CODE-SWITCHING AND PREDICTABILITY OF MEANING IN DISCOURSE

Mark Myslín — *University of California, San Diego*
Roger Levy — *University of California, San Diego*

What motivates a fluent bilingual speaker to switch languages within a single utterance? We propose a novel discourse-functional motivation: less predictable, high information-content meanings are encoded in one language, and more predictable, lower information-content meanings are encoded in another language. Switches to a speaker's less frequently used, and hence more salient, language offer a distinct encoding that highlights information-rich material that comprehenders should attend to especially carefully. Using a corpus of natural Czech-English bilingual discourse, we test this hypothesis against an extensive set of control factors from sociolinguistic, psycholinguistic, and discourse-functional lines of research using mixed-effects logistic regression, in the first such quantitative multifactorial investigation of code-switching in discourse. We find, using a Shannon guessing game to quantify predictability of meanings in conversation, that words with difficult-to-guess meanings are indeed more likely to be code-switch sites, and that this is in fact one of the most highly explanatory factors in predicting the occurrence of code-switching in our data. We argue that choice of language thus serves as a formal marker of information content in discourse, along with familiar means such as prosody and syntax. We further argue for the utility of rigorous, multifactorial approaches to sociolinguistic speaker-choice phenomena in natural conversation.*

Keywords: code-switching, bilingualism, discourse, predictability, audience design, statistical modeling

1. Introduction. In an early sketch of language contact, André Martinet (1953:vii) observed that in multilingual speech, choice of language is not dissimilar to the 'choice[s] among lexical riches and expressive resources' available in monolingual speech. In code-switching situations, multilingual speakers are faced with a continual choice between roughly meaning-equivalent alternatives from each language. What governs this choice when meanings can be expressed equally well in either of two languages? One line of explanation is that in both monolingual and multilingual contexts, choices between distinct linguistic forms have informative functions in the larger, interactive discourse context. Many of these functions have to do with the flow of information between participants: for example, important, less predictable, or conversationally confrontational meanings might be marked by a distinct or more extensive linguistic encoding (e.g. Fox & Thompson 2010, Jaeger 2010, Karrebaek 2003, among many others). When multiple languages are available, each one may serve as a distinct encoding of this kind. One factor governing the choice between languages, then, might be a need to signal meanings that are less predictable in context and thus carry more information. We hypothesize that there is a tendency for these less predictable, high information-content meanings to be encoded in one language, and for more predictable, lower information-content meanings to be encoded in another language. In this way, switches to a speaker's less frequent, and hence more salient, language offer a distinct encoding that serves to highlight information-rich material that must be especially carefully attended to.

* We are grateful to audiences at Deutsche Gesellschaft für Sprachwissenschaft 2012, the Santa Barbara Cognition and Language Workshop 2012, and the Linguistic Society of America 2013 annual meeting; members of the UCSD Computational Psycholinguistics Lab; and Mary Bucholtz, Vic Ferreira, Stefan Gries, Florian Jaeger, Andy Kehler, Marianne Mithun, Pieter Muysken, and an anonymous referee for insightful feedback on this work. We also thank our speakers and participants for their invaluable contributions to this research. Any remaining errors and omissions are our own. This work was supported by a Jacob K. Javits Fellowship and NSF Graduate Fellowship to MM, and by NSF grant 0953870 and an Alfred P. Sloan Research Fellowship to RL.

The status of code-switching as a speaker choice, as well as its potential correlation with the information content of meanings, is illustrated in the following instance of Czech-English code-switching from a speech community in California. The speaker is persuading his bilingual interlocutor not to go out with a particular woman.

(1) Tady vidiš že ona je **in need**.
  'Here you see that she is **in need**.'
(2) A potřebuje **entertainment**.
  'And she needs **entertainment**.'

The concept of NEED is expressed in both English and Czech, by the same speaker, in two consecutive clauses. One contextual property that differs between the two tokens, however, is the amount of relevant prior information given in the discourse. In 1, the concept of NEED, encoded in English, is being mentioned for the first time and thus represents a highly informative, discourse-new predicate. In 2, in contrast, the concept, now expressed in Czech as *potřebuje* 'needs', has just been mentioned in the immediately preceding clause and so does not carry new information. The new piece of information in 2, namely the object of NEED, is expressed in English as *entertainment*. In both clauses, then, low information-content material is encoded in Czech, and high information-content material is encoded in English, regardless of the particular concept being expressed. This pattern is consistent with our hypothesis that language choice in code-switching is a formal marker of information content, with switches to the less frequent—and thus more salient—language (here, English) serving as a cue to less predictable meanings that comprehenders must attend to especially carefully.

The article has three objectives. The first is to develop a formal account of code-switching and information content. We build on discourse-functional explanations in which choices between forms carry out conversational functions such as marking information-structural status or simply 'importance' of certain material (e.g. Karrebaek 2003:431). To make these conceptualizations of information concrete and testable, we employ MEANING PREDICTABILITY as a reflection of a word's information content in context: the less predictable a word's meaning, the more information that word carries, and the higher its probability of receiving a distinct encoding by means of a code-switch. Information-theoretic metrics derived from predictability (Shannon 1948) correlate with other speaker choices from phonetics to (morpho)syntax and discourse (Aylett & Turk 2004, Bell et al. 2003, Genzel & Charniak 2002, Jaeger 2006, 2010, Komagata 2003, Levy & Jaeger 2007, Mahowald et al. 2013, Piantadosi et al. 2011, Qian & Jaeger 2012, Tily et al. 2009, Tily & Piantadosi 2009). A more complete description of this approach is given in §3.4.

The second objective of the article is to test this meaning-predictability account of code-switching against multiple control factors inspired by insights from several disciplines. Sociolinguistic, discourse-functional, and psycholinguistic traditions offer potentially compelling explanations of code-switching, but these generally do not systematically consider multiple factors in code-switching. Using multifactorial statistical techniques, we investigate for the first time the respective contributions of an extensive, cross-disciplinary range of factors long hypothesized to inform code-switching.

The third objective of the article is to bridge a methodological gap in existing code-switching research between observational and experimental methods, by analyzing a naturalistic data set of spontaneous speech using rigorous statistical methods. On the one hand, many observational studies to date have focused on small numbers of individual instances of code-switching rather than making statistical generalizations about code-switching or the speech community under investigation. In experimental settings, on the other hand, code-switching behavior is markedly different than in its natural discourse habitat (discussed below in §4) and may be further distorted by exposure to probability distributions that are unusual in natural language, such as uniform distributions resulting from balanced designs, rather than, for example, Zipfian distributions more typical to naturalistic use (Jaeger 2010). We thus argue for rigorous corpus-driven approaches to code-switching research, building on similar methodological advances in monolingual settings (e.g. Gahl 2008, Gries et al. 2005, Jaeger 2010, Szmrecsanyi 2005, Tagliamonte 2006, Wasow 2002).

The article is structured as follows. We first survey the sociolinguistic, psycholinguistic, and discourse-functional control factors in our analysis (§2), and then build on discourse-functional insights to propose a meaning-predictability account of code-switching (§3). The data set of spontaneous discourse is introduced in §4, and §5 describes an experiment to estimate the predictability of words in conversation. The results of the logistic regression model testing the predictability account against control factors are then presented (§6), followed by a discussion of their generalizability (§7) and then a conclusion (§8).

2. DEFINING CODE-SWITCHING. We adopt a definition of code-switching as the alternation of multiple languages within a single discourse, sentence, or constituent (e.g. Poplack 1980) by FULLY PROFICIENT MULTILINGUALS.[1] We focus exclusively on contexts where switching is a true SPEAKER CHOICE between alternatives with (near-)equivalent truth-conditional meaning: in other words, there is no dependence between the language of a particular word and the literal state of affairs communicated by it. One hallmark of this situation is reference to the same object by the same speaker in different languages, implying that differences in proficiency or meaning in either language are not at play. To be sure, language choice may be imbued with metaphorical and social meaning (e.g. Gumperz & Hymes 1986), and indeed this is one of the factors discussed below that are hypothesized to govern choices between truth-conditionally equivalent forms. This assumption of (near-)equivalence in truth-conditional meaning is implicit in most code-switching research, although some researchers make it explicit by comparing language choice in code-switching to synonym choice in monolingual speech (Gollan & Ferreira 2009, Martinet 1953, Moreno et al. 2002, Sridhar & Sridhar 1980).

3. WHY CODE-SWITCH? In this section, we introduce the existing sociocultural, psycholinguistic, and discourse-functional explanations to be evaluated alongside our meaning-predictability proposal in answering the question: why switch between languages when the truth-conditional meanings offered by each are essentially equivalent?

3.1. SOCIOCULTURAL FACTORS. In sociocultural approaches, language switching is a resource that can be used to construct identity, modulate social distance and affiliation, and carry out interspeaker accommodation (Beebe & Giles 1984). For example, code-switching itself may be the unmarked choice for a community in which speakers maintain affiliation with two different socioethnic groups simultaneously (Myers-Scotton 1993b). However, these accounts generally do not make explicit, word-by-word predictions of language choice—and it is indeed antithetical to some of these approaches to assume fixed, predictable functions of code-switching not individually constructed in the

[1] Following this definition, we employ CODE-SWITCHING as a blanket term for switches both within and between utterances. Although CODE-MIXING is sometimes used for intrasentential switching, consensus is not widespread on the term's precise meaning and the theoretical distinctions it may make (see discussion in Matras 2009). Therefore, we simply refer to all of these phenomena as *code-switching*.

local context of each switch (Bailey 2000). Nevertheless, if code-switching is a tool to signal affiliation with social groups, code-switching patterns should depend in part on the participants present and their social affiliations. For example, young, English-dominant speakers may be expected to switch to Spanish more often when older, Spanish-dominant speakers are present and the younger speakers wish to accommodate them or show affiliation. Participant constellation—the social makeup of the group of participants present in a discourse episode—is thus a testable factor affecting language choice in these sociocultural approaches.

3.2. Psycholinguistic factors. Psycholinguistic approaches to code-switching, in contrast, traditionally treat language choice as a largely automatic function of speaker-internal production circumstances, unaffected by discourse-functional goals or conscious control. Most models of bilingual production parallel standard models of monolingual production, in which messages are first formulated before passing through a stage of lexical (lemma) selection followed by morphophonological encoding and finally articulation (e.g. Levelt 1999; see Ferreira & Slevc 2007 and Ferreira 2010 for reviews of such models). These models assume that bilinguals have a single conceptual store shared by both languages, and that language selection takes place later during the lexical selection phase of production, either through higher activation of a lemma in one language, or through failure to inhibit the lemma in that language (for discussion, see e.g. Costa & Santesteban 2004, Marian 2009). In this section, we review the factors that may affect lexical activation (or inhibition) in each language, beginning with baseline lexical accessibility before turning to contextual and syntactic factors.

Baseline lexical accessibility. A common intuition is that a speaker will choose the language in which the desired word first comes to mind (e.g. Gollan & Ferreira 2009). All else being equal, then, lexical selection among multiple languages is subject to each (language-specific) lemma's baseline accessibility—how easily it can be retrieved from the lexicon for production, irrespective of context. Since higher word frequency and shorter length each increase accessibility (D'Amico et al. 2001, Forster & Chambers 1973), multilingual speakers may be more likely to use the language in which the relevant word is shorter or more frequent (Heredia & Altarriba 2001).

A related word-inherent property is the way its meaning is stored in the bilingual lexicon. In the standard models of bilingual production described above, bilinguals first access meanings from a single semantic system, and subsequently choose a language during lexical selection. An alternative view is that the semantic system is only partially shared across languages: nouns are stored in a common system, but verbs and other words reside in language-specific parts of the semantic system, since these words elicit slower and less consistent associations across languages (Marian 2009, Van Hell & de Groot 1998). This makes nouns more 'portable', or switchable, a prediction that is consistent with observations that they are the word class most frequently code-switched (e.g. Myers-Scotton 1993a) and borrowed (Muysken 2000), followed by verbs and then other parts of speech. Nouns are thus predicted to be code-switched most often, followed by verbs and then by other words.

Similarly, concreteness and imageability, in addition to part of speech, affect lexical accessibility in the bilingual lexicon. Concrete, highly imageable words such as *tiger* are translated faster and elicit more reliable crosslinguistic priming (Van Hell & de Groot 1998) than abstract words such as *liberty*, suggesting that concrete words are more integrated in the bilingual lexicon than abstract words. Because of this tighter integration, concrete words' translation equivalents are more likely to be coactivated in production than abstract words' translation equivalents, predicting greater probability of code-switching for concrete, imageable words than for abstract words (Marian 2009).

LEXICAL AND SYNTACTIC CONTEXTUAL FACTORS. In addition to the above properties of the event of a single word's production, properties of the context also affect bilingual lexical activation and thus the probability of code-switching. One of these is language-specific LEXICAL COHESION: Munoa (1997) and Angermeyer (2002) observe that lexical items often persist in their original language of mention, even if the embedding stretch of discourse is in a different language. This persistence of language choice may serve to bolster cohesive ties to previous mentions (Angermeyer 2002) and/or result from automatic priming, in which activation of language-specific lemmas facilitates subsequent productions in the same language (Kootstra et al. 2009). Thus, words are likely to reoccur in their language of most recent mention.

Another contextual factor in language choice is TRIGGERING. Trigger words, such as the proper noun *California*, may be stored in completely shared representations across language systems. When a trigger is produced, it increases the activation of the second language, thereby increasing the probability that the next word is a code-switch (Clyne 1991, 2003, Riehl 2005). Trigger words comprise three types: proper nouns, phonologically unintegrated loanwords from the second language, and bilingual homophones. This last category consists of words from different languages that are pronounced identically, such as Dutch *smal* 'narrow' and English *small* (Clyne 2003:164). Broersma and de Bot (2006) and Broersma (2009) revise the original triggering hypothesis to take entire clauses, rather than bigrams, as speech planning units, and indeed observe facilitation of code-switching if a trigger word is present anywhere in the clause rather than just immediately adjacent the potential switch site.

