Examining the effect of predictability on code-switching: A production experiment

Yanting Li, Gregory Scontras, and Richard Futrell
{yantil5, g.scontras, rfutrell} @uci.edu
Department of Language Science, 3151 Social Sciences Plaza
University of California, Irvine
Irvine, CA 92697-5100

Abstract

Corpus studies suggest that less predictable words are more likely to be code-switched in bilingual communication, yet experimental evidence remains limited, partially due to the difficulty of eliciting voluntary code-switches with a controlled linguistic context. In this study, we develop a production paradigm that reliably elicits Mandarin–English code-switches and allows for precise manipulation of word predictability through numeral classifiers. Using this paradigm, we show converging evidence that both contextual predictability and baseline word frequency influence the likelihood of code-switching. This experimental paradigm provides a valuable tool for future studies of bilingual speech production.

Introduction

When bilingual language users communicate, they sometimes code-switch, that is, switch from one language to another (Gumperz, 1977; Poplack, 1980; Solorio et al., 2014; Kroff & Dussias, 2023). Such switches can happen within a sentence, usually referred to as *intra-sentential* code-switching (Poplack, 1980). Corpus studies show that word predictability influences the switching point: less predictable words tend to be switched more (Myslín & Levy, 2015; Bhattacharya & van Schijndel, 2026; Calvillo, Fang, Cole, & Reitter, 2020). Multiple explanations have been proposed for this predictability effect, making it difficult to disentangle their contributions using corpus data alone. However, experimental evidence remains limited, partially due to the difficulty of eliciting natural code-switches with a controlled linguistic context.

To address this gap, the current study presents an experimental paradigm designed to elicit voluntary code-switches while maintaining control over the linguistic input and output, so it is possible to test the effect of specific factors (e.g. word predictability) on determining the intra-sentential switching point. We apply this paradigm to Mandarin–English bilinguals, manipulating noun predictability via the preceding classifier: a generic classifier compatible with many nouns versus a specific classifier compatible with a more restricted set. Nouns following the generic classifier are thus less predictable than those following a specific classifier. Our **research question** is whether less predictable nouns (i.e., those following the generic classifier) are more likely to be produced in English, resulting in a code-switch, given a sentential context that is (primarily) in Mandarin.

To preview the result, participants code-switched 40.77% of the time, confirming the effectiveness of the paradigm in eliciting voluntary code-switching. The predictability manipulation by classifier type did not turn out to be significant: we do not find more switching after generic classifiers than after specific ones. However, a post-hoc analysis transforming the categorical classifier type to a more continuous entropy measurement shows the result in the expected direction: when entropy over nouns following a classifier is high, that is, when the noun is less predictable, a code-switch is more likely to happen. Additionally, baseline word frequency turns out to be a significant predictor of code-switching: less frequent nouns in Mandarin have a higher code-switching rate. These findings provide behavioral evidence consistent with the existing literature and pave the way for experimentally disentangling the processes underlying this effect.

Background

Many factors have been identified to influence bilinguals' decisions to code-switch (or not). Socioculturally, code-switching can project certain intended identity or persona (Poplack, 1980; Heller, 1988; Myers-Scotton, 1993a, 1993b; Myers Scotton & Ury, 1977). Conversely, code-switching can be deliberately avoided in societies where it is stigmatized as lack of language proficiency (Heredia & Altarriba, 2001) or a sign of laziness (Grosjean, 1982). Meanwhile, code-switching behaviors can be influenced by the interlocutor's identity (Poplack, 1980; Y. Li, Yang, Scherf, & Li, 2013; Blanco-Elorrieta & Pylkkänen, 2015) and language proficiency (Myslín & Levy, 2015). The speaker's proficiency also plays a role (Poplack, 1980; Gollan & Ferreira, 2009; Poplack, 1980; Backus, 2003).

Within a discourse, code-switching can be adopted to quote others' speeches (Gumperz, 1982; Barredo, 1997), reiterate a message (Gumperz, 1982; Myers-Scotton, 1993b), maximize the saliency of discourse markers (De Rooij, 2000), introduce new topics (Barredo, 1997), and more.

