distributed, and gradient architecture implemented in a shared vector-space model that employs compression through routinization and dimensionality reduction. We discuss the emergence of such representations and their function in the minds of bilinguals. This architecture aims to be consistent with empirical results on bilingual cognition and memory representations in computational cognitive architectures.

Keywords: typological proximity, bilingualism, computational modeling, parallel architectures, vector space models

INTRODUCTION

The concept of typological proximity/distance has long been a useful one in language science, but despite its intuitiveness on many levels, it remains maddeningly difficult to measure in any large-scale sense. Part of the problem, we argue, is that its development and consequences at the diachronic vs. the synchronic levels have not yet been sufficiently articulated. Diachronically, a great deal of attention has long been paid to the evolution of grammars, from sound change to morphosyntax (Fedzechkina et al., 2012), and historical linguistics has made enormous contributions to our understanding of language, and provides (among other things) ways of understanding typological distance as instantiated in language phylogeny. However, in most cases, our only evidence of phylogenetic relationships are the synchronic correspondences among putatively related languages, meaning that diachronic measures of typological distance are generally

based on synchronic correspondences between languages. Frequently, lexical overlap forms the basis for these classifications, but where this fails, as in Papuan and Oceanic languages, researchers have attempted to make classifications on the basis of shared grammatical features (Dunn et al., 2005). The fundamental problem here is that the researcher must decide what grammatical features are to be used.

Moreover, typological relatedness in the synchronic sense plays an important role in understanding phenomena associated with bilingualism, including second language acquisition, language transfer, attrition, and code-switching, and in this domain, both genetic relationships among languages and also proximity due to convergent evolution are important. What is needed, therefore, is a general concept of typological proximity that can serve as a foundation for a metric that is independent of the source of that proximity, and one that is not based on arbitrary decisions made by the researcher (see e.g., also similar criticisms directed at the generative notion of “parameter” by Newmeyer, 2004, 2005). Specifically, to the extent that any human language can be situated within a common space of possible languages implies that typological distance is measurable synchronically as well, regardless of its source.

This more synchronic conceptualization of typological proximity has played a larger role in second language acquisition research and related sub-disciplines, both explicitly (in various instantiations of the idea of contrastive analysis, going back at least to Lado, 1957), and implicitly (Recchia et al., 2010). This research, too, has tended to focus on specific shared features or families of features, with the general intuition that second language learning proceeds more easily where there is overlap, and that contrast presents more challenges (though partial overlap may present the greatest challenges, e.g., Flege, 2007). The impact of correspondence and contrast between two grammars in second language acquisition is, however, just a specific instantiation of much more general questions about how two or more grammars are instantiated in the multilingual mind, questions that have garnered increasing attention in recent years (Grosjean, 1989; Cook, 1992, 1995; Kecskes, 1998; Roeper, 1999; Kecskes and Papp, 2000; Hall et al., 2006; Braunmüller, 2009; Amaral and Roeper, 2014; Grohmann, 2014; Cook and Wei, 2016, and references therein).

The major intuition in this research is that grammars are not instantiated side-by-side in the multilingual mind, but they are integrated into a single, compound system. What we argue here is that this integration provides a useful way of conceptualizing, and even measuring, the typological proximity of language pairs. It can thus fill the gap in understanding the direct, synchronic relatedness of two grammars, independent of the diachronic histories that brought them to that point. Moreover, via a second important intuition, that language change grows out of synchronic variation, and that the representation of variation can be thought of as a form of multilingualism, this synchronic view of typological proximity can be integrated with the diachronic one, leading to a more comprehensive view of how this proximity arises, and how it shapes the competence and usage of individual language users. Of course, the idea that multilingualism contributes to language change (thus contributing to linguistic

relatedness) has long been acknowledged, particularly in the subdiscipline of contact linguistics (Thomason and Kaufman, 1992 is but one substantial example). But the original focus there was on whether and how specific structures at various levels of linguistic description can pass from one language into another, whereas our proposal is much more comprehensive: that by conceptualizing the language knowledge of a multilingual (or a monolingual, counting variation) as a single grammatical system, and comparing the result to a coordinate system, where both languages are represented side by side, but independently, we can gain vital insight into the notion of typological distance.

Here we introduce the core aspects of an algorithm which can measure typological proximity/distance between languages. Importantly, our primary focus here is on modeling typological proximity in the bilingual mind, which requires the inclusion of a common, combined grammar that we will discuss below. The key to all of this, of course, is to approach a well-developed understanding of what it means for two languages to be “integrated” in one mind. Here we discuss the fundamental ontology of an integrated grammar and how typological similarities and differences can be accounted for in a clear and systematic way.

A HOLISTIC VIEW - INTEGRATED GRAMMARS

Research in cognitive neuroscience over the past three decades has provided a cascade of evidence that both languages are, to various degrees, simultaneously active in the mind of bilinguals (e.g., Hartsuiker et al., 2004; Deuchar, 2005; Pickering and Ferreira, 2008; Coppock, 2010; Hsin et al., 2013; Kroll and Gollan, 2014; Melinger et al., 2014; Starreveld et al., 2014; de Groot, 2016). Such research has gathered steam since initial proposals from pioneers such as Grosjean (1989) who advanced a holistic view of language, language development, and language use in bilinguals. The impact of this body of research issues significant challenges to research on modeling techniques that seek to better understand the emergence of grammar in individuals, and to an extent, our species. These findings have a profound impact on the (generative) models that we impose on the grammatical competence of multilinguals, as suggested by de Bot (2004), Hall et al. (2006), and Roeper (1999). Cook (2016, section 1.4) lists three primary premises regarding the role of the multi-competent native speaker:

Premise 1: Multi-competence concerns the total system for all languages (L1, L2, Ln) in a single mind and their inter-relationships.

Premise 2: Multi-competence does not depend on monolingual native speakers.

Premise 3: Multi-competence affects the whole mind, i.e., all language and cognitive systems, rather than language alone.

In the remainder of this article, we will focus primarily on the first two of Cook's premises, while acknowledging that we agree with the third and final premise, but will not address it directly due to space constraints (see e.g., Jarvis and Pavlenko, 2008). The initial

charge to treat bilingual grammars on par with monolingual grammars, i.e., as natural/authentic grammars, led to proposals such as the *Null Hypothesis* (Mahootian, 1993), which banned the postulation of constraints and representations that were strictly unique to bilingual grammars. In spite of these advances, work on bilingualism—especially research on the language of late bilinguals—tends to be “deficit”-oriented (Ortega, 2014), i.e., with the focus on differences between target outputs being the result of some sort of competence or production deficit of one of the source grammars.

This perspective can challenge the validity of treating both grammars in the mind of an individual as “natural languages.” In our integrated perspective, we adopt Ortega’s (2016, pp. 50–51) proposal—following initial proposals by Cook (2012, 2016)—that “linguistic competencies and indeed language itself are dynamic and they change at multiple time scales, including over the lifespan, as the function of actual use (Beckner et al., 2009; de Bot et al., 2013).” Of equal importance, the influence of one source grammar upon another need not be unidirectional; much research (e.g., Kecskes and Papp, 2000; Cook, 2003; Flege, 2007; and others) provides evidence that such influence is bidirectional. Finally, again as pointed out by Ortega (2016, p. 51), “language is part of cognition and, as such, cognition and language influence and affect each other (Langacker, 2008; Pulvermüller, 2013; Bylund and Athanasopoulos, 2014).” Here we sketch out the key underpinnings of an integrated cognitive architecture, while remaining true to Mahootian’s (1993) *Null Hypothesis* that bans the inclusion of features, operations, and constraints that are unique to bilingual grammars.

In the remainder of this article we take a bold step forward in attempting to unite these observations about the nature of multi-competence with current cognitive models and linguistic theorizing. Building upon an initial proposal by Hsin (2014), which we will explicate in more detail in the next section, we call for an integrated view of grammatical competence in the bilingual mind. To be clear, our adoption of an integrated grammar should not be confused with previous attempts in the generative tradition to come to terms with the simultaneous acquisition of grammar in bilingual children. In this literature, there are two dominant positions; the FUSED or UNIFIED DEVELOPMENT HYPOTHESIS (Volterra and Taeschner, 1978; Taeschner, 1983) and the ISOLATION HYPOTHESIS (Meisel, 1990). According to the former, the initial state would consist of a unified, or “common” grammar, which, over time, bilingual children would begin to gradually differentiate into (largely) separate source grammars. Hybrid representations found in code-mixing served as the initial empirical support for this hypothesis¹. In contrast, the latter hypothesis also draws on code-mixing data, but builds upon the observation that although there exists a high degree of lexical items from both source

grammars in such hybrid representations, the amount of cross-linguistic influence from both syntactic systems is relatively scarce. Meisel’s (1990) proposal that simultaneously developing grammars remain (mostly) isolated from one another thus contrasts with the former proposal. Hsin (2014, pp. 6–7) introduces a third option, which she calls the INTEGRATION HYPOTHESIS which “embodies an account in which bilingual children indeed begin with the same basic endowments [...] as monolingual children, and [...] where the two languages diverge with respect to a particular syntactic rule, the grammar responds by duplicating, or splitting, the constraint that is not satisfied for both languages.” We demonstrate below that these degrees of freedom are necessary in coming closer to an accurate, working model of multilingual competence.

Remaining consistent with the general theme of this *Frontiers* volume, we then explore how a model that adopts some version of the INTEGRATION HYPOTHESIS can accurately model the typological proximity (and, conversely, distance) between entire linguistic systems. As we discuss below, what is needed is a model that extends beyond the traditional notion of (innate) parameters (a concept that Cook, 1991, already began to adjust in his initial proposal of multicompetence), as suggested in the ongoing research carried out in the Principles and Parameters model (P&P, Chomsky, 1982) and beyond. Recent theorizing has sought to eliminate the reliance on such parameters for a number of reasons, opting instead for “realization options” (Boeckx, 2016, p. 90; also see Roeper, 2016 for similar ideas)². To briefly clarify this point, operations in the Narrow Faculty of Language (Hauser et al., 2002) are reduced significantly to notions of Merge, (possibly) recursion, and another subset of locally-defined operations (such as Agree and c-command) (see e.g., Chomsky et al., 2017 for a detailed overview of the current state of this research program). The generative component of this model is relatively unrestricted and unconstrained when compared with previous instantiations of the P&P-framework, where elements that were previously interpreted as catalysts for syntactic operations (e.g., Case, wh-movement, etc.), now become realization options external to the computational systems (i.e., at the hands of “external” interfaces). Under such assumptions, traditional “parameters” exist outside of the Narrow Faculty of Language (Hauser et al., 2002) and cross-linguistic variation is thus relegated to “third factor” considerations (Chomsky, 2005). We welcome this development for a number of reasons, most notably, because it presents a platform to unite theorizing traditionally thought to be unique to generative inquiry to a larger body of cognitive science. In the third section of this report, we discuss how these recent developments can be integrated into an emergent model of

¹As a point of clarification, our definition of hybrid representations is a cover term for linguistic outputs that contain elements from multiple source grammars (e.g., lexical items, morphological units, syntactic elements, etc.). This is not to be confused with more narrow definitions of hybridity provided by scholars such as Aboh (2015), which applies directly to creoles and mixed languages. Although we do not directly discuss how these grammars would fit into our model due to space considerations, it does appear that this would not pose significant difficulties.