A third contextual factor in language choice is language-internal collocational strength between words. Backus (2003) argues that sequences of words that often cooccur in one language are accessed as units and are therefore unlikely code-switch sites. Thus a code-switch from, say, English to Spanish within a strong collocation such as *all over the place* (e.g. *all over* ***el lugar***) is less likely than a code-switch within a weaker collocation such as *all over the city* (e.g. *all over* ***la ciudad***).

The final contextual factor in the probability of code-switching that we examine is syntactic dependency distance.[2] In an extension of DEPENDENCY LOCALITY THEORY (Gibson 1998, 2000), Eppler (2011) provides evidence from spontaneous German-English code-switching that the greater the number of intervening words between a potentially code-switched word and its syntactic governor, the more difficult it is to track the (language-specific) dependency due to memory constraints, and therefore the less likely the word is to match its syntactic governor in language choice. Together with other contextual factors outlined above, as well as inherent properties of a word's lexical accessibility, these factors reflect a broad set of speaker-internal psycholinguistic production circumstances that may inform code-switching behavior.

3.3. DISCOURSE-FUNCTIONAL FACTORS. In the final class of explanations of code-switching, discourse-functional approaches, code-switching serves to signal contrasts between portions of speech. In other words, switches are CONTEXTUALIZATION CUES in

[2] Another syntactic class of models of code-switching specifies grammatical constraints on the possibility of switching (e.g. Joshi 1982, Myers-Scotton 2002, Poplack 1980). We do not discuss these models in detail, since our investigation concerns motivations for switching GIVEN THAT IT IS GRAMMATICALLY POSSIBLE. However, because even these grammatical accounts stipulate some exceptions, a probabilistic implementation would be a natural future extension to these models.

the sense of Gumperz (1982:131), with wide-ranging discourse functions such as clarification, emphasis, or qualification of information (e.g. marking some material as a parenthesis, personal comment, or reported speech, even in a language other than that of the original speech; Auer 1995:120, Gumperz 1982:79, Zentella 1997). For example, de Rooij (2000) observes that discourse markers occur predominantly in French in a Shaba Swahili-French code-switching data set and argues that this strategy functions to increase the salience of these discourse markers, since they occur in the less frequent, and thus more salient, language in this community.

One key class of discourse functions of code-switching centers explicitly on the information status of concepts. One function within this class is the signaling of new discourse topics (Munoa 1997, Zentella 1997). Munoa (1997) reports that new topics can be signaled by Spanish noun phrases in otherwise Basque clauses, as in the following example in which the question of restroom availability in various venues is under discussion.

(3) Fabrika baten ere da **un servicio al público**.
 ‘A factory is also **a public service**.’

PUBLIC SERVICE is introduced with a code-switch to Spanish, and the conversation subsequently turns to examples and characteristics of public services, rather than continuing with the topic of restrooms. Since, as Munoa argues, PUBLIC SERVICE could easily have been expressed in Basque, the code-switch is best explained as functioning to mark a new topic.

Code-switching may also serve as a strategy to contrast topic and focus elements in discourse. Romaine (1989:162) collects cases of code-switch boundaries corresponding to topic-focus boundaries, including French-Russian (Timm 1978), English-Spanish and English-Hindi (Gumperz 1982:79), and Hebrew-English (Doron 1983). Ritchie and Bhatia (2004) contribute an additional Hindi-English example. Consider Gumperz’s (1982) English-Hindi case.

(4) Bina veṭ kiye ap a gəe?
 ‘Without waiting you came?’

(5) Nehĩ, I came to the bus stop **nau bis pəččis pər.**
 ‘No, I came to the bus stop **about nine twenty-five**.’

The speaker in 5 first reprises the topic of COMING in English before switching to Hindi for the focused TIME OF ARRIVAL, thus demarcating information status through language choice. Code-switching may also conspire with other topic-marking strategies: Franceschini (1998) reports cases of fronted, topicalized noun phrases spoken in Swiss German with Italian predicates, while Nishimura (1989) observes topic elements in Japanese accompanied by the usual topic-marking particle *wa*, but followed by English comments.

Karrebaek (2003) asks whether particular languages must be stably associated with topic or focus within discourse episodes. In her data, Turkish and Danish are interchangeable in their status as topic-marker or focus-marker, suggesting that in some cases it is code-switching itself that carries out the topic-focus marking function, rather than language-specific associations. Karrebaek concludes that it is simply the ‘important’ discourse information that receives a language encoding different from its immediate context (2003:431). In summary, a variety of observed cases suggests that language choice marks information status of concepts.

Important questions remain, however, about the systematicity of the correlation between information structure and language choice. First, because these arguments have, to our knowledge, exclusively been made on the basis of individual tokens of code-switching, it is unclear whether the correlation is reliable even within single speakers, let alone across entire speech communities. As we argue below, if code-switching is to serve discourse functions for the benefit of comprehenders, some systematicity would help them learn and draw the right inferences about the information-structural functions of switches (§8.1).

Second, and perhaps more crucial from a theoretical point of view, it is also unclear whether any precise informational principle unites the studies above; instead information status (variously construed) is argued for on a case-by-case basis within each study, and large numbers of more ambiguous examples are excluded from the analyses (e.g. Karrebaek 2003:431). As a result, evidence for the correlation between information status and language choice is limited to a small number of examples selected for ease of subjective information-structural analysis. An alternative approach is to adopt a more precise operationalization of information that can be straightforwardly tested across entire data sets. In the next section, we argue for predictability of meanings as such a metric.

3.4. Meaning predictability and speaker-choice phenomena. In line with the discourse-functional accounts of code-switching described above, a ubiquitous intuition in accounts of speaker-choice phenomena is that some content is more important or informative than other content, and it is this disparity that governs choices between alternant linguistic forms: important material receives the 'more explicit, more distinct, or more extensive encoding' (Karrebaek 2003:431; see also e.g. Givón 1985:206). In order to test these accounts, however, we need an explicit, objective operationalization of importance or informativity. In this section, we argue for predictability of meanings as a useful metric of information, since (i) it is a building block in many theories of information structure, (ii) it is objectively measurable, and (iii) it correlates with a wide range of speaker-choice phenomena. We discuss these properties in turn below.

First, numerous theories of information structure characterize information in terms of predictability, starting from the intuition that the more predictable some content is, the less (new) information it contains. Classic information-structural distinctions such as topic vs. focus and given vs. new have long been cast in these terms: Prince (1981), following Halliday (1967), Halliday and Hasan (1976), and Kuno (1972, 1978, 1979), includes a predictability dimension in her definition of given information, classifying information as given if the speaker assumes the hearer can predict it. Topic-focus structure has also been defined in terms of predictability: according to Lambrecht (1994:6), topic and focus refer to the relative predictability of the relations between propositions and their elements in a given discourse situation. Although we certainly do not propose to reduce all information-structural categories to predictability, it is clearly relevant to information-structural distinctions that have been claimed to inform language choice in code-switching.

A second attractive property of predictability is that it is objectively measurable, through, for example, cloze methodology (Taylor 1953), and it interfaces naturally with the mathematical framework of information theory (Shannon 1948). Here information is inversely related to probability in context: the less probable certain material is, the higher its information content. Formally, the information content I (or surprisal; Hale 2001, Levy 2008) of the meaning m of a unit of an utterance is the logarithm-transformed inverse of the probability of m in context.

(6) $I(m) = \log_2 \frac{1}{P(m|\text{context})}$

Surprisal is positively correlated with human processing difficulty (Demberg & Keller 2008, Smith & Levy 2013), providing evidence that comprehenders are sensitive to predictability of meanings.

Finally, not only does predictability affect comprehension, but it also affects production: when multiple grammatical options are available, speakers choose to convey less predictable meanings with distinct or more extensive encodings, thus distributing information uniformly across the linguistic signal to the extent possible (UNIFORM INFORMATION DENSITY; Jaeger 2006, 2010, Levy & Jaeger 2007). Phonetic duration and articulatory detail are reduced for meanings with high predictability, at the level of both syllables (Aylett & Turk 2004) and words (Bell et al. 2003, Tily et al. 2009). Word lengths, too, are optimized such that the more predictable a meaning in context, the shorter the word conveying it (Mahowald et al. 2013, Piantadosi et al. 2011). Speakers choose contractions over full variants when the meanings conveyed are more predictable (Frank & Jaeger 2008). Referring-expression choice is similarly correlated with surprisal, so that pronouns are chosen over noun phrases for more predictable referents (Tily & Piantadosi 2009). In syntax, optional *that* is mentioned when the upcoming complement or relative clauses are least expected, thus distributing the surprisal associated with these clause onsets across an additional word and minimizing peaks in information density (Jaeger 2006, 2010, Wasow et al. 2011). Komagata (2003) argues that word order is also sensitive to a preference for a uniform distribution of information. At the discourse level, the more contextual information that precedes a sentence, the greater the sentence's unpredictability in isolation, suggesting that information is distributed uniformly across discourses (ENTROPY RATE CONSTANCY; Genzel & Charniak 2002, Qian & Jaeger 2012). Predictability, in sum, is not only relevant to information structure and objectively measurable, but also affects speaker choices in language production.

What underlies this correlation between unpredictability of meanings and more extensive or distinct encodings? On the assumption that message transmission is a noisy channel between interlocutors, one explanation is that speakers take their interlocutors' knowledge state into account and choose more extensive encodings to allow more detailed processing of unpredictable meanings that have an inherently higher risk of miscommunication (AUDIENCE DESIGN; see discussion in Jaeger 2010). In the next section, we relate this correlation between predictability and speaker choice to the functional, information-based motivations for code-switching introduced in §3.3.

4. A MEANING-PREDICTABILITY ACCOUNT OF CODE-SWITCHING. Predictability of meanings is correlated with speaker choice, so that less predictable, more informative meanings receive a more extensive or distinct encoding. Code-switching is a choice that allows for these distinct encodings. If code-switching indeed serves to highlight important information, less predictable, more informative meanings should be code-switch sites. In other words, the less predictable a meaning, the more likely a code-switch.

What is the communicative function of choosing a distinct language encoding for less predictable information? The strategy may be motivated by audience design: speakers choose more salient encodings in order to highlight less expected information and potentially minimize risk of miscommunication. The distinct encoding available through a language switch may direct comprehender attention to less predictable material, thus serving as a comprehension cue analogous to morphemic topic markers or topicalization through syntactic fronting. This cue can be made even more salient if the DIRECTION of the switch is taken into account: since code-switchers generally use only one language for the majority of words (Grosjean 1997, Myers-Scotton 1993a), words from a speaker's less frequently used language offer a more salient encoding by virtue of relative rarity (an argument related to the one by de Rooij (2000); see our §3.3). This leads to a more specific prediction: a switch to a speaker's less frequent and therefore more salient language may alert comprehenders to high information content that must be especially carefully attended to.

One supporting mechanism for this process may be the phonological distinctiveness of alternant languages in code-switching. Code-switching is characterized by total alternation not only between grammatical systems but also between phonological systems (e.g. Grosjean & Miller 1994, Sankoff & Poplack 1981). Distinctive phonology of a code-switched word may therefore serve as a low-level cue of encoding difference even before the word is completed by the speaker and fully processed by the comprehender. Suggestive evidence is provided by a gating study by Li (1996): participants were asked to guess (code-switched or non-code-switched) words on the basis of increasingly long fragments of the word, and guesses converged on the correct language before the full word was correctly identified. Further, anticipatory phonetic signatures of impending code-switches appear on some words immediately preceding switch boundaries, and comprehenders may be sensitive to these markers (Piccinini 2012, Weiss et al. 2009). In this way, phonological cues of other-language encoding are available at multiple points before an initial code-switched word is fully processed, potentially alerting comprehenders to allocate more attention in anticipation of an unpredictable meaning.

However, one may counter that a language switch itself is costly to process, and may consume whatever extra resources are needed to process an already difficult-to-predict meaning. Some studies report this kind of language switch cost in comprehension: for example, Proverbio, Leoni, and Zani (2004) observe longer reaction times and increased N400 amplitudes for code-switched words in a sensibility-judgment task in reading. A number of factors mitigate such switch costs, however. First, task effects are relevant: in AUDITORY comprehension tasks, which better replicate the natural conversational locus of code-switching than do reading tasks, code-switched words in sentential contexts are recognized as quickly as non-code-switched words (Li 1996) (possibly thanks to the phonological cues discussed above). Second, discourse-level context facilitates processing of code-switching. Chan, Chau, and Hoosain (1983) report that reading times for entire mixed-language passages were the same as those for equivalent monolingual passages. Third, accurate expectations for upcoming code-switches are likely to reduce switch costs. Moreno, Federmeier, and Kutas (2002) argue that the enhanced late positivity (LPC) they observe for code-switched words is reduced when comprehenders find the switch less unexpected. Indeed, when switch locations become predictable, LPC switch costs are not observed at all (Proverbio et al. 2004). Thus switch costs appear to be reduced or absent in auditory processing, rich discourse contexts, and situations in which switches are relatively predictable, supporting code-switching as a viable comprehension cue.

In sum, communicative principles underlying both discourse-functional accounts and meaning-predictability accounts suggest that more informative, less predictable meanings should receive a distinct or more extensive encoding. In multilingual situations, a speaker's lesser-used language offers this more salient encoding and may serve as a comprehension cue to direct attention to less expected meanings. This account of code-switching thus predicts that words conveying unpredictable meanings should be code-switch sites.

5. Data. The data for this study consist of three hours of spontaneous Czech-English conversation among five proficient bilinguals of Czech heritage living in California. Two of the bilingual speakers, ages fifty-five and sixty, were monolingual in Czech until immigrating to the United States in their early thirties and learning English, but remain Czech dominant. A third speaker, thirty-three, was monolingual in Czech until moving to the United States and beginning English acquisition at age five and subsequently becoming English dominant. The final two speakers, twenty and twenty-six, were born in the United States and are English-dominant but have used Czech in family interaction and occasional socializing with Czech friends in the United States since childhood. Participants gave a blanket consent to have their conversations recorded at unannounced intervals during a two-month period, and were therefore unaware of specific recording times until after the fact. Following the two-month period, each participant had the option of reviewing the recordings and requesting deletion of any portion thereof.

