Why and where to switch: understand code-switching using RSA models

Penny Pan (dpan3@stanford.edu)
Symbolic Systems Program
Stanford, CA 94305 USA

Amir Zur (amirzur@stanford.edu)
Department of Computer Science, Department of Linguistics
Stanford, CA 94305 USA

Abstract

When talking to another person, a bilingual speaker sometimes will alternate between the two languages and switch to the other language for certain words or phrases. Since code-switching is mostly spontaneous, what triggers the switch? In this paper, we focus on the switch of single nouns and noun phrases. We argue that the cost of the switch, which correlates to certain linguistic properties of the words, the communicative goals of the speaker, and the syntactic structure can trigger the switch. We formalize this idea using RSA models of language production and comprehension, and predict which language will the speaker choose to say the word in a given situation and which word is more likely to be switched.

Keywords: Computational modeling; code-switching; pragmatics

Introduction

Code-switching is a common phenomenon among bilinguals, who frequently change between the two languages during the conversation. It can occur in a cued situation where different languages are preferred under different circumstances or with different interlocutors. In addition, bilinguals also alternate between the two languages voluntarily either during a conversation, where one sentence is followed by a sentence in another language, or within a single sentence, where certain words are switched (Gardner-Chloros, 2009; de Bruin, Samuel, & Duñabeitia, 2018). Yet, switching from one language to another in a sentence seems to be less optimal than staying within one language. Specifically, the switch of language seems to be more costly even when the speaker voluntarily chooses to switch, possibly due to the inhibition of the non-target language (de Bruin et al., 2018). At the same time, the listener also needs to be able to process the switched utterance quickly and without being "surprised" and "delayed" by trying to retrieve the word in another language. In addition, the speaker may risk that the listener is not familiar with the language that the speaker switches to. Although code-switching seems to be cognitively burdensome, bilinguals still frequently mix the two languages and even prefer to switch in certain circumstances. Indeed, recent studies have suggested that the switch is sometimes beneficial (de Bruin et al., 2018) and may not be as effortful to produce (Johns & Steuck, 2021) and process (Gosselin & Sabourin, 2021). Then, why does code-switching occur? Specifically, when is code-switching is more likely to occur and what triggers the speaker to switch?

Previous studies have analyzed code-switching from different perspectives, including sociolinguistics, pragmatics, and syntax (Gardner-Chloros, 2009). To start with, there may be certain linguistic factors that contribute to the switch. Gollan and Ferria (2009) suggest that the accessibility of the words may affect the selection of the language. For example, words with higher frequency can be accessed more easily and thus are more likely to be code-switched. In addition, the speaker also tends to voluntarily switch the language if the switch can improve the accuracy of the expression (Gollan & Ferreira, 2009). Moreover, different syntactic theories attempt to explain why only certain switches are plausible. For instance, the Matrix Language Frame approach (Myers-Scotton, 1997) suggests that the Matrix language provides the syntactic and morphological frame, whereas the Embedded language (which is usually language in which the word is switched into) provides the lexicon. On the contrary, generative theories, including the Constraint Free Model as proposed by MacSwan (2008), attempt to use a uniform framework to describe both monolingual and bilingual code-switched structure (Parafita Couto & Gullberg, 2019).

Thus, the current study focuses on the intra-sentential switch of nouns and suggests that the switch can be predicted by certain properties of the word, the pragmatic goal of the speaker, and the syntactic structure. We first examine the production of a single noun, similar to a naming task in a reference game: When the bilingual speaker has two words that are equally accurate in referring the given object, which language will the speaker choose. Then, we test how the word choice will be affected by the communicative goal of the speaker when certain terms are ambiguous in one language. Then, we extend to the switch within noun phrases and examine the effect of syntactic structure on the switch.