²As pointed out by a *Frontiers* reviewer, there is also ongoing work to refine the notion of parameter in generative theorizing (e.g., Baker, 2001, 2009; Fábregas et al., 2015; Eguren et al., 2016; Biberauer and Roberts, 2017). Two particular recalcitrant issues concerning the notion of parameter that are of relevance to our model are: (1) the grain-size of parameters (e.g., Westergaard, 2013), and (2) the vertical and horizontal interaction of different sets of parameters (e.g., Biberauer and Roberts, 2017; Putnam, 2017). Here we adopt an agnostic approach to the notion of parameter, recognizing that in a multi-dimensional grammar space capturing the interaction and competition of these units is of primary importance.

language acquisition, such as that proposed and developed by MacWhinney (2005, 2008).

In the sections that follow, we flesh out our proposal of the general cognitive architecture that underlies a multi-competence language faculty. The fourth section of our report lays out the conceptual motivation and foundation for our model, while the fifth and final section advances a novel sketch of the core desiderata that would be deployed in such a system.

DYNAMIC INTEGRATION

If we are to move beyond the monolingual biases discussed by Cook (2012, 2016) and Ortega (2016) in an attempt to develop a cognitive architecture, we need to approach such an endeavor with our own set of axioms:

Axiom 1: Mental representations and their sub-components are lossy and gradient by nature. The reliability and stability of representations can be affected by myriad factors such as proficiency, working memory constraints, and activation/usage³.

Axiom 2: These mental representations only exhibit temporary “resting periods” or, *attractor states*, although these states may often be extremely stable and long-lasting.

Axiom 3: Parametric variation is no longer (primarily) tied to parameters licensed in a narrow computational faculty (i.e., the narrow syntax), and are now external from this core architecture.

Our first axiom shares many similarities with Cook’s Premise 1 to the extent that both assume the competence of bi/multilinguals to be an amalgamation of all contributing source grammars. The very existence of mental representations is of paramount importance in understanding and modeling cognition, as explained by Kühn and Cruse (2005, pp. 344–345):

By means of these representations, the behavior can be uncoupled from direct environmental control. This enables the organism, for example, to respond to features of the world that are not directly present, to use past experiences, to shape present behavior, to plan ahead, to manipulate the content, etc. (Cruse, 2003). All of these instances characterize a special feature of language called ‘displacement’ (Hockett, 1960). Therefore we conclude that these mental representations form an essential prerequisite to explaining how organisms can behavior in a cognitive way.

Importantly, in becoming a unified linguistic system⁴, this amalgamation must cope with varying degrees and

concentrations of correspondence between the source systems. This is what leads us to Axiom 1, where features that distinguish similar patterns across two source systems may play a lesser or greater role in representation and processing, depending on the usefulness of the commonalities.

The second Axiom grows out of an important fact about bilingualism, which is that usage patterns change over time. People may become bilingual at different times as well as to different degrees, and the balance of usage may shift toward or away from any given (source) language. Bilingualism thus demands a notion of grammar and mental representation that is generally stable, but underlyingly dynamic, much more clearly than monolingualism, where the underlying dynamism is much less apparent.

Lastly, concerning Axiom 3, the responsibility of the grammar is to generate environments where this unified grammar network can establish instances of congruency. This occurs in monolingual grammars, e.g., in the piecemeal acquisition of structures whose generality is not grasped by the child until later (Yang, 2002). Extending this reasoning to bilingual systems, the relatedness of structures in each language must be reflected in the representational resources at the core of the multilingual system (i.e., bilinguals’ knowledge of overlap in their systems does not merely stem from metalinguistic reflection). This in turn implies a novel view of typological relatedness as the degree to which congruency can be established across languages in a combined system.

We view the establishing of congruency and the architecture that it takes place in to be dynamic and emergent, but importantly, this does not eliminate the need for formal theorizing. On the contrary, as we argue here, this view of the cognitive architecture underlying the language faculty strongly supports the integration of competing linguistic information from multiple source grammars at designated points in the grammatical structure. In summary, the consequence of these axioms, and the integrated view we take here, for typological distance is that the kinds, amount, and degree of overlap or correspondence between source grammars is expected to strongly shape the way that the integration plays out, and that a global understanding of typological relatedness falls out from the way bilinguals use that overlap to build an integrated, multicompetent language system.

At this juncture, it will be useful to visit some of the data that motivates this view, and certain research topics where this kind of reasoning is developing, which in turn will provide guidance for where to seek further evidence and test the predictions that will grow out of a more formalized approach to the notions of overlap and equivalence in these aforementioned ways. As examples, we discuss evidence from the literature on code-switching, cross-linguistic structural priming, typological/genetic relationships, bilingual speech, and L2 acquisition.

An obvious domain where these notions of overlap and equivalence take center stage is in code-switching. Although code-switched utterances are frequently analyzable as relying on one source grammar, leading to a strong role in code-switching research for the idea that one grammar is in play at a time (e.g., Myers-Scotton, 2001), this is not always the case, and it has proven difficult—if not impossible—to derive absolute rules and

³Although the concept of lossy representations is commonly associated with constructionist and usage-based approaches (e.g., Goldberg, Forthcoming; Lau et al., 2016), it has also played a role in shaping generative approaches (e.g., Featherston, 2005, 2007; Pater, 2009; Goldrick et al., 2016a,b; Putnam and Klosinski, 2017).

⁴An interesting point about two languages in the mind of bilinguals is that they may merge into a unified system at different time scales. It may occur in the mind of an individual who is becoming bilingual (whether simultaneous or not), but it also occurs at the level of a bi- or multilingual speech community. Something like this reasoning already appears in research on pidgins and creoles, but what we discuss here is broader than that, and could apply to any situation where there is language contact.

strict constraints to account for these data. To address this point, Goldrick et al. (2016a,b) employ a probabilistic grammar model known as Gradient Symbolic Computation (GSC; Smolensky et al., 2014) that shares many affinities with other earlier versions of Harmonic Grammar (HG; Legendre et al., 1990; Smolensky and Legendre, 2006). To account for the fact that both grammars are active to various degrees in the mind of a bilingual, Goldrick et al. propose a calculation that determines the strength of each contributing source grammar as well as a “common” grammar (which is consistent with Cook’s Premise 1 discussed above). These values then interact with higher-level cognitive symbols (i.e., violable constraints), which evaluate input-output candidate representations to determine a Harmony profile for each pair. Importantly, the yielded Harmony value of each input-output pair contributes to the probability of occurrence of each representation relative to one another. In other words, every representation that has a non-zero probability of occurrence is essential for computing the probability of a particular form in relation to all possible forms. Hybrid representations containing various lexical and grammatical elements may differ to the extent that they include elements from source grammars (although it is a ubiquitous assumption that one of the source grammars functions as the matrix/dominant language in switches). Putnam and Klosinski (2017) extend the initial work of Goldrick et al. by investigating two different types of code-switches that vary with respect to the degree that both grammars contribute to hybrid representation. They make the distinction between MIXES and BLENDS (see also Chan, 2008, 2009), illustrated below in (1) and (2):

- (1) Mix: Welsh-English Determiner Phrase (Parafita Couto and Gullberg, 2016, p. 855):

y Belgian loaf
det^W Belgian^E loaf^E
‘the Belgian loaf’

- (2) Blend: Verb + Adverb English-Japanese (Nishimura, 1986, p. 139)

We bought about two pound gurai kattekita no.
we bought about two pounds about bought TAG
‘We bought about two pounds.’

Mixes are hybrid representations that consist of lexical items from both/multiple source grammars but appear to only follow one particular source grammar for structural purposes. In the mix-example in (1) above, the determiner phrase (DP) contains a mixture of lexical elements from both source grammars, but, crucially, only the English order of Det(eterminer)-Adj(ective)-N(oun) appears (Welsh: Det-N-Adj). In contrast, blends are representations where elements of both source grammars appear in the representation [i.e., the Verb + Adverb – Adverb + Verb orderings in the English-Japanese blend in (2)]⁵. The

work of Goldrick et al. (2016a,b) and Putnam and Klosinski (2017) provide a working metric to determine the activation and gradient nature of bi/multilingual representations. Importantly, this approach is consistent with an integrationist approach to the bilingual cognitive architecture as well as the Null Hypothesis (Mahootian, 1993) and Cook’s (2016) Premise 1 (listed above) for the following reasons: First, at no point do they assume that both grammars are truly either fused/united or isolated from one another. The activation and strength of representation (which could also be construed to be an analog for proficiency) of each grammar contributes to determine the value of a shared “common” grammar at a given point in time. It is crucial to reiterate that the value of this “common” grammar can be altered at a given time and over the course of a longer period of time due to a variety of mitigating factors; e.g., priming effects, lack of activation/usage, etc. Second, at no point does this model require features, constraints, or axioms that are solely unique to bi/multilingual cognition and its representations.

Additional evidence that forces us to revisit and better define the notions of overlap and equivalence comes from psycholinguistic data on syntactic priming effects (e.g., Bock, 1986; Branigan and Pickering, 1998; Bernolet et al., 2007; Schoonbaert et al., 2007). This work can provide valuable constraints pertaining to the nature of grammatical representations: how categorical they are, what is their granularity, and what are the mechanisms for general implicit (non-declarative) and procedural memories shared with those storing lexico-syntactic information. A model formulating syntactic storage within a hybrid symbolic/sub-symbolic cognitive architecture (Reitter et al., 2011) has seen several empirical predictions borne out (e.g., Kaschak et al., 2011; Segaert et al., 2016), including that such priming is modulated by the long-term activation (frequency) of syntactic information in the same way in L1 and in L2 speakers (Kaan and Chun, 2017). This lends credence to joint representational mechanisms (i.e., hybrid symbolic/subsymbolic representations), regardless of age of acquisition. As a case in point, Jacob et al. (2016) conducted two cross-linguistic priming experiments with L1 German-L2 English speakers where they investigated both the role of constituent order and level of embedding in cross-linguistic structural priming. The results of these experiments showed significant priming effects in connection with two factors: (i) whenever both languages shared the same constituent order, and (ii) when both languages were identical with regard to level of embedding.