Finally, linguistic factors play an important role determining where switches occur within the utterances. There are word-specific factors including length, frequency, part-of-speech (Myslín & Levy, 2015; Calvillo et al., 2020; Bhattacharya & van Schijndel, 2026; Poplack, 1980) and word meaning (Sapir, 1929; Otheguy & García, 1993; Clyne, 1967; Grosjean, 1982; Green & Wei, 2014; Y. Li, Scontras, & Futrell, 2024; Beatty-Martínez, Navarro-Torres, & Dussias, 2020). Other factors are embedded in the context. Some known triggers of code-switching include cognates (i.e.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

translation equivalents with overlapping lexical form across languages, e.g. Dutch-English *boek-book*) and previous instances of code-switching (Kootstra, Van Hell, & Dijkstra, 2012; Myslín & Levy, 2015; Fricke & Kootstra, 2016; Clyne, 2003; Broersma & De Bot, 2006; Kootstra et al., 2012; Fricke & Kootstra, 2016). The context also affects the predictability of the word, which is found to be another contributing factor to code-switching. This is where our main interest lies.

Predictability and Code-switching

Corpus studies have revealed that code-switched words tend to be less predictable given their linguistic context. Myslín and Levy (2015) obtained the predictability of meaning (PoM) of the code-switched words in Czech–English spoken conversations through a cloze test, measuring how likely the meaning (either in Czech or English) can be guessed by another group of bilinguals given audio context. Their results show that the PoM is lower at code-switched words, compared to non-code-switched words. Similarly for Chinese–English written communication, Calvillo et al. (2020) compared the 5gram surprisal (i.e. negative log probability; Hale, 2016; Levy, 2008) of the Chinese translation equivalent of a code-switched English word with that of a matching non-code-switched Chinese word in a similar syntactic environment. The results also show that the higher the surprisal, that is, the lower the predictability, the more likely that the word is code-switched. This effect was replicated by Bhattacharya and van Schijndel (2026) using surprisal values estimated by large language models (LLMs). They also extended the analysis to a spoken corpus and found the same pattern.

To account for the predictability effect, Myslín and Levy (2015) proposed that code-switching can function as a signal to the audience, alerting them to upcoming content that is unpredictable and therefore highly informative, serving audience-design purposes—an account supported by evidence that listeners can anticipate upcoming switches (Myslín & Levy, 2015; Tomić & Valdés Kroff, 2022). Bhattacharya and van Schijndel (2026) provided additional support for this account. Calvillo et al. (2020), on the other hand, proposed a speaker-oriented possibility, suggesting that less predictable words are more difficult to retrieve, leaving fewer cognitive resources available to inhibit the non-target language and thus increasing the likelihood of code-switching.

In contrast to the robust results in corpus analysis, behavioral evidence is relatively limited. The only study that we are aware of is conducted by de Bruin and Shiron (2024), where Bulgarian–English bilinguals were instructed to read a carrier sentence out loud, and finish it by naming a picture in either language. The carrier sentence can either make the picture more predictable ("*He milks*" + picture of a cow) or not ("*She holds*" + picture of a fork). Their findings show that switches are less likely when the carrier sentence is predictive of the picture, providing converging evidence with the corpus studies above. However, due to the constraints around sentence complexity, naturalness, the need to avoid cognates, and so on, pictures that appeared in predictable contexts in the experiment (e.g., cow) never appeared in the unpredictable contexts, and vice versa (de Bruin & Shiron, 2024, p. 1118). Because of this, it is difficult to tease apart contextual predictability from item-level properties, which to some extent limits the strength of causal inferences that can be drawn.

With the aim of better controlling the meaning of predictive vs. unpredictive contexts, Mandarin–English communication provides a useful testing ground. We turn to this case next.

Mandarin Classifiers

Mandarin is an obligatory-classifier language, which means that classifiers are required in numeral–noun constructions (C. N. Li & Thompson, 1981; N. N. Zhang, 2013). While we can say "five teachers" in English, in Mandarin we have to say *wǔ gè lǎoshī* 五个老师 "five $\text{CLF}_{\text{generic}}$ teachers", or *wǔ wèi lǎoshī* 五位老师 "five $\text{CLF}_{\text{respectable person}}$ teachers". In total, there are 105 frequently-used individual classifiers[1], including one generic classifier *gè* 个, and 104 specific classifiers (Ma, 2015) that apply in a more restricted fashion, often with a limited set of nouns that share certain semantic features (C. N. Li & Thompson, 1981; Srinivasan, 2010). Table 1 provides some examples of specific classifiers and the type of nouns that they are compatible with.

As illustrated in Fig 1, most nouns are compatible with the generic *gè* 个, although some are only compatible with specific classifiers (C. N. Li & Thompson, 1981; N. N. Zhang, 2013). A noun can also be compatible with both the generic and one or more specific classifiers, like the word *teacher* in the example earlier. These nouns are the focus of the present study, because their contextual predictability can be systematically manipulated through the choice of classifier. Specif-

[1]There are also classifiers for aggregate measures (e.g., a pile), containers (e.g., a bottle), or standard measurements (e.g., a pound), and they can co-occur with a relatively flexible group of nouns. These classifiers are not considered in this study.