The three hours of Czech-English conversation in the final data set are distributed as follows. One hour, representing three different conversations, consists of interaction between all of the speakers. Another hour (four conversations) is limited to the two older speakers only. The final hour (three conversations) consists of one-on-one interactions between the youngest speaker and each of the older speakers (approximately half an hour each).

The data were collected and transcribed by the first author for an unrelated project prior to the formulation of the current research question, following the methods in Du Bois et al. 1993. Each line consists of one INTONATION UNIT (IU), a sequence of words produced under a single, coherent intonational contour (Chafe 1987, 1994). Intonation units are perceptual units distinguished through (i) pitch-resets, (ii) final-word lengthening, (iii) intensity changes, (iv) pauses, and/or (v) changes in voice quality. Although IUs are defined with respect to these perceptual auditory features and not syntactic features, they generally emerge as approximate clause-equivalents and are, according to Chafe, cognitive units in discourse, each containing no more than one new idea (1987:32). Shenk (2006) shows IUs to be relevant units in code-switched discourse, finding that 96% of code-switches in a one-hour corpus of Spanish-English code-switching correspond to IU boundaries.

5.1. LANGUAGE DISTRIBUTION WITHIN IUS. Of the 3,201 IUs comprising the current data set, approximately 52% are monolingual Czech IUs, 25% are monolingual English IUs, and 23% are mixed-language IUs (see Table 1). Czech → English switches are by far the more common switch type, and switching typically occurs late in the IU: the most frequent switch is a single, final-word switch to English.

MONOLINGUAL		MIXED-LANGUAGE		
CZE	ENG	CZE → ENG	ENG → CZE	MULTIPLE SWITCHES
1,668	796	601 (494)	24 (13)	112

TABLE 1. Intonation units comprising the corpus. Quantities in parentheses are the subsets of the relevant IU type that are characterized by single, final-word code-switches.

5.2. ITEMS FOR ANALYSIS. The distribution of code-switches in the corpus is given in Table 1. The current analysis, however, focuses on a particular class of speaker and code-switch. Only those IUs produced by the two older speakers are included as critical items, since these speakers are the most fully proficient bilinguals and are therefore the least likely to switch languages for reasons of incomplete proficiency in either language. Further, they have equivalent language backgrounds and are each fully proficient in Czech, while the younger speakers vary much more dramatically in their Czech proficiencies. This is reflected in one way by the proportion of monolingual Czech IUs produced by each speaker: for the older speakers, it ranges from 71–74%, and for the younger speakers, it ranges from 12–40%. The older speakers are also the most prolific intrasentential code-switchers, together producing 81% of the mixed-language IUs in the corpus, and are the only speakers participating in all ten conversations.

A final point of homogeneity among the older speakers is code-switch position within IUs. To quantify this, we can define a normalized IU-position metric as in 7, so that 0 corresponds to initial words and 1 to final words, with all words equidistant from each other.

(7) $$\text{IU position} = \frac{\text{word number} - 1}{\text{number of words in IU} - 1}$$

For the older speakers, the median IU position of English words was consistently 1 (interquartile ranges (IQRs): 0.39 and 0.50), whereas younger speakers had medians of 0.67, 0.78, and 1.0 (IQRs = 0.60, 0.50, 0.38). In other words, the older speakers have a strong and consistent tendency toward one type of code-switch: a final, single-word switch from Czech to English. These are the switches investigated here.

We investigate the relative contributions of meaning predictability and other control factors to the propensity to code-switch by posing the following question of our data: when a fully bilingual speaker has produced an IU entirely in Czech from the first through the penultimate word, how likely is she to produce the final word in English? Therefore, our crucial items for analysis were all older-speaker IUs that begin in Czech and either (i) feature a final, single-word switch to English (SWITCH ITEMS, $n = 253$) or (ii) do not contain any code-switch (NONSWITCH ITEMS, $n = 472$). To confirm that the final word of a given nonswitched IU was in principle switchable, we asked the original speakers to replace the final Czech word in their own utterances with a single-word switch to English, and in all cases the speakers found this possible. We further verified that none of these potential switch sites violated hypothesized grammatical constraints on switching (Myers-Scotton 2002, Poplack 1980). In other words, ALL AND ONLY items with an actual or potential single-word IU-final code-switch to English were considered. Each speaker contributed roughly equivalent proportions of switch and nonswitch items. Precise counts and examples are provided in Table 2.

ITEM TYPE	DESCRIPTION	N (N_{Spkr1})	EXAMPLE
SWITCH	Czech IU with final single-word switch to English	253 (127)	A potřebuje **entertainment**. CONJ need.3SG ‘And she needs **entertainment**.’
NONSWITCH	Monolingual Czech IU	472 (197)	Ona se na tebe bude lepit. 3SG.F REFL on 2SG FUT cling ‘She will cling to you.’

TABLE 2. Items for analysis. Total numbers of each item type are provided, as well as (in parentheses) the subset of these attributed to speaker 1.

6. METHODS. To test for an effect of meaning predictability while controlling for other factors known to affect code-switching, we employ binary logistic regression. The dependent variable is presence (1) or absence (0) of code-switching (that is, whether each item is a switch item or a nonswitch item as described above), and the independent variables include meaning predictability and ten control factors. Operationalization of these control factors, which were introduced in §3, is described with respect to the current data set in §6.1 below; §6.2 describes how we estimate meaning predictability in discourse. Table 3 below summarizes all factors in the logistic regression.

6.1. Control factors.

Participant constellation. Since speakers may code-switch in order to accommodate other participants' preferences or establish affiliation with various social groups, code-switching behavior may vary as a function of the participants present in a given conversational episode. In the current data set, participant constellation has two levels reflecting the presence or absence of younger, United States-born participants in the conversation. More code-switching to English on the part of the older speakers is expected when any younger participant is present.

Baseline lexical accessibility. Speakers are expected to choose the language where the relevant word is more accessible. We determined accessibility levels for the final word in each switch and nonswitch item, since these are the potential switch sites (see §5.2). Since all code-switches are to English, greater accessibility of the final word in English predicts the item to be a switch item, and greater accessibility of the final word in Czech predicts the item to be a nonswitch item.

The first operationalization of accessibility was word frequency.[3] The general frequencies of attested final words in their original language were compared with the frequencies of their translation equivalents in the other language—that is, the frequencies of the words that would have been spoken had the speaker made the opposite language choice in each case. Translation equivalents were determined in consultation with the original speakers by reviewing the transcripts of the conversations and asking speakers what they would have said had they chosen the opposite language for the final word of each item. Frequencies per million for each word and translation equivalent were determined using the CELEX database for English (Baayen et al. 1995) and the SYN2010 portion of the Czech National Corpus for Czech (Hajič 2004, Jelinek 2008, Petkevič 2006, Spoustova et al. 2007). Frequencies of the original and translation-equivalent words were highly correlated: $r(725) = 0.92$, $p < 0.001$, providing evidence that the translation equivalents are reasonable. In order to compare frequency-based lexical accessibility of attested words to their translation equivalents in the linear model, the log-transformed relative frequency ratio r (Damerau 1993) was computed for each item.

(8) $$r = \log\left(\frac{\text{English relative frequency}}{\text{Czech relative frequency}}\right)$$

Thus, greater relative-frequency ratios reflect greater accessibility in English and predict occurrence of English (that is, switch) items.[4]

Accessibility was also operationalized as word length in syllables, with the expectation that speakers should prefer the language in which the relevant word is shorter and thus more easily produced. Syllable count was determined for English words again using CELEX, and for Czech simply by counting the number of orthographic vowels. A length difference score was computed by subtracting each item's Czech syllable count from its English syllable count. Here a smaller difference score predicts switching to English, since smaller difference scores imply longer, and thus less accessible, Czech words.

A final suite of accessibility metrics captures ease of code-switching generally and does not depend on direct between-language competition in the way that frequency and

[3] Of course, it is theoretically possible that the individual variables tested here each have their own effect on code-switching behavior, rather than truly reflecting broader phenomena such as accessibility.

[4] We also computed a difference score by subtracting each word's Czech frequency from its English translation equivalent's frequency, but using this as our frequency metric did not change any qualitative results of the logistic regression in §7.

length above do. These include IMAGEABILITY, CONCRETENESS, and PART OF SPEECH. More imageable and concrete words are argued to share more semantic features across languages in the bilingual lexicon, and thus to more readily lend themselves to code-switching (Marian 2009). Similarly, nouns are the most easily transferable part of speech between languages, followed by verbs and then other parts of speech. Part of speech was annotated manually for each switch and nonswitch item. Imageability and concreteness along a 100–700 scale for each item were determined by merging available norming databases: for imageability, Altarriba et al. 1999, Coltheart 1981, Friendly et al. 1982, Stadthagen-Gonzalez & Davis 2006; and for concreteness, Altarriba et al. 1999, Coltheart 1981, Friendly et al. 1982. Where multiple databases reported different values for an item, these were simply averaged.

LEXICAL CONTEXTUAL FACTORS. Two lexical contextual factors were taken into account. First, speakers may be more likely to code-switch if they have just produced a trigger word (proper noun, phonologically unintegrated loanword, or bilingual homophone; see §3.2). For each switch and nonswitch item, the presence of this kind of triggering was coded in a three-level factor capturing the various levels of the triggering hypothesis: none, for cases where there is no trigger word in the clause containing the potential code-switch; CLAUSE TRIGGER, for cases with a trigger present anywhere in this clause; and IMMEDIATE TRIGGER, for cases where a trigger occurs just prior to the potential switch. For example, 9 contains a trigger—the proper noun *Vista*, a city name—in the clause containing the potential code-switch, but the trigger does not immediately precede the potential switch site (*daleko* 'far'). Thus it contains a clause trigger.

(9) Vista už je daleko.
'Vista by now is far.'

The trigger in 10, the proper noun *Huckabee*, in contrast, is an immediate trigger, since it directly precedes the potential switch site *babka* 'lady'.

(10) A nebo mám jít za Huckabee babka?
'Or should I go to the Huckabee lady?'

All words falling into any of the three trigger-word categories described in §3.2 were manually coded as triggers (see Appendix A for a complete list). Immediate triggers are predicted to result in more switching than clause triggers, and clause triggers are predicted to result in more switching than no trigger.

The second lexical contextual factor was lexical cohesion. Speakers may converge on a particular language for certain referents, regardless of the embedding language of each mention of the referent. The factor LEXICAL COHESION encodes the most recently used language for the critical word in each switch and nonswitch item. For each potentially switched word, we determined whether the word or its translation equivalent (§6.1) had already occurred at some point in the current conversation; if so, we encoded the language of the word's most recent mention before the potential switch site (CZECH or ENGLISH), and if not, recorded NONE. Continuity is expected, so that an English most-recent mention predicts a word to be spoken in English (that is, be a code-switch) and a Czech most-recent mention predicts another Czech instance (that is, a nonswitch). This factor also helps control for symbolic cultural associations in which certain referents are overwhelmingly associated with a particular language within a speech community, as well as speaker-specific idiosyncratic preference for a given word to be realized in a particular language.

Syntactic contextual factors. The final class of control factors consists of collocational strength and dependency distance. The greater the collocational strength of a pair of words within a single language, the more likely those words are to be accessed as a unit, and the more likely they are to be produced in the same language (Backus 2003). For each potential code-switch, we compute the monolingual (Czech) collocational strength between (i) the word immediately preceding the potential switch and (ii) either the nonswitched Czech word, or the Czech translation equivalent of the switched English word. High values indicate strong Czech unitary status of the two words and predict that no switch to English will be made for the second word.

As our measure of collocational strength between word$_1$ and word$_2$, we employ a metric from associative learning theory, ΔP, defined as follows.[5]

(11) $\Delta P_{2|1} = P(w_i = word_2 | w_{i-1} = word_1) - P(w_i = word_2 | w_{i-1} \neq word_1)$

$\Delta P_{2|1}$ is the probability of an outcome ($word_2$) given that a cue ($word_1$) is present, minus the probability of that outcome given that the cue is absent. When these probabilities are the same, there is no covariation, and $\Delta P = 0$. As the presence of the cue increases the likelihood of the outcome, ΔP approaches 1, and as it decreases the likelihood, ΔP approaches −1. For the current study, ΔP was computed using relative frequency estimation from the SYN2010 portion of the Czech National Corpus (Hajič 2004, Jelinek 2008, Petkevič 2006, Spoustova et al. 2007). For each potentially code-switched word and the word immediately preceding it, we computed both rightward and leftward ΔP (how strongly a word predicts a collocate to the right or to the left, respectively) and entered two operationalizations of collocational strength into the logistic regression: (i) rightward ΔP, which captures a directional, sequential planning view of production, and (ii) the maximum of rightward and leftward ΔP, which treats bigrams as planning units and ignores directionality.[6]

The second syntactic control factor is dependency distance. Longer dependency distances between a word and its syntactic governor may increase the probability that the word is code-switched (Eppler 2011). A dependency distance factor therefore reflects the (hand-annotated) number of words from the final word of each potential switch IU to its syntactic governor (so that a word whose governor is adjacent to it would be coded as 1), following the coding principles described in Eppler 2011.[7] Further, because the dependency distance hypothesis is undefined for words that are their own syntactic governors, a binary variable syntactic governor captures whether the potentially switched word is the head word of the sentence and thus its own governor (and dependency distance was arbitrarily coded as 0 in these cases), allowing the logistic regression model to fit an arbitrary effect for head words of sentences that is separate from the effect of dependency distance.[8] These predictors complete the set of control factors included in the model.

[5] We thank Stefan Gries for suggesting this metric.

[6] Substituting pointwise mutual information for the ΔP measures did not change the qualitative results of our analysis.

[7] It was theoretically possible that a critical final word and its translation equivalent would correspond to different syntactic governors. However, probably due in large part to the fact that the translation equivalents were determined by presenting the original speakers with the original (and thus relatively constraining) speech strings leading up to the potential code-switches, this mismatch was never observed.

[8] An alternative method would be, in these cases, to set dependency distance to the mean of dependency distance in other cases. This would increase orthogonality to the variables but would not affect correlation (which is already 0), and it would come at the cost of transparency of model interpretation, because it would change the meaning of the intercept term. Fitting the model with this alternative parameterization did not change any qualitative results.

PREDICTOR	DESCRIPTION	DISTRIBUTION
Social distance/affiliation		
Participant constellation	Younger participants present?	yes: 79%, no: 21%
Baseline accessibility		
Frequency	Log English-to-Czech freq. ratio	*mean* = 0.68, *SD* = 1.8, *range* = [−6.22, 7.60]
Length	Syllables: English minus Czech	*mean* = −0.61, *SD* = 0.96, *range* = [−3, 4]
Imageability	Norming database ratings	*mean* = 451, *SD* = 127, *range* = [183, 668]
Concreteness	Norming database ratings	*mean* = 421, *SD* = 130, *range* = [143, 680]
Part of speech	Noun, verb, or other	noun: 42%, verb: 34%, other: 24%
Lexical context		
Trigger	Trigger word preceding?	immediately: 1%, in clause: 4%, none: 95%
Lexical cohesion	Word's previous mention	English: 15%, Czech: 18%, none: 67%
Syntactic context		
Rightward collocation	Rightward ΔP with prev. word	*mean* = 0.03, *SD* = 0.10, *range* = [−0.03, 0.99]
Maximum collocation	Left/right max ΔP with prev. word	*mean* = 0.03, *SD* = 0.11, *range* = [−0.02, 0.99]
Dependency distance	Distance in words to governor	*mean* = 1.51, *SD* = 0.81, *range* = [0, 6]
Syntactic governor	Word is its own governor	yes: 20%, no: 80%
Information content		
Meaning unpredictability	1 − (proportion of correct guessers)	*mean* = 0.64, *SD* = 0.36, *range* = [0, 1]

TABLE 3. Predictors in the logistic regression. For continuous variables, mean, standard deviation, and range are reported, and for categorical predictors, proportions of each level (prior to centering and standardizing) are reported. Imageability and Concreteness include values from the final iteration of imputation (see §7.1 and Appendix B).

6.2. ESTIMATING PREDICTABILITY OF MEANINGS IN NATURAL DISCOURSE. The variable of primary theoretical interest is the predictability of the meanings conveyed by potentially code-switched words. For reasons of practicality, predictability estimation for speaker-choice phenomena often makes use of *n*-gram models to calculate probabilities of events in context (Frank & Jaeger 2008, Jaeger 2006, Levy & Jaeger 2007). However, *n*-grams are not suited for our study, since they are unlikely to capture the discourse-level information structure that we hypothesize to influence speaker choice of language. Instead, we used a novel variant of the Shannon guessing game to estimate meaning predictability (Shannon 1951). In the original experiment, a participant was asked to guess entire passages of printed English letter by letter, and must have correctly guessed the current letter before moving on to the next letter, with the assumption that the more guesses required for a given letter, the more information carried by that letter. We build on recent adaptations of the Shannon game that have correlated unpredictability with linguistic variation at the word level: Manin (2006) asked participants to guess missing words in literary passages and found that unpredictability was positively correlated with word length, while in Tily & Piantadosi 2009, participants instead guessed upcoming referents, with the result that more-predictable referents tended to be encoded by pronouns rather than full noun phrases.

6.3. A SHANNON GAME FOR CONVERSATION. In order to estimate the predictability of the meanings of potentially code-switched words in the corpus, we adapted the Shannon game methodology for auditory discourse context. Participants listened to the ten conversational episodes comprising the Czech-English code-switching corpus and were asked to guess missing, IU-final words. Since the property of interest was the predictability of language-independent MEANINGS, participants could guess meanings using either language. The predictability of each item would then be estimated based on the rate of correct guesses.

Participants. A new set of eleven bilingual guessers, approximately demographically and sociolinguistically matched to the original speakers, was recruited at a Czech-American cultural event in the city where the original speakers reside. Like the original speakers, all guessers were native speakers of Czech born in Czechoslovakia, had begun learning English in early adulthood, and had been living in the United States for several years at the time of participation. This information is summarized in Table 4.

	N	MEAN AGE (*SD*)	MEAN AGE ENG ACQ (*SD*)
Speakers (of critical items; §5.2)	2	58 (3.5)	31 (3.5)
Guessers	11	45 (14.0)	30 (9.0)

Table 4. Demographics of participants in the guessing-game experiment.

Materials. For each of the ten conversational episodes, each critical IU-final word (see §5.2) was replaced by an auditory tone cueing participants to guess the missing word. We predetermined the set of correct responses for each item as the originally attested word and its translation equivalent in the other language (see §6.1).

Procedure. In a web experiment, participants listened to each conversational episode and were asked to submit guesses in either language for missing words. They could not move on until either they had correctly guessed the missing word or its translation equivalent, or they had submitted six incorrect guesses. Participants could replay the current item, as well as up to two items preceding it, as many times as they wished. After guessing an item correctly or submitting six guesses, they heard the complete IU with the original missing word now intact, followed by the next part of the discourse up to the next critical item. In this way, participants had access to the entire episode of natural discourse in making their guesses. This procedure is exemplified in Figure 1. In an exit survey, we asked participants what they thought was being investigated and then what kinds of words they found easiest and hardest to guess.

6.4. Predictability study results.

Exit survey. While no participant was explicitly aware of the exact experimental manipulation, several did spontaneously mention language encoding when asked about easiest and hardest words. For easy words, in addition to 'short words', 'simple words', 'repetitions', 'common expressions', and 'words with previously understood meanings', two participants mentioned 'Czech words' and no participant mentioned 'English words'. For difficult words, conversely, participants offered 'long words', 'new ideas', and 'slang', and the two participants who mentioned 'Czech words' as easy to guess mentioned 'English words' as difficult to guess. No participant offered 'Czech words' as a response for difficult-to-guess items.

Predictability of critical items. Turning to the quantitative results, guessing was completed for forty-nine individual conversations (reflecting all ten unique conversations in the corpus), totaling 3,458 sets of guesses, where a set is defined as all of the guesses given by a single participant for a single IU. For each IU, the proportion of participants who had correctly guessed the word within six attempts was computed. Consistent with intuitions expressed in the exit survey, switch items (those that had been spoken in English) were more difficult to guess than nonswitch items (those that had been spoken in Czech). On average, the meanings of switch items were correctly

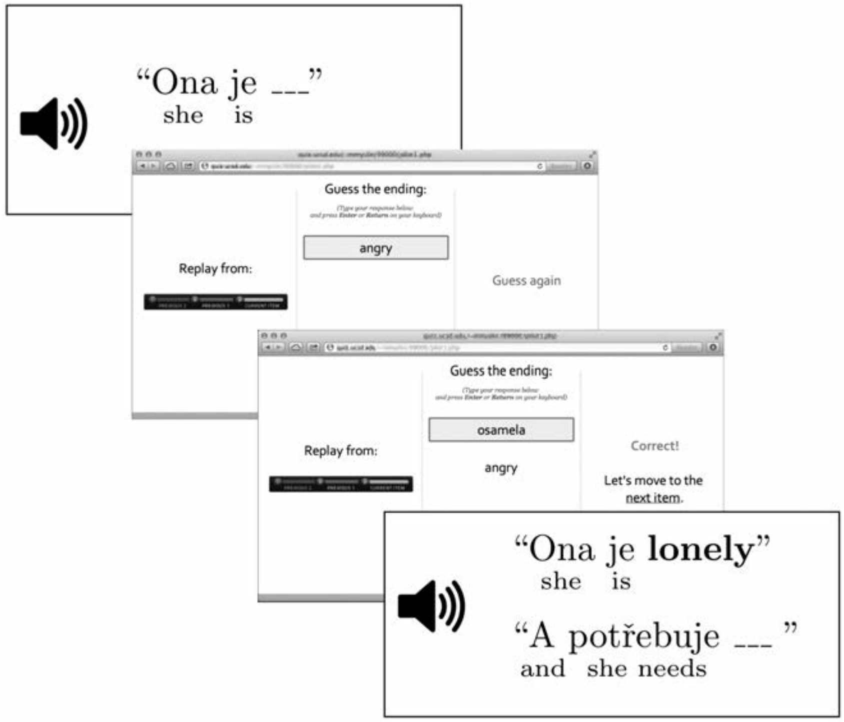

FIGURE 1. Example of guessing-game procedure: (i) audio prompt with final word removed, (ii) incorrect guess, (iii) correct guess, and (iv) repetition of audio prompt with missing word now intact and continuation to next missing word.

guessed by 25.4% of participants (*SD* = 34.4%), while the meanings of nonswitch items were correctly guessed by 41.9% of participants (*SD* = 35.8%). These results are broken down into cumulative probability by guess number in Figure 2.

For inclusion as a factor in the binary logistic regression, we defined the UNPREDICTABILITY U of the meaning m of each (potentially) code-switched word as the difference between 1 and the proportion of participants who had correctly guessed the word within six attempts, among those who had provided guesses for the IU.[9]

(12) $U(m) = 1 - P(\text{guessed})$

Thus, items that were correctly guessed by most or all participants have U near or at 0, indicating low information content, and items that were correctly guessed by very few or no participants have U near or at 1, indicating high information content. Observed values for U ranged from 0 to 1, with mean 0.64 (*SD* = 0.36); in other words, some critical meanings were correctly guessed by all participants and some were not correctly guessed by any participants, and the average item was correctly guessed by just over half of the participants.

CODE-SWITCH EXPECTATION. The Shannon game data allow for the investigation of one additional quantity of interest. Accurate comprehender expectations of upcoming code-switches may mitigate so-called processing 'switch costs', which increases the plausibility of the hypothesis that code-switches are not unequivocally burdensome to

[9] Following Manin (2006), we selected this metric since there is no straightforward way to compute surprisal (§3.4) for items for which no correct guess was ever submitted.

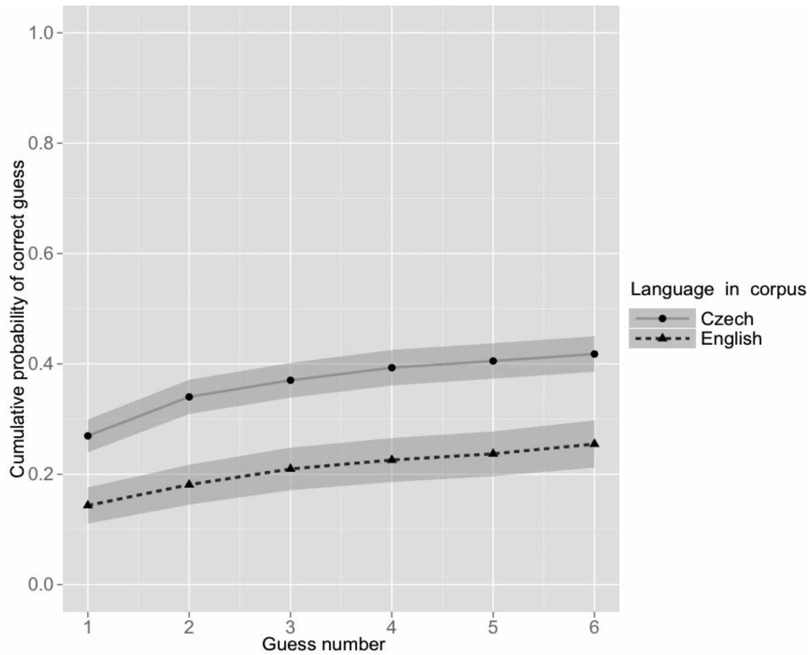

FIGURE 2. Cumulative probability of correct guess (as proportion of participants correctly guessing item) by guess number, conditionalized on original language of mention of each item.

process and may indeed serve as a useful comprehension cue (§4). The language in which a guesser expects the next word to occur may be inferred through the language of her first guess: if she expects an English word, she may be more likely to submit her first guess in English. Consequently, if comprehenders are correctly anticipating language choice, IUs that indeed ended with a switch to English should have higher proportions (across participants) of first guesses submitted in English. The percentage of English first guesses for items that were spoken in English was 38.4 (SD = 25.7), and the proportion of English first guesses for items that were spoken in Czech was 23.6 (SD = 23.2), a significant difference according to a t-test: $t(473) = 7.7$, $p < 0.0001$. That is, although there was an overall baseline trend for guesses to be given in Czech (71%), there was a reliable pattern above and beyond this baseline for guesses to be given in the language in which the item was spoken.

7. MULTIFACTORIAL RESULTS. We tested effects of predictability of meaning on language choice through a procedure similar to that used in other research on predictability effects on speaker choice (Jaeger 2006, 2010), first developing a parsimonious logistic regression model (Agresti 2002) of the effects of control factors and then assessing the predictive value of meaning predictability on code-switching behavior above and beyond the effects of the control predictors. We describe our modeling procedure in more detail (§7.1) and then report results of the control factors (§7.2). We test meaning-predictability effects separately for each individual speaker (§7.3) and then investigate the strength of evidence for generalizability of these effects beyond the speakers studied here (§7.4).

7.1. MODELING PROCEDURE. We first provide a brief overview of our modeling procedure; full details are given in Appendix B. All predictors were centered and standardized so that categorical variables had a mean of 0 and a difference of 1 between levels, and continuous variables had a mean of 0 and a standard deviation of 0.5. Because values for two of our control factors, imageability and concreteness, were not available for roughly 20% of cases, we estimated these values on the basis of the other factors in the data set using multiple imputation (Harrell 2001).[10]

We first developed a parsimonious model of our control factors against which to subsequently evaluate the effect of meaning predictability. To develop this model, we used a genetic algorithm (Calcagno & de Mazancourt 2010) to search efficiently through the space of possible models including up to two-factor interactions, optimizing for the Bayesian information criterion (BIC). Since both the genetic algorithm and multiple imputation are stochastic processes, we repeated the entire modeling routine ten times, and in the sections below report results from the final iteration. We observed no qualitative differences over these ten runs in results relating to meaning predictability.

This model-selection process allowed us to explore a large space of possible interaction terms among the base predictors, checking for any effects that could explain away the meaning-predictability effect in our data. As Harrell (2001) and others have described, however, model selection can have negative consequences for type I error and for interpretation of the coefficients associated with predictors operated upon by model selection. While this concern does not apply to meaning predictability, since it was not included in the model-selection process, it could warrant caution in interpreting the results of the control predictors. However, in this case we are reasonably confident in the general pattern of control factor results: a model including all base control predictors, plus the significant interactions identified by the model-selection process, resulted in virtually the same set of significant control factors as did model selection. Results of this model are reported in Appendix C.

In §§7.2 and 7.3, we investigate meaning-predictability effects separately for each speaker by fitting a model with control factors selected by the genetic algorithm, and an individual meaning-predictability parameter for each speaker (but no random effects). A result summary for this model (reflecting the last of ten iterations of the entire routine) is reported in Table 5. No signs of substantial collinearity were present in the final model; all correlations between fixed effects were very low (all $|r| < 0.25$). We address the issue of generalizability across speakers in §7.4.

7.2. Control factor results. The results of each of the control factors are discussed in turn below. The response variable was coded as 0 for Czech/nonswitch, and 1 for English/switch.

Participant constellation. Consistent with sociolinguistic accounts of code-switching as a tool to modulate social affiliation and accommodate interlocutor preferences, the older speakers code-switched to English less often when the younger, and less Czech-dominant, speakers were not present ($\beta = -1.06$, $z = -4.0$, $p < 0.0001$ in the final model).

Frequency. The relative-frequency ratio between Czech words and their English equivalents was not selected by the genetic algorithm for inclusion in the final model; thus there is no evidence that speakers choose the language in which a word is more frequent. This result provides a first empirical test of this hypothesis (Heredia & Altarriba

[10] To ensure that none of our modeling results depend crucially on our decision to impute missing values for imageability and concreteness, we also fit a version of the model discussed in this section omitting these two factors completely; no qualitative results relating to the remaining factors changed.

2001). Interestingly, it is consistent with studies of the effect of frequency on optional *that*-mention, where the effect is weak or absent (Ferreira & Dell 2000, Jaeger 2010).

	PARAMETER ESTIMATES		WALD'S TEST		LIKELIHOOD RATIO TEST	
PREDICTOR	Coef. β	*SE*(β)	*Z*	p_z	χ^2	*p*
Participant constellation =older.speakers.only	−1.06	0.27	−4.0	< 0.0001	16.9	< 0.0001
Length	−0.76	0.21	−3.6	< 0.01	13.5	< 0.01
Concreteness	0.90	0.26	3.5	< 0.001	6.0	< 0.05
Part of speech=verb	−2.30	0.34	−6.8	< 0.0001	109.9	< 0.0001
=noun	0.53	0.26	2.0	< 0.05		
Lexical cohesion=prev.English	0.92	0.28	3.3	< 0.0001	28.9	< 0.0001
=prev.Czech	−1.12	0.28	−3.2	< 0.01		
Length : Syntactic governor	2.23	0.82	2.7	< 0.01	7.7	< 0.01
Participant constellation =older.speakers.only : Concreteness	−1.47	0.47	−3.1	< 0.01	9.5	< 0.01
Unpredictability : Speaker =speaker.1	0.61	0.33	1.9	0.06	45.7	< 0.001
=speaker.2	1.76	0.30	5.8	< 0.0001		
Speaker=speaker.1	0.38	0.21	1.8	0.07	12.3	< 0.05

TABLE 5. Result summary for the final model (for the last of ten iterations of the multiple imputation process): coefficient estimates β, standard errors *SE*(β), Wald's *z*-score and its significance level, contribution to likelihood χ^2 and its significance level. The response variable was coded as Czech/nonswitch = 0 and English/switch = 1. Predictors were centered and standardized so that numeric variables had a mean of 0 and *SD* of 0.5, and categorical variables had a mean of 0 and a difference of 1 between levels. Baselines are Participant constellation=all.participants, Part of speech=other, Lexical cohesion=no.prev.mention, and Speaker=speaker.2.

WORD LENGTH. The difference in number of syllables between the English and Czech equivalents of a word was a reliable predictor of language choice, with speakers generally opting for the shorter alternative ($\beta = -0.76$, $z = -3.6$, $p < 0.01$). This is consistent with an account in which shorter words are more accessible (D'Amico et al. 2001) and thus more likely to be selected for production.

IMAGEABILITY. Imageability was not selected for inclusion by the genetic algorithm. This result reflects a first empirical test of another largely untested hypothesis—in this case, that the role of imageability in TRANSFER of structures between languages is equivalently relevant in CODE-SWITCHING (Marian 2009).

CONCRETENESS. Consistent with the hypothesis (Marian 2009) that concrete words' semantic representations are more tightly linked across languages, leading to easier switching, more concrete words were more likely to be code-switched ($\beta = 0.90$, $z = 3.5$, $p < 0.001$).

PART OF SPEECH. We replicate a robust finding: nouns are the most likely words to be switched (e.g. Marian 2009, Muysken 2000, Myers-Scotton 1993a). With words that are neither nouns nor verbs as the baseline level of Part of speech, nouns are more likely to be code-switched ($\beta = 0.53$, $z = 2.0$, $p < 0.05$), and verbs less ($\beta = -2.30$, $z = -6.8$, $p < 0.0001$).

LEXICAL COHESION. Consistent with findings that referents continue to occur in the same language throughout a discourse episode (Angermeyer 2002, Munoa 1997), prior mention of a referent in English strongly predicted subsequent mention in English ($\beta = 0.92$, $z = 3.3$, $p < 0.0001$), while a prior Czech mention predicted subsequent Czech mention ($\beta = -1.12$, $z = 2.7$, $p < 0.01$), relative to a baseline where there is no prior mention of the referent.

Triggering. Triggers (proper nouns, phonologically unintegrated loanwords, and bilingual homophones) are claimed to be stored in shared representations across language systems, which increase the activation of the second language. Consequently, words immediately following a trigger, or following a trigger within a single clause, were predicted to be code-switch sites (Broersma 2009, Broersma & de Bot 2006, Clyne 1991). However, this factor did not reach significance. One potential explanation is the low variability in the data for this factor, making statistical power difficult to achieve (see Table 3 for summary statistics).

Dependency distance. Dependency distance to a word's syntactic governor was not a significant factor in language choice. The hypothesis has elsewhere been tested on only a single German-English data set so far (Eppler 2011), and its nonsignificance here may reflect its specificity to a particular speech community or language pair. As in the case of triggering, however, this could also be the result of the low variability in the data for this factor.

Collocational strength. Finally, collocational strength was also not a significant predictor: code-switches are no less likely given strong collocational association with the preceding word, with either rightward or maximum ΔP. This result thus reflects the first quantitative test of the hypothesis. Once more, however, variability in the data is low for this factor.

Length : Syntactic governor. Two unpredicted interactions emerged from model selection by the genetic algorithm. First, Syntactic governor interacted with Length: when the potentially switched word was its own syntactic governor, the tendency to choose the language with the shorter variant was weaker ($\beta = 2.23$, $z = 2.7$, $p < 0.01$). We investigated this interaction by reparameterizing the model with separate length parameters for self-governors and non-self-governors. For non-self-governors, the tendency to choose the language with the shorter word was significant ($\beta = -0.76$, $z = -3.60$, $p < 0.001$), whereas for self-governors, the trend was in the opposite direction and was only marginally significant ($\beta = 1.50$, $z = 1.86$, $p = 0.06$). This result was not predicted, but it is consistent with uniform information density, which holds that speakers encode less predictable material with longer forms (§3.4). If we assume that governors contain more information than nongovernors, it is reasonable not to prefer short encodings (assuming also that this information is independent of what is captured by our unpredictability metric). Note also that we neither predicted nor observed a main effect of Syntactic governor: language choice was not directly affected by whether a word was its own governor.

Participants : Concreteness. In the second unpredicted interaction, speakers are especially likely to code-switch concrete words when the younger speakers are present ($\beta = -1.47$, $z = -3.1$, $p < 0.01$). The direction of this interaction is not surprising: the presence of English speakers magnifies the older Czech speakers' existing tendencies to switch to English.

In summary, well-established effects were replicated in the current statistical model (Part of speech and Lexical cohesion), as well as Participant constellation, Word length, and Concreteness, despite their simultaneous inclusion for the first time in a multifactorial analysis of naturalistic data. Those factors that were not statistically significant were either previously untested in code-switched discourse (Frequency, Imageability) or statistically less well supported in the code-switching literature and low in variability in the current data set, limiting statistical power (Triggering, Dependency distance, Collocational strength). Given this general validation of previous code-switching research as well as the current data and model, we now turn to the variable of primary theoretical interest: UNPREDICTABILITY of the meanings of potentially code-switched words.

7.3. MEANING-PREDICTABILITY EFFECTS. As predicted, for each speaker Unpredictability emerged as a significant factor in language choice: greater unpredictability of meaning was associated with increased probability of code-switching (speaker 1: $\beta = 0.61$, $z = 1.9$, $p = 0.06$; speaker 2: $\beta = 1.76$, $z = 5.8$, $p < 0.0001$). This tendency is stronger for one speaker than the other; a likelihood ratio test justifies the current model's speaker-specific Unpredictability parameters rather than a single Unpredictability parameter plus a Speaker parameter ($\chi^2_{\Delta(\Lambda)}(1) = 5.1$, $p < 0.05$). Relative to a model with no Unpredictability factor, both a model with a single Unpredictability factor and a model with a Speaker parameter and speaker-specific Unpredictability parameters are more explanatory ($\chi^2_{\Delta(\Lambda)}(1) = 37.2$, $p < 0.0001$; $\chi^2_{\Delta(\Lambda)}(3) = 45.7$, $p < 0.0001$, respectively); in either case, Unpredictability makes the second-largest contribution to the model's overall likelihood, surpassed only by Part of speech ($\chi^2_{\Delta(\Lambda)}(1) = 148.8$, $p < 0.0001$). Thus, not only does each speaker in this data set tend to code-switch at points of high meaning unpredictability, but, on the basis of model likelihood, this is actually one of the most highly explanatory predictors of switching behavior.

7.4. SPEAKER-SPECIFIC EFFECTS AND GENERALIZABILITY. Section 7.3 shows that our data set contains evidence for our hypothesized effect of meaning predictability on code-switching in each individual speaker—this evidence was highly significant for speaker 2 and, at $p = 0.06$, marginally significant for speaker 1. To the extent that our central research question is viewed as whether THESE PARTICULAR SPEAKERS in THIS PARTICULAR SPEECH COMMUNITY show evidence for this effect—an interpretation that would be natural in, for example, some research traditions in sociolinguistics—our result is relatively strong. However, an alternative interpretation of our research question is equally natural. Suppose that we view these speakers as a random sample of the overall population of all Czech-English bilingual code-switchers and assume that individuals in this larger overall population vary idiosyncratically in the relationship between meaning predictability and code-switching behavior. Under these assumptions, what is the strength of evidence in our data set that the AVERAGE EFFECT of meaning predictability on code-switching is in the direction we hypothesized—that is, how strong is our evidence that the effect we observe in our two speakers generalizes to the wider population of Czech-English bilingual code-switchers?

Clark (1973) and Barr and colleagues (2013) have argued that this type of question needs to be addressed by a statistical test in which idiosyncratic variability in sensitivity to the theoretically critical predictor, here meaning predictability, must be included in the null hypothesis. Such a test can be carried out by using a mixed-effects logistic regression model (Baayen et al. 2008, Jaeger 2008) with a random by-speaker slope for the effect of meaning predictability on code-switching behavior.[11] Following Barr and

[11] Another potential source of idiosyncratic variability is at the ITEM level: in our case, certain meanings may have different code-switching behaviors. Jaeger (2006, 2010) addressed by-item variability in corpus studies using random by-item slopes. In our case, such models on our full data set failed to converge, but a model that was fit only to the subset of our data containing items (meanings) that occurred exactly once (i.e. a data set in which item variability is not a concern since observations are independent from each other at the item level) yielded the same qualitative results as models fit to the full data set. This partial data set included 488 of the 725 total critical utterances in the full data set, and in the resulting model, the significant control

colleagues (2013), we use a likelihood ratio test comparing models differing only in the presence versus absence of a FIXED effect of meaning predictability; both the null- and alternative-hypothesis models contain by-speaker random intercepts and random slopes for meaning predictability (jointly normally distributed with unconstrained covariance matrix) and all control factors used in the single-level logistic regression model reported in §§7.2 and 7.3.[12] In the two models, results of control factors were virtually identical, but the magnitude of random by-speaker effects were larger in the null-hypothesis model (Table 6). The likelihood ratio test found the alternative-hypothesis model (fixed effect AND by-speaker slopes) to be significantly more explanatory than the null-hypothesis (by-speaker slopes only) model in seven of the ten iterations of our overall modeling routine (see §7.1; $\chi^2_{\Delta(\Lambda)}(1) = 4.54$, $p = 0.03$ in the final iteration, and $p \leq 0.07$ in all ten iterations).[13] These results suggest that the effect of meaning predictability on code-switching generalizes beyond the current speakers; we return to this issue in the general discussion.

MODEL	RANDOM EFFECT	VARIANCE	*SD*	CORRELATION
With fixed effect of unpredictability	(intercept)	0.30	0.55	
	Unpredictability	0.53	0.73	−0.61
Without fixed effect of unpredictability	(intercept)	0.79	0.89	
	Unpredictability	0.73	0.86	−0.79

TABLE 6. Random speaker-effects results for two versions of the model in the final iteration of the modeling routine: one with a fixed effect of unpredictability, and one without (see §7.4).