Computational Models

To understand the effect of the linguistic, pragmatic, and syntactic factor, we formalize our hypotheses using the Rational Speech Act (RSA) framework (Goodman & Frank, 2016). In this framework, the speaker reasons about the knowledge of the listener and selects the utterance, whereas the listener in turn processes the utterance and infers about the speaker's intention based on their word choice. Since code-switching is a type of communication strategy among bilinguals, this framework seems to capture the recursive probabilistic reasoning between the speaker and the listener when selecting the code in order to achieve the common communicative goal. Although lack of empirical data, we propose three RSA models, one for each hypothesis, to predict which language the speaker will choose and where the speaker will code-switch within the phrase, if the switch occurs.

Linguistic Model

The linguistic model seeks to predict the variance in frequency of code-switching across different lexical items. It hypothesizes that cost of utterance, introduced in (Frank, 2016), can model a speaker's ability to retrieve and produce the utterance within spoken conversation. An utterance with a lower cost of production would be more likely to be used by a speaker, regardless of the language it belongs to. A speaker would code-switch from L1 to L2 when they are generally more comfortable with lexical items from L1, but wish to describe a particular object for which a lexical item from L2 is either better suited or more readily produced.

The model considers a literal listener, which uses a semantic meaning function, $\mathcal{L}(u,m)$, to evaluate the meaning m of utterance u. It outputs a pragmatic speaker distribution over utterances u given the intended meaning m.

$$P_{L_0}(m \mid u) \propto \mathcal{L}(u,m) \cdot P(m) \tag{1}$$

$$P_{S_1}(u \mid m) \propto \exp(\alpha \cdot \log P_{L_0}(m \mid u) + C(u)) \tag{2}$$

To compute the cost of a particular utterance, $C(u)$, we use a bilingual corpus of Mandarin-English speakers (Calvillo, Fang, Cole, & Reitter, 2020). Table 1 below presents the cost of utterance used in the RSA model, calculated as the negative log frequency of the utterance in the bilingual corpus.

Table 1: Cost of utterance in Linguistic RSA.

Utterance, u	Cost, $C(u)$
house	6.694
fangzi	7.085
money	10.91
qian	10.22
furniture	10.91
jiaju	7.397

Model predictions The linguistic RSA model makes predictions about utterances used in a reference game in which the speaker communicates one of the following objects to the listener: *house* (described by "house" or "fangzi"), *money* (described by "money" or "qian"), and *furniture* (described by "furniture" or "jiaju"). The RSA model predicts that the speaker would choose the least costly of the most informative utterances when communicating with the listener.

Figure 1: Linguistic RSA model predictions for utterances produced by speaker when referring to the object *house* (top), *money* (middle), and *furniture* (bottom).

Note that in the case of *house*, the speaker has a slight preference towards the English utterance, since it is more frequent in the speaker's corpus (and hence has a lower cost). However, in the case of *money*, the preference is for the Mandarin utterance, "qian", which has a lower cost. Our model predicts that words such as "furniture", which rarely appear in the Mandarin-English corpus - likely because of their length - are unlikely to be code-switched at all.

Model discussion In order to evaluate the linguistic model, we can compare the predicted rate of code-switching to a rate parsed from the spoken utterances catalogued by (Calvillo et al., 2020), as in the table below.

Table 2: Comparison of Model Predictions.

Referent	CS Rate from Model	CS Rate from Corpus
house	0.596	0.767
money	0.334	0.500
furniture	0.029	0.040

The linguistic RSA model predicts to a reasonable accuracy the relative tendency of a lexical item to be code-switched, especially in the cases of the *house* and *furniture* referents. Although the model fails to capture that "money" and "qian" seem to be code-switched an equal amount according to the corpus, the size of the data is small - comprising of 6 total examples (as opposed to 25 for *furniture* and 103 for *house*) - and so it seems that the rate of code-switching of a word is dependent on its cost of utterance, regardless of the language to which it belongs.