Specific CLF	Compatible noun features	Sample nouns
张 *zhāng*	objects with flat, sheet-like surface	photo, paper, table, bed
头 *tóu*	relatively large-sized animals	ox, pig, elephant, rhino
台 *tái*	machines, devices and appliances	computer, phone, fridge
位 *wèi*	respectable people and occupations	teacher, scientist, veteran

Table 1: Examples of specific individual classifiers in Mandarin.

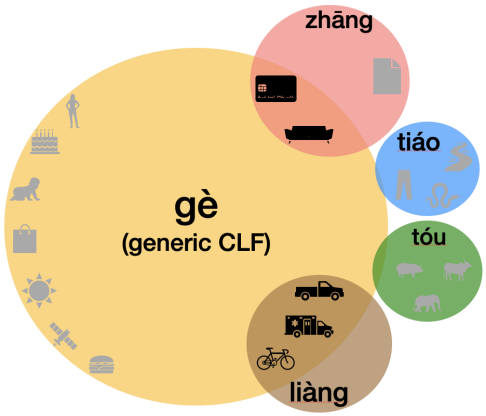

Figure 1: A conceptual illustration of generic and specific classifiers and the ranges of nouns with which they are compatible, as indicated by the size of the circles (not drawn to scale). Selected images shown in black are nouns that are compatible with both types of classifiers, whereas images shown in grey are nouns that are compatible with only one type. Our study focuses on the former.

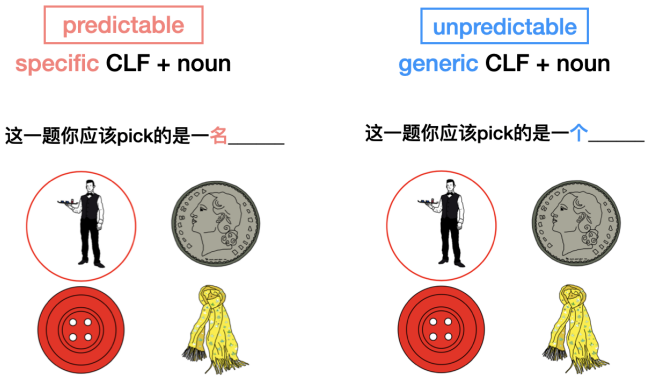

Figure 2: Experiment conditions. The predictable condition (left) has a preamble ending with a specific classifier, whereas the unpredictable condition (right) has a preamble ending with the generic *gè* 个. Both classifiers are compatible with the target picture that is circled. All preambles are semantically neutral. The one in this example reads: "In this trial, what you should pick is a _____".

ically, the generic *gè* 个 can precede a wide range of nouns, making the following noun difficult to predict. In contrast, a specific classifier substantially narrows down the set of possible following nouns: for example, with wèi 位 as a classifier, the following noun is constrained to denote a respected profession, such as teacher, scientist and veteran. As a result, the noun *lǎoshī* 老师 "teacher" is much more predictable following the specific than the generic classifier.

Current Study We test the effect of predictability on the code-switching of nouns, with predictability manipulated through classifiers (Fig 2). **Our hypothesis** is that nouns following the generic classifier *gè* 个 (which are therefore less predictable) are more likely to be produced in English, compared to nouns following a specific classifier.

Method

We design a production experiment that elicits voluntary code-switching by Mandarin–English bilinguals in contexts that allow for controlled classifier manipulation. We analyze the language used to produce nouns following the two types of classifiers and assess whether English were used more frequently after the generic classifier.

Design

The experiment aims to give bilinguals a *communicative goal*, and elicit *voluntary* code-switching with *controlled linguistic input*. Here is how we achieve these goals:

Communicative goal Participants are asked to complete a brief language background questionnaire at the beginning of the experiment. Afterwards, they are told that the system is pairing them up with another Mandarin–English bilingual with similar language background. They are then told that the two participants need to cooperate in a picture matching game, in which one serves as the speaker and the other serves as the listener. On each trial, both the speaker and the listener will see four pictures, but only the speaker sees which is the target. To win points, the listener must select the correct target picture, and the speaker must communicate the relevant information, making the task communicative in nature. However, it is important to point out that this "partner" does not actually exist. To make this communicative goal more real, participants get to experience the task both as a speaker and as a listener during practice, although they are always assigned the role of a speaker in the test phase.