This kind of cross-linguistic effect also extends to morphosyntactic features, such as gender. In a visual world study conducted by Morales et al. (2016), Italian-Spanish bilinguals and Spanish monolinguals listened to sentences in Spanish while viewing an array of pictures, one of which was the target in the sentence. The objects in the experiment depicted elements that either shared the same gender in both languages, or were mismatched with respect to gender assignment. Bilinguals looked less at the target object when its Italian gender mismatched its Spanish gender, suggesting that, to some degree, both gender features were active as the sentence was interpreted, even though the experiment was exclusively

⁵Technically speaking, in the model we propose here, there is no distinction between blends and mixes as suggested by Putnam and Klosinski (2017). From the perspective of a multi-dimensional architecture, the difference between the two examples above reduces to the number of levels where the two source grammars do (not) overlap.

in Spanish. Additional studies by Paolieri et al. (2010a,b) offer further evidence of morphosyntactic interactions in bilingual grammars, and research by Malt et al. (2015) demonstrates similar phenomena in the domain of semantics.

In the realm of bilingual speech, the mapping of phonetic space to speech sounds has long been an active area of inquiry (e.g., Flege, 2003; Best and Tyler, 2007; Gonzalez and Lotto, 2013), but recent evidence suggests integration of phonological systems at more abstract levels. Carlson et al. (2016) tested fluent, early Spanish-English bilinguals on a perceptual illusion related to Spanish phonotactics. Specifically, word-initial /s/-consonant clusters are prohibited in Spanish, and are obligatorily repaired by prepending an [e]. Presented with acoustic stimuli beginning with the illicit clusters, Spanish speakers tend to perceive an illusory [e], but this effect was lessened in Spanish-English bilinguals, and more so if they were dominant in English. Thus, properties of a second language can lead to more veridical perception of certain sound sequences. Similarly, properties of a speaker's L1 can confer an advantage in L2 speech perception, compared to native speakers, as seen in a study by Chang and Mishler (2012) on Korean-English bilinguals' perception of word-final unreleased stops in English. Word-final stops are obligatorily unreleased in Korean, but optionally so in English, which Chang and Mishler linked to a measurable advantage in perceiving the place of articulation in the absence of a stop release.

Considering these empirical issues together, an ideal architecture must account for the gradient nature of knowledge in the form of mental representations, which is sensitive to the possible overlap of grammatical information from two or more (competing) source grammars. These mental representations consist of multiple levels of linguistic information, which leads to the potential of both vertical and horizontal overlap and conflict. In addition to establishing and declaring the (typological) (dis)similarities of both source grammars, this architecture must also establish equivalence amongst categories and constraints in the conjoined "common grammar." In summary, and agreeing with Cook (2016, p. 18) once again, "the mental representation of language is a complex system with all sorts of internal and external relationships; it may be quite arbitrary to divide a bilingual system into separate areas, modules, and subsystems, that can be called languages in the plural."

The variable, dynamic nature of these mental representations is the result of an architecture that embraces the fact that competition amongst these factors is the norm rather than the exception. The final mental representations are thus conditioned and shaped by both internal and external factors that operate perhaps on different time frames and exhibit unique developmental histories. Such is the nature of a dynamic system, whose core attributes are listed by de Bot (2016, pp. 126–30):

- Sensitive dependence on initial conditions
- Complete interconnectedness
- Non-linearity in development
- Change and development through internal reorganization and interaction with the environment
- Systems are constantly changing

- Dependence on internal and external resources
- Systems may show chaotic variation over time
- Development is conceived of as an iterative process

According to the integrationist perspective taken here, in addition to the gradient nature of knowledge in the form of mental representations we also adopt these conditions. Importantly, as explained by MacWhinney (2005, p. 191), "What binds all of these systems together is the fact that they must all mesh in the current moment. One simple view of the process of meshing is that cues combine in an additive manner (Massaro, 1987) and that systems are partially decomposable (Simon, 1969)." An attractive outcome of viewing the language development, maintenance, and activation/usage of bilinguals as a dynamic system is that it stands to bring generative models more in line with emergent (Kirby, 1999) and Bayesian (Culbertson, 2010) approaches to the development of grammar systems.

The shift toward a dynamic system with gradient representations raises questions concerning the compatibility that such a model might share with currently existing frameworks. Again, here we seek to outline how much these current frameworks can handle these important architectural adjustments. In our view, representations that are "partially decomposable" (Simon, 1969) are best interpreted as distributed knowledge that combines to deliver complex representations. There are multiple ways to postulate how these complex representations come into existence, from the use of declarative and violable constraints (van Oostendorp et al., 2016; Putnam, 2017) to those that employ an architecture of grammar with an invariant computational syntax (Kandybowicz, 2009; Lohndal, 2013; Boeckx, 2014, 2016; Grimstad et al., 2014; Alexiadou et al., 2015; Riksem, 2017). Questions regarding the difficulty in arriving at the proper definitive set of universal parameters and the inability to determine if and how these constraints could combine to deliver complex representations had emerged in the work of Newmeyer (2004, 2005) and has led to a reappraisal of the role of the traditional notion of parameters (see e.g., Fábregas et al., 2015; Eguren et al., 2016). What the majority of these recent proposals have in common is the move from parameters to features and cues that are either distributed across multiple levels or realized as associations that are united with a particular combination of derivational units (as is the case in Distributed Morphology, DM). The result from this exploration is that these mental representations consist of multiple levels and simultaneously display complex and atomic natures (cf. Quine, 1940). Under such assumptions, both lexical items (= lexicon) as well as more complex units (= syntacticon) (a la Emonds, 2000) are generated in similar fashion. Gallego (2016, p. 157) suggests that such an approach, i.e., one where items in a lexicon and syntacticon (i.e., the storage of fused units—chunks—typically larger than a lexical item) exhibit a dual atom-complex nature, must address the following questions:

- Q1 : What is the set of morphosyntactic features {F} that UG provides?
- Q2 : How do these features bundle to form LIs (= lexical items)?

Q3 : Why is LI-internal structure opaque to computation?

Gallego (2016, pp. 157–158) advances a system,

where syntax recycles complex sound-meaning pairings as brand new units of computation, creating a loop between syntax and the lexicon (roughly as in Starke 2010). A way to conceive of this relation would be the warp the standard “Y Model” into what we would call a “U-Turn Model” collapsing the pre-syntactic lexicon and the post-syntactic interpretive components into a unique interface that would communicate with other cognitive modules (the C-I and S-M systems).

The proposal of such a U-Turn Model enables communication between coexistent systems that is “highly reminiscent of the syntagmatic-paradigmatic distinction” (Gallego, 2016, p. 158). Furthermore, such an architecture is similar in scope and design to other proposals in the literature such as those put forward by Uriagereka (2008) and Stroik and Putnam (2013). As pointed out by Stroik and Putnam, such an architecture supposes that the Faculty of Human Language is situated *within* the performance system, which resonates with connectionist models of cognitive processes and their neurobiological implementation. Neural networks consist of interwoven neural links that “criss-cross in a three-dimensional curved grid structure” and “this grid structure contains highly ordered neural overlaps” (Stroik and Putnam, 2013, pp. 14–15; see also Ramachandran, 2011; Weeden et al., 2012). We maintain that in this overlapped architecture, the common grammar shares a unified semantic representation, which connected with other levels of linguistic information (i.e., phonology, morphology, and syntax). Two additional points are in order here as well: First, appeal to a U-turn-type architecture does not implicate that (strictly) modular models are preferred over parallel ones. On the contrary, as discussed by van Oostendorp et al. (2016), a parallel architecture with limited degrees of overlap among the sub-domains of grammatical knowledge is fully capable of deriving modularity. Second, O’Donnell (2015) shows that not only is the overlap (to some degree) expected between domains of grammatical knowledge, but also that the notions of storage and computation should not be viewed as completely separate entities. In fact, the notion of *fragmented grammars* that he advances in his treatment of probabilistic parsing shares significant overlap with the integrated approach we develop here. We can translate this idea to other representations of grammar as well: a system involving some ranked “soft” constraints that are violable can explain empirical data showing when and why judgments of acceptability are graded (Keller, 2000; Haegeman et al., 2014)⁶.

This brings us once again to the notion of cross-linguistic proximity and the challenge of capturing and measuring this heuristic without the aid of traditional parameters. The move

away from traditional parameters toward e(xternal)-parameters raises interesting challenges for the ontology of a model (see e.g., Putnam, 2017; Putnam et al., 2017 for an overview). The challenge for finding proximity and congruence between two source grammars requires a multi-dimensional search. This situation is once again a bit more complex in the bilingual mind, where the separation of individual source grammars is essentially not possible. The notion of proximal and distal has been a cornerstone in research on cross-linguistic influence (CLI) in sequential L2 acquisition. For example, Kellerman (1995, p. 125) states:

At its simplest, the L1 can be seen as a direct cause of erroneous performance, especially where such performance is shown to vary systematically among learners with different L1 backgrounds...

The concept of “error” is difficult to define in the context of L2 (and in bilingual language production *a priori*), given that other factors such as language mode, cognitive load, and other extraneous factors can impact linguistic output. Once again, one of the primary culprits in this line of thinking is “monolingual bias,” that (i) “there are separate language systems that have an impact on each other,” and (ii) “languages exist as stable entities in our brain” (de Bot, 2016, p. 133). The very existence of joint representations (i.e., of competence) also address this problem, which reduces notions of language contact since Weinreich (1953) to (non-)facilitative “transfer” as a fluid continuum of bidirectional influence (e.g., Schmid and Köpcke, 2007; Seton and Schmid, 2016).

What is more, it is unclear at present how much of an aid or a hindrance similar linguistic information can be. McManus (2015) shows that the high degree of overlap between the aspectual and tense systems in English and French can pose a challenge for the acquisition of French as an L2, and the notion that substantial but incomplete overlap presents greater challenges to learners than more distant correspondences has long been current in the literature on bilingual phonetics (Flege, 2003; Best and Tyler, 2007). With respect to research on code-switching, establishing congruence across categories appears to be essential in hybrid outputs (Deuchar, 2005). To date, it is unclear what role typological proximity may play in L1 attrition, although current research hopes to provide some insight into this matter (Schwarz, in progress). In the next section, we take on the task of providing a detailed overview of the fundamental components of our model.