7.5. SUMMARY OF MULTIFACTORIAL RESULTS. A wide variety of monofactorial explanations of code-switching behavior were operationalized and included in the logistic regression model, and previously reported results were largely replicated for the first time in a multifactorial analysis. Even taking these control factors into account, unpredictability of meaning emerged as a significant predictor of code-switching and was indeed the second most explanatory variable in the model, following part of speech. This correlation was reliable within individual speakers, and there is also evidence that it generalizes to other speakers. A separate analysis revealed that comprehenders correctly anticipate language choice in code-switched discourse.

8. GENERAL DISCUSSION. Our primary objective was to test the hypothesis that multilingual speakers code-switch words that carry a high amount of information in discourse, based on the predictability of these words' meanings. On the basis of a corpus of spontaneous Czech-English conversation, this pattern was indeed reliably observed and in fact emerged as a key explanatory factor in code-switching behavior. This is consistent with the claim that code-switches to a speaker's less frequent, and hence more

---

factors as well as the fixed effect of meaning unpredictability remained significant according to likelihood ratio tests ($\chi^2_{\Delta(\Lambda)}(1) = 4.89$, $p = 0.03$ for meaning unpredictability).

An alternative method pursued by Jaeger (2006, 2010) presents bootstrapping with random replacement of speaker clusters to adjust for anti-conservativity with regard to speaker intercepts and slopes for all predictors.

[12] In our case it is important not to use the Wald *z* statistic often used to assess statistical significance in generalized mixed-effects models. The *z* statistic is computed conditional on a point estimate of the random-effects covariance matrix, without taking into account the uncertainty in the true value of this matrix (Baayen et al. 2008:396). Because our data are categorical and we have a small number of speakers, this uncertainty is considerable and would lead to anti-conservative inference; the likelihood ratio test is not susceptible in the same way.

[13] Adding both the fixed-effect and by-speaker slopes SIMULTANEOUSLY also resulted in a significantly more explanatory model than one with no effects of unpredictability and by-speaker random intercepts: $\chi^2_{\Delta(\Lambda)}(1) = 39.22$, $p < 0.0001$ in the final iteration, and $p < 0.0001$ in all ten iterations.

salient, language offer a distinct encoding that serves to highlight meanings of low predictability in discourse.

The article had three subsidiary objectives. The first was to relate this account to discourse-functional accounts of code-switching and to other speaker-choice phenomena predicated on information and predictability. The second goal was to investigate for the first time the relationships between a cross-disciplinarily motivated set of hypothesized factors in language choice. The third objective was to bridge a methodological gap in code-switching research by analyzing spontaneous natural data with rigorous statistical modeling. We discuss each of these objectives in turn.

8.1. Meaning predictability and language choice. On the discourse-functional accounts of code-switching described in §3.3, language choice serves to highlight important information in conversation (de Rooij 2000, Gumperz 1982, Karrebaek 2003, Romaine 1989). We showed that language choice is indeed correlated with one formal operationalization of importance or information content, namely the predictability of meaning in context. This underscores the status of code-switching as a speaker choice, since not only is it essentially independent of truth-conditional meaning in the cases we consider (§2), but its correlation with predictability of meanings also is similar to that of other speaker choices such as, for example, optional complementizer mentioning or referring expression type (§3.4).

In this sense, the code-switching patterns described here add to a long-observed correlation between marked forms and marked meanings. In the current case of code-switching, the markedness of the form comes not from its complexity, but from its frequency: less expected meanings are conveyed in the less frequently used language. The pattern is analogous to other cases in which an equally complex but less frequent form is selected for a marked meaning, such as word-order freezing in languages with free word order (e.g. Lee 2003, Tomlin 1986) or topicalization in English-like languages (e.g. Chafe 1976, Halliday 1967, Prince 1984).

Why should marked forms correlate with marked meanings? More specifically, why would the choice to code-switch be sensitive to meaning predictability? Our explanation is in line with audience-design accounts of production, in which speakers take their interlocutors' knowledge state into consideration and make choices that minimize risk of miscommunication or processing burden. Switching to the less frequent, and therefore more salient, language encoding functions to highlight new or important information that comprehenders must attend to especially carefully. In other words, the change in linguistic form is a comprehension cue alerting comprehenders to allocate more attentional resources to the current word since its information content is high, a strategy that may facilitate processing or reduce risk of miscommunication. The formal properties of the switch itself may not be especially difficult to process in real-world circumstances: studies of code-switching using auditory and discourse contexts find few or no switch costs (§4), and we find that comprehender expectations of switch sites are accurate, easing processing (§§4 and 6.4).

One critique of audience-design accounts, however, is that their predictions are sometimes indistinguishable from speaker-centric accessibility accounts. For example, Fukumura and van Gompel (2012) argue that speakers do not use their addressee's discourse model when choosing between producing a pronoun or a definite noun phrase, and instead choose on the basis of the referent's accessibility in their own memory or discourse model (which is often nevertheless highly correlated with their interlocutor's model). The case of code-switching, however, bears on this debate without being straightforwardly subject to this critique. Unlike with phenomena such as referring-expression choice, *that*-mentioning, and others in which speakers ultimately REDUCE THEIR OWN EFFORT (§3.4), switching languages may not, all things being equal, be LESS COSTLY for a speaker than staying in the same language (although see e.g. Gollan et al. 2014 for factors that mitigate switch costs in production). In other words, on average, it may not be efficient from a speaker-centric perspective to switch languages for unpredictable meanings. But further work is needed to explicitly investigate potential interactions of audience design and speaker-internal factors in code-switching.

Of course, for speakers' tendency to code-switch for unpredictable meanings to benefit listeners, listeners must know that this tendency exists. This knowledge could be acquired in several ways. First, a comprehender could rapidly learn the correlation for the current speaker, even if they are members of different speech communities. Consistent with this hypothesis, although a majority of participants in the guessing game had not previously interacted with the speakers, in the exit survey several of them did spontaneously suggest that code-switched English words were the hardest to guess, and no guesser suggested that nonswitched Czech words were the hardest to guess. Second, this knowledge could be tacitly shared among members of a particular speech community. A final possibility is that the correlation is general across all code-switching behavior, and comprehenders are unlikely to encounter speakers who do not exhibit it. In our study, we provide evidence that the correlation between language choice and predictability is consistent within individual speakers, lending plausibility to at least the first scenario above. We also provide some evidence that the tendency may be a general property of a larger population from which our speakers were drawn, suggesting that the correlation may indeed extend within or beyond speech communities (a hypothesis also supported by the wide range of languages for which similar phenomena have been reported; §3.3); more work is necessary to investigate this possibility.

Other questions of generalizability concern the particular mapping of languages to information content. We have suggested that it is the less frequent language that encodes less predictable meanings, consistent with a related argument by de Rooij (2000). In principle, however, it is possible that other factors underlie the particular mapping of language to meaning predictability, or that no specific mapping is necessary, if it is simply distinctiveness of encoding that is relevant to comprehension (as suggested by Karrebaek (2003)). However, this distinctiveness may hold only when there is a substantial asymmetry between the frequency of use of languages; it is an empirical question whether code-switching can become 'too unmarked' to function in the way argued here.[14] This motivates future work on prevalence and exchangeability of languages in the correlation with meaning predictability.

8.2. MULTIPLE FACTORS IN CODE-SWITCHING. The second objective of this article was to test the meaning-predictability account of code-switching against control factors from multiple disciplines, as well as to investigate the interrelationships among these factors. This approach allows for an expanded view of why multilinguals code-switch: since previous studies did not involve multifactorial statistical analysis, little could be said about relative effect sizes, potential interactions, and epiphenominality. Reassuringly, most previously proposed factors in code-switching were replicated in the current study, and indeed, one of the most widely cited constraints on code-switching, part of speech, made the largest contribution to the model's likelihood.

The multifactorial approach does, however, reveal an interesting split between discourse-related and speaker-internal motivations for code-switching. Factors explicitly

[14] We thank a referee for pointing this possibility out to us.

focusing on internal lexical accessibility, such as frequency, imageability, concreteness, and triggering, were in general less explanatory than discourse-related factors such as lexical cohesion, participant constellation, and meaning predictability. Since these most explanatory factors are inherently tied to conversational context, this result underscores the importance of attention to context-specific production circumstances and thus the utility of rich discourse data in understanding code-switching in its natural habitat.

This result also highlights a more far-reaching theoretical consequence of current methodological practice in psycholinguistics. Evidence for the effects of inherent properties of linguistic forms (frequency, animacy, and so on) on language production primarily comes from decontextualized tasks such as isolated word production. However, in naturally occurring, contextualized speech, these factors have a markedly weaker, if at all significant, effect (Jaeger 2006, 2010). As we discuss in the next section, therefore, a shift toward more sophisticated methodologies is needed in order to understand the critical influence of linguistic context on speakers' preferences during language production.

8.3. Rigorous approaches to natural discourse. The third objective of this article was to combine the ecological validity of naturalistic discourse data with a rigorous quantitative methodology. This is especially important given that code-switching behavior differs in crucial ways between laboratory paradigms and spoken discourse, the natural locus of code-switching (§4). We believe the approaches developed in the article are useful in two ways. First, the specific problem of estimating meaning predictability in natural conversation is not straightforward, and it is our hope that the auditory conversational Shannon game developed here will inspire other rigorous yet ecologically valid approaches. Second and more generally, we hope the combination of methodologies applied here advances the use of converging evidence in psycholinguistic research; Gries and colleagues (2005) and Wulff and colleagues (2009), among others, argue for the value of this practice in correctly interpreting evidence from any individual methodology.

9. Conclusions. This study investigated a central question in code-switching research: what motivates speakers to switch languages when truth-conditional meaning is equivalent in both languages? Based on a corpus of spontaneous Czech-English conversation, we tested accounts from sociolinguistic, discourse-functional, and psycholinguistic research traditions simultaneously using mixed-effects logistic regression. Most of these previously reported monofactorial effects were replicated in the model, although support was in general stronger for factors of code-switching related to higher-order discourse contexts than speaker-internal production circumstances. The key exception was the part of speech of potentially switched words, a widely cited constraint on code-switching that made the largest contribution to our model's likelihood.

Novel evidence was provided in support of the predictability of a word's meaning as a determinant of code-switching, such that speakers code-switch at points of high information content. Meaning predictability, as estimated through a Shannon guessing game developed for our Czech-English corpus, made the second largest contribution to the model's likelihood. These results are consistent with an audience-design view of code-switching: switches to a speaker's less frequent language offer a distinct and potentially more salient encoding for meanings of low predictability; thus one reason speakers alternate between meaning-equivalent forms may be to highlight certain material as particularly important or informative. In this way, along similar lines as other speaker-choice phenomena such as prosody and word order, code-switching may be understood as a formal linguistic marker of information content in discourse.

Appendix A: Trigger words

Trigger words (§§3.2 and 6.1) were coded manually by the first author. All proper nouns were coded as trigger words; these (anonymized) were *Nick*, *Huckabee*, *Vista*, *Simba*, *Van Nostrand*, *Vincent*, *Jack*, *Michael*, *Chip*, *Valley West*, and *Nováková*. Phonologically unintegrated loanwords were also coded as triggers and included *trolleycar* and *bar*. Finally, all tokens of the single bilingual homophone in the corpus, the discourse marker *no* 'well', were also coded as triggers.

Appendix B: Details on modeling procedures

In this appendix, we provide details on the logistic regression models presented in §7. We first discuss our missing-value imputation procedures, then turn to our model-selection routine, and finally discuss how we tested the critical meaning-predictability factor against these factors and investigated its generalizability beyond the speakers included in this study. The complete process is summarized in Figure A1. Note that statistics from the final iteration of these procedures are presented in §7.

B1. Multiple imputation. Because imageability and concreteness ratings were not available for 155 and 192 items, respectively, of the 725 items in our data set, we estimate these values from the other predictors and the response variable using multiple imputation (Harrell 2001). We use the aregImpute() function in the R package Hmisc for multiple imputation (Harrell & Dupont 2012, R Development Core Team 2012). This function imputes missing values of predictors $X_1 \dots X_m$ by first estimating via multiple regression the conditional distribution of a given predictor X_i given the remaining predictors $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m, Y$ (where Y is the dependent variable), and then drawing values of X_i observed elsewhere in the data set that have high conditional probability based on the regression model. The function cycles through the set of predictors so that each predictor is successively used as the response variable in the regression-based imputation process, akin to the Markov chain Monte Carlo method of Gibbs sampling. We use the aregImpute() default of three 'burn-in' cycles through our predictor set, and then cycle through our predictor set j more times, each cycle producing a set of imputed values for missing observations in our data. We set $j = 10$, resulting in ten 'complete' versions of our data set $D_1 \dots D_{10}$, each with potentially different values imputed for missing imageability and concreteness ratings.

There are two 'top-level' settings that govern the behavior of aregImpute(). The nk setting determines the number of knots used for the spline basis for continuous predictors in the imputation process, with nk = 0 forcing linearity and nk = 3 allowing restricted cubic splines. The match setting determines how imputed values of X_i are selected from the conditional distribution given $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m, Y$: weighted draws from previously observed values with multinomial probabilities monotonically descending in distance from the predicted mean of the regression, whereas closest deterministically selects the previously observed value closest to the predicted mean. To determine which combination of settings to use, we tried the four logically possible combinations of values for these two arguments and found that nk = 3 and match = "weighted" yielded the lowest squared error in prediction of nonmissing values within our data set. We therefore used these settings to impute missing values as described in the preceding paragraph.

B2. Selecting control factors. We now turn to selecting significant control factors against which to test unpredictability of meaning. Since multiple imputation as implemented in aregImpute() (see §B1 above) is a stochastic process, it gives us multiple (in our case, ten) 'complete' versions of our data set, each with a unique set of estimates of missing values. This means that for each data set, a different linear model may be selected as best by a model-selection routine, and there is currently no general consensus on variable selection with multiply imputed data (for discussion, see Wood et al. 2008). We therefore follow the heuristic developed in this section.