Controlled linguistic input Similar to de Bruin and Shiron (2024), participants are presented with a preamble on each trial, and are instructed to read it out loud before naming the target picture. Fig 2 demonstrates the two conditions: in the predictable condition (left), the preamble ends with a specific classifier; in the unpredictable condition (right), the preamble ends with the generic classifier. Both classifiers are compatible with the target noun, which is circled in red. The preambles are semantically neutral, so that the predictability of the noun is manipulated solely by the type of classifier. Because of this, it is important that participants read the preambles word by word. To encourage careful reading, this requirement is incorporated into the communicative goal: participants are informed that the listener will occasionally be asked to select the correct preamble they heard, and a wrong answer will result in point deduction.

Voluntary code-switching Participants are told to use either language to name the target picture. To encourage code-switching, in addition to introducing the "partner" with similar language background, we also use mixed-language instructions throughout the experiment to set up an environment conducive to code-switching. Bonus points for fast completion are added to incentivize participants to use whichever word comes to mind first; a timer is displayed at the top of the screen to increase time pressure.

The experiment has six parts:

- **Language background questionnaire**: The questionnaire is adapted from the LEAP-Q questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007) to collect information about participants' Mandarin and English proficiency, as well as their daily language use.
- **Task instruction**: This part briefly demonstrates the task.
- **Practice trials**: Participants get to play the game both as a speaker and as a listener for a couple of trials.
- **Test trials**: There are 30 trials, each with one target picture, three randomly selected distractors, and a randomly selected preamble. Each participant sees each target only once, in either the predictable or unpredictable condition.
- **English picture naming trials**: The task includes 30 trials using the same 30 target pictures. Participants name the pictures in English to ensure they know the English names; Mandarin knowledge is not tested, as all target participants are native speakers.
- **Post-experiment feedback**: This includes a short list of questions asking about participants' code-switching experience and their attitude towards code-switchin behavior.

Participants

100 Mandarin–English bilinguals living in English speaking countries were recruited through Prolific at $12/hour. Among these, we excluded participants whose native language was not Mandarin (N=20), those who produced more than five instances of incompatible classifier–noun combinations (N=2), and those who used only one language to name all pictures in the test trials due to a lack of evidence that the code-switching manipulation affected their responses (N=16).

The remaining 62 participants (16 male) ranged from 21-63 years old with a mean of 32.87 (SD=9.49). Their mean age of acquisition (AoA) for English is 9.08 years (SD=3.93), and the mean length of time that they have spent in an English-speaking country ranged from 0.5-39 years with a mean of 11.46 (SD=8.15). Their self-rated language proficiency and daily language use are summarized in Table 2. Overall, their Mandarin proficiency is at ceiling, whereas their English proficiency is a little lower. They use slightly more English than Mandarin on average. At the end of the experiment, participants reported a moderate frequency of code-switching (1=never, 10=always; M=4.56 SD=2.44). Attitudes were relatively tolerant (1 = "I try to avoid it," 10 = "I switch whenever I want to"; M=5.90, SD=3.41).

	Mandarin	English
Speaking	9.24 (1.39)	7.65 (1.53)
Understanding speech	9.61 (0.88)	7.95 (1.64)
Reading	9.23 (1.62)	8.27 (1.46)
No. of hours per day	7.06 (5.20)	7.67 (4.37)

Table 2: Means (and SDs) of self-rated proficiency and daily language use (N=62).

Materials

We use a set of 30 target pictures and 138 fillers. 26 targets and all fillers are from the MultiPic database (Duñabeitia et al., 2018), and the remaining 4 targets are from Google Images. Each target picture is compatible with the generic classifier and at least one specific classifier. 19 unique classifiers are used, including the generic *gè* 个 and 18 specific classifiers. Five semantically neutral preambles are constructed to increase variability. To encourage code-switching, each preamble had a unilingual Mandarin version and a bilingual version containing a single instance of code-switching. The preambles are presented in a random order on each trial. To make sure that the classifier manipulation truly changes word predictability, we calculated $p(\text{noun} \mid \text{preamble} + \text{CLF}_{\text{generic}})$ and $p(\text{noun} \mid \text{preamble} + \text{CLF}_{\text{specific}})$ for each noun using a transformer-style LLM (Z. Zhang et al., 2021) and confirmed that the former is indeed smaller than the latter through a paired t-test ($t = -13.246$, $p < 0.001$).

Procedure

Participants recruited from Prolific.com were directed to the experiment website, where they went through the six parts described earlier. During the test trial, each participant saw all 30 target pictures, each appearing in one condition only. During the English picture naming trials, participants saw the target picture only. The target pictures appeared in randomized order. The experiment lasted around 15 minutes.