However, the linguistic RSA model is starkly limited by its assumptions. Although the correlation between cost of utterance and code-switching is notable, the cost is computed by the utterance's frequency in the bilingual corpus, which results in slightly circular reasoning - the model's explanation for the tendency of a word to be produced in code-switched speech is the word's tendency to be produced in speech. Further work with the cost-based RSA could be done to distinguish between a word's cost of retrieval and its frequency in code-switched utterances, perhaps using separate corpora for L1 and L2.

Despite giving predictions for which words are more likely to be code-switched generally, the linguistic RSA model fails to predict where within a sentence the word would be code-switched. The model also fails to take into account the context of the conversation (e.g. the relationship between the listener and the speaker, the underlying language for the conversation, or the speaker's intentions), and thus cannot predict the circumstances under which the phenomenon of code-switching might arise. To account for the underlying language of the conversation, we construct a Language-Prior model, detailed in the **Appendix**. In order to model the speaker-listener relationship, we develop a sociolinguistic model, which considers the speaker's goals as well as the linguistic knowledge of the listener.

Sociolinguistic Models

Similar to the trade-off of being accurate and being kind in polite speech (Yoon, Tessler, Goodman, & Frank, 2020), bilinguals sometimes face the similar dilemma: When the word in L1 is easier for the speaker to produce and less likely to lead to ambiguity, but the listener is more proficient in L2, which language will the speaker choose? If the speaker wants to show that they are being considerate about the linguistic ability of the listener, they will choose the one that is more familiar to the listener, namely the word in L2. On the other hand, if the speaker wants to be informative and avoid ambiguity, then saying the word in L1 is optimal.

Hence, in order to capture the cooperative relationship between the listener and the speaker, we include the pragmatic listener as well the as pragmatic speaker and build the sociolinguistic model. In this model, the literal listener, L_0 interprets the literal semantic meaning of the utterance, as it is previously defined in equation 1. The pragmatic speaker, S_1, has two goals: one is to allow the listener to identify the correct object, and the other is to make the listener feel good by using the word in the language that the listener is comfortable with. The trade-off between the informative and pro-social goal of the speaker is capture weighting parameter ϕ. Here, we assume the familiarity of an utterance is defined by $F(u)$ at the lexical level, instead of at the language level. Hence, it is possible that the speaker is more familiar with a word in L1, in comparison to its translation equivalence in L2, but another word in L2. The familiarity score shown in Table 3, is based on approximation, and the future pilot study needs to be conducted to derive an more accurate familiarity score. Then, given the utterance, pragmatic listener, L_1, not only reasons about the state of the world but also infers the goal of the speaker.

$$U_{S_1}(u \mid m, \phi) \propto \phi \cdot \log P_{L_0}(m \mid u) + (1-\phi) \cdot F(u) \quad (3)$$

$$P_{S_1} \propto \exp(\alpha \cdot U_{S_1}(u \mid m, \phi) + C(u)) \quad (4)$$

$$P_{L_1}(m, \phi \mid u) \propto P(m) \cdot P(\phi) \cdot P_{S_1}(u_l \mid m, \phi) \quad (5)$$

In addition, similar to the self-representational utility proposed by Yoon et al. (2020), the pragmatic speaker in the sociolinguistic model also signals their goal. In particular, the probability of choosing a non-target word should be higher, which indicates that the conflict between the informative and social utilities. Thus, as shown in (6), the speaker weights each of the three goals by the parameter ω. Altogether, given the object and the weights of the goal, the speaker infers about the state of the pragmatic listener L_1 and selects an utterance.

$$\begin{aligned} U_{S_2}(u \mid m, \phi, \omega) \propto\ & \omega_i \cdot \log P_{L_1}(m \mid u) + \omega_s \cdot F(u) \\ & + \omega_p \cdot \log P_{L_1}(\phi \mid u) \end{aligned} \quad (6)$$

$$P_{S_2}(u \mid m, \omega) \propto \exp(\alpha \cdot U_{S_2}(u \mid m, \phi, \omega) + C(u)) \quad (7)$$

Table 3: Word familiarity score of the listener and cost of utterance in sociolinguistic model.

Referent	Utterance, u	Familiarity, $F(u)$	Cost, $C(u)$
house	house	0.9	7.08
	fangzi	0.3	6.69
lease	lease	0.5	7.35
	zuyue	0.8	10.07
rent	rent	0.2	12.01
	zuyue	0.8	10.07

Model predictions With a similar reference game as in the Linguistic Model, we use the model to predict the word choice of the speaker. Specifically, when given a set of objects, one of the term "zuyue" is ambiguous polysemy, which can refer to either *lease* or *rent*. However, compared to its translation equivalents for both objects in English, "zuyue" is familiar to the listener.

The sociolinguistic model predicts the preference of the unambiguous but possible unfamiliar term when the speaker is trying to signal to the listener that they are being informative. On the contrary, the speaker is more likely to select the ambiguous utterance that is more familiar to the listener, when they want to appear to be social. Additionally, it also suggests that

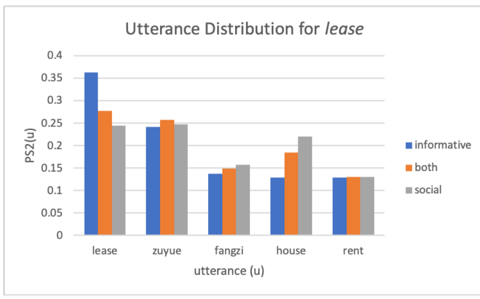

Figure 2: Sociolinguistic RSA model (with the self-presentational goal of the user prediction) for the object *lease*. In the "informative" condition, $\omega_i = 0.5, \omega_s = 0.01, \omega_p = 0.49$; in the "both" condition , $\omega_i = 0.3, \omega_s = 0.3, \omega_p = 0.4$; in the "social" condition, $\omega_i = 0.2, \omega_s = 0.4, \omega_p = 0.4$

On the other hand, as shown in Figure 3, the model predicts similar patters when the speaker does not project their goal to the listener view themselves. In contrast to Figure 2, one major difference is that the overall probability that selecting a non-target term decreases. Moreover, the ambiguous term "zuyue" is slightly more likely to be produced when the speaker is being informative than when being formal[1].

[1]Here, we set the speaker's utility parameter α to 1. However, when $\alpha = 2$, "zuyue" has the lowest probability score when the speaker aims to appear to be both informative and pro-social (i.e. with all of the three goals) than when the speaker only aims to be social (i.e. with the pro-social and self-presentation goal) and when only aims to be informative. This pattern is the opposite of that shown in Figure 2, where "zuyue" is most likely to be used in the "both" condition. The overall decrease in the probability of using the ambiguous term can be explained by the increase in speaker's utility weight, yet without experimental data, we are not sure which one actually captures the actual phenomenon, and the selection of parameter values is arbitrary, which may leads to the inverted pattern. In addition, it is also possible that there may be other factors to be considered.

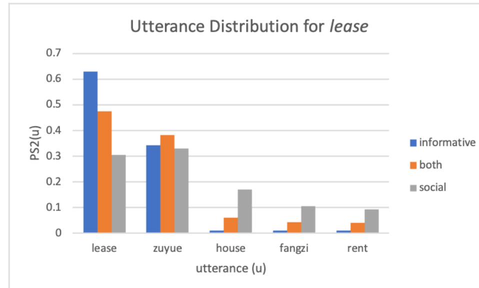

Figure 3: Sociolinguistic RSA model (without the self-presentational goal of the user) prediction for the object *lease*. $\phi = 0.9, 0.2, 0.5$ when the speaker is being informative, social, or both, respectively.

Model discussion This socilinguistic model captures the conflict between the goal of being clear and being considerate of the listener's linguistic knowledge. Meanwhile the model predicts that the speaker is more likely to choose a non-target utterance to signal the predicament when the most accurate word is not the easiest one for the listener to comprehend.

Yet, although the model can predict the general pattern of the language choice, we lack empirical data to fit the model and compare the two models. In particular, the self-representational utility seems to make a difference in the speaker's word choice, yet we are not sure to what degree bilingual speakers are using it to indicate the trade-off between being accurate and being considerate.

