Creolization versus code-switching: An agent-based cognitive model for bilingual strategies in language contact

Charles Torres
University of California, Irvine
charlt4@uci.edu

Weijie Xu
University of California, Irvine
weijie.xu@uci.edu

Yanting Li
University of California, Irvine
yantil5@uci.edu

Richard Futrell
University of California, Irvine
rfutrell@uci.edu

Abstract

Creolization and code-switching are closely related contact-induced linguistic phenomena, yet little attention has been paid to the connection between them. In this paper, we propose an agent-based cognitive model which provides a linkage between these two phenomena focusing on the statistical regularization of language use. That is, we identify that creolization as a conventionalization process and code-switching as flexible language choice can be optimal solutions for the same cognitive model in different social environments. Our model postulates a social structure of bilingual and monolingual populations, in which a set of agents seek for optimal communicative strategy shaped by multiple cognitive constraints. The simulation results show that our model successfully captures both phenomena as two ends of a continuum, characterized by varying degrees of regularization in the use of linguistic constructions from multiple source languages. The model also reveals a subtle dynamic between social structure and individual-level cognitive constraints.

1 Introduction

Creolization and code-switching, despite being two distinct linguistic phenomena, share notable similarities in many aspects. For example, both occur in a social situation that involves language contact, where multiple linguistic communities encounter and engage in communication; both involve speakers processing linguistic signals encoded in different languages; and, most importantly, both result in utterances with a mixture of linguistic representations from multiple source languages.

However, creolization and code-switching are often studied with different theoretical and empirical focus (Muysken, 2013). Creolization, on one hand, is often treated as a population-level phenomenon, drawing attention primarily from historical and sociolinguists whose main goal is to depict the change and evolution of linguistic representations as a function of social movements (Thomason and Kaufman, 1988; Mufwene, 2011).[1] On the other hand, research on code-switching leans towards the individual level, with joint force from psycho- and contact linguistics to understand the psychological mechanism adopted by bilingual speakers (Green and Abutalebi, 2013; Green and Wei, 2014) as well as the grammatical and structural configurations constraining code-switching (Myers-Scotton, 1997; Muysken, 2000; Poplack, 1980).

In the current study, we aim to bring together these two closely related phenomena, focusing on the statistical aspect of language use (Bybee, 2006; Reali and Griffiths, 2009). In short, our use of the term *creolization* mainly focuses on its convention-formation nature, while the use of the term *code-switching* mainly focuses on the flexibility of language choice. In this sense, we consider creolization and code-switching as two ends of a continuum, varying in the degree of statistical regularization in the use of linguistic constructions from multiple source languages. Crucially, we identