

Brown, Esther L. (pre-print version). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. *International Journal of Bilingualism*, Special issue "Gauging convergence on the ground: Code-switching in the community ", Rena Torres Cacoullos and Catherine E. Travis (eds.). To appear (2015) Vol. 19.4

The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production

Esther L. Brown

University of Colorado, Boulder

Abstract

Missing from the body of literature on contact induced phonological influence are studies that examine language variation as it occurs in speech production among members of a speech community. This study uses a corpus of naturally occurring Spanish/English code-switched discourse to determine whether cross-language phonological effects are evident in the data. Specifically, 2,629 tokens of word-initial /d/ were analyzed in spontaneous interactions to identify the linguistic factors that condition the variable reduction (unreduced [d], reduced [ð]/Ø) of /d-/ in Spanish words. Cognate words (*doctor*) were found to reduce significantly less often than non-cognate words (*después* ‘after’). Additionally, a significant effect is found for a novel, contextually informed measure that estimates words’ proportion of use in online contexts promoting reduction (FFC – Frequency in a Favorable Context). The greater a word’s prior exposure to online contexts promoting reduction, the greater the likelihood of reduced articulations. Indeed, this work argues that the distinction between cognates and non-cognates in fact emerges through this cumulative effect of significantly different patterns of use in discourse. Cognate /d/ words are used overall (considering speakers’ use of both English and Spanish) less often in contexts that promote reduction than non-cognate words. As a result of the diminished net exposure to reducing environments, per usage-based grammar, the lexical representations of cognate words have strengthened non-reduced exemplars ([d]). The distinct rates of variation for the word categories thus emerge from distinct usage patterns. This paper proposes such a focus on usage patterns within naturally occurring speech for phonological analyses within contact linguistics.

Keywords: discourse context frequency, usage-based phonology, bilingualism, code-switching, cognates

1. Multidisciplinary approaches to cross-linguistic phonological influence

Research into bilingualism has been undertaken from a variety of perspectives— ethnographic, historical, social, psychological, and linguistic—all with different methodologies. Indeed it is the ‘eclectic methodology’ drawn from various approaches that in some ways gives the field of contact linguistics its strength (Winford, 2003, p. 9). Over the last few decades there has been an ‘unprecedented upsurge’ of interest into linguistic analyses of bilingualism (Leikin, Schwartz, & Tobin, 2012, p. 1), including the psycholinguistics of bilingualism (Grosjean & Li, 2013), in which the quest to model how bilinguals process and store multiple languages takes center stage.

During this same time frame, usage-based approaches to linguistic analyses have been in the ascendancy (Backus, 2012). Within a usage-based framework to language (and, hence, to bilingualism), social interactions and generalized cognitive processes are argued to be responsible for the structure and knowledge of language(s). Beckner et al. (2009) note that recent “research in the cognitive sciences has demonstrated that patterns of use strongly affect how

language is acquired, is used, and changes” (p. 1-2). It is precisely these ‘patterns of use’ that will be the primary focus of this current work, and the mechanism by which they can “affect” the cognitive representations of language (Bybee, 2010, p. 12), and hence linguistic outcomes (or ‘changes’), that are commonly noted in situations of language contact (e.g., convergence, interference, borrowing).

This work analyzes variable realizations of word-initial /d-/ in Spanish within an oral corpus of spontaneous Spanish/English code-switched discourse (described below). Realizations of word-initial /d/ provide apt data to explore cross-linguistic phonological influence because articulations differ across Spanish and English, with Spanish forms showing considerable variability in degree of reduction. Significant differences in /d-/ realizations between cognate and non-cognate words are found, with cognate words showing less reduction, which may be interpreted as evidence for cross-language phonological influence.

However, by employing an underutilized usage-based linguistic variable (FFC – *Frequency in a Favorable Context*) that estimates lexicalized effects of patterns of use in discourse (cf. Bybee, 2002), this work argues that the source of the cognate effect in these data emerges out of distinct usage patterns of the word classes (cognate vs. non-cognate). Non-cognate words appear more often in phonetic contexts that condition reduction and, unlike cognates, lack exemplar connections to non-reduced forms sharing phonological and semantic similarity (*i.e.* cognates).

Usage patterns within naturally occurring speech have been heretofore neglected for phonological analyses within contact linguistics. Yet the quantification of lexical effects of exposure to online (discourse) contexts promoting reduction is shown in this work to significantly predict variable realizations of words. A primary aim of this study, therefore, is to advocate for inclusion of this more contextually informed frequency measure.

With few exceptions (Bullock & Toribio, 2009; Grosjean & Miller, 1994; Olson, 2012, 2013), articulation in code-switched production remains largely unexplored and phonological analyses of spontaneous interactions are virtually absent from the literature. This lack of attention is noteworthy given the view of code-switching as a possible cause of innovation and propagation of change in language contact (Backus, 2005, p. 316). The present variationist analysis of variable realizations of /d/ in words in Spanish/English code-switched discourse redresses this lacuna, adding much-needed, corpus-based evidence to the pool of linguistic phenomena attributed to contact and contributing “comparative data from on-going (synchronic) contact situations” with which to build and test theories of bilingual lexical representation and speech production (Backus, Dogruoz, & Heine, 2011, p. 750).

2. A variationist usage-based approach

2.1 A usage-based approach to bilingual data

A usage-based approach assumes that language is shaped by the way in which it is used. As Bybee (2010, p. 1) explains,

the structural phenomena we observe in the grammar of natural languages can be derived from domain-general cognitive processes as they operate in multiple instances of language use. The processes to be considered are called into play in every instance of language use; it is the repetitive use of these processes that has an impact on the cognitive representation of language and thus on language as it is manifested overtly.

For phonological processes, a certain amount of online reduction is expected in production due to neuromuscular requirements of speaking (Raymond, Dautricort, & Hume, 2006). This has the potential to impact lexical representation because when a word is realized with a reduced form, the reduction impacts the word's exemplar cluster or cloud by increasing the number of reduced exemplars or by strengthening existing reduced exemplars (Bybee, 2012, p. 216).

These exemplar clouds do not exist in isolation, however, but rather are conjectured to be embedded in a highly organized network of lexical connections based upon phonological and/or semantic overlap (Bybee, 2001). It is through this network of lexical connections that word-specific patterns of articulation can spread to other lexical items (Bybee, 2002, p. 272). Under the assumption that both the language processing dynamics and the lexical representations of bilinguals are qualitatively similar to those of monolinguals, the theoretical and methodological advances of the usage-based approach to linguistics, albeit built largely on monolingual data, will be applied equivalently to bilingual data here. In this vein, the bilingual is viewed not dissimilarly from a monolingual (Grosjean, 1989), a view that is gaining traction (cf. Runnqvist, Strijkers, Alario & Costa, 2012, p. 850). As such, just as knowledge and use of inflectionally and derivationally related forms in one language have been argued to impact pronunciation variations of related forms via networks of lexical associations (Bybee, 2001; Brown, 2011), this work argues that phonologically and semantically related forms across languages are open to comparable types of paradigmatic peer pressure. Bilingualism, then, is a special case of variable use, where languages may be separated ideologically, but not psycholinguistically (Franceschini, 2011; Kroll & Bialystok, 2013), and the organized network of lexical connections is not language independent (Brown & Harper, 2009).

Figuring prominently in experiments investigating bilingual lexical representation are analyses of cognates, "translations with a similar meaning, phonology and orthography" (de Groot, 1995, p. 167). By studying the phonetic outputs for specific phones (e.g. /t/), studies show that while speakers are adept at avoiding massive interference from the other language (Costa, Santesteban, & Caño 2005, p. 137), subtle encroachments of one language on another in phonetic space do occur. Such permeability may be most noticeable in cognates (e.g.; Amengual, 2012; Torres Cacoullos & Ferreira, 2000; Jacobs, 2007). As Bybee (2001, p. 26) observes, "the strength of association between [lexical] items with identical or similar features may vary according to the number and nature of the features...". In this way, cognates which share phonological, semantic, and to some extent orthographic similarity, may be posited to establish a stronger interlingual connection than non-cognates which do not share such similarities. The prediction that follows, therefore, is that word-initial /d/ will be realized more often as [d] (as opposed to [ð] or Ø) in cognates than non-cognates.

2.2 Discourse context frequency

As has been noted, patterns of language use are likely to provide key evidence to help "shape future models of the multilingual lexicon" (Goral, Levy, Opler & Cohen, 2006, p. 244). A primary goal of this study, therefore, is to outline precise patterns of use in bilingual speech production. There is precedent for examining 'patterns of use'. For example, in an examination of variable rates of word final /-t, -d/ deletion in English, Bybee (2002) proposes lower deletion rates in regular past tense forms disfavoring reduction stem from the discourse pattern of past tense forms compared to other words, as they occur in discourse significantly more often in pre-vocalic contexts (a non-reducing context), as in verb-particle combinations, such as *lived in* or *looked at*. Thus, the lexical representation for the past tense forms will reflect fewer deleted

forms overall, which reflects the contexts of use in discourse. The stored experiences, then, predict less subsequent deletion of past tense /-t,-d/ compared to other /-t,-d/ words, even when considered in identical contexts. The hypothesis is, then, that *in addition to* the online effect of phonetic context during articulation, there is a cumulative (lexicalized) effect of experience in specific discourse contexts which affects pronunciation of that word (Bybee, 2002).

Hall, Cheng, and Carlson (2006) note that “what gives rise to the differences in language knowledge are the particular circumstances within which an individual experiences and uses language” (p. 230). The approach taken in the current work embodies and substantiates this assertion by demonstrating a cumulative (lexicalized) effect of patterns of use, in particular, discourse context frequency, which has been shown to be a powerful factor in variation and change (Brown & Raymond, 2012; Raymond & Brown, 2012).

2.3 The dependent variable: Spanish /d-/ realizations

Spanish has two primary allophones of /d/; the voiced, dental stop [d], and the voiced, dental approximant or fricative [ð] (Barrutia & Schwegler, 1994, p. 114-120). The voiced, dental stop [d] is prescriptively preferred in a post-pause, post-nasal, or post-/l/ (voiced, lateral, alveolar) context as can be seen in example (1). The second principle allophone [ð] is realized in all post-vocalic contexts and post-consonantal contexts not described for [d], such as that found in example (2).

(1)

Bartolomé	<i><X creo X> que nos costó como dos mil <u>d</u>ólares a cada uno entonces.</i>	‘I think it cost us each about a thousand dollars.’	[02 La marina, 0:37:10-0:37:14] ¹
-----------	--	--	--

(2)

Susan	<i>me preguntó si quería bailar y <u>le dije</u>,</i>	‘He asked me if I wanted to dance and I told him,’	[01 El abuelo, 0:36:58-0:37:00]
-------	---	--	---------------------------------

The realizations and allophonic distribution of /d/ exhibit “extensive regional variability” (Amastae, 1989, p. 170; Waltermire, 2010), as well as “a certain amount of variability within a given dialect or idiolect” (Cole, Hualde, & Iskarous, 1999, p. 2). Articulations of /d/ vary considerably, ranging from “a complete stop to a vocalic glide” (Cole et al., 1999, p. 2), to outright deletions [Ø], though such deleted segments may be stigmatized (Barrutia & Schwegler, 1994, p. 120). This high degree of variability is evident in the New Mexican data. As such, the phonological contexts illustrated in (1) and (2) are viewed in terms of probabilities of contexts favoring or disfavoring stop [d] vs. reduced realizations such as [ð] or [Ø]), as opposed to obligatory contexts for rule application (Cole et al., 1999, p. 2). Multivariate analysis is employed here to tease apart the possible independent effects of the many previously identified linguistic factors constraining the variation (stress, word frequency, etc.).

Whereas the Spanish phone /d/ has a dental articulation, the English phone /d/ is alveolar, and, more notably, the languages differ in the number of allophones of /d/ in word-initial position. The English phoneme /d/ has just one variant, which is realized as a stop word-initially

¹ All examples given are from New Mexico Spanish-English Bilingual corpus (NMSEB, Torres Cacoullous & Travis In Preparation) and are reproduced verbatim from the transcripts (see transcription conventions in Appendix 2 in the Introduction to the volume). Within brackets is the recording number, name and time stamp. In examples where Spanish and English are used, the original appears on the left, and the translation on the right, with stretches of speech originally produced in English appearing in italics.

(Barrutia & Schwegler, 1994, p. 114), whereas it is the fricative variant, not the stop, which is more common in Spanish (Barrutia & Schwegler, 1994, p. 114; Teschner, 2000, p. 96). Thus, Spanish /d/ words can be articulated as a stop [d] or a fricative [ð] (more commonly [ð]), while an English word will have just one stop variant [d] (e.g.; *dentist* [d], *dentista* [ð]/[d]). Consequently, if a Spanish speaker were to demonstrate an interlingual phonological influence from English, we could predict reinforcement of the stop [d] articulation in /d/ tokens to the detriment of the fricative variant [ð].

3. Data and Coding of Linguistic Factors

The data used for this study was extracted from a corpus of naturally-occurring speech in the New Mexico Spanish-English Bilingual (NMSEB) corpus (Torres Cacoullós & Travis, in preparation). The NMSEB corpus is comprised of recorded sociolinguistic interviews between bilingual Spanish-English New Mexicans and in-group members (for a detailed overview of the speakers and the corpus, see the introduction to this volume and Travis & Torres Cacoullós, 2013). All of the speakers use both Spanish and English regularly and naturally in their speech. This study is based on 18 recordings, representing approximately 180,000 words spoken by 20 speakers over 17.5 hours.

All instances of word-initial /d/ in Spanish were extracted for analysis and tokens were coded auditorily as either occlusive ([d]) or reduced ([ð], Ø). Perception of /d/ spirantization in Spanish can generally be said to be “auditorily fairly obvious” (Port & Leary, 2005, p. 953). Nevertheless, coding as reduced or retained for several tokens was not interpretable at times due to noise or multiple speakers talking at once, and these tokens were not considered for analysis. Proper nouns were not included nor was the one token of non-prevocalic /d/ (*droga*). As the only preconsonantal token, this singleton was excluded. Additionally, the 67 tokens of the one lexical item *donde* ‘where’ are not included in the analysis since the archaism *onde*, with no initial /d-/, is common in New Mexican Spanish (Bills & Vigil, 2008, p. 15). A total of 236 tokens were thus not included for analysis.

All word types are counted separately and are not conflated into paradigmatically related forms (e.g., *digo*, *dije*, *dijera*, etc. counted individually and not simply as forms of *decir*). Although within a usage-based approach to language inflectionally and derivationally related words are presumed to form a network of phonological and semantic associations (Bybee, 2001, 2010), each word registers effects of use individually (e.g. Gahl, 2008). A total of 2629 tokens of /d/ words were extracted in 197 different word types.

To verify that these bilinguals’ English follows the monolingual norm of stop realization of /d/, a sample of tokens of word-initial /d/ in English were also extracted from the first six interviews (N = 1370). These words exhibited minimal variation (12 occurrences of reduced [ð], appearing across three word types; *didn’t*, *don’t*, *days*).

Unlike studies suggesting an effect of proximity to switch site on phonetic outputs in elicited bilingual speech production (e.g.; Bullock & Toribio, 2009; Olson, 2012), the rate of reduced and unreduced forms of /d/ was found not to be significantly affected by factors coded to explore effects of code-switching (distance from switch as measured in number of words, presence of switch in the Intonation Unit (IU)), nor was there a phonetic priming effect of the realization of the previous word-initial /d/, whether the previous word was Spanish or English (Brown, 2013). These effects are thus not considered further in the current analysis. Each token was coded for the following independent variables (predictors):

Contextual and Lexical Factors

1. Previous phonetic context: The previous phonological environment, as determined by the discourse context, was coded as either a context favoring stop articulation (preceding pause when utterance initial², preceding /n/, and preceding /l/) or as a context favoring a reduced articulation of /d/ (all other post-consonantal and post-vocalic contexts).
2. Following vowel type: Previous studies of voiced stop realizations in Spanish find that the quality of flanking vowels (Cole et al., 1999) can significantly affect articulations of the consonants. Thus, we code for the following vowel class immediately adjacent to the word-initial /d/ [front (/i/, /e/) and non-front (/a/, /o/, /u/)]. The diphthong /j/ (*dieron* ‘they gave’) is grouped with the front vowels, and the diphthong /w/ (*duele* ‘it hurts’) is grouped with the back vowels for the multivariate analysis.
3. Word frequency per million. To investigate whether lexical frequency is correlated with phonological reduction (cf. Bybee, 2012, p. 214-215), the token frequency of each word type containing a word-initial /d/ was calculated in the oral portion of the *Corpus del español* online (Davies, 2002-); (5,113,249 words). There are large discrepancies in token frequency values of the words (frequency per million of *de* ‘of, from’ = 46,472, frequency per million of *debías* ‘you should’ < 1). The effect of frequency has been argued to be operative above a threshold at which cumulative experience with words can affect lexical representations and, as such, the data were discretized into high and low frequency groups (Erker & Guy, 2012). High frequency words were arbitrarily set at those with a frequency per million greater than 100 (45 types) and low frequency words as those with a frequency of < 100 per million (153 types).
4. Stress: Previous work has determined that stress can play a role in realizations of /d/ (Eddington, 2011), with unstressed syllables typically exhibiting more reduced forms (and less articulatory energy) than stressed syllables. All tokens were coded as to whether the syllable containing the word-initial /d/ carried lexical stress (*dice* ‘says’) or did not (*decimos* ‘we say’).
5. Cognate status: Coding cognate status of Spanish words for this project was based upon subjective assessment of degree of phonological, orthographic, and/or semantic overlap to their English equivalent. Though degree of ‘cognateness’ can be said to be highly gradient (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010), the present analysis categorizes words into two categories; cognates (phonologically and semantically similar), non-cognates (dissimilar). Three native Spanish speakers with near-native English proficiency were asked to rate the list of 198 word types as cognate with English or not. Words deemed phonologically and semantically similar by (at least) two out of three consultants were categorized as cognates (see Appendix). These comprise a total of 37 word types (N = 227) (e.g., *decidieron* ‘they/you plural decided’, *doble* ‘double’, *diferente* ‘different’, *diciembre* ‘December’, *doctor* ‘doctor’). A total of 161 word types

² “Utterance initial” is defined as anything in which a speaker starts a turn, or if their IU is preceded by a pause.

were classified as non-cognate (N = 2402) (e.g., *dice* ‘you (formal), s/he say(s)’, *después* ‘after’, *dientes* ‘teeth’, *domingo* ‘Sunday’).³

FFC – Frequency in a Favorable Context

The presence of an effect from discourse context frequency has been tested on word-initial /s-/ reduction in New Mexican Spanish by Brown (2006) and Raymond and Brown (2012), who label the effect *frequency in a favorable context*, or FFC. Reduction (aspiration [h], deletion Ø) of /s/ in this variety is favored by a preceding non-high vowel (/e,a,o/), and disfavored by preceding high vowel, consonant or pause. Words differ significantly in their frequency of use in this reducing discourse context [e.g.: (*la, una*) *señora* vs. (*el, un*) *señor*] and hence are presumed to differ in the number (or strength) of reduced exemplars stored in memory, which predicts likelihoods of reduced articulations. Raymond and Brown (2012) establish that FFC significantly constrains synchronic variation of /s-/ realizations, independent from and in addition to, other linguistic factors known to influence articulation (stress, phonetic context, word frequency).

Indeed, even phenomena traditionally given a language contact explanation, such as the case of [f] > [h] > Ø in Spanish of Latin FV- words (Menéndez Pidal, 1968, p. 198–233; Penny 1972, 1991), when reexamined with this more precise, contextually informed measurement of FFC, can be shown to be internally motivated. Brown and Raymond (2012) examine the diachronic change and show how the modern standard Spanish outcome of *f-* and *h-* words reflects usage patterns and cumulative exposure to reducing contexts (FFC). Forms more often preceded by a non-high vowel are more apt to have lost the initial consonant [FACTUS > *hecha* ‘done’ (with 53% of occurrences post non-high vowel) vs. *fecha* ‘date’ (39% of uses)]. Further, FFC better predicts modern lexical distribution of FV- words than transmission history (prestige borrowings or *cultismos* vs. orally transmitted words).

Turning to /d/-initial words, the allophonic distribution discussed in Section 2.3 highlights the important role of online articulation effects stemming directly from the phonetic context in which a /d-/ word is embedded. For instance, a preceding /l/ is said to promote stop articulations of the /d/, presumably due to fact that /l/ assimilates a dental articulation which gives both phones the same place of articulation (Eddington, 2011, p. 15). These phonetic contexts, then, [the non-reducing (#, /l/, /n/, /m/ ___) and the reducing environments] significantly predict the realization of the /d-/ during production. FFC measures a word’s ratio of occurrence in a phonetic context conditioning reduction.

Each word’s instances of use in a reducing context is calculated in the oral portion of the *Corpus del español* (Davies, 2002-). This includes all contexts other than post-nasal, post-lateral, and post-pause. The estimates for use in post-pausal position is operationalized in the online corpus as tokens occurring immediately following punctuation (.,?!:;-¿¡). The total frequency per million in the reducing context is divided by the overall frequency per million of that word. The FFC calculation, then, is the number of tokens in a reducing context/total number of tokens. The

³ Two highly frequent word types [*de* ‘of, from’ (N = 863), *del* ‘of / from masc. sing. the’ (N = 100)] make up a large portion of the non-cognate forms. In separate analysis without the *de* and *del* tokens the same linguistic factor groups are significant.

value is expressed as a percentage; 100, the lexical item categorically occurs in a context conducive to reduction and 0, the word never occurs in a reducing context.

A comparison of the nouns *domingo* ‘Sunday’ and *diferencia* ‘difference’ illustrates this method. Examples listed in (3a), taken from Davies (2002-) are tokens of *domingo* used in contexts favorable to reduction of word-initial /d/, and examples listed in (3b) are uses of *domingo* in contexts unfavorable to reduction. Similarly, tokens of *diferencia* are used in contexts favorable to reduction (4a) and unfavorable to reduction (4b).

(3a) *Domingo*: Favorable context (N = 250)

El próximo **domingo** cumple sesenta años [*Entrevista (ABC)*]
El primer **domingo** solitario [*Entrevista (ABC)*]
Libre completamente sábado **y domingo**. [*Habla Culta: Buenos Aires*]
¿verdad?, cada **domingo** [*Habla Culta: Caracas: M12*]
si hoy es **domingo**? [*Habla Culta: Santiago: M43*]

(3b) *Domingo*: Unfavorable context (N = 258)

En relación **al domingo** pasado [*Entrevista (PRI)*]
regresábamos **el domingo** en la tarde [*Habla Culta: Lima: M15*]
di a luz **un domingo** y [*Habla Culta: Madrid: M19*]
no libré ningún **domingo** [*España Oral: ADEB022A*]
ir en la catedral el **, domingo** a misa, es encontrar, [*Habla Culta: Lima: M12*]

(4a) *Diferencia*: Favorable context (N = 619)

sino que **la diferencia** solamente es [*Habla Culta: San Juan (PR)*]
¿qué **diferencia** hay entre la que se hizo [*Habla Culta: San Juan (PR)*]
Con una **diferencia** cada vez mayor [*España Oral: ADEP009A*]
no sentí mayor **diferencia** [*Habla Culta: Lima: M7*]
no hay más **diferencia** sustancial [*Entrevista (ABC)*]

(4b) *Diferencia*: Unfavorable context (N = 34)

pero hay una sutil **diferencia** entre [*Habla Culta: México: M9*]
La **gran diferencia** que encontré [*Habla Culta: Santiago: M49*]
Y notaban **diferencia** en España [*España Oral: CCON031B*]
se manifiesta con **n diferencia** de propuestas [*Entrevista (Zedillo)*]
no sentí... la.. **. diferencia** en la cuestión [*Habla Culta: México: M21*]

The FFC measure can be determined from the proportion of instances of use for each token in contexts favoring reduction; or each word’s frequency in a favorable environment. As can be seen in examples (3) and (4), the FFC measure is not dependent upon specific lexical items or specific phones but rather reducing environments generally. Table 1 summarizes the FFC values for each lexical item. The noun *domingo* is less often used (49%) in contexts that favor the approximant allophone ([ð]). Conversely, the lexical item *diferencia* is almost never found in a post-pause, post /l/, or post-nasal context, giving it an FFC measure of 95%. In almost every instance of its use, the phonological context in which *diferencia* is uttered predicts the approximant ([ð]) rather than the stop ([d]) articulation. FFC measures the cumulative effect on the lexical representation of repeated utterances in such contexts.

Table 1: FFC value for word-initial /d/ environments (Davies, 2002-)

	Word frequency	N in Favorable Environment (not post-pause, post-lateral, post-nasal)	FFC
<i>Domingo</i>	508	250	250/508 = 49
<i>Diferencia</i>	653	619	619/653 = 95

This measure can be taken as an approximation of exemplar make-up based upon experience with a lexical item, and, importantly, this measure is distinct from the probability of occurring in a certain phonetic context (e.g. post-nasal, post-lateral, post-pause). This differentiates the FFC measurement from other measures of probability (e.g. Hume & Mailhot, in press; Jurafsky et al., 2001) in which the dependent variable (the predicted linguistic form) and the probabilistic measure predicting the linguistic outcome are derived from the same context. The FFC value of a lexical item remains constant whether it is being analyzed in a post-nasal environment, for instance, or a post-vocalic environment.⁴

The reliability of an FFC value depends on the number of examples. A full 98% of the FFC values were calculated based upon 10 or more examples in the *Corpus del español*, with only 68 tokens having fewer than 10 occurrences in the online corpus. Only an additional 100 tokens (4% of the data) were calculated based upon 20 or fewer examples. Multivariate analyses excluding FFC values for these tokens return the same results. For the multivariate analysis the data were divided into high and low FFC groups; words used in a context for reduction in 75% or more of their instances are labeled high FFC (140 types) and words occurring in such contexts less than 75% of the time are considered low FFC words (58 types).

Bilingual FFC – Frequency in a Favorable Context

FFC has not previously been explored for bilingual data. For cognate words in these Spanish-English bilingual data, the Bilingual FFC calculation considers the effects of use stemming from English. To do this, instances of use of the cognate in English are entered into the FFC calculation (as outlined above) as examples of non-reducing contexts (Brown & Raymond, in preparation). English frequencies per million are taken from an online *Corpus of Contemporary American English* (COCA, Davies, 2008-). Formulated in this way, the FFC measurement incorporates gradient frequency effects of English forms as can be appreciated in Table 2.

Table 2: Frequency per million of cognate and non-cognate words in reducing contexts

Word	Spanish N	N in reducing context	English N	FFC
<i>Dan</i> ‘3 p.p. give’	188	179	n.a.	179/188 = 95
<i>Dinero</i> ‘money’	234	143	n.a.	143/234 = 61
<i>Dirección</i> ‘direction’	83	77	60	77/143 = 54
<i>Dólar</i> ‘dollar’	13	5	42	5/55 = 9

⁴ For an application of FFC to non-phonological processes, see Brown and Rivas (2012) who find that non-standard pluralization in the Spanish presentative [*haber* ‘there(be)’ + NP] is favored with nouns often used as subjects (or those with a high FFC, so to speak), independent from and in addition to the linguistic constraints known to condition variation of *haber*.

This FFC calculation is a ‘stand in’ for knowing exactly the experiences speakers have in production and perception. Ideally we would know a speaker’s precise experiences with words and the make-up of the exemplar representation. Based upon the effects we know to operate during articulation, FFC attempts to estimate probabilistically this experience.

The following section presents the results of analyses conducted on the 2629 coded tokens extracted from the New Mexican bilingual discourse.

4. Results

In order to determine which linguistic factor groups significantly constrain the variable realizations of /d-/ in the NMSEB corpus while simultaneously considering independent contributions of each linguistic factor group, the data were submitted to a multivariate analyses using Varbrul (Rand & Sankoff, 2001) (Guy, 1993; Sankoff, 1988). The findings are summarized in Table 3. Of the six factor groups analyzed, selected as significant were four: the phonological contexts flanking the word-initial /d/ (previous and following), cognate status and FFC. Factors are listed in the order of their magnitude of effect as determined by their order of selection in the multivariate analyses, as well as by the range in probability of individual factors. These probabilities reflect the degree to which the linguistic constraint relatively favors (closer to 1) or disfavors (closer to zero) a stop realization.

Table 3: Multivariate analysis of factors contributing to reduced articulations [ð], Ø (vs. non-reduced forms [d]) in NMSEB corpus (non-significant factors within [])

	Probability	% [ð]/Ø	N	% data
N = 2629, Input = .66, rate = 64%				
Preceding phonetic context				
Favorable (not post-pause, post-nasal, post-lateral)	.62	76	2030	77
Unfavorable (post-pause, post-nasal, post-lateral)	.16	25	599	23
Cognate status				
Non-cognate	.54	68	2402	91
Cognate	.19	25	227	9
Following phonetic context				
Front vowel (/i/, /e/) & glide /j/	.53	67	2307	88
Central & back vowels (/a/, /o/, /u/) and glide /w/	.28	43	322	12
FFC – Frequency in Favorable Context				
High FFC	.53	71	2106	80
Low FFC	.30	37	523	20
Stress				
Unstressed	[.51]	67	1591	60
Stressed	[.49]	60	1038	40
Word Frequency				
High	[.51]	65	2178	83
Low	[.47]	59	451	17

The factor group ‘preceding phonetic context’ has the greatest relative magnitude of effect on realizations of word-initial /d/ in Spanish. If the word-initial /d/ is preceded by a nasal, lateral, or a pause, reduction is highly disfavored (probability .16) and /d/ reduces at a rate of 25%. The factor group ‘following phonetic context’ also significantly constrains realizations of word-initial /d/. For words with a following front vowel (*dinero* ‘money’, *demás* ‘rest’), reduction rates are higher (67%) and reduced articulations are favored (probability .53).

In addition to the significant contribution to variation made by both previous and following phonological context, the analysis also selected cognate status as significant. These word categories (cognate, non-cognate) behave differently with regard to initial /d/ reduction. Non-cognate words are significantly ($p = 0.0000$, $X^2 = 167.2588$) more likely to reduce a word-initial /d/ (68%) than a cognate word (25%). Cognates, then, are more likely to be articulated with a stop /d/ than with a reduced /d/ in line with predictions regarding potential influence of English. That is, even when bringing other factors under statistical control, cognates (e.g. *depende* ‘depend’) are less likely to reduce than non-cognates (e.g. *doloroso* ‘painful’).

A model of cascaded interactivity between languages (e.g., Costa, Santesteban, & Caño, 2005) might predict such a result. Yet, even in the face of such positive evidence, disagreements persist with regard to whether cognate effects “are due to differences in the way words are represented in the lexicon (e.g., shared morphemes for cognates), or rather are due to a more general property of the speech production system (e.g., cascade/interactivity dynamics)” (Costa, Santesteban, & Caño, 2005, p. 99). In fact, this present study proposes another source of such effects which is outlined below: significantly different usage patterns.

4.1 Unraveling the cognate effect: the role of discourse context frequency (FFC)

The pattern of variation in the data reflects lexicalized effects, as is indicated by significant contributions to variation of FFC. As summarized in Table 3 above, the previously untested linguistic factor FFC significantly predicts realizations of word-initial /d/ in the NMSEB corpus in line with previous usage-based research. The likelihood of word-initial /d/ reduction reflects the degree to which the lexical item is overall used in discourse contexts favoring reduction. That is, words with a high FFC favor reduction of the word-initial /d/, while words with a low FFC disfavor reduction. If the Spanish /d/ word has a high FFC (e.g.; *dije* ‘I said’), reduction is favored with a probability of .53. Reduction of /d/ occurs overall with these words at a rate of 71%. The probability for /d/words with a low FFC (e.g.; *día* ‘day’) is .30 indicating reduction is strongly disfavored for this group of words. These words reduce overall at a rate of 37%.

This innovative probabilistic measure (FFC) is found to significantly predict variable realizations of word-initial /d/ while controlling for multiple factors known to affect pronunciation, in support of research reporting significant effects of cumulative measures of discourse context frequency on phonological reduction (Brown & Raymond, 2012; Bybee, 2002; Raymond & Brown, 2012). The cumulative experience speakers have with words is manifest in the pronunciation variation apparent in natural speech production. Through repeated use in reducing (or non-reducing contexts), the exemplar representation of words shifts in accordance to the usage patterns; an effect easily captured in the “detail-preserving episodic memory” characteristic of exemplar theory (Mendoza-Denton, 2004, p. 443-444).

There is variability in the discourse patterns, nevertheless. For example, some cognates often occur in reducing environments (e.g.; *defender* ‘to defend’, FFC = 89; *disciplina*

‘discipline’, FFC = 99), while others exhibit very infrequent use in discourse contexts favoring reduction (e.g.; *distrito* ‘district’, FFC = 14, *dólar* ‘dollar’, FFC = 36). Low FFC cognates reduce significantly ($p = .0000$, $X^2 = 28.88289$) less often (12%) than high FFC cognates (43%), as is summarized in the last column of Table 4. Logically, there is also variation in the non-cognate words with regard to discourse distribution patterns (e.g.; *déjame* ‘let me/leave me’, FFC = 17, *divertido* ‘fun’, FFC = 91), with high FFC words exhibiting significantly ($p = .0000$, $X^2 = 103.2255$) more reduction (73%) than low FFC words (46%). As a result, when modeling variation, FFC has the same direction of effect across these word categories – with higher FFC words reducing more readily.

Table 4: % /d-/ reduction for cognate and non-cognate words with high and low FFC tokens (Ns)

Non-cognate	Previous fav	Previous unfav	Following front	Following back	Unstressed	Stressed	high freq	low freq	Overall
High FFC	81 (1673)	32 (335)	74 (1867)	56 (141)	71 (1333)	76 (675)	73 (1701)	71 (307)	73 (2008)
Low FFC	61 (236)	25 (158)	54 (237)	35 (157)	55 (185)	39 (209)	51 (294)	34 (100)	46 (394)
Total N	78 (1909)	30 (493)	72 (2104)	45 (298)	69 (1518)	67 (884)	70 (1995)	62 (407)	68 (2402)
Cognate	Previous fav	Previous unfav	Following front	Following back	Unstressed	Stressed	high freq	low freq	Overall
High FFC	49 (82)	13 (16)	44 (81)	35 (17)	50 (56)	33 (42)	44 (64)	41 (34)	43 (98)
Low FFC	33 (39)	2 (90)	11 (122)	14 (7)	6 (17)	13 (112)	12 (119)	10 (10)	12 (129)
Total N	44 (121)	4 (106)	25 (203)	29 (24)	40 (73)	18 (154)	23 (183)	34 (44)	25 (227)

Table 4 also makes evident that a larger proportion of non-cognate words (84%, $N = 2008/2402$) have a high FFC than the proportion of cognate words with a high FFC (43%, $N = 98/227$), close to double. On average, cognate /d/ words are used in Spanish less often in contexts that promote reduction (FFC = 62) than non-cognate words (FFC = 79)⁵. As a result of the diminished net exposure to reducing environments, per usage-based grammar, the lexical representations of cognate words have strengthened non-reduced exemplars ([d]). Conversely, non-cognate words with overall increased use in phonetic contexts promoting reduction have lexical representations with an increased number or increased activation of reduced exemplars ([ð]/ Ø).

The analysis of the data set reveals FFC as a factor capable of predicting variation of /d-/ realizations. Variable realizations of words, in addition to reflecting probabilistic outcomes indicative of online contextual pressures during articulation, also reflect lexicalized effects indicative of cumulative experience with that word (significant FFC effects). However, the FFC values represented in Tables 3 and 4 solely reflect discourse patterns in Spanish for cognate and non-cognate words and disregard potential impacts from knowledge and use of English. If bilingual lexical representation allows for interactivity between cognates, the FFC calculation based exclusively on Spanish usage patterns might less accurately reflect exemplar representations of cognates. Independent analysis of just cognate tokens ($N = 227$) suggests this to be the case.

⁵ The same result is found for /d/ initial words examined in a corpus of Puerto Rican Spanish (Brown & Raymond in preparation). If English use and experience is added into the model of the knowledge (and lexical representations) of these New Mexican speakers with the factor ‘Bilingual FFC’, the disparity between cognate and non-cognate forms becomes even more pronounced (with the average FFC of cognates falling to 28).

The left side of Table 5 summarizes a repetition of the variable rule analysis exclusively of cognate tokens (to the exclusion of non-cognates). It can be noted that although the individual probabilities for high and low FFC tokens suggest a direction of effect as seen for the data set as a whole, FFC is not selected as significantly constraining the variation of the cognate tokens. Nevertheless, the FFC value for each cognate, estimating the proportion of reduced and non-reduced exemplars stored in memory, is based on usage patterns in one language. The FFC value does not incorporate any information from the English lexical item that also forms part of speakers' experience and that shares phonological and semantic ties with Spanish words.

How can we operationalize or quantify the potential effect from English in the usage-based framework adopted here? If the FFC value is replaced with the Bilingual FFC value, one that incorporates the potential impact for paradigmatic peer pressure from English via the network of exemplar connections, the results of the variable rule analysis are distinct, as is summarized on the right side of Table 5. When the cumulative experience factor (FFC) incorporates a value considering the relative impact of English non-reduced forms on the exemplar representation in Spanish, and hence becomes a Bilingual FFC measure, this group is selected as significantly constraining variation of the cognate forms. High Bilingual FFC tokens strongly favor reduction and low Bilingual FFC tokens disfavor reduction.

Table 5: Independent multivariate analyses of factors contributing to reduced articulations [ð], Ø (vs. non-reduced forms [d]) of cognates in the NMSEB corpus (non-significant factors within []), with FFC based only on Spanish tokens of occurrence, and FFC including Spanish and English (bilingual FFC)

	Cognates N = 227; Input = .16 (overall rate = 25% reduction)				Cognates N = 227; Input = .15 (overall rate = 25% reduction)			
	Prob	% red	N	% data	Prob	% red	N	% data
Preceding phonetic context								
Favorable	.79	43	121	53	.81	43	121	53
Unfavorable	.18	3	106	47	.16	3	106	47
FFC				Bilingual FFC				
High FFC	[.58]	42	98	43	.91	69	13	6
Low FFC	[.44]	11	129	57	.47	22	214	94

Other factor groups included in analysis (Following vowel class, Stress, Frequency) were not significant.

Future cognate analyses with more tokens are needed to confirm this result. Nevertheless, for the data as a whole, the results of the variable rule analysis are also improved through the quantification and incorporation of experience with another language (in this case, English), providing the best model for the data (as determined by log likelihood, see Guy, 1993, p. 246-247). These results are summarized in Table 6. Again preceding and following phonetic contexts significantly constrain variation of word-initial /d/. Importantly, however, there is no significant contribution to the model by cognate status once the FFC factor is quantified so as to include the potential impact of English. The Bilingual FFC effect, which attempts to provide a more complete (bilingual) picture of the knowledge and use stored in speakers' exemplar representations, significantly constrains variation; high FFC tokens favor reduction (69%) with a probability of .55 and low FFC tokens disfavor reduction (22%) with a probability of .18.

Table 6: Multivariate analysis of factors contributing to reduced articulations [ð], Ø (vs. non-reduced forms [d]) in New Mexico Spanish English Bilingual corpus (non-significant factors within []), FFC including experience with Spanish and English (bilingual FFC)

	Probability	% [ð]/Ø	N	% data
N = 2629, Input = .66, rate = 64%				
Preceding phonetic context				
Favorable (not post-pause, post-nasal, post-lateral)	.62	76	2030	77
Unfavorable (post-pause, post-nasal, post-lateral)	.16	25	599	23
Bilingual FFC				
High FFC	.55	69	2345	89
Low FFC	.18	22	284	11
Following phonetic context				
Front vowel (/i/, /e/) & glide /j/	.54	67	2307	88
Central & back vowels (/a/, /o/, /u/) and glide /w/	.26	43	322	12
Cognate status				
Non-cognate	[.51]	68	2402	91
Cognate	[.37]	25	227	9
Stress				
Unstressed	[.51]	67	1591	60
Stressed	[.49]	60	1038	40
Word frequency				
High	[.51]	65	2178	83
Low	[.50]	59	451	17

The explanation for how the FFC measurement provides a better measure of reduction than cognate status may be an artifact of the measure itself. By combining several independent measures (word frequency in combination with extra-lexical preceding phonological environment), the FFC factor captures more detail than the category of cognate status per se, and hence may more accurately model reduction (cf. Jurafsky, Bell, Gregory & Raymond, 2001, p. 233). However, more than being a methodological refinement, by suggesting that phonological cognate effects in speech production are lexicalized effects of patterns of use, this work provides a theoretically consistent explanation of the source of the effect. Language patterns emerge through use. Cognates do not have a privileged role in the bilingual's lexicon (Duñabeita, Perea, & Carreiras, 2010); their distinct rates of variation emerge from distinct usage patterns.

Discussion

Missing from the body of literature on contact induced phonological influence are studies that examine language variation as it occurs in speech production among members of a speech community. This study uses a corpus of naturally occurring Spanish/English code-switched discourse to test the notion of whether cross-linguistic phonological influence is evident in the

data and to add to our understanding of bilingual lexical representation. Code-switched data was ideal to examine this question because although it is known that sound systems are not impermeable to external influences (Sankoff, 2002, p. 644-649), few attempts have been made to see what, if any, effects there are on variable phonological processes when alternating between languages.

The object of study examined here has been the variable realizations of word-initial /d/ in Spanish. Hypothesized influence from English would be evidenced in increased use of the variant [d] in Spanish (as opposed to the [ð]) due to its similarity to (and support from) English. The examination of variable rates of /d/ reduction in word-initial position revealed that words sharing phonological and semantic overlap with English (cognates) are significantly less likely to be realized with a reduced initial consonant. This is perhaps an ‘unsurprising’ result given the fact that phonological convergence is typically most evident precisely where two languages are most congruent (cf. Bullock & Gerfen, 2004). This result can be taken as evidence that languages are interconnected lexically. The interconnectivity, however, is not generalized to a phonemic level. That is, the effects felt cross-linguistically do not apply uniformly to all /d/ words (decreased rate of reduction generally), but rather variation is found to be lexically specific.

The explanation for the significant differences evident between cognate and non-cognate words, however, is approached from within a usage-based framework. In lieu of arguing that cognates have exceptional status in the lexicon (compared to non-cognate words) *a priori*, this work argues that the distinction between cognates and non-cognates emerges through the cumulative effect of significantly different patterns of use in discourse. Viewing bilingual language production in this way, as a specific case of variable use, predicts an outcome by which knowledge and use of one language can have predictable effects on the knowledge and use of the other language of a bilingual.

The exemplars are organized into a network of connections relating forms that are phonologically and semantically similar (Bybee, 2001, p. 29). The lexical connections are argued to be gradient and to reflect the degree of form/meaning overlap between lexical entries. Words with a high degree of phonological and semantic similarity share stronger lexical connections than words lacking such similarities. These connections represent the form/meaning overlap from which morphology is emergent (Bybee, 1999, p. 224). For instance, Bybee (2001, p. 152-153) highlights the lexical connections with an example of the participle suffix in Spanish /-ado/. Importantly, as Bybee notes for the participle, “instances of –ado are associated with one another and can have an effect on one another”. What is the nature of this influence? Rates of reduction of intervocalic /d/ in this suffix are higher than rates of /d/ reduction in the same intervocalic context outside of the participle, attributable to the higher token frequency of the /-ado/ suffix. Thus, a word with low token frequency (such as *cenizar* ‘to turn to ash’, with a frequency per million of .04), but with the /-ado/ suffix, will exhibit higher /d/ reduction than the lexical frequency would otherwise predict as a result of the lexical associations with the highly reducing participle form. The strong form/meaning overlap of morphemes, therefore, allows for the possibility of influence.

This predicts, consequently, via the same mechanism, the likelihood of mutual lexical influence between other forms with strong phonological and semantic overlap. For instance, the masculine and feminine forms of the noun *doctor* in Spanish (*doctor*, *doctora*) would have strong lexical connections due to the high degree of similarity in form and meaning, as is illustrated in Figure 1. What does this predict for the bilingual lexicon? Precisely the type of cognate effect made apparent in the quantitative analysis of these data. Non-reduced forms

typical of English pronunciation for this phonological variable ([d]) are closely associated cognitively in the lexicons of bilingual speakers to cognate pairs. Such associations bolster the strength of non-reduced exemplars in Spanish.

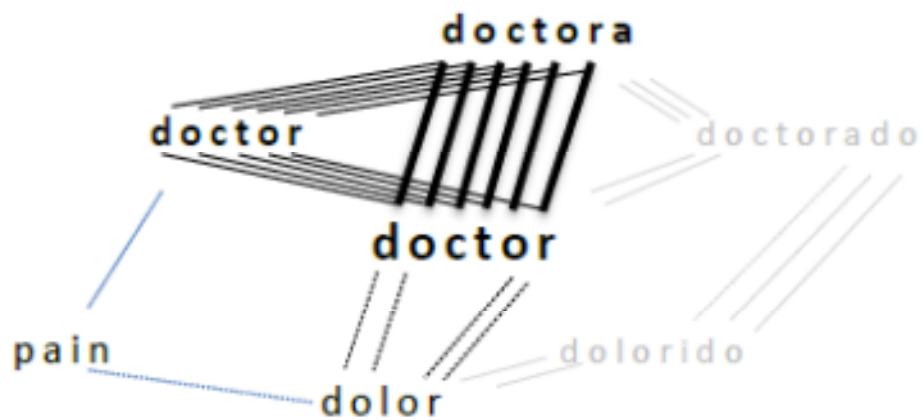


Figure 1: Bilingual lexical representation (cf. Bybee, 2001, p. 22-25)

Cognates, then, reduce less often overall than non-cognates (due to their different exposure to phonetic environments conditioning reduction). The cross-linguistic phonological influence from English to Spanish in this case is lexically specific. Bybee (2012) highlights the interrelatedness of general articulatory routines which are built up from lexically specific articulatory routines, and notes that “while individual words have specific routines associated with them, their use activates the more general routines as well” (p. 218). Thus, the preponderance of stop [d] articulations in cognates affects not just the cognate’s exemplar cloud, but also the “exemplar cloud at the more general level of the articulatory routine” (Bybee, 2012, p. 218). Such influence could predict moderate slowing in the trajectory of /d/ lenition in Spanish (Penny, 1991, p. 68-69) compared to non-contact varieties, though there is no evidence in the current data for such a change.

In sum, this study presents the results of a relatively untested notion regarding the effect of cumulative exposure to reducing discourse contexts (FFC). The novel application of this linguistic variable to the analysis of variation in bilingual speech reveals that usage patterns better account for variation than labels such as cognate, and, in fact, provides an explanation as to the source of such effects in speech production. Significant differences exist in usage patterns between word categories generally. Such patterns predict that for words used frequently in online contexts promoting reduction (*ie*, non-cognates), the likelihood of producing a reduced form of the word increases. These reduced articulations increase the number (and/or strength) of the reduced exemplars stored for that word. Such patterns yield different strengths of reduced/non-reduced forms in exemplar clouds averaged across the categories. Hence the cognate effect is submitted to be a secondary effect of usage patterns, and, even in situations of language contact, the primacy of language internal sources of change is highlighted.

References

Amastae, J. (1989). The intersection of s-aspiration/deletion and spirantization in Honduran Spanish. *Language Variation and Change*, 1, 169-183.

- Ameel, E., Malt, B. C., Storms, G., & Van Assche, F., (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language*, 60, 270-290.
- Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, 15, 517-530.
- Backus, A. (2005). Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism*, 9, 307-340.
- Backus, A. (2012). A usage-based approach to borrowability. *Tilburg Papers in Culture Studies*. http://www.tilburguniversity.edu/upload/f22a84e9-e3f0-41e4-bb7f-be84fad343d8_tpcs%20paper27.pdf (accessed 04/20/2013).
- Backus, A., Doğruöz, A.S., & Heine, B. (2011). Salient stages in contact-induced grammatical change: Evidence from synchronic vs. diachronic contact situations. *Language Sciences*, 33, 738-752.
- Barrutía, R., & Schwegler, A. (1994). *Fonética y fonología españolas*. New York: John Wiley & Sons.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language Is a Complex Adaptive System: Position Paper. *Language Learning*, 59, 1-26.
- Bills, G., & Vigil, N. (2008). *The Spanish Language of New Mexico and Southern Colorado: A Linguistic Atlas*. Albuquerque: University of New Mexico Press.
- Brown, E.K. (2011). Paradigmatic peer pressure: Word-medial, syllable initial /s/ lenition in Dominican Spanish. In S. Alvord (Ed.), *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology* (pp. 46-58). Somerville, MA: Cascadilla Proceedings Project.
- Brown, E.L. (2006). The effects of discourse context on phonological representation in memory. Paper presented at the 80th Annual Meeting of the Linguistic Society of America (LSA), Albuquerque, NM. January 5-8.
- Brown, E.L., (2013). Probabilistic cross-language phonological influence. Paper presented at the 24th Conference on Spanish in the United States and 9th Conference on Spanish in Contact with Other Languages, McAllen, TX. March 6-9.
- Brown, E.L., & Harper, D. (2009). Phonological evidence of crosslinguistic exemplar connections. *Studies in Hispanic and Lusophone Linguistics*, 2, 257-274.
- Brown, E.L., & Raymond, W.D. (2012). How discourse context shapes the lexicon: Explaining the distribution of Spanish f- / h- words. *Diachronica*, 29, 139-161.
- Brown, E.L., & Raymond, W.D. (in preparation). Contextual frequency effects in variation, change, and contact.
- Brown, E.L., & Rivas, J. (2012). Grammatical relation probability: How usage patterns shape analogy. *Language Variation and Change*, 24, 317-341.
- Bullock, B.E., & Gerfen, C. (2004). Phonological convergence in a contracting language variety. *Bilingualism: Language and Cognition*, 7, 95-104.
- Bullock, B.E., & Toribio, A.J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. In L. Isurin, D. Winford, & K. de Bot (Eds.), *Multidisciplinary approaches to code switching* (pp. 189-206). Amsterdam: John Benjamins.
- Bybee, J. (1999). Usage-based phonology. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan & K. Wheatly (Eds.), *Functionalism and formalism in linguistics* (pp. 212-242). Amsterdam: John Benjamins.
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language variation and change*, 14, 261-90.
- Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Bybee, J. (2012). Patterns of lexical diffusion and articulatory motivations for sound change. In M. Josep-Solé & D. Recasens (Eds.), *The Initiation of Sound Change: Perception, Production, and Social Factors* (pp. 211-234). Amsterdam: John Benjamins.
- Cole, J., Hualde, J.I., & Iskarous, K. (1999). Effects of Prosodic and Segmental Context on /g/-Lenition in Spanish. In O. Fujimura, B.D. Joseph, and B. Palek (Eds.), *Proceedings of the Fourth International Linguistics and Phonetics Conference* (pp.575-589). Thessaloniki: University Studio Press.
- Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and language*, 94, 94-103.
- Davies, M. (2002-) Corpus del Español (100 million words, 1200s-1900s). Available online at <http://www.corpusdelespanol.org>.
- Davies, M. (2008-) The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at <http://corpus.byu.edu/coca/>.
- de Groot, A.M.B. (1995). Determinants of bilingual lexicosemantic organisation. *Computer Assisted Language Learning*, 8, 151-181.
- de Groot, A.M.B. (2011). *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York: Psychology Press
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62, 284-301.
- Duñabeita, J.A., Perea, M., & Carreiras, M. (2010). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental Psychology*, 57, 98-107.
- Eddington, D. (2011). What are the contextual phonetic variants of /B, ð, G/ in colloquial Spanish? *Probus*, 23, 1-19.
- Erker, D., & Guy, G. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, 88, 526-557.
- Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. & Halle, P. 2008. Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics*, 36, 649-663.
- Franceschini, R. (2011). Multilingualism and Multicompetence: A Conceptual View. *The Modern Language Journal*, 9, 344-355.
- Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84, 474-496.
- Goral, M., Levy, E., Obler, L.K., & Cohen, E. (2006). Cross-language lexical connections in the mental lexicon: Evidence from a case of trilingual aphasia. *Brain and Language*, 98, 235-247.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3-15.
- Grosjean, F., & Miller, J.L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science*, 5, 201-206.
- Grosjean, F., & Li, P. (2013). *The Psycholinguistics of Bilingualism*. Oxford: Wiley-Blackwell.
- Guy, G. (1993). The quantitative analysis of linguistic variation. In D. R. Preston (Ed.), *American dialect research* (pp. 223-249). Amsterdam: John Benjamins.

- Hall, J., Cheng, A., & Carlson, M. (2006). Reconceptualizing Multicompetence as a Theory of Language Knowledge. *Applied Linguistics*, 27, 220-240.
- Hume, E., & Mailhot, F. (In press). The role of entropy and surprisal in phonologization and language change. In A. Yu, (Ed.), *Origins of Sound Patterns: Approaches to Phonologization*. Oxford University Press.
- Jacobs, A. (2007). Segmental production in bilingual speech: A psycholinguistic approach. Unpublished dissertation. University of Pennsylvania.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic Relations Between Words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229-254). Amsterdam: Benjamins.
- Kroll, J. & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25, 497-514.
- Leikin, M., Schwartz, M., & Tobin, Y. (2012). Current Issues in Bilingualism: A Complex Approach to a Multidimensional Phenomenon. In M. Leikin, M. Schwartz, & Y. Tobin (Eds.), *Current Issues in Bilingualism: Cognitive and Socio-linguistic Perspectives* (pp. 1-19). New York: Springer.
- Mendoza-Denton, N. (2004). Sociolinguistic extensions of exemplar theory: Comments on Flege, Khatib, and Darcy, Peperkamp and Dupoux. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 443-454). Berlin: Mouton de Gruyter.
- Menéndez Pidal, R. (1968). *Orígenes del Español*. 6a Edición. Madrid: Espasa-Calpe.
- Olson, D. (2012). The phonetics of insertional code-switching: Suprasegmental analysis and a case for hyper-articulation. *Linguistic Approaches to Bilingualism*, 2, 439-457.
- Olson, D. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41, 407-420.
- Penny, R.J. (1972). The Reemergence of /f/ as a Phoneme in Castilian. *Zeitschrift für Romanische Philologie*, 88, 463-482.
- Penny, R.J. (1991). *A History of the Spanish Language*. Cambridge: Cambridge University Press.
- Port, R., & Leary, A. (2005). Against formal phonology. *Language*, 81, 927-964.
- Rand, D., & Sankoff, D. (2001). GoldVarb: A variable rule application for Macintosh. Toronto: University of Toronto, Department of Statistics.
- Raymond, W. D., Dautricourt, R. & Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18, 55-97.
- Raymond, W.D. & Brown, E.L. (2012). Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In S. Th. Gries & D. S. Divjak (Eds.), *Frequency Effects in Language. Vol 2: Learning and Processing* (pp. 35-52). The Hague: Mouton de Gruyter.
- Runnqvist, E., Strijkers, K., Alario, X., & Costa, A. (2012). Cumulative semantic interference is blind to language: Implications for models of bilingual speech production. *Journal of Memory and Language*, 66, 850-869.
- Sanchez, T. (2008). Accountability in Morphological Borrowing: Analyzing a linguistic subsystem as a sociolinguistic variable. *Language Variation and Change*, 20, 225-253.
- Sankoff, D. (1988). Variable rules. In U. Ammon, N. Dittmar, & K. J. Mattheier (Eds.), *Sociolinguistics: An international handbook of the science of language and society* (pp. 984-997). New York: Walter de Gruyter.

- Sankoff, G. (2002). Linguistic outcomes of language contact. In J.K. Chambers, P. Trudgill, & N. Schilling-Estes (eds.), *The Handbook of Language Variation and Change* (pp. 638-668). Oxford: Blackwell Publishing.
- Teschner, R.V. (2000). *Camino oral: Fonética, fonología y práctica de los sonidos del español*. 2nd Edition. McGraw-Hill.
- Torres Cacoullous, R. & Ferreira, F. (2000). Lexical frequency and voiced labiodental-bilabial variation in New Mexican Spanish. *Southwest Journal of Linguistics*, 19, 1-17.
- Torres Cacoullous, R. & Travis, C.E. (In preparation). New Mexico Spanish / English Bilingual (NMSEB) corpus, National Science Foundation 1019112/1019122, <http://nmcode-switching.la.psu.edu/>
- Travis, C.E. & Torres Cacoullous, R. (2013). Two languages, one effect: structural priming in spontaneous code-switching.
- Waltermire, M. (2010). Variants of intervocalic /d/ as markers of sociolinguistic identity among Spanish-Portuguese bilinguals. *Spanish in Context*, 7, 279-304.
- Winford, D. (2003). *An Introduction to Contact Linguistics*. Oxford: Blackwell Publishing

Appendix: Responses per word given by three consultants indicating cognate with English (N = 2629)

word	N	defendiendo	3	desde	0	diferencias	3	doctor	3
da	0	deja	0	deseabas	2	diferente	3	doctores	3
daba	0	dejaba	0	desfendía	3	diferentes	3	dólar	3
dábamos	0	dejaban	0	desierto	3	diffcil	2	dólares	3
daban	0	dejado	0	desmayar	0	diga	0	dolía	0
dabas	0	déjalo	0	desollábanos	1	dígame	0	dolían	0
dado	0	déjame	0	desollé	1	digan	0	dolió	0
daga	3	dejamos	0	despachaba	0	digas	0	dolor	0
dale	0	dejan	0	despacio	0	digo	0	dolorosos	0
dame	0	dejar	0	despeinada	0	dije	0	domingo	0
damos	0	dejarne	0	despenada	0	dijera	0	domingos	0
dan	0	dejaron	0	despidiendo	0	dijeran	0	dompeaba	3
dando	0	dejas	0	despidió	0	dijeras	0	dompeaban	3
dándole	0	dejé	0	después	0	dijeron	0	donden	0
dar	0	déjenme	0	destapaba	1	dijimos	0	dóndequiera	0
daré	0	dejo	0	detener	0	dijiste	0	dormía	0
darle	0	dejó	0	detenida	0	dijo	0	dormíamos	0
darles	0	del	0	detrás	0	dile	0	dormían	0
das	0	delante	0	detuve	0	dime	0	dormida	0
de	0	demás	0	di	0	dinerito	0	dormido	0
dé	0	demasiado	0	día	2	dinero	0	dormir	0
debajo	0	demen	0	días	2	dio	0	dos	0
debe	0	démenlo	0	dibujo	0	Dios	0	doscientas	0
deben	0	den	0	dice	0	dirección	3	doy	0
debía	0	denme	0	dicen	0	diremos	0	duele	0
debían	0	depende	3	dices	0	dirían	0	duelen	0
debías	0	deponder	3	dicho	0	disciplina	3	dulce	0
decía	0	deprimido	3	dichos	0	dispare	1	dulces	0
decíamos	0	derecha	0	diciembre	3	dispuesto	0	dura	0
decían	0	derechita	0	diciendo	0	distantemente	2	duramos	0
decíanos	0	derecho	0	diciéndole	0	diste	0	durante	0
decidieron	3	derrite	0	diecinueve	0	distrito	3	duraron	0
decimos	0	desabrochaba	0	diecisiete	0	divertíanos	0	duras	0
decir	0	desague	1	dientes	1	divertido	0	duraznitos	0
decirle	0	desaparecer	3	dentista	3	divirtiéndolo	0	durmiendo	0
decirlo	0	desapareces	3	diera	0	divorciado	3	durmieron	0
decirte	0	desaparecías	2	dieran	0	dizque	0	duro	0
dedo	0	desaparecieron	2	dieron	0	doble	3	duró	0
dedos	0	desbarataron	0	diez	0	doce	0		
defender	3	descendidos	3	diferencia	3	docena	3		