Despite being able to model an interaction between a pragmatic speaker and a pragmatic listener, the sociolinguistic model fails to fully capture the relationship between the speaker and listener - that is, it does not capture the interaction between two pragmatic participants *over time*. One possibility to account for this is a repeated reference game model, explained and analyzed in the **Appendix**. This model adapts the speaker's and listener's prior assumptions on the other's language preference as the reference game - a model for a conversation - is continued for multiple iterations.

Moreover, the current model only considers the word familiarity of the listener, yet it is possible that the speaker also has a prior preference of words. In particular, de Bruin et al. (2018) suggest that the individual differences in lexical access may predict the difference in language choice among bilinguals. Additionally, similar to the Linguistic Model, the current model only focuses on the use of a single word and fails to predict where within a phrase or sentence will the switch.

Syntactic Model

The syntactic RSA model attempts to explain the relationship between syntax and code-switched morphemes. There are two prevailing theories regarding the syntactic constraints on code-switching, the Matrix Language Frame Model (MLF), introduced by (Myers-Scotton, 1997), and the Constraint-Free Model, introduced by (MacSwan, 2008). While the MLF model generalizes rules on the interactions between the underlying Matrix Language and the code-switched Embedded Language based on universal features of syntax, the Constraint-Free model claims that constraints on code-switched statements are solely dictated by the mixed grammars of the two languages. Our syntactic RSA model employs the Constraint-Free model, mostly for the sake of simplicity, since our L1 and L2 are pre-defined as Mandarin and English.

The model defines a literal listener, L_0, which interprets an utterance by its semantic meaning. It then defines an incremental speaker, S_1, modeled after (Cohn-Gordon, Goodman, & Potts, 2018), which constructs a phrase utterance word by word, selecting the most informative utterance at each step. However, the speaker's utterance is constrained by the lexicon and syntax of their spoken language (either English or Mandarin), l. This utterance is in turn interpreted by the pragmatic listener L_1, which infers both the speaker's intended object m and spoken language l incrementally. Lastly, the model outputs the distribution for a social speaker S_2, which generates statements corresponding to the speaker's social goal of conveying their preferred language, and the speaker's informative goal of conveying their target object to the pragmatic listener, modulated by the parameter ϕ.

$$L_0(m \mid c,u) \propto [c+u](m) \tag{8}$$

$$S_1(u \mid c,m,l) \propto L_0(m \mid c,u,l) \tag{9}$$

$$L_1(l,m \mid c,u) \propto S_1(u \mid c,m,l) \tag{10}$$

$$\begin{aligned} S_2(u \mid c,m,l) \propto \phi \int L_1(l,m \mid c,u)dl \\ + (1-\phi)\int L_1(l,m \mid c,u)dm \end{aligned} \tag{11}$$

The syntactic RSA model generates phrases in the genitive case, of the form "N_1 of N_2" in English and "N_1 de N_2" in Mandarin. Furthermore, the meaning of the utterances depends on their syntax: in the English version, "N_1 of N_2" conveys that N_2 possesses N_1, while in Mandarin, "N_1 de N_2" conveys that N_1 possesses N_2. We predict that the different semantic interpretation caused by the dependence on a language's syntax will in turn mean that code-switching will depend both on the intended meaning of the utterance as well as on the syntax governing it.

Model predictions The syntactic model produces an interesting phenomenon, dependent both on language syntax and intended meaning: the most likely utterance for an informative speaker is code-switched so that the syntax of the language places the most salient part of the utterance foremost. Furthermore, the salient item of the utterance is produced in the same language as its governing syntax - to signal that it is either the possessor or the possession - while the less salient item is produced in the speaker's preferred language.

For example, consider a setting with the following objects: *house's lease*, *lease's house*, *furniture's lease*, *lease's furniture*, *furniture's money*, *money's furniture*. Below are the top six utterances produced by a pragmatic, informative, English speaker ($l = \text{English}, \phi = 0.9$), trying to convey the item *house's lease*:

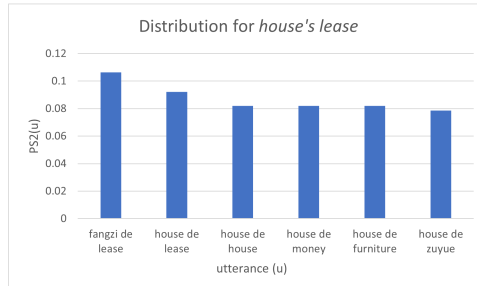

Figure 4: Syntactic RSA model prediction for the object *house's lease*, where $l = \text{English}, \phi = 0.9$.

Note that, of the six objects in front of the listener, four of them share the attribute *lease*, while only two share the attribute *house*. This imbalance leads the incremental speaker to prefer to use an utterance that would first signal *house*, by either using "fangzi" or "house." Since in the referent, *house's lease*, the attribute *house* is the possessor, only the Mandarin syntax, using "de", allows for an informative utterance which begins with "house". Hence, despite preferring English, the informative speaker would opt to use Mandarin syntax in this context. Furthermore, note that the most preferred utterance, "fangzi de lease", code-switches the lexical item for *house* into Mandarin, despite the speaker's preference for English. This is because the word "fangzi" alone signals that "de" is likely to follow, and hence signals both that "house" is the key attribute of the target reference, and that it is the possessor, rather than the possession.

In turn, consider the same context and the same speaker, this time attempting to identify the item *lease's house*:

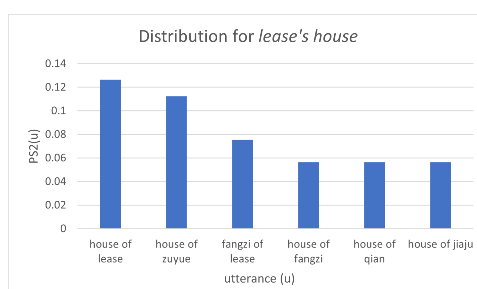

Figure 5: Syntactic RSA model prediction for the object *lease's house*, where $l = \text{English}, \phi = 0.9$.

Similarly to *house's lease*, the distinguishing feature of *lease's house* is the attribute *house*; however, in this context, *house* is the possession. Hence, in order to produce an utterance which signals that the salient feature is the possession, the speaker prefers to use the English syntax. Furthermore, in order to signal that they are using an English syntax, the speaker's most preferred utterance begins with "house", in contrast to the previous context, in which the produced utterance begins with "fangzi." Note that the latter, less salient item of the genitive-case statement is code-switched to the speaker's preferred language (here, English in both cases).

Model discussion The syntactic model proposes an interesting hypothesis on the dependence of code-switching on syntax: speakers choose to code-switch into the language whose syntax most readily allows for emphasis on the salient features of an object - and furthermore, speakers use a lexicon of the same language as the code-switched syntax. On the other hand, less salient features of an object would be described in a speaker's more preferred language, and not be code-switched. This hypothesis best agrees with the Constraint-Free Model of code-switching (MacSwan, 2008), since it stipulates that code-switching directly depends on the interaction between the grammars of L1 and L2, and not a universal feature of syntax.

The limitation of such a hypothesis is that it is difficult to generalize. The interaction between Mandarin and English genitive-case statements is unique to these two languages. It is also not altogether captured by our defined syntax: in Mandarin, "de" has more uses than genitive-case statements; in English, one can use both "N_1 of N_2" as well as "N_1's N_2" to refer to both directions of possession. In order to make meaningful predictions of code-switching given the syntactic RSA model, the full mixed syntax of two unique languages must be defined.

Additionally, similar to the sociolinguistic RSA model, another limitation of the syntactic RSA model is that there is currently no experimental data to support its hypothesis of the dependence of code-switching on both context and syntax. The example reference game could serve as a useful quantitative evaluation of the syntactic RSA model - although the current set of objects should be revised so that genitive-case phrases are semantically meaningful in both directions (while *house's lease* is a plausible object, *lease's house* is not as sensible).

Discussion

In this paper, we build three different RSA models to explore the code-switching phenomenon among bilinguals, from the linguistic, sociolinguistic, and syntactic perspectives. The cost-based Linguistic Model predicts that words with higher frequency is more likely to be selected and switched into. Yet given that it only focuses on the production cost of the speaker, this model fails to captures the recursive reasoning between the speaker and the listener in trying to achieve the communicative goal in the conversation. Hence, we propose the Sociolinguistic Model, which predicts that the goal of speaker affects the word choice. The speaker can either being informative and accurate in referring to the correct object, or being pro-social and considerate about the linguistic ability of the listener, or both, and signal to the listener about their goal. Last, we construct the Syntactic Model to predict the occurrence of switch within a possessive noun phrase. The results indicates that the switch can be predicted by the syntax of the two languages, such that the language whose structure that emphasizes on the salient feature of the object and the lexicon from that given language will be preferred.

Nonetheless, because we do not have empirical data to fit the models and compare them, the next step will be to conduct experiments and collect behavioral data, not only to infer the parameter values but also to test and compare the models. In addition, there are still limitations in each model. As aforementioned, all three models predict the switch in a simple setting and simulate one single interaction between the speaker and listener. Yet, the choice of code is dependent on the speaker-listener familiarity, and sometimes the speaker and the listener may switch only a few "random" words and ultimately finalize on using one language or a few "habitually-switched" words after multiple interactions. Likewise, at a higher level, the switch of certain phrases is commonly agreed and accepted by bilinguals, which is similar to the formation of conventions (R. D. Hawkins et al., 2021; D. R. Hawkins, Liu, Goldberg, & Griffiths, 2021). Hence, the future study can use a hierarchical Bayesian model to capture the iterative process in code-switching.

As suggested by Gardner-Chloros (2009), code-switching is a multifaceted phenomenon. Current paper is an initial attempt in building computational models to understand the why bilinguals code switch. Thus, besides the linguistic, sociolinguistic/pragmatic, and syntactic factors proposed and explored in the current paper, additional factors can be considered to understand the choice of the language and the switch of the code, including the effect of social identity associated with the code, the influence of the language environment (which can be captured by the prior preference of one language over the other).

References

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).

Cohn-Gordon, R., Goodman, N. D., & Potts, C. (2018). *An incremental iterated response model of pragmatics.*

de Bruin, A., Samuel, A. G., & Duñabeitia, J. A. (2018). Voluntary language switching: When and why do bilinguals switch between their languages? *Journal of Memory and Language*, *103*, 28-43. doi: https://doi.org/10.1016/j.jml.2018.07.005

Frank, M. C. (2016, Sep). *Rational speech act models of pragmatic reasoning in reference games.*

PsyArXiv. Retrieved from psyarxiv.com/f9y6b doi: 10.31234/osf.io/f9y6b

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge university press.

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.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818–829.

Gosselin, L., & Sabourin, L. (2021). Lexical-semantic processing costs are not inherent to intra-sentential code-switching: The role of switching habits. *Neuropsychologia*, *159*, 107922.

Hawkins, D. R., Liu, I., Goldberg, A., & Griffiths, T. (2021). Respect the code: Speakers expect novel conventions to generalize within but not across social group boundaries. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 43).

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2021). *From partners to populations: A hierarchical bayesian account of coordination and convention.*

Johns, M. A., & Steuck, J. (2021). Is codeswitching easy or difficult? testing processing cost through the prosodic structure of bilingual speech. *Cognition*, *211*, 104634.

MacSwan, J. (2008). Unconstraining codeswitching theories. *Papers from the Regional Meetings, Chicago Linguistic Society*, *44*, 151-168.