MODEL: CORE COMPONENTS

We put forward a dynamic model of linguistic representations that shares representations between languages that change over time in response to experience. We note that the encoding strategy in our model is not specific to language. Rather, it is an instantiation of general cognitive mechanisms that encode either declarative knowledge or procedural programs. As a consequence, proximity between languages is determined by how frequently shared representations are used in processing each language. We contrast the concept of this inherent proximity

⁶A Frontiers reviewer raises the question as to whether or not there are fundamental differences between the “soft” constraints that we mention here versus those found in Optimality Theory. Although they share the trait that they are violable in nature, the constraints we refer to above can combine with one another, resulting in additive gang-up effects as argued for in Harmonic Grammar (Pater, 2009).

from such proximity that is the result of a cultural-evolutionary process, which, of course, has resulted in more or less co-representation between any two given languages.

Typological distance is a result of divergence in a vector space that is spanned by the representations that define syntactic operations in each language. *Compression* describes how shared representations form between those grammar representations associated with each language. In cognitive psychology, *chunking* is an example of such a compression operation. Similar to chunking, we assume that these grammatical representations are built up empirically; yet, they are probabilistic (symbolic and subsymbolic), and they dynamically change as language is used. Compression facilitates the efficient encoding of constructions in each language, with a construction being represented by a vector in space. The *cosine* metric is a common method to characterize the distance (angle) between two vectors in representational space. Then, the mean distance between constructions situated in this shared vector space describes typological similarity. In other words, if languages are represented in similar areas of this space, they are deemed similar, and facilitate and interfere with each other. If languages end up in more distinct clusters, they are typologically more distant, and may interact to a lesser degree.

With this model, we embed linguistic representations in a more general program of distributed (semantic) representations that have been empirically successful in describing human memory, in a psychological sense (Landauer and Dumais, 1997; Jones and Mewhort, 2007) but also language, in an engineering context (Mikolov et al., 2013). All of these approaches define some way to compress representations—often, from an initial vector space with several hundred thousand dimensions into a vector space with, e.g., 300 dimensions. This compression, also called *dimensionality reduction*, achieves generalization of the acquired representations while preserving much of their distinctiveness.

In the following, we discuss four candidate algorithms for compression that can form part of the model. Not all of them make different predictions, but they represent different cognitive mechanisms that have different neuropsychological correlates. All of them share the idea of *compression*, in that they make storage of representations more efficient over space, and/or over access time, and all four could result in a loss of information (i.e., all are *lossy*).

Chunking

In light of limited memory resources, humans apply an effective technique to recognize commonly used combinations or sequences of signals, storing them as a single, declarative memory item. For example, the sequence **BBCPHDCIA** might be stored not as nine letters (exceeding most people's working memory capacity), but as three well-known acronyms, becoming an easily storable three-item sequence. Chunking has been found at many levels, from perceptual/sensory information to high-level reasoning. Efficient memory encoding, using chunking strategies, has been shown to be a hallmark of expertise (a classic of cognitive psychology: Chase and Simon, 1973). Chunks may capture lexicalized sequences of words, or they may bind related ideas. It bears repeating, that an appeal to chunking does not

necessarily come at the exclusion of a minimalist model of syntax/computation. As pointed out by Adger (2013, p.c.), the idea that the initial stages of language acquisition begin with a limited, yet invariant narrow syntax and then eventually move toward a system of chunked representations is not inconsistent with some versions of minimalist theorizing. Whether this assertion can be upheld is beyond the central claim of this paper; however, we would like to point out that the decomposition of these complex units (i.e., chunks) in order to determine the degree of typological similarity requires a compression-unpacking algorithm discussed here (see e.g., Christiansen and Chater, 2016).

Routinization

Chunking has its equivalent in learning procedures and sequences. A repeatedly successful sequence of cognitive operations may be combined into a larger one (Anderson, 2013). This principle may apply to goal-oriented actions in the same manner as to linguistic phrases, so that syntax can be represented as a system of routines (Jackendoff, 2002). Within a parallel architecture of grammar, distributed units of information (call them *features*) can become routinized within particular levels of grammars as well as in combination with others (with the aid of functional mapping). As particular combinations of language (on multiple levels) are more frequently used/activated, the units as a whole become easier to generate and comprehend, thus facilitating efficiency as well as reducing entropy in the prediction of immediately preceding units. These units become highly routinized. One interesting consequence of this interpretation of the generation and routinization of linguistic chunks is that it blurs the clear distinction between elements that are exclusively regarded as stored elements of declarative knowledge vs. those that are generated as the result of computational operations (cf. O'Donnell's (2015) notion of *fragmented grammars*).

Distributed Representations and Declarative Memory

In a distributed model meaning is represented as a composition of weighted references to other meanings, to episodic experiences, or in arbitrary feature space. The *distributional hypothesis* states that words that appear in the same context share (some) meaning. So, we begin with a feature space that is composed of as many dimensions as there are contexts (practically, documents or paragraphs in text). Then, each word is represented as a vector of binary values that describes which contexts the word occurs in, defining the meaning of a word in terms of its usage. Consequently, similar meanings are then represented in nearby locations, or *embeddings*, in this vector space.

The semantic space is optimized in order to maintain a unique representation of meanings while simultaneously ensuring computational efficiency. Throughout language use, it can also be gradually optimized to improve understanding or producing language in context by using predictions about related meanings. Earlier forms of vector space models, such as *Latent Semantic Analysis* (Deerwester et al., 1990), apply a

mathematical operation that reduces the dimensionality of such spaces systematically.

Modern architectures (e.g., Mikolov et al., 2013) are optimized to actually predict a word given its context, i.e., its left and right neighbors. Thus, while these representations can capture some local syntactic regularities, they are not designed to represent syntax more generally. However, it is easy to see that the algorithm that reduces dimensionality and thus determines the encoding is a form of compression: It allows for a more efficient representation of meaning. The representational principles associated with distributed semantic encoding are not limited to the domain of semantics. Syntactic knowledge has rich stochastic ties to semantic representations, and can be seen as configural constraints that display a mix of regularities and exceptions. Distributed representations may well be a neurologically and psychologically plausible framework for syntactic knowledge, and it is a technically realistic candidate (Kelly et al., 2013, 2017). At lexical, syntactic, and morphological levels, the overlap in semantic space and joint compressibility of lexicons associated

with two languages determine their mutual facilitation. As we discuss below in the section Testing Our Model, we suggest that these levels exist in parallel with their semantic counterparts occupying another layer of parallel structure (i.e., there exists only one shared semantic/conceptual structure).

We provide an example of joint representation in compressed vector spaces in **Figure 1**. Here, a 1-million-word corpus of parallel Romanian and English newspaper texts was used (Mihalcea and Pedersen, 2003). A semantic space was obtained from a term-document matrix (sample of 1,500 words, 1,050 documents per language), which associates each term with the documents (or paragraphs) it occurs in, and their frequencies. This space, thus, characterizes word meanings in terms of their co-occurrences. A recent, high-performing dimensionality reduction technique that has a neural implementation was used (t-Distributed Stochastic Neighbor Embedding, van der Maaten and Hinton, 2008) to produce a two-dimensional vector space shared by the two languages. **Figure 1** shows the words and their locations in space. Note that a plausible model of such

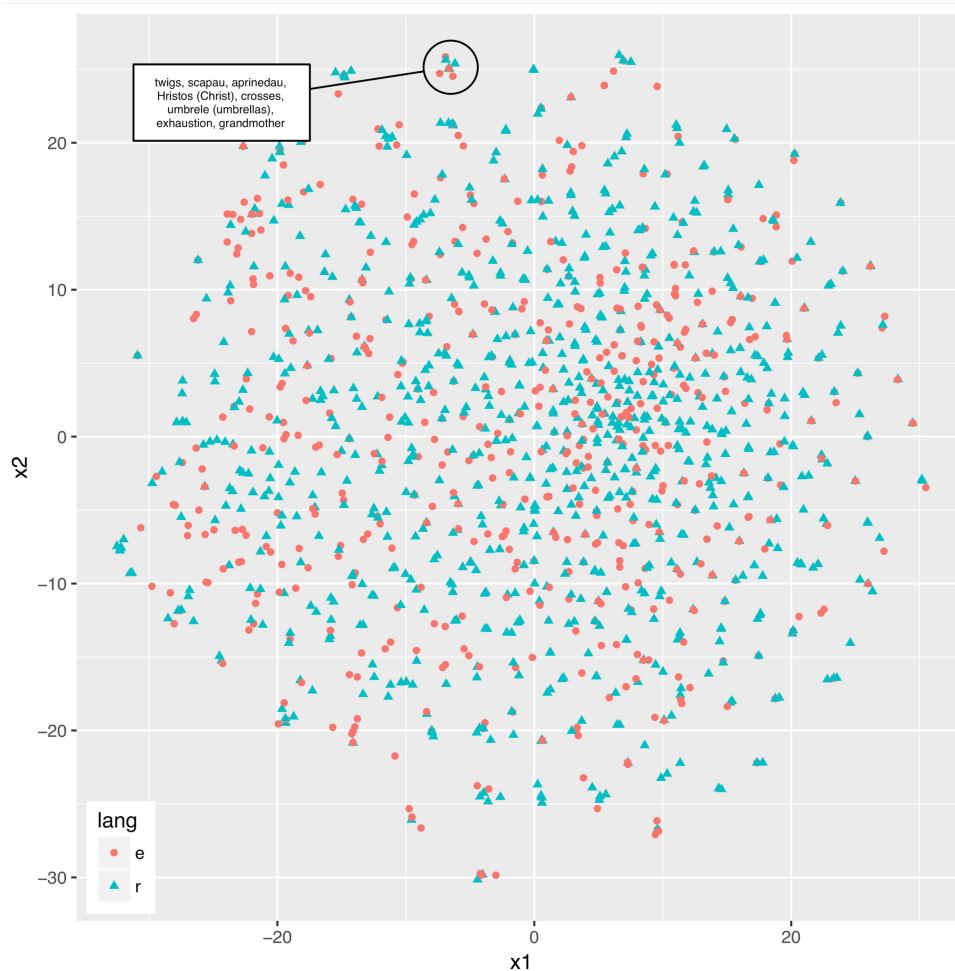


FIGURE 1 | Two languages sharing the same lexical-semantic space. Distributed semantic representations for 1,500 word samples were acquired from a parallel Romanian ("r") and English ("e") newspaper corpus and reduced to a two-dimensional vector space using T-SNE for demonstration purposes.

representations will use on the order of 300 dimensions rather than two, and it will result in regions of shared and regions of language-separate semantic-syntactic representations. Such a model would resemble models such as Kantola and van Gompel (2011) and de Bot (1992), in which related representations are connected (such that they can influence each other in processing), and unrelated representations are not, but in our model, the connected representations would be shared, as in Hartsuiker et al. (2016). However, the multi-dimensionality of the representational space would permit sharedness to be gradient. Although for the immediate purposes of illustration **Figure 1** only provides a two-dimensional representation, there is clear room for expansion of vector space that could include other level of grammatical information (e.g., morphology and syntax).

Compression Algorithms from Computer Science

Compression has its equivalent in computer science. In lossless compression, arbitrary but not random sequences can be represented by replacing frequent subsequences with references to an *ad hoc* table (Ziv and Lempel, 1978). A commonly used example of this principle (in an improved version) would be the popular *Zip* program. Lossy compression allows for differences between the target and the representation retrieved from memory, which is typically used where psychophysics or cognitive phenomena will prevent humans from perceiving the differences, such as in sound (e.g., *MP3*) or vision (e.g., *JPEG*). In terms of grammatical representations, our proposed model posits that compression of sequences of linguistic representations or of grammatical knowledge is a representation of encoding of grammar, including that shared between languages.

Grammatical representations start with declarative representations, which are routinized as a result of their use (see, e.g., Reitter et al., 2011 for such model, and Anderson, 2013 for a cognitive architecture that describes this routinization process). Repeated use of a sequence of memory retrievals leads to their compilation into fast routines that do not require memory retrievals. This process combines short subsequences first, and throughout repeated use, the resulting new chunks are combined again. This mirrors commonly used compression algorithms (Ziv and Lempel, 1978).

Grammatical encoding in the model depends on distributed representations to account for semantics, and it is possible that distributed (neural) representations can account for (some) syntactic knowledge as well. We propose an account of compression that combines this symbol-level compression and the notion of compression of representational spaces. At first sight, compression through routinization (psychology) or lookup tables (computer science) would be applicable to representations of syntactic procedures, while compression of semantic spaces would apply to distributed semantic representations. We propose that semantic and syntactic spaces are represented jointly. Compression at the symbolic level, akin to chunking, is lossless, while compression of representational spaces is lossy.

The architecture we discuss is compatible with accounts of “Shared Syntax,” and with empirical data that shows cross-linguistic priming (e.g., English-Spanish, Hartsuiker et al., 2004), as noted in the previous section. An instantiation of the

architecture will need to explicate how syntax is represented and how it is compressed; this would yield predictions for the facilitatory and inhibitory effects of an L2 on an L1 in language performance, a point which we turn to in the following section.

TESTING OUR MODEL

Admittedly, the programmatic model put forward here does not yet constitute a fully implemented model. The principles on which our proposal is founded nonetheless make testable predictions. For example, there is debate over whether structural priming effects are best accounted for through models in which bilinguals’ syntactic representations are shared vs. separate (Bernolet et al., 2007; Kantola and van Gompel, 2011; Hartsuiker et al., 2016). At first blush, our model shares a salient affinity with Hartsuiker et al.’s shared syntax model, but by integrating two grammars through lossy compression, we actually bring together both the separateness and sharedness of syntactic representations. We thereby aim to reconcile apparently conflicting findings by predicting when results will support a shared vs. separate (but interacting) syntax model.

As a second example, representational similarity as predicted by the account of distributed representations (and possibly more symbolic chunking) leads to observable behavior, such as facilitation of jointly represented constructions through syntactic priming and the difficulties encountered when attempting to rapidly compress grammatical information where typologically-contrastive information is present. As a point of illustration, consider the following hybrid representation reported by Karabag (1995) (cited by Treffers-Daller, 2017; German appears in regular font, Turkish in *italics*, doubled elements are underlined):

- (3) Deutschland muß mit dies-en Hippie-ler-le *ba?-a*
 Germany must mit this-dat Hippie-pl-instr. head-dat

çik-ma-si gerek-iyor.
 leave-nom-3sg must-Pr.Prog-Ø
 ‘Germany must cope with these hippies.’

In the example above (3), there are two instances of doubling, one involving the doubling of modal verbs (G: *muß*/T: *gerek*) and two adposition elements (G: *mit*/T: *le*). Congruence is established in the common grammar with respect to dative/instrumental case. Although German does not license independent morphosyntactic forms of instrumental case, it is subsumed as a sub-function of dative case in this language. In this structure, the lexical item *Hippie* is double-marked with dative/instrumental case, which we predict is likely due to the difficulty encountered in the common grammar to rapidly compress structural information (i.e., syntax) when one of the source grammars is a fusion-language (German) and the other an agglutinating-one (Turkish). To further illustrate this point, we provide a sketch of a formal analysis of this structure in (4) below making use of the Simpler Syntax framework (Culicover and Jackendoff, 2005).

- (4) Phonology: /mit dies-en Hippie-ler-le ba?-a çik-ma-si gerek-iyor /
 Morphology: [G: Dative/instrumental_dies-en / T: N-le]
 Syntax: [VP₁ V(modal) [Det-N(plural) V(leave)] V(modal) VP₂]
 Semantics: [Progressive(COPE([INSTRUMENTAL/DATIVE(with) [Hippies; DEF]]))]

Working from the bottom up, we see unified semantic content; i.e., “X must cope with these hippies” in this representation. With respect to the syntax, we propose two separate verb phrases (VP₁ and VP₂, respectively) that overlap where the modal verbs from each respective language appear at the edge of each VP based on the preference associated between the source grammar home of the verb and the VO- vs. OV-preference⁷. The lexical verb *gerek* “to leave” serves as the anchor of these overlapping VPs (see e.g., Chan, 2008, 2009, 2015; Goldrick et al., 2016a for similar arguments). We propose that there is an additional level of structural (i.e., syntactic) overlap with respect to the determiner/noun phrase (D/NP), with the noun *Hippie* serving as the anchor. Morphological marking indicating the dative/instrumental case on this noun appears both on an independent determiner (G: *dies-en* “these”) and as an agglutinating morpheme (T: *-le*). Crucially, the appearance of such hybrid units are dependent on some degree of congruence (i.e., the semantic representation and the recognition that dative case in German and Turkish instrumental case are approximate equivalents) as well as some degree of typological divergence (i.e., the fact that German is typologically classified as a fusion-language, whereas Turkish is agglutinating) in the source grammars contributing to the integrated, common grammar. Admittedly, models based on separate representations for each language could also account for data such as these, but we argue that an integrated model does so more naturally and, more importantly, it can do so in a way consistent with how bilinguals produce and process such structures in real time.

A third and final example of the role of typological relatedness in determining the possibility of doubled elements in the outputs of bilinguals comes from Austin’s (2015, 2017) research on the acquisition of Differential Object Marking (DOM) and preverbal complementizers in the speech of young, bilingual Basque-speaking children in contact with Spanish. These two particular languages contrast in significant ways, with Euskara, the Basque language, marking DOM-effects with a dative verbal suffix *zu-*, and Spanish realizing DOM-effects by means of a preverbal marking *a* [compare (5a) and (5b) below; both from Austin, 2015]:

(5a) Basque DOM⁸

Nik zuri entzun di- zu- t
 Erg1sg Dat2sg hear Abs3sg- Dat2sg- Erg1sg
 ‘I have heard you-Dat.’

⁷We thank Ray Jackendoff (p.c.) for engaging in discussion with us about this analysis.

⁸Austin (2015) notes that the Basque DOM effect in (5a) where the dative suffix appears is common in the dialect of Basque spoken in the Spanish Basque country but is not in the Basque spoken in France.

- (5b) Spanish DOM
 He visto **(a)* mi hija.
 Have-1sg seem DOM my daughter
 ‘I have seen my daughter.’

A second structural trait that distinguishes Basque and Spanish concerns the order of constituents in a clause; i.e., Basque is head-final language, whereas Spanish adheres to a head-initial ordering of constituents [compare (6a) and (6b); both from Austin, 2015]:

- (6a) Basque (head-final)
 Guk liburu asko irakurri dugu
 We-Erg book a lot read Aux-Abs3sg-Erg1sg
 ‘We have read a lot of books.’
- (6b) Spanish (head-initial)
 Nosotros hemos leído muchos libros
 We-Nom have-Nom1sg read many books
 ‘We have read many books.’

One of Austin’s key research questions focused on the appearance (or lack thereof) of preverbal complementizers in the speech of monolingual and bilingual Basque-speaking (and -acquiring) adults and children. To broadly summarize her findings, Austin’s study revealed the somewhat unexpected result that monolingual children produce more instances of DOM than bilingual children. With respect to the use of preverbal complementizers, Austin (2015, p. 10) provides the following rationale for her findings:

The use of pre-verbal complementizers presents a very different developmental pattern. These forms are used exclusively by four bilingual children⁹ between the ages of 2;08 and 3;02, and were never produced by monolingual children or adults. Five bilingual children in this age range never used them at all, and their production does not seem to be correlated with their MLU in Basque. [...] I understand these utterances as a temporary relief strategy which may be used by some bilingual children when they are confronted with a construction that they have not yet acquired, following proposals by Gawlitzek-Maiwald and Rosemary (1996) and Bernardini and Schlyter (2004).

Example (7) below (from Austin, 2015) illustrates this non-target structure, where a non-target preverbal complementizer appears:

- (7) zergatik badoa eskuelara
 why-Comp go-Abs.3sg school-to
 ‘Because s/he goes to school.’

Interpreting Austin’s findings through the lens of the integrated model of bilingual language and cognition that we adopt here, this relief strategy may likely be the result of elements from both source grammars simultaneously competing for a finite space of representation in syntactic structure. Under such conditions of typological contrast, the elements from

⁹Out of a total of 20 bilingual children who participated in the study.

both grammars with occasionally appear together, resulting in hybrid code-mixing. Of particular interest, these structures are only found in developing bilingual grammars, and crucially not in the speech of adults or monolingual children¹⁰. In summary, Austin's findings and suggested explanations are largely consistent with both ours and Muysken's (2013), where the linguistic output of bilinguals is the result of a (complex) optimization process.