For each of the data sets $D_1 \dots D_{10}$, we used a model-selection routine to determine the best set of control factors. We used a genetic algorithm (R package glmulti; Calcagno & de Mazancourt 2010) to search efficiently through the space of possible logit models that include some subset of base control predictors (including Speaker) and their pair-wise interactions, optimizing for BIC.[15] For each of the ten data sets, the algorithm was run three times, and of the three resulting 'best' models, the one with the lowest BIC was selected as the final model for that data set. The most frequent of these models, selected for six of the ten data sets, was as follows.

(A1) response ~ Lexical cohesion + Participants + Part of speech + Length + Concreteness + Length : Syntactic governor + Participants : Concreteness

[15] The entire routine described in this appendix was also performed with the Akaike information criterion, with no qualitative change in results for the relationship between meaning predictability and code-switching.

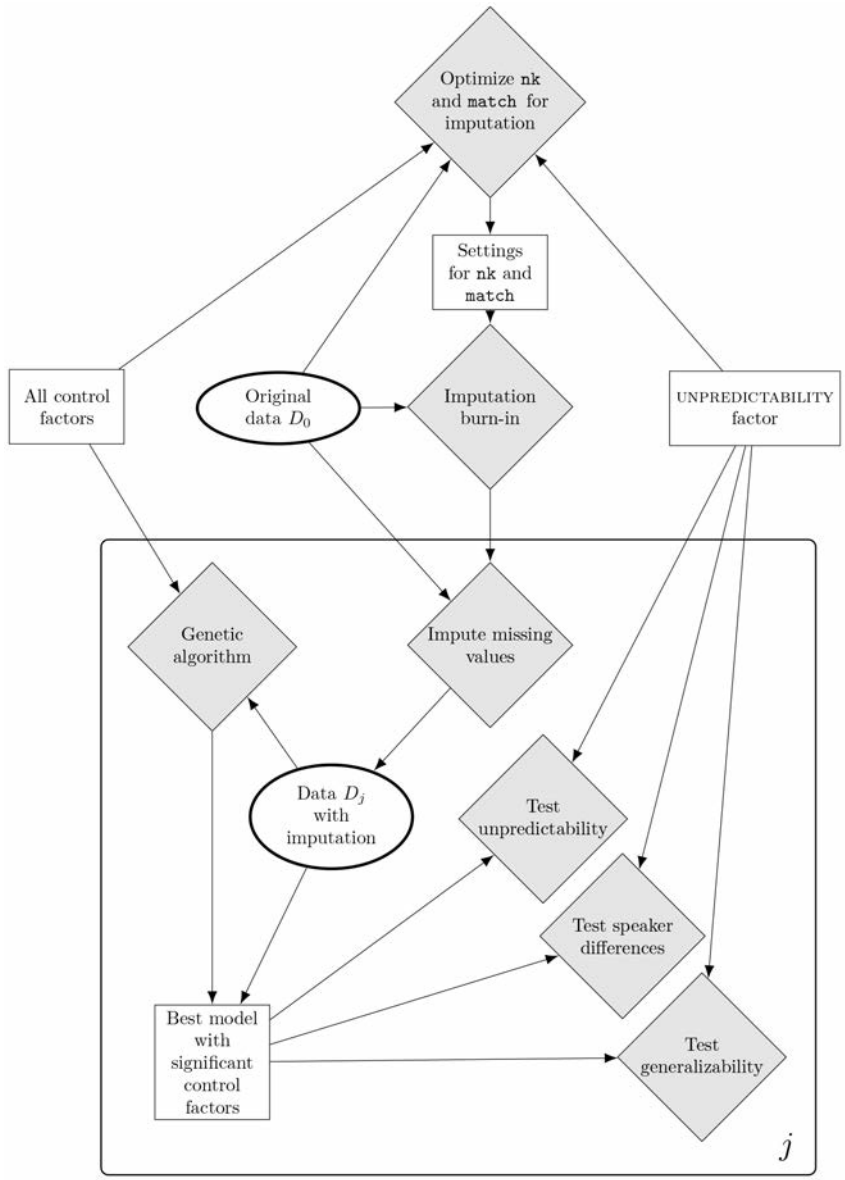

FIGURE A1. Modeling workflow. Procedures within the *j* plate are repeated ten times.

An identical model, but without Concreteness and the two interactions, was selected for the remaining four data sets.

B3. TESTING MEANING-PREDICTABILITY EFFECTS. With the set of control factors now finalized for each data set, we proceeded to test effects of meaning predictability by comparing the best control model to one including the Unpredictability factor using a likelihood ratio test. For all ten data sets, this factor significantly improved explanatory power.

We also investigated individual speaker effects of meaning predictability. First, we test whether the speakers in our data set differ from each other in how strongly meaning predictability affects language choice: in each data set, on the basis of a likelihood ratio test, including separate speaker parameters for meaning unpredictability significantly increases explanatory power over including a single unpredictability parameter and a speaker parameter, suggesting that these speakers do indeed differ in their preferences.

Finally, we investigate the generalizability of the meaning-predictability effect beyond the individual speakers in this data set by comparing, for each data set, (i) a mixed-effects model with by-speaker random slopes AND a fixed effect of meaning predictability, and (ii) the same model without this fixed effect (§7.4) using the lme4 package in R (Bates et al. 2012). The complete set of *p*-values over all ten replicates for the fixed effect of meaning predictability was: 0.06, 0.07, 0.03, 0.03, 0.03, 0.07, 0.03, 0.03, 0.03, 0.03.

APPENDIX C: COMPLETE CONTROL FACTOR MODEL

	PARAMETER ESTIMATES		WALD'S TEST		LIKELIHOOD RATIO TEST	
PREDICTOR	Coef. β	*SE*(β)	*Z*	p_z	χ^2	*p*
Participant constellation =older.speakers.only	−0.99	0.27	−3.60	< 0.001	14.50	< 0.001
Frequency	0.02	0.19	0.09	0.93	0.10	0.93
Length	−0.69	0.21	−3.30	< 0.001	11.50	< 0.001
Imageability	−0.01	0.48	−0.03	0.98	0.00	0.98
Concreteness	0.58	0.49	1.19	0.23	4.50	0.11
Part of speech=verb	−1.57	0.37	−4.17	< 0.0001	45.70	< 0.0001
=noun	0.62	0.26	2.45	< 0.05		
Trigger=immediate	−0.95	1.16	−0.82	0.41	1.87	0.39
=in.clause	0.48	0.47	1.02	0.31		
Lexical cohesion=prev.English	0.67	0.26	2.57	< 0.05	26.95	< 0.0001
=prev.Czech	−1.28	0.34	−3.72	< 0.001		
Rightward collocation	1.74	2.26	0.77	0.44	2.11	0.15
Maximum collocation	−1.98	2.59	−0.76	0.44	2.33	0.13
Dependency distance	−0.33	0.25	−1.30	0.19	1.78	0.18
Syntactic governor	−2.10	0.81	−3.60	< 0.01	7.69	< 0.01
Length : Syntactic governor	3.13	1.21	2.60	< 0.01	3.88	< 0.05
Participant constellation =older.speakers.only : Concreteness	−0.95	0.47	−1.99	< 0.05	9.50	< 0.01
Speaker=speaker.1	0.36	0.20	1.80	0.07	3.20	0.07

TABLE A1. Result summary for a model including all base control factors and the two interactions selected during the model-selection process (for the last of ten iterations of the multiple imputation process): coefficient estimates β, standard errors *SE*(β), Wald's *z*-score and its significance level, contribution to likelihood χ^2 and its significance level. The response variable was coded as Czech/nonswitch = 0 and English/switch = 1. Predictors were centered and standardized so that numeric variables had a mean of 0 and a standard deviation of 0.5, and categorical variables had a mean of 0 and a difference of 1 between levels. Baselines are Participant constellation=all.participants, Part of speech=other, Trigger=none, Lexical cohesion=no.prev.mention, and Speaker=speaker.2.

REFERENCES

AGRESTI, ALAN. 2002. *Categorical data analysis*. Hoboken, NJ: John Wiley and Sons.

ALTARRIBA, JEANETTE; LISA M. BAUER; and CLAUDIA BENVENUTO. 1999. Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods* 31.578–602. DOI: 10.3758/BF03200738.

ANGERMEYER, PHILIPP SEBASTIAN. 2002. Lexical cohesion in multilingual conversation. *International Journal of Bilingualism* 6.361–93. DOI: 10.1177/13670069020060040101.

AUER, PETER. 1995. The pragmatics of code-switching: A sequential approach. *One speaker, two languages*, ed. by Lesley Milroy and Pieter Muysken, 115–36. Cambridge: Cambridge University Press.

AYLETT, MATTHEW, and ALICE TURK. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47.31–56. DOI: 10.1177/00238309040470010201.

BAAYEN, R. HARALD; DOUG DAVIDSON; and DOUGLAS BATES. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59.390–412. DOI: 10.1016/j.jml.2007.12.005.

Baayen, R. Harald; Richard Piepenbrock; and Leon Gulikers. 1995. The CELEX lexical database [CD-ROM]. Philadelphia: Linguistic Data Consortium.

Backus, Ad. 2003. Units in code switching: Evidence for multimorphemic elements in the lexicon. *Linguistics* 41.83–132. DOI: 10.1515/ling.2003.005.

Bailey, Benjamin. 2000. Social/interactional functions of code switching among Dominican Americans. *Pragmatics* 10.165–93. DOI: 10.1075/prag.10.2.01bai.

Barr, Dale J.; Roger Levy; Christoph Scheepers; and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.255–78. DOI: 10.1016/j.jml.2012.11.001.

Bates, Douglas; Martin Maechler; and Bin Dai. 2012. lme4: Linear mixed-effects models using s4 classes R package version 0.999999-0. Online: http://cran.r-project.org/package=lme4.

Beebe, Leslie, and Howard Giles. 1984. Speech-accommodation theories: A discussion in terms of second-language acquisition. *International Journal of the Sociology of Language* 46.5–42. DOI: 10.1515/ijsl.1984.46.5.

Bell, Alan; Daniel Jurafsky; Eric Fosler-Lussier; Cynthia Girand; Michelle Gregory; and Daniel Gildea. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America* 113.1001. DOI: 10.1121/1.1534836.

Broersma, Mirjam. 2009. Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition* 12.447–62. DOI: 10.1017/S1366728909990204.

Broersma, Mirjam, and Kees de Bot. 2006. Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition* 9.1–13. DOI: 10.1017/S1366728905002348.

Calcagno, Vincent, and Claire de Mazancourt. 2010. glmulti: An R package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software* 34.1–29. DOI: 10.18637/jss.v034.i12.

Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and topic*, ed. by Charles N. Li, 25–55. New York: Academic Press.

Chafe, Wallace. 1987. Cognitive constraints on information flow. *Coherence and grounding in discourse*, ed. by Russell Tomlin, 21–51. Amsterdam: John Benjamins.

Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chan, Mun-Chee; Helen Chau; and Rumjahn Hoosain. 1983. Input/output switch in bilingual code switching. *Journal of Psycholinguistic Research* 12.407–16. DOI: 10.1007/BF01067622.

Clark, Herbert H. 1973. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 12.335–59. DOI: 10.1016/S0022-5371(73)80014-3.

Clyne, Michael. 1991. *Community languages: The Australian experience*. Cambridge: Cambridge University Press.

Clyne, Michael. 2003. *Dynamics of language contact*. Cambridge: Cambridge University Press.

Coltheart, Max. 1981. The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology* 33A.497–505. DOI: 10.1080/14640748108400805.

Costa, Albert, and Mikel Santesteban. 2004. Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language* 50.491–511. DOI: 10.1016/j.jml.2004.02.002.

Damerau, Fred. 1993. Generating and evaluating domain-oriented multi-word terms from texts. *Information Processing & Management* 29.433–47. DOI: 10.1016/0306-4573(93)90039-G.

D'Amico, Simonetta; Antonella Devescovi; and Elizabeth Bates. 2001. Picture naming and lexical access in Italian children and adults. *Journal of Cognition and Development* 2.71–105. DOI: 10.1207/S15327647JCD0201_4.

de Rooij, Vincent A. 2000. French discourse markers in Shaba Swahili conversations. *International Journal of Bilingualism* 4.447–66. DOI: 10.1177/13670069000040040401.