Myers-Scotton, C. (1997). *Duelling languages: Grammatical structure in codeswitching*. Oxford University Press.

Parafita Couto, M. C., & Gullberg, M. (2019). Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism*, *23*(2), 695–714.

Yoon, E. J., Tessler, M. H., Goodman, N. D., & Frank, M. C. (2020). Polite speech emerges from competing social goals. *Open Mind*, *4*, 71–87.

Appendix

Here we will discuss two other models considered: a language prior model, and a repeated reference game model. The language prior model is an extension of the cost-based model, in the Linguistic Model section of the paper. The repeated reference game model is an extension of the Sociolinguistic Model, attempting to capture speaker-listener familiarity over time.

The cost-based Linguistic Model is limited by its failure to capture information about the context of the conversation: the relationship between the speaker and the listener, the preferred language of the listener, the social goal of the speaker, and the underlying context of the conversation (for example, is it taking place in an L1-speaking community or L2-speaking community?). The language prior model attempts to address the latter concern by providing a shared underlying language prior, which the speaker draws from when producing utterances.

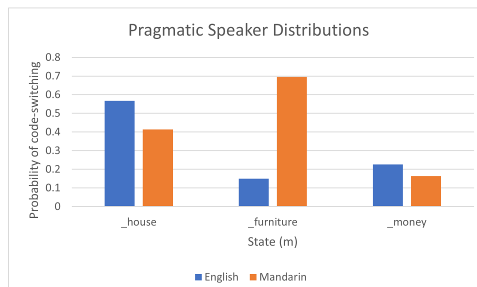

Figure 6: Language-prior RSA model run with a listener more familiar with English (0.8) and a speaker more familiar with Mandarin (0.7). Note that the speaker generally attempts to accommodate the listener's stronger preference for English. However, when the cost of utterance is too large (see Table 1), cost has a stronger impact language prior.

The language prior model hence explains the code-switching phenomenon as an interaction between the language underlying the conversation, and the speaker's ability to retrieve or generate utterances within that language. It thus predicts that the speaker's preference would be modulated by the context - the speaker would be more likely to produce words that match the language which the listener expects to hear.

The Sociolinguistic Model captures inference made both by the speaker about the listener and by the speaker about the listener. However, it does not consider interpretations made by the conversation participants over time. Drawing upon the repeated reference game RSA model (R. D. Hawkins et al., 2021), we consider a repeated familiarity model, in which the speaker and the listener iteratively infer the utterances which the other participant is most familiar with. This also averts the issue of defining our own familiarity scores for utterances, thus reducing the number of parameters which the model is dependent on.

Perhaps unsurpisingly, we find that the repeated familiarity model predicts that over time, speakers and listeners converge towards specific utterances, thus converging on a specific language. One interpretation of this is that as the speaker and listener know more about each other, they are more likely to code-switch into the other's more familiar language. Yet an equally valid interpretation of this model is that as the speaker and listener get to know more about each other, they become less likely to code-switch, and instead speak in their mutually comfortable language.

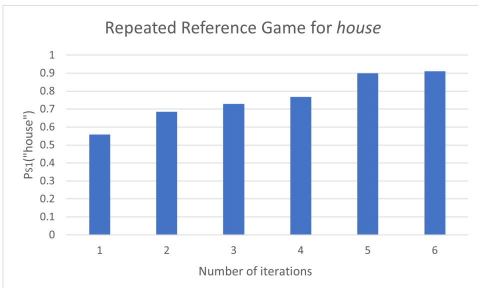

Figure 7: A repeated reference game run with the referent *house* for six iterations. Note that, as the game is repeated, the speaker and listener establish each others' familiarity with the word "house", and opt to use it increasingly more frequently (over 0.90 in iteration 6).

Future work could combine the repeated familiarity model with the incremental inference model, in order to generate utterances that are longer than a single word. This could provide better data to support either the hypothesis of increased or decreased code-switching over time, by inspecting whether the utterances produced are homogeneous or heterogeneous in their lexical items.