Data Analysis

Among the trials produced by the 62 participants, trials are excluded if the participant failed to name the target in the English picture naming trial (indicating a lack of familiarity with the English name), or produced an incompatible classifier–noun pair, resulting in 1,636 valid trials.

Our primary analysis is a mixed-effects logistic regression model predicting the language of the noun (English vs. Mandarin) using the type of classifier produced (generic vs. specific), with random intercepts and slopes for classifier type by speaker and noun:

```
language ~ CLF type + (1 + CLF type || speaker)
           + (1 + CLF type || noun)
```

Although specific classifiers are generally compatible with fewer nouns than the generic classifier, there remains substantial variability across specific classifiers in the number of nouns they license. To obtain a more granular measure than the categorical distinction between classifier types, we compute the entropy over nouns following each classifier:

$$H(N \mid c, \text{clf}) = -\sum_{n \in N} p(n \mid c, \text{clf}) \log p(n \mid c, \text{clf}) \quad (1)$$

Here N is a set of 71,586 nouns drawn from HowNet (Qi et al., 2019), c is a fixed semantically neutral context *wǒ kàn jiàn le yī* 我看见了一 "I saw a", and clf is a classifier. The probabilities are calculated using the same LLM used earlier (Z. Zhang et al., 2021). Higher post-classifier entropy reflects greater uncertainty about the upcoming noun and thus provides a continuous measure of noun predictability. We replace the categorical `CLF type` in the above mixed-effects logistic regression model with this continuous `entropy` value, and run another model in the post-hoc analysis.

In addition, since baseline frequency of a noun also influences its predictability, we test whether baseline frequency can predict code-switching rates of the target nouns using a linear regression model.

Results

First, we check whether the experimental paradigm was effective in eliciting code-switching. As shown in Fig. 3, participants produced more nouns in Mandarin (left two bars) than in English (right two bars), but a substantial 40.77% were produced in English, indicating that the task successfully elicited voluntary code-switching.

Moving on to our main analysis, the effect of classifier type on language choice did not turn out to be significant ($p = 0.318$). However, replacing the categorical `CLF type` with the continuous `entropy` measurement in the regression model revealed an effect in the expected direction (Table 3). That is, when entropy over nouns following a classifier is high, i.e. when we are less certain about what noun will be following the classifier, the noun is more likely to be produced in English, resulting in a code-switch.

Examining noun-specific code-switching rates, we observed substantial variability, ranging from 21.57% to 61.40%. Linear regression analysis revealed a negative correlation with frequency: less frequent nouns in Mandarin were more likely to be switched to English ($p < 0.05$, Fig 4). The same pattern was obtained using English word frequency ($p < 0.05$), consistent with the strong correlation between Mandarin and English frequencies ($p < 0.001$). However, frequency differences between the two languages were not statistically significant ($p = 0.29$).

	Estimate	Std. Error	z score	p-value
Intercept	9.339	4.203	2.222	0.0263*
entropy	-1.206	0.573	-2.104	0.0354*

Table 3: Mixed-effects logistic regression results predicting noun language (English = 1, Mandarin = 2) from post-classifier entropy. As entropy increases, nouns are more likely to be produced in English.

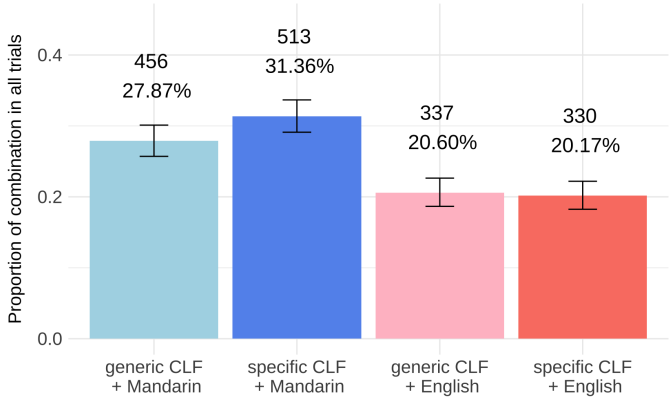

Figure 3: Proportion of trials in which each classifier–noun combination was produced. Bars show the mean proportion across all trials, separated by classifier type (generic vs. specific) and noun language (Mandarin vs. English). Error bars indicate binomial confidence intervals. Numbers above the bars report the raw count and the corresponding percentage.

Discussion

In this work, we examined the effects of three predictability-related predictors on code-switching, including two context-dependent measures (classifier type and post-classifier entropy), as well as a baseline, context-independent measure (word frequency).

Although the manipulation of generic vs. specific classifier type did not appear to influence code-switching of the following nouns, our entropy-based measure of predictability computed over the distribution of nouns following each classifier showed the expected pattern: nouns following classifiers with **higher post-classifier entropy** (and therefore with lower predictability) are more likely to be produced in English. This finding is consistent with the previous corpus analyses (Myslín & Levy, 2015; Calvillo et al., 2020; Bhattacharya & van Schijndel, 2026) and behavioral study (de Bruin & Shiron, 2024), and extends this line of work by providing behavioral evidence from a new language pair. Complementing this finding, we also observed a **clear effect of word frequency**: less frequent words in Mandarin are more likely to be switched to English. Despite previous corpus studies showing that a frequency effect often disappears once in-context predictability is accounted for (Myslín & Levy, 2015; Calvillo et al., 2020; Bhattacharya & van Schijndel, 2026), frequency and predictability are tightly correlated, so the frequency effect is consistent with the broader

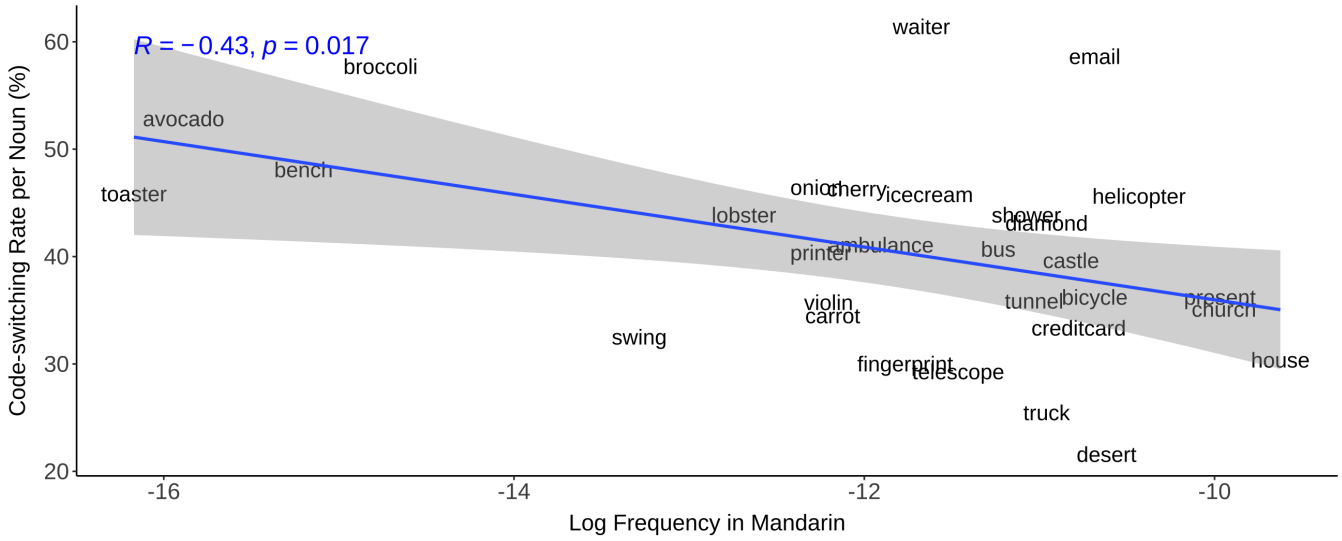

Figure 4: Relationship between log word frequency in Mandarin and code-switching rate with the solid line representing the linear regression fit.

predictability-based account.

Importantly, our experimental paradigm effectively and efficiently elicits **voluntary code-switches** in **communicative contexts** while maintaining **tight control** over the meaning of predictive vs. unpredictive contexts, which can be readily adapted for future studies of code-switching production. Code-switching is a highly complex phenomenon, shaped by multiple interacting linguistic, cognitive, and social factors, as reviewed in the Background section. Disentangling the contributions of these factors is therefore difficult if we use observational or corpus-based approaches alone. By maintaining control over the linguistic context while preserving a meaningful communicative goal, this paradigm provides a powerful tool for isolating specific factors and allows us to examine their effects on code-switching more rigorously.

Limitations

One potential reason for the absence of a classifier type effect is that the preamble and pictures were presented simultaneously. Participants may therefore have begun processing the picture before reaching the end of the preamble and encountering the classifier. If so, the observed baseline frequency effect becomes more understandable, although the fact that post-classifier entropy still remains significant after baseline word frequency is controlled for suggests that context still played a role. To examine this further, the display of pictures could be delayed until participants have fully processed the preamble, as done in de Bruin and Shiron (2024). While the short pause between the production of preamble and the naming of picture is slightly less natural, it likely allows the classifier to more fully influence noun predictability.

Another challenge lies in selecting target nouns, which must be: 1) compatible with both types of classifiers, 2) imageable, and 3) have the potential to be code-switched in Mandarin–English communication. One possible way to relax criterion 1) is to use alternative manipulations for word predictability. For instance, de Bruin and Shiron (2024) used verb-object relationships, which could provide a larger set of target nouns because multiple objects can plausibly follow a given verb, and a given object can plausibly follow many verbs. However, adopting this change would mean giving up on the tight control of meaning in our current design.

Conclusion

In this study, we behaviorally tested whether word predictability influences code-switching during bilingual speech production. Using classifier-based manipulations in Mandarin–English communication, we showed that less predictable nouns are more likely to be code-switched to English. These findings converge with previous corpus-based and behavioral evidence while extending to a new language pair under controlled experimental conditions. More broadly, our results support the view that predictability plays a systematic role in bilingual lexical choice and demonstrate the value of experimental approaches that isolate influencing factors, like predictability, in the study of code-switching.

References

Backus, A. (2003). Units in code switching: Evidence for multimorphemic elements in the lexicon. *Linguistics*, *41*(1), 83–132.

Barredo, I. M. (1997). Pragmatic functions of code-switching among basque-spanish bilinguals. *Proceedings of Actas do i Simposio Internacional sobre o Bilingismo*, 528–541.

Beatty-Martínez, A. L., Navarro-Torres, C. A., & Dussias, P. E. (2020). Codeswitching: A bilingual toolkit for opportunistic speech planning. *Frontiers in Psychology*, *11*, 1699. doi: https://doi.org/10.3389/fpsyg.2020.01699

Bhattacharya, D., & van Schijndel, M. (2026). Code-switching in text and speech challenges information-theoretic speaker design. *Language and Cognition*, *18*, e31. doi: https://doi.org/10.1017/langcog.2026.10074

Blanco-Elorrieta, E., & Pylkkänen, L. (2015). Brain bases of language selection: MEG evidence from Arabic-English bilingual language production. *Frontiers in Human Neuroscience*, *9*, 27.

Broersma, M., & De Bot, K. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition*, *9*(1), 1–13.

Calvillo, J., Fang, L., Cole, J., & Reitter, D. (2020). Surprisal predicts code-switching in Chinese–English bilingual text. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 4029–4039). doi: https://aclanthology.org/2020.emnlp-main.330.pdf

Clyne, M. G. (1967). *Transference and triggering: Observations on the language assimilation of postwar German-speaking migrants in Australia*. Martinus Nijhoff.

Clyne, M. G. (2003). *Dynamics of language contact: English and immigrant languages*. Cambridge University Press.

de Bruin, A., & Shiron, V. (2024). Putting language switching in context: Effects of sentence context and interlocutors on bilingual switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(7), 1112. doi: https://doi.org/10.1037/xlm0001309

De Rooij, V. A. (2000). French discourse markers in Shaba Swahili conversations. *International Journal of Bilingualism*, *4*(4), 447–466. doi: https://doi.org/10.1177/13670069000040040401

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six european languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816. doi: https://doi.org/10.1080/17470218.2017.1310261

Fricke, M., & Kootstra, G. J. (2016). Primed codeswitching in spontaneous bilingual dialogue. *Journal of Memory and Language*, *91*, 181–201. doi: https://doi.org/10.1016/j.jml.2016.04.003

Gollan, T. H., & Ferreira, V. S. (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*(3), 640. doi: https://doi.org/10.1037/a0014981

Green, D. W., & Wei, L. (2014). A control process model of code-switching. *Language, Cognition and Neuroscience*, *29*(4), 499–511. doi: https://doi.org/10.1080/23273798.2014.882515

Grosjean, F. (1982). Code-switching and borrowing. In *Bilingual : Life and reality* (pp. 51–62). Harvard University Press.

Gumperz, J. J. (1977). The sociolinguistic significance of conversational code-switching. *RELC Journal*, *8*(2), 1–34. doi: https://doi.org/10.1177/003368827700800201

Gumperz, J. J. (1982). Conversational code switching. In *Discourse strategies* (pp. 59–99). Cambridge University Press.

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, *10*(9), 397–412. doi: https://doi.org/10.1111/lnc3.12196

Heller, M. (1988). Strategic ambiguity: code-switching in the management of conflict. In M. Heller (Ed.), *Codeswitching : Anthropological and sociolinguistic perspectives* (pp. 51–62). De Gruyter Mouton. doi: https://doi.org/10.1515/9783110849615

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science*, *10*(5), 164–168. doi: https://doi.org/10.1111/1467-8721.00140

Kootstra, G. J., Van Hell, J. G., & Dijkstra, T. (2012). Priming of code-switches in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition*, *15*(4), 797–819. doi: https://doi.org/10.1017/S136672891100068X

Kroff, J. R. V., & Dussias, P. E. (2023). Production, processing, and prediction in bilingual codeswitching. *Psychology of Learning and Motivation*, *78*, 195–237. doi: https://doi.org/10.1016/bs.plm.2023.02.004

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. doi: https://doi.org/10.1016/j.cognition.2007.05.006

Li, C. N., & Thompson, S. A. (1981). Simple declarative sentences. In *Mandarin Chinese: A functional reference grammar* (pp. 104–113). Berkeley and Los Angeles, California: University of California Press.

Li, Y., Scontras, G., & Futrell, R. (2024). On the communicative utility of code-switching. In *Proceedings of the Society for Computation in Linguistics 2024* (pp. 343–349). Retrieved from https://aclanthology.org/2024.scil-1.25.pdf

Li, Y., Yang, J., Scherf, K. S., & Li, P. (2013). Two faces, two languages: An fMRI study of bilingual picture naming. *Brain and Language*, *127*(3), 452–462. doi: https://doi.org/10.1016/j.bandl.2013.09.005

Ma, A. (2015). *Hanyu geti liangci de chansheng yu fazhan [The emergence and development of Chinese individual classifiers]*. China Social Sciences Press.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Language*, *50*(4), 940–967. doi: https://doi.org/10.1044/1092-4388(2007/067)

Myers Scotton, C., & Ury, W. (1977). Bilingual strategies: The social functions of code-switching. *International Journal of the Sociology of Language*, *13*, 5–20. doi: https://doi.org/10.1515/ijsl.1977.13.5

Myers-Scotton, C. (1993a). Common and uncommon

ground: Social and structural factors in codeswitching. *Language in Society*, *22*(4), 475–503. doi: https://doi.org/10.1017/S0047404500017449

Myers-Scotton, C. (1993b). *Social motivations for codeswitching: Evidence from Africa*. Oxford University Press.

Myslín, M., & Levy, R. (2015). Code-switching and predictability of meaning in discourse. *Language*, 871–905. Retrieved from https://www.jstor.org/stable/24672251

Otheguy, R., & García, O. (1993). Convergent conceptualizations as predictors of degree of contact in US Spanish. In A. Roca & J. M. Lipski (Eds.), *Spanish in the united states: Linguistic contact and diversity* (pp. 135–154). De Gruyter, Inc.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: toward a typology of code-switching. *Linguistics*, *18*, 581–618. doi: https://doi.org/10.1515/ling.1980.18.7-8.581

Qi, F., Yang, C., Liu, Z., Dong, Q., Sun, M., & Dong, Z. (2019). *OpenHowNet: An open sememe-based lexical knowledge base.* Retrieved from https://arxiv.org/abs/1901.09957

Sapir, E. (1929). The status of linguistics as a science. *Language*, *5*(4), 207–214. doi: https://doi.org/10.2307/409588

Solorio, T., Blair, E., Maharjan, S., Bethard, S., Diab, M., Ghoneim, M., ... Fung, P. (2014, October). Overview for the first shared task on language identification in code-switched data. In M. Diab, J. Hirschberg, P. Fung, & T. Solorio (Eds.), *Proceedings of the First Workshop on Computational Approaches to Code Switching* (pp. 62–72). Doha, Qatar: Association for Computational Linguistics. Retrieved from https://aclanthology.org/W14-3907/ doi: 10.3115/v1/W14-3907

Srinivasan, M. (2010). Do classifiers predict differences in cognitive processing? A study of nominal classification in Mandarin Chinese. *Language and Cognition*, *2*(2), 177–190. doi: https://doi.org/10.1515/langcog.2010.007

Tomić, A., & Valdés Kroff, J. R. (2022). Expecting the unexpected: Codeswitching as a facilitatory cue in online sentence processing. *Bilingualism: Language and Cognition*, *25*(1), 81–92. doi: https://doi.org/10.1017/S1366728921000237

Zhang, N. N. (2013). *Classifier structures in Mandarin Chinese*. Walter de Gruyter.

Zhang, Z., Han, X., Zhou, H., Ke, P., Gu, Y., Ye, D., ... Sun, M. (2021). CPM: A large-scale generative chinese pre-trained language model. *AI Open*, *2*, 93–99. doi: https://doi.org/10.1016/j.aiopen.2021.07.001