Demberg, Vera, and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109.193–210. DOI: 10.1016/j.cognition.2008.07.008.
Doron, Edit. 1983. On a formal model for code-switching. *Proceedings of the Texas Linguistic Forum*, 35–59.
Du Bois, John W.; Stephan Schuetze-Coburn; Susanna Cumming; and Danae Paolino. 1993. Outline of discourse transcription. *Talking data: Transcription and coding in discourse research*, ed. by Jane A. Edwards and Martin D. Lampert, 45–89. Hillsdale, NJ: Lawrence Erlbaum.
Eppler, Eva Duran. 2011. The dependency distance hypothesis for bilingual code-switching. *Proceedings of the International Conference on Dependency Linguistics*, Barcelona.
Ferreira, Victor S. 2010. Language production. *Wiley Interdisciplinary Reviews: Cognitive Science* 1.834–44. DOI: 10.1002/wcs.70.
Ferreira, Victor S., and Gary S. Dell. 2000. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40.296–340. DOI: 10.1006/cogp.1999.0730.
Ferreira, Victor S., and L. Robert Slevc. 2007. Grammatical encoding. *The Oxford handbook of psycholinguistics*, ed. by M. Gareth Gaskell, 453–70. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780198568971.013.0027.
Forster, Kenneth I., and Susan M. Chambers. 1973. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12.627–35. DOI: 10.1016/S0022-5371(73)80042-8.
Fox, Barbara A., and Sandra A. Thompson. 2010. Responses to *wh*-questions in English conversation. *Research on Language & Social Interaction* 43.133–56. DOI: 10.1080/08351811003751680.
Franceschini, Rita. 1998. Code-switching and the notion of code in linguistics: Proposals for a dual focus model. *Code-switching in conversation: Linguistic perspectives on bilingualism*, ed. by Peter Auer, 51–75. London: Routledge.
Frank, Austin, and T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. *Proceedings of the 30th annual conference of the Cognitive Science Society (CogSci08)*, 939–44.
Friendly, Michael; Patricia Franklin; David Hoffman; and David Rubin. 1982. The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods* 14.375–99. DOI: 10.3758/BF03203275.
Fukumura, Kumiko, and Roger P. G. van Gompel. 2012. Producing pronouns and definite noun phrases: Do speakers use the addressee's discourse model? *Cognitive Science* 36.1289–1311. DOI: 10.1111/j.1551-6709.2012.01255.x.
Gahl, Susanne. 2008. *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84.474–96. DOI: 10.1353/lan.0.0035.
Genzel, Dmitriy, and Eugene Charniak. 2002. Entropy rate constancy in text. *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 199–206.
Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68.1–76. DOI: 10.1016/S0010-0277(98)00034-1.
Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, ed. by Alec Marantz, Yasushi Miyashita, and Wayne O'Neil, 95–126. Cambridge, MA: MIT Press.
Givón, Talmy. 1985. Iconicity, isomophism, and non-arbitrary coding in syntax. *Iconicity in syntax*, ed. by John Haiman, 187–220. Amsterdam: John Benjamins.
Gollan, Tamar H.; Daniel Kleinman; and Christine E. Wierenga. 2014. What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General* 143.2167–95. DOI: 10.1037/a0038006.
Gollan, Tamar H., and Victor S. Ferreira. 2009. Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35.640–65. DOI: 10.1037/a0014981.

Gries, Stefan Th.; Beate Hampe; and Doris Schnefeld. 2005. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16.635–76. DOI: 10.1515/cogl.2005.16.4.635.

Grosjean, François. 1997. Processing mixed language: Issues, findings and models. *Tutorials in bilingualism: Psycholinguistic perspectives*, ed. by Annette de Groot and Judith Kroll, 225–54. Mahwah, NJ: Lawrence Erlbaum.

Grosjean, François, and Joanne Miller. 1994. Going in and out of languages: An example of bilingual flexibility. *Psychological Science* 5.201–6. DOI: 10.1111/j.1467-9280.1994.tb00501.x.

Gumperz, John. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Gumperz, John, and Dell Hymes. 1986. *Directions in sociolinguistics: The ethnography of communication*. Oxford: Basil Blackwell.

Hajič, Jan. 2004. *Disambiguation of rich inflection: Computational morphology of Czech*, vol. 1. Prague: Karolinum Charles University Press.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the 2nd meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 1–8. Online: http://www.aclweb.org/anthology/N/N01/N01-1021.pdf.

Halliday, M. A. K. 1967. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3.199–244. DOI: 10.1017/S0022226700016613.

Halliday, M. A. K., and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Harrell, Frank. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer.

Harrell, Frank, and Charles Dupont. 2012. Hmisc: Harrell miscellaneous. R package version 3.9-3. Online: http://cran.r-project.org/web/packages/Hmisc/index.html.

Heredia, Roberto, and Jeanette Altarriba. 2001. Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science* 10.164–68. DOI: 10.1111/1467-8721.00140.

Jaeger, T. Florian. 2006. *Redundancy and syntactic reduction in spontaneous speech*. Stanford, CA: Stanford University dissertation.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59.434–46. DOI: 10.1016/j.jml.2007.11.007.

Jaeger, T. Florian. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology* 61.23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Jelinek, Tomas. 2008. Nové značkování v Českém národním korpusu. *Naše řeč* 91.13–20.

Joshi, Aravind. 1982. Processing of sentences with intra-sentential code-switching. *Proceedings of the International Conference on Computational Linguistics (COLING '82)* 9.145–50. Online: http://www.aclweb.org/anthology/C82-1023.

Karrebaek, Martha. 2003. Iconicity and structure in codeswitching. *International Journal of Bilingualism* 7.407–41. DOI: 10.1177/13670069030070040401.

Komagata, Nobo. 2003. Contextual effects on word order: Information structure and information theory. *Modeling and using context*, ed. by Varol Akman, Paolo Bouquet, Richmond Thomason, and Roger Young, 190–203. Dundee: Springer.

Kootstra, Gerrit; Janet van Hell; and Ton Dijkstra. 2009. Two speakers, one dialogue: An interactive alignment perspective. *Multidisciplinary approaches to code switching*, ed. by Ludmila Isurin, Donald Winford, and Kees de Bot, 129–60. Amsterdam: John Benjamins.

Kuno, Susumu. 1972. Functional sentence perspective. *Linguistic Inquiry* 3.269–320.

Kuno, Susumu. 1978. Generative discourse analysis in America. *Current trends in textlinguistics*, ed. by Wolfgang U. Dressler, 275–94. Berlin: De Gruyter.

Kuno, Susumu. 1979. Lexical and contextual meaning. *Linguistic Inquiry* 5.469–77.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge: Cambridge University Press.

Lee, Hanjung. 2003. Prominence mismatch and markedness reduction in word order. *Natural Language and Linguistic Theory* 21.617–80. DOI: 10.1023/A:1024198104514.

Levelt, Willem J. M. 1999. Models of word production. *Trends in Cognitive Sciences* 3.223–32. DOI: 10.1016/S1364-6613(99)01319-4.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106.1126–77. DOI: 10.1016/j.cognition.2007.05.006.
Levy, Roger, and T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems 19 (NIPS 2006)*. Online: http://papers.nips.cc/paper/3129-speakers-optimize-information-density-through-syntactic-reduction.
Li, Ping. 1996. Spoken word recognition of code-switched words by Chinese-English bilinguals. *Journal of Memory and Language* 35.757–74. DOI: 10.1006/jmla.1996.0039.
Mahowald, Kyle; Evelina Fedorenko; Steven T. Piantadosi; and Edward Gibson. 2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* 126.313–18. DOI: 10.1016/j.cognition.2012.09.010.
Manin, Dmitrii. 2006. Experiments on predictability of word in context and information rate in natural language. *Journal of Information Processes* 6.229–36.
Marian, Viorica. 2009. Language interaction as a window into bilingual cognitive architecture. *Multidisciplinary approaches to code switching*, ed. by Ludmila Isurin, Donald Winford, and Kees de Bot, 161–88. Amsterdam: John Benjamins.
Martinet, André. 1953. Preface. *Languages in contact*, by Uriel Weinreich. The Hague: Mouton.
Matras, Yaron. 2009. *Language contact*. Cambridge: Cambridge University Press.
Moreno, Eva; Kara Federmeier; and Marta Kutas. 2002. Switching languages, switching *palabras* (words): An electrophysiological study of code switching. *Brain and Language* 80.188–207. DOI: 10.1006/brln.2001.2588.
Munoa, Inma Barredo. 1997. Pragmatic functions of code-switching among Basque-Spanish bilinguals. *Proceedings of Actas do i Simposio Internacional sobre o Bilingismo*, Vigo, Spain, 528–41.
Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
Myers-Scotton, Carol. 1993a. *Duelling languages*. Oxford: Clarendon.
Myers-Scotton, Carol. 1993b. *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon.
Myers-Scotton, Carol. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. New York: Oxford University Press.
Nishimura, Miwa. 1989. The topic-comment construction in Japanese-English code-switching. *World Englishes* 8.365–77. DOI: 10.1111/j.1467-971X.1989.tb00675.x.
Petkevič, Vladimír. 2006. Reliable morphological disambiguation of Czech: Rule-based approach is necessary. *Insight into the Slovak and Czech corpus linguistics*, ed. by Maria Simkova, 26–44. Bratislava: Veda.
Piantadosi, Steven; Harry Tily; and Edward Gibson. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108.3526–29. DOI: 10.1073/pnas.1012551108.
Piccinini, Page. 2012. The gradient production of Spanish-English code-switching. Paper presented at the 68th annual meeting of the Linguistic Society of America, Portland, OR.
Poplack, Shana. 1980. 'Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL': Toward a typology of code-switching. *Linguistics* 18.581–618. DOI: 10.1515/ling-2013-0039.
Prince, Ellen F. 1981. Toward a taxonomy of given-new information. *Radical pragmatics*, ed. by Peter Cole, 223–56. New York: Academic Press.
Prince, Ellen F. 1984. Topicalization and left-dislocation: A functional analysis. *Annals of the New York Academy of Sciences* 433.213–25. DOI: 10.1111/j.1749-6632.1984.tb14769.x.
Proverbio, Alice; Giuliana Leoni; and Alberto Zani. 2004. Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia* 42.1636–56. DOI: 10.1016/j.neuropsychologia.2004.04.013.
Qian, Ting, and T. Florian Jaeger. 2012. Cue effectiveness in communicatively efficient discourse production. *Cognitive Science* 36.1312–36. DOI: 10.1111/j.1551-6709.2012.01256.x.
R Development Core Team. 2012. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Online: http://www.R-project.org/.

Riehl, Claudia Maria. 2005. Cognitive processes in bilinguals: Impacts of mental processes and language awareness. *Proceedings of the Fourth International Symposium on Bilingualism*, 1945–57.

Ritchie, William C., and Tej K. Bhatia. 2004. Social and psychological factors in language mixing. *The handbook of bilingualism and multilingualism*, 2nd edn., ed. by Tej K. Bhatia and William C. Ritchie, 336–52. Chichester: John Wiley & Sons. DOI: 10.1002/9781118332382.ch15.

Romaine, Suzanne. 1989. *Bilingualism*. Oxford: Blackwell.

Sankoff, David, and Shana Poplack. 1981. A formal grammar for code-switching. *Research on Language and Social Interaction* 14.3–45. DOI: 10.1080/08351818109370523.

Shannon, Claude. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27.379–423.

Shannon, Claude. 1951. Prediction and entropy of printed English. *Bell System Technical Journal* 30.50–64.

Shenk, Petra Scott. 2006. The interactional and syntactic importance of prosody in Spanish-English bilingual discourse. *International Journal of Bilingualism* 10.179–205. DOI: 10.1177/13670069060100020401.

Smith, Nathaniel J., and Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128.302–19. DOI: 10.1016/j.cognition.2013.02.013.

Spoustová, Drahomíra; Jan Hajič; Jan Votrubec; Jan Krbec; and Pavel Květoň. 2007. The best of two worlds: Cooperation of statistical and rule-based taggers for Czech. *Workshop on Balto-Slavonic Natural Language Processing*, 67–74. Online: http://www.aclweb.org/anthology/W07-1709.

Sridhar, Shikaripur N., and Kamal K. Sridhar. 1980. The syntax and psycholinguistics of bilingual code mixing. *Canadian Journal of Psychology* 34.407–16. DOI: 10.1037/h0081105.

Stadthagen-Gonzalez, Hans, and Colin Davis. 2006. The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods* 38.598–605. DOI: 10.3758/BF03193891.

Szmrecsanyi, Benedikt. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1.113–50. DOI: 10.1515/cllt.2005.1.1.113.

Tagliamonte, Sali. 2006. 'So cool, right?': Canadian English entering the 21st century. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 51.309–31. DOI: 10.1353/cjl.2008.0018.

Taylor, Wilson. 1953. Cloze procedure: A new tool for measuring readability. *Journalism Quarterly* 30.415–33.

Tily, Harry; Susanne Gahl; Inbal Arnon; Neal Snider; Anubha Kothari; and Joan Bresnan. 2009. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition* 1.147–65. DOI: 10.1515/LANGCOG.2009.008.

Tily, Harry, and Steven Piantadosi. 2009. Refer efficiently: Use less informative expressions for more predictable meanings. *Proceedings of the Workshop on the Production of Referring Expressions (PRE-CogSci 2009): Bridging the gap between computational and empirical approaches to reference*. Online: http://pre2009.uvt.nl/pdf/tilypiantadosi.pdf.

Timm, Lenora. 1978. Code-switching in *War and peace*. *Aspects of bilingualism*, ed. by Michel Paradis, 302–15. Columbia, SC: Hornbeam.

Tomlin, Russell. 1986. *Basic word order: Functional principles*. Wolfeboro, NH: Croom Helm.

Van Hell, Janet, and Annette M. B. de Groot. 1998. Disentangling context availability and concreteness in lexical decision and word translation. *The Quarterly Journal of Experimental Psychology* 51.41–63. DOI: 10.1080/713755752.

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford, CA: CSLI Publications.

Wasow, Thomas; T. Florian Jaeger; and David Orr. 2011. Lexical variation in relativizer frequency. *Expecting the unexpected: Exceptions in grammar*, ed. by Simon Horst and Wiese Heike, 175–96. Berlin: De Gruyter.

WEISS, DANIEL J.; CHIP GERFEN; and AARON D. MITCHEL. 2009. Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? *Language Learning and Development* 5.30–49. DOI: 10.1080/15475440802340101.

WOOD, ANGELA M.; IAN R. WHITE; and PATRICK ROYSTON. 2008. How should variable selection be performed with multiply imputed data? *Statistics in Medicine* 27.3227–46. DOI: 10.1002/sim.3177.

WULFF, STEFANIE; NICK ELLIS; UTE RÖMER; KATHLEEN BARDOVI-HARLIG; and CHELSEA LEBLANC. 2009. The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *The Modern Language Journal* 93.354–69. DOI: 10.1111/j.1540-4781.2009.00895.x.

ZENTELLA, ANA CELIA. 1997. *Growing up bilingual: Puerto Rican children in New York*. Oxford: Blackwell.

Department of Linguistics
UC San Diego
9500 Gilman Drive #0108
La Jolla, CA 92093-0108
[mmyslin@ucsd.edu]
[rlevy@ucsd.edu]

[Received 14 November 2013;
revision invited 14 March 2014;
revision received 9 July 2014;
accepted 22 September 2014]