As we have discussed throughout this article, grammatical knowledge—especially in the case of the bilingual mind—is best understood as a multi-dimensional, multi-vector space. In order to combat the need to produce and comprehend grammatical information with a high degree of efficiency, compression is applied whenever possible. One potential strategy to avoid the loss of (important) information, due in no small part to the lossy information within symbolic chunks, is to represent structural information twice. Although hybrid representations are commonly found in both typologically similar and dissimilar languages (see e.g., Braunmüller, 2009 and his work on code-switching among Danish-German and Danish-Faroese bilinguals), we suggest the presence of doubled elements in both developmental grammars and in simultaneous code-switching data represent solid evidence in favor of the dual activation of elements from both source grammars. Here we make the prediction that the difficulty to compress linguistic information into a common bilingual grammar consisting of source grammars that differ on at least one level of linguistic information, will lead to a higher degree of doubled structures. A preliminary survey of the nascent literature on doubled-elements in code-switches supports our hypothesis (see e.g., Chan, 2009, 2015; Goldrick et al., 2016a, and references therein). In contrast, we anticipate that doubled elements in hybrid representations will be far less likely in outputs when the source grammars exhibit higher degrees of (near) typological overlap. For example, we predict that it would be less probable to find doubled adpositions where both source grammars license prepositions [i.e., English-German **mit with der Seife* (with-G with-E the-dat soap)]¹¹. What is yet to be determined is how much overlap across which particular levels of grammatical information represents important thresholds for any particular increase in the appearance of such forms; however, such hypotheses are indeed testable through the analysis of

existing corpus data as well as experimental research with code-switching populations.

Finally, by implementing the compression algorithm in cognitively plausible ways (Anderson, 2013), our model aims to explain the various phenomena associated with bilingualism, as well as second language learning as grounded in the ways that general cognitive mechanisms interact with linguistic experience. With regard to second language acquisition in particular, a compression algorithm makes specific predictions about how learners perform the integration of new and existing linguistic knowledge, including differences based on specific language pairs or individual differences between learners.

CONCLUSION

Although sufficient evidence exists supporting the importance of typological similarity and distance in the acquisition, spontaneous speech, and attrition of bilingual grammars across the lifespan, deriving a working definition of this concept with predictive power has been a challenge in both generative and cognitive models of language. Here we make the case for a multi-dimensional, multi-vector network and a hybrid symbolic/sub-symbolic cognitive framework, which we deem to be necessary to model linguistic representations. In our view, this approach leads to a more accurate view of typological relatedness in both stored/routinized elements (i.e., lexical items and larger chunks) and their interaction with one another. Modeling bilingual grammar through the lens of this architecture, as we propose here, enables us to establish a depiction of the reality of dueling grammars and the routinization, chunking, and compression operations that take place in establishing congruence among elements of these grammars. There are many diagnostic tools that allow us to evaluate such a model, including code-switching phenomena, typological relatedness as evidenced by facilitation of L2 acquisition, or resistance to attrition. The model put forward here can be extended and adapted into various models, such as, but certainly not limited to, exo-skeletal frameworks of grammar where traditional parameters have been externalized from any sort of universal, or narrow computational system.

AUTHOR CONTRIBUTIONS

MP was the lead author for this manuscript, developing the outline and core ideas for this manuscript. DR contributed to the overall editing and added valuable insights from cognitive science and computational applications of language science. He was the primary author of section Model: Core Components. MC contributed throughout the paper to strengthening the overall argumentation of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Matthew A. Kelly for his helpful comments and inspiration regarding vector spaces. We would also like to recognize and thank Ray Jackendoff, Lara Schwarz, and Jeanine Treffers-Daller for suggestions and comments; additionally, we would like to acknowledge critical comments

¹⁰It is important to recognize here that although typological dissimilarity apparently plays a critical role here, at this point we cannot claim with any degree of certainty which linguistic and cognitive factors may be at play here. One possible candidate would be inhibitory control; if this were a dominant factor, it would indicate that these bilingual grammars have representations that are retrievable to varying degrees, but the lack of inhibitory control leads to doubled structures. Potential evidence in favor of this hypothesis come from work by Lipski (2009, 2014) on code-mixing in dysfluent (or low-proficiency) bilinguals. We leave this for future research and thank Nuria Sugarra (p.c.) for an insightful discussion with us on these matters.

¹¹Upon closer inspection, Babel and Pfänder's (2014) hypothesis that typology does not play a dominant role in "copying" phenomena does not pose a serious challenge for our model. Their argument that "speakers' perceptions of differences or similarities between languages are crucial to their development and change" (p. 254) is an appeal to the importance of congruence between simultaneously active grammars. Furthermore, in their proposal they do not address representations that contained doubled elements.

provided by colleagues at Rutgers University on this topic, in particular, Jennifer Austin, Nuria Sagarra, and Liliana Sánchez. DR acknowledges funding from the National Science Foundation (PAC/RI 1734304), and all three authors acknowledge seed

funding from the Social Science Research Institute (SSRI) at Penn State. Finally, we would like to thank the College of Information Sciences and Technology (IST) at Penn State for providing funds to cover publication costs.

REFERENCES

- Aboh, E. O. (2015). *The Emergence of Hybrid Grammars: Language Contact and Change*. Cambridge: Cambridge University Press.
- Adger, D. (2013). *A Syntax of Substance*. Cambridge, MA: MIT Press.
- Alexiadou, A., Lohndal, T., and Áfarli, T. A., Grimstad, M. B. (2015). "Language mixing: a distributed morphology approach," in *NELS 45, Proceedings of the forty-fifth Annual Meeting of the North East Linguistic Society*, ed Ö. Bui (Cambridge, MA: MIT), 25–38.
- Amaral, L., and Roeper, T. (2014). Multiple grammars and second language representation. *Second Lang. Res.* 30, 3–36. doi: 10.1177/0267658313519017
- Anderson, J. R. (2013). *The Architecture of Cognition*. New York, NY: Psychology Press.
- Austin, J. (2015). Transfer and contact-induced variation in child Basque. *Front. Psychol.* 5:1576. doi: 10.3389/fpsyg.2014.01576
- Austin, J. (2017). "The role of defaults in the acquisition of Basque ergative and dative morphology," in *The Oxford Handbook of Ergativity*, eds J. Coon, D. Massam, and L. de Mena Travis (Oxford: Oxford University Press), 647–664.
- Babel, A., and Pfänder, S. (2014). "Doing copying: why typology doesn't matter to language speakers," in *Congruence in Contact-Induced Language Change: Language Families, Typological Resemblance, and Perceived Similarity*, eds J. Besters-Dilger, C. Dermarkar, S. Pfänder, and A. Rabus (Berlin: Mouton de Gruyter), 239–257.
- Baker, M. (2001). *The Atoms of Language*. New York, NY: Basic Books.
- Baker, M. (2009). Language universals: abstract but not mythological. *Behav. Brain Sci.* 32, 448–449. doi: 10.1017/S0140525X09990604
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language as a complex adaptive system: positional paper. *Lang. Learn.* 59, 1–26. doi: 10.1111/j.1467-9922.2009.00533.x
- Bernardini, P., and Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: the Ivy hypothesis. *Bilingualism Lang. Cogn.* 7, 49–69. doi: 10.1017/S1366728904001270
- Bernolet, S., Hartsuiker, R. J., and Pickering, M. J. (2007). Shared syntactic representations in bilinguals: evidence for the role of word-order repetition. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 931–949. doi: 10.1037/0278-7393.33.5.931
- Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 13–34.
- Biberauer, T., and Roberts, I. (2017). "Parameter setting," in *The Cambridge Handbook of Historical Syntax*, eds A. Ledgeway and I. Roberts (Cambridge: Cambridge University Press), 134–162.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387. doi: 10.1016/0010-0285(86)90004-6
- Boeckx, C. (2014). *Elementary Syntactic Structures*. Cambridge: Cambridge University Press.
- Boeckx, C. (2016). "Considerations pertaining to the nature of logodiversity," in *Rethinking Parameters*, eds L. Eguen, O. Fernández-Soriano, and A. Mendikoetxea (Oxford: Oxford University Press), 64–104.
- Branigan, H. P., and Pickering, M. J. (1998). Syntactic priming in written production: evidence for rapid decay. *Psychon. Bull. Rev.* 6, 635–640. doi: 10.3758/BF03212972
- Braunmüller, K. (2009). "Converging genetically related languages: Endstation code mixing?" in *Convergence and Divergence in Language Contact Situations*, eds K. Braunmüller and J. House (Amsterdam: John Benjamins), 53–69.
- Bylund, E., and Athanasopoulos, P. (2014). Linguistic relativity in SLA: towards a new research program. *Lang. Learn.* 64, 952–985. doi: 10.1111/lang.12080
- Carlson, M. T., Goldrick, M., Blasingame, M., and Fink, A. (2016). Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism Lang. Cogn.* 19, 939–954. doi: 10.1017/S1366728915000334
- Chan, B. H.-S. (2008). Code-switching, word order and the lexical/functional category distinction. *Lingua* 118, 777–809. doi: 10.1016/j.lingua.2007.05.004
- Chan, B. H.-S. (2009). "Code-switching between typologically distinction languages," in *The Cambridge Handbook of Linguistic Code-Switching*, eds B. Bullock and A. J. Toribio (Cambridge: Cambridge University Press), 182–198.
- Chan, B. H.-S. (2015). Portmanteau constructions, phrase structure, and linearization. *Front. Psychol.* 6:1851. doi: 10.3389/fpsyg.2015.01851
- Chang, C. B., and Mishler, A. (2012). Evidence for language transfer leading to a perceptual advantage for non-native listeners. *J. Acoust. Soc. Am.* 132, 2700–2710. doi: 10.1121/1.4747615
- Chase, W. G., and Simon, H. A. (1973). "The mind's eye in chess," in *Visual Information Processing*, ed W. G. Chase (New York, NY: Academic Press), 215–281.
- Chomsky, N. (1982). *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (2005). Three factors in language design. *Linguist. Inq.* 36, 1–22. doi: 10.1162/0024389052993655
- Chomsky, N., Gallego, Á., and Ott, D. (2017). Generative grammar and the faculty of language: insights, questions, and challenges. *Catalan J. Linguist.*
- Christiansen, M. H., and Chater, N. (2016). *Creating Language: Integrating Evolution, Acquisition, and Processing*. Cambridge, MA: MIT Press.
- Cook, V. (2012). "Multi-competence," in *The Encyclopedia of Applied Linguistics*, ed C. A. Chapelle (New York, NY: Blackwell-Wiley), 3768–3774.
- Cook, V. (2016). "Premises of multi-competence," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 1–25.
- Cook, V. J. (1991). The poverty-of-the-stimulus argument and multicompetence. *Second Lang. Res.* 7, 103–117. doi: 10.1177/026765839100700203
- Cook, V. J. (1992). Evidence for multicompetence. *Lang. Learn.* 42, 557–591. doi: 10.1111/j.1467-1770.1992.tb01044.x
- Cook, V. J. (1995). Multi-competence and the learning of many languages. *Lang. Cult. Curric.* 8, 93–98. doi: 10.1080/07908319509525193
- Cook, V. J. (2003). *Effects of the Second Language on the First*. Clevedon: Multilingual Matters.
- Cook, V., and Wei, L. (eds.) (2016). *The Cambridge Handbook of Linguistic Multi-Competence*. Cambridge: Cambridge University Press.
- Coppock, E. (2010). Parallel grammatical encoding in sentence production: evidence from syntactic blends. *Lang. Cogn. Process.* 25, 38–49. doi: 10.1080/01690960902840261
- Cruse, H. (2003). The evolution of cognition - a hypothesis. *Cogn. Sci.* 27, 135–155. doi: 10.1016/S0364-0213(02)00110-6
- Culbertson, J. (2010). *Learning Biases, Regularization, and the Emergence of Typological Universals in Syntax*. Ph.D. dissertation, Johns Hopkins University.
- Culicover, P., and Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.
- de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Appl. Linguist.* 13, 1–24.
- de Bot, K. (2004). The multilingual lexicon: modeling selection and control. *Int. J. Multiling.* 1, 17–32. doi: 10.1080/14790710408668176
- de Bot, K. (2016). "Multi-competence and dynamic/complex systems," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 125–141.
- de Bot, K., Lowie, W. M., Thorne, S., and Verspoor, M. H. (2013). "Dynamic systems theory as a comprehensive theory of second language development," in *Contemporary Perspectives on Second Language Acquisition*, eds M. del Pilar García Mayo, M. J. Gutiérrez Mangado, and M. Martínez-Adrián (Amsterdam: John Benjamins), 167–189.

- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *J. Am. Soc. Inform. Sci.* 41:391.
- de Groot, A. M. B. (2016). "Language and cognition in bilinguals," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 248–275.
- Deuchar, M. (2005). Congruence and Welsh-English codeswitching. *Bilingualism Lang. Cogn.* 8, 255–269. doi: 10.1017/S1366728905002294
- Dunn, M., Terrill, A., Reesink, G., Foley, R. A., and Levinson, S. C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science* 309, 2072–2075. doi: 10.1126/science.1114615
- Eguren, L., Fernández-Soriano, O., and Mendoikoetxea, A. (eds.) (2016). *Rethinking Parameters*. Oxford: Oxford University Press.
- Emonds, J. (2000). *Lexicon and Grammar: The English Syntacticon*. New York, NY; Berlin: Mouton de Gruyter.
- Fábregas, A., Mateu, J., and Putnam, M. (eds.) (2015). *Contemporary Linguistic Parameters*. London: Bloomsbury.
- Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: some wh-constraints in German. *Lingua* 115, 1525–1550. doi: 10.1016/j.lingua.2004.07.003
- Featherston, S. (2007). Data in generative grammar: the stick and the carrot. *Theor. Linguist.* 33, 269–318. doi: 10.1515/TL.2007.020
- Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17897–17902. doi: 10.1073/pnas.1215776109
- Flege, J. E. (2003). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, eds N. O. Schiller and A. S. Meyer (Berlin: Mouton de Gruyter), 319–355.
- Flege, J. E. (2007). Language contact in bilingualism: phonetic system interactions. *Lab. Phonol.* 9, 353–382. Available online at: <https://www.degruyter.com/view/product/20932>
- Gallego, Á. J. (2016). "Lexical items and feature bundling: consequences for microparametric approaches to variation," in *Rethinking Parameters*, eds L. Eguren, O. Fernández-Soriano, and A. Mendoikoetxea (Oxford: Oxford University Press) 133–169.
- Gawlitze-Maiwald, I., and Rosemary, T. (1996). Bilingual bootstrapping. *Linguistics* 34, 901–926. doi: 10.1515/ling.1996.34.5.901
- Goldberg, A. (Forthcoming). *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton: Princeton University Press.
- Goldrick, M., Putnam, M., and Schwarz, L. (2016a). Co-activation in bilingual grammars: a computational account of code mixing. *Bilingualism Lang. Cogn.* 19, 857–876. doi: 10.1017/S1366728915000802
- Goldrick, M., Putnam, M., and Schwarz, L. (2016b). The future of code mixing research: integrating psycholinguistics and formal grammatical theories. *Bilingualism Lang. Cogn.* 19, 903–906. doi: 10.1017/S1366728916000390
- Gonzalez, K., and Lotto, A. J. (2013). A barfi, un pafri: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychol. Sci.* 24, 2135–2142. doi: 10.1177/0956797613486485
- Grimstad, M. B., Lohndal, T., and Åfarli, T. A. (2014). Language mixing and exoskeletal theory: a case study of word-internal mixing in American Norwegian. *Nordlyd* 41, 213–237. doi: 10.7557/12.3413
- Grohmann, K. K. (2014). Towards comparative bilingualism. *Linguist. Approach. Bilingualism* 4, 337–342. doi: 10.1075/lab.4.3.06gro
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain Lang.* 36, 3–15. doi: 10.1016/0093-934X(89)90048-5
- Haegeman, L., Jiménez-Fernández, Á. L., and Radford, A. (2014). Deconstructing the Subject Condition in terms of cumulative constraint violation. *Linguist. Rev.* 31, 73–150. doi: 10.1515/tr-2013-0022
- Hall, J. K., Cheng, A., and Carlson, M. T. (2006). Reconceptualizing multicompetence as a theory of language knowledge. *Appl. Linguist.* 27, 220–240. doi: 10.1093/applin/aml013
- Hartsuiker, R. J., Beerts, S., Loncke, M., Desmet, T., and Bernolet, S. (2016). Cross-linguistic structural priming in multilinguals: further evidence for shared syntax. *J. Mem. Lang.* 90, 14–30. doi: 10.1016/j.jml.2016.03.003
- Hartsuiker, R. J., Pickering, M. J., and Velkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychol. Sci.* 15, 409–414. doi: 10.1111/j.0956-7976.2004.00693.x
- Hauser, M. D., Chomsky, N., and Tecumseh Fitch, W. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569
- Hockett, C. (1960). The origin of speech. *Sci. Am.* 203, 88–96. doi: 10.1038/scientificamerican0960-88
- Hsin, L. (2014). *Integrated Bilingual Grammatical Architecture: Insights from Syntactic Development*. Ph.D. dissertation, Johns Hopkins University.
- Hsin, L., Legendre, G., and Omaki, A. (2013). "Priming cross-linguistic interference in Spanish-English bilingual children," in *Proceedings of the 37th Annual Boston University Conference on Language Development*, eds S. Baiz, N. Goldman, and R. Hawkes (Somerville, MA: Cascadia Press), 165–177.
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York, NY: Oxford University Press.
- Jacob, G., Katsika, K., Family, N., and Allen, S. E. M. (2016). The role of constituent order and level of embedded in cross-linguistic structural priming. *Bilingualism Lang. Cogn.* 20, 269–282. doi: 10.1017/S1366728916000717
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. New York, NY; London: Routledge.
- Jones, M. N., and Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* 114, 1–37. doi: 10.1037/0033-295X.114.1.1
- Kaan, E., and Chun, E. (2017). Priming and adaptation in native speakers and second-language learners. *Bilingualism Lang. Cogn.* doi: 10.1017/S1366728916001231. Available online at: <https://www.cambridge.org/core/journals/bilingualism-language-and-cognition/article/priming-and-adaptation-in-native-speakers-and-secondlanguage-learners/9A45D19C3FD53DA92623E254DED8F78E>
- Kandybowicz, J. (2009). Externalization and emergence: on the status of parameters in the Minimalist Program. *Biolinguistics* 3, 93–98. Available online at: <http://www.biolinguistics.eu/index.php/biolinguistics/article/view/59/96>
- Kantola, L., and van Gompel, R. P. (2011). Between-and within-language priming is the same: evidence for shared bilingual syntactic representations. *Mem. Cogn.* 39, 276–290. doi: 10.3758/s13421-010-0016-5
- Karabag, I. (1995). "Türkisch-deutsche Sprachmischung bei Rückkehrern: Ein Arbeitsbericht," in *Zwischen den Sprachen: Sprachgebrauch, Sprachmischung, und Sprachfähigkeiten türkischer Rückkehrer aus Deutschland*, eds J. Treffers-Daller and H. Daller (Istanbul: The Language Center, Bogaziçi University), 133–155.
- Kaschak, M. P., Kutta, T. J., and Jones, J. L. (2011). Structural priming as implicit learning: cumulative priming effects and individual differences. *Psychon. Bull. Rev.* 18, 1133–1139. doi: 10.3758/s13423-011-0157-y
- Kecskes, I. (1998). The state of L1 knowledge in foreign language learners. *Word* 49, 321–340. doi: 10.1080/00437956.1998.11432476
- Kecskes, I., and Papp, T. (2000). *Foreign Language and Mother Tongue*. London: Routledge; Psychology Press.
- Keller, F. (2000). *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. Ph.D. dissertation, University of Edinburgh.
- Kellerman, E. (1995). Crosslinguistic influence: transfer to nowhere? *Annu. Rev. Appl. Linguist.* 15, 125–150. doi: 10.1017/S0267190500002658
- Kelly, M. A., Reitter, D., and West, R. L. (2017). "Degrees of separation in semantic and syntactic relationships," in *Proceedings of 15th International Conference on Cognitive Modeling* (Warwick).
- Kelly, M. A., Blostein, D., and Mewhort, D. J. K. (2013). Encoding structure in holographic reduced representations. *Can. J. Exp. Psychol.* 67, 79–93. doi: 10.1037/a0030301
- Kirby, S. (1999). *Function, Selection, and Innateness: The Emergence of Language Universals*. Oxford: Oxford University Press.
- Kroll, J. F., and Gollan, H. (2014). "Speech planning in two languages: what bilinguals tell us about language production," in *The Oxford Handbook of Language Production*, eds M. Goldrick, V. Ferreira, and M. Miozzo (Oxford: Oxford University Press), 165–181.
- Kühn, S., and Cruse, H. (2005). Static mental representations in recurrent neural networks for the control of dynamic behavioural sequence. *Connect. Sci.* 17, 343–360. doi: 10.1080/09540090500177638
- Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.

- Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240.
- Langacker, R. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Lau, J. H., Clark, A., and Lappin, S. (2016). Grammaticality, acceptability, and probability. *Cogn. Sci.* 41, 1202–1241. doi: 10.1111/cogs.12414
- Legendre, G., Miyata, M., and Smolensky, P. (1990). "Harmonic Grammar: a formal multi-level connectionist theory of well-formedness: theoretical foundations," in *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society* (Cambridge, MA: Lawrence Erlbaum), 388–395.
- Lipski, J. (2009). 'Fluent dysfluency' as congruent lexicalization: a special case of radical code-mixing. *J. Lang. Contact* 2, 1–39. doi: 10.1163/000000009792497742
- Lipski, J. (2014). Spanish-English code-switching among low-fluency bilinguals: towards an expanded typology. *Sociolinguist. Stud.* 8, 23–55. doi: 10.1558/sols.v8i1.23
- Lohndal, T. (2013). Generative grammar and language mixing. *Theor. Linguist.* 39, 215–224. doi: 10.1515/tl-2013-0013
- MacWhinney, B. (2005). The emergence of linguistic form in time. *Connect. Sci.* 17, 191–211. doi: 10.1080/09540090500177687
- MacWhinney, B. (2008). "A unified model," in *Handbook of Cognitive Linguistics and Second Language Acquisition*, eds P. Robinson and N. C. Ellis (New York, NY: Routledge), 341–371.
- Mahootian, S. (1993). *A Null Theory of Code-Switching*. Ph.D. dissertation, Northwestern University.
- Malt, B. C., Li, P., and Pavlenko, A. (2015). What constrains simultaneous mastery of first and second language word use? *In. J. Bilingualism* 20, 684–699. doi: 10.1177/1367006915583565
- Massaro, D. (1987). *Speech Perception by Ear and Eye*. Hillsdale, NJ: Lawrence Erlbaum.
- McManus, K. (2015). L1-L2 differences in the acquisition of form-meaning pairings: a comparison of English and German learners of French. *Can. Modern Lang. Rev.* 71, 51–77. doi: 10.3138/cmlr.2015.2070
- Meisel, J. (1990). *Two First Languages: Early Grammatical Development in Bilingual Children*. Dordrecht: Foris.
- Melinger, A., Branigan, H. P., and Pickering, M. J. (2014). Parallel processing in language production. *Lang. Cogn. Neurosci.* 29, 663–683. doi: 10.1080/23273798.2014.906635
- Mihalcea, R., and Pedersen, T. (2003). "An evaluation exercise for word alignment," in *Proceedings of the HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation And Beyond*, Vol. 3 (Edmonton, AB: Association for Computational Linguistics), 1–10.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). "Distributed representations of words and phrases and their compositionality," in *NIPS'13 Proceedings of the 26th International Conference on Neural Information Processing Systems* (Lake Tahoe, NV), 3111–3119.
- Morales, L., Paolieri, D., Dussias, P. E., Kroff, J. R. V., Gerfen, C., and Bajo, M. T. (2016). The gender congruency effect during bilingual spoken-word recognition. *Bilingualism Lang. Cogn.* 19, 294–310. doi: 10.1017/S1366728915000176
- Muysken, P. (2013). Language contact outcomes as the results of bilingual optimization strategies. *Bilingualism Lang. Cogn.* 16, 709–730. doi: 10.1017/S1366728912000727
- Myers-Scotton, C. (2001). The matrix language frame model: development and responses. *Trends Linguist. Stud. Monogr.* 126, 23–58. doi: 10.1515/9783110808742.23
- Newmeyer, F. (2004). Against a parameter-setting approach to language variation. *Linguist. Var. Yearb.* 4, 181–234. doi: 10.1075/liv.4.06new
- Newmeyer, F. (2005). *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*. Oxford: Oxford University Press.
- Nishimura, M. (1986). "Inter-sentential code-switching: the case of language assignment," in *Language Processing in Bilinguals: Psycholinguistic and Neuropsychological Perspectives*, ed J. Vaid (Hillsdale: Lawrence Erlbaum Associates), 123–143.
- O'Donnell, T. (2015). *Productivity and Reuse in Language: A Theory of Linguistic Computation and Storage*. Cambridge, MA: MIT Press.
- Ortega, L. (2014). "Ways forward for a bi/multilingual turn in SLA," in *The Multilingual Turn: Implications for SLA, TESOL, and Bilingual Education*, ed S. May (New York, NY: Routledge), 32–53.
- Ortega, L. (2016). "Multi-competence in second language acquisition: inroads into the mainstream?" in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 50–76.
- Paolieri, D., Cubelli, R., Macizo, P., Bajo, T., Lotto, L., and Job, R. (2010a). Grammatical gender processing in Italian and Spanish bilinguals. *Q. J. Exp. Psychol.* 63, 1631–1645. doi: 10.1080/17470210903511210
- Paolieri, D., Lotto, L., Morales, L., Bajo, T., Cubelli, R., and Job, R. (2010b). Grammatical gender processing in romance languages: evidence from bare noun production in Italian and Spanish. *Eur. J. Cogn. Psychol.* 22, 335–347. doi: 10.1080/09541440902916803
- Parafita Couto, M. C., and Gullberg, M. (2016). *Code-Switching within the Noun Phrase - Evidence from Three Corpora*. University of Leiden and Lund University.
- Pater, J. (2009). Weighted constraints in generative linguistics. *Cogn. Sci.* 33, 999–1035. doi: 10.1111/j.1551-6709.2009.01047.x
- Pickering, M. J., and Ferreira, V. S. (2008). Structural priming: a critical review. *Psychol. Bull.* 134, 427–459. doi: 10.1037/0033-2909.134.3.427
- Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* 17, 458–470. doi: 10.1016/j.tics.2013.06.004
- Putnam, M. (2017). Feature reassembly as constraint satisfaction. *Linguist. Rev.* 34, 533–567. doi: 10.1515/tlr-2017-0010
- Putnam, M., and Klosinski, R. (2017). The good, the bad, and the gradient – The role of 'losers' in code-switching. *Linguist. Approach. Bilingualism*. doi: 10.1075/lab.16008.put. [Epub ahead of print].
- Putnam, M., Perez-Cortes, S., and Sánchez, L. (2017). "Language attrition and the feature reassembly hypothesis," in *The Oxford Handbook of Language Attrition*, eds B. Köpke and M. S. Schmid (Oxford: Oxford University Press).
- Quine, W. V. (1940). *Mathematical Logic*. New York, NY: W.W. Norton and Company.
- Ramachandran, V. S. (2011). *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes Us Human*. New York, NY: W.W. Norton and Company.
- Recchia, G. L., Jones, M. N., Sahlgren, M., and Kanerva, P. (2010). "Encoding sequential information in vector space models of semantics: comparing holographic reduced representation and random permutation," in *Proceedings of the 32nd Cognitive Science Society*, eds S. Ohlsson and R. Catrambone, 865–870.
- Reitter, D., Keller, F., and Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cogn. Sci.* 35, 587–637. doi: 10.1111/j.1551-6709.2010.01165.x
- Riksem, B. R. (2017). Language mixing and diachronic change: american Norwegian noun phrases now and then. *Languages*. Available online at: <http://www.mdpi.com/2226-471X/2/2/3>
- Roeper, T. (1999). Universal bilingualism. *Bilingualism Lang. Cogn.* 2, 169–186. doi: 10.1017/S1366728999000310
- Roeper, T. (2016). Multiple grammars and the logic of learnability in second language acquisition. *Front. Psychol.* 7:14. doi: 10.3389/fpsyg.2016.00014
- Schmid, M., and Köpke, M. S. (2007). "Bilingualism and attrition," in *Language Attrition: Theoretical Perspectives*, eds B. Köpke, M. S. Schmid, M. C. J. Keijzer, and S. Dostert (Amsterdam: John Benjamins), 1–7.
- Schoonbaert, S., Hartsuiker, R. J., and Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: evidence from syntactic priming. *J. Mem. Lang.* 56, 153–171. doi: 10.1016/j.jml.2006.10.002
- Segaert, K., Wheeldon, L., and Hagoort, P. (2016). Unifying structural priming effects on syntactic choices and timing of sentence generation. *J. Mem. Lang.* 91, 59–80. doi: 10.1016/j.jml.2016.03.011
- Seton, B., and Schmid, M. S. (2016). "Multi-competence and first language attrition," in *The Cambridge Handbook of Linguistic Multi-Competence*, eds V. Cook and L. Wei (Cambridge: Cambridge University Press), 338–354.
- Simon, H. (1969). *Sciences of the Artificial*. Cambridge, MA: MIT Press.
- Smolensky, P., Goldrick, M., and Mathis, D. (2014). Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition. *Cogn. Sci.* 38, 1102–1138. doi: 10.1111/cogs.12047

- Smolensky, P., and Legendre, G. (2006). *The harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, Vol. 1–2. Cambridge, MA: MIT Press.
- Starreveld, P. A., De Groot, A. M. B., Rossmark, B. M. M., and Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: evidence from word production in sentence context. *Bilingualism Lang. Cogn.* 17, 258–276. doi: 10.1017/S1366728913000308
- Stroik, T., and Putnam, M. (2013). *The Structural Design of Language*. Cambridge: Cambridge University Press.
- Taeschner, T. (1983). *The Sun Is Feminine: A Study on Language Acquisition in Bilingual Children*. Berlin: Springer Verlag.
- Thomason, S. G., and Kaufman, T. (1992). *Language Contact, Creolization, and Genetic Linguistics*. Berkeley; Los Angeles: University of California Press.
- Treffers-Daller, J. (2017). *Turkish-German Code-Switching Patterns Revisited: What Naturalistic Data Can(not) Tell Us*. University of Reading.
- Uriagereka, J. (2008). *Syntactic Anchors: On Semantic Restructuring*. Cambridge: Cambridge University Press.
- van der Maaten, L. J. P., and Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605. Available online at: <http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf>
- van Oostendorp, M., Putnam, M., and Smith, L. C. (2016). “Intersecting constraints: why certain constraint-types overlap while others don’t,” in *Optimality-Theoretic Syntax, Semantics, and Pragmatics: From Uni- to Bidirectional Optimization*, eds G. Legendre, M. Putnam, H. de Swart, and E. Zaroukian (Oxford: Oxford University Press), 33–54.
- Volterra, V., and Taeschner, T. (1978). The acquisition and development of language by bilingual children. *J. Child Lang.* 5, 311–326. doi: 10.1017/S0305000900007492
- Weeden, V. J., Rosene, D. L., Wang, R., Dai, G., Mortazavi, F., Hagmann, P., et al. (2012). The geometric structure of the brain fiber pathways. *Science* 335, 1628–1634. doi: 10.1126/science.1215280
- Weinreich, U. (1953). *Languages in Contact, Findings, and Problems*. New York, NY: Linguistics Circle of New York.
- Westergaard, M. (2013). “The acquisition of linguistic variation: parameters vs. micro-cues,” in *In Search of Universal Grammar: From Old Norse to Zoque*, ed T. Lohndal (Amsterdam: John Benjamins), 275–298.
- Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.
- Ziv, J., and Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Trans. Inform. Theory* 24:530. doi: 10.1109/TIT.1978.1055934

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Putnam, Carlson and Reitter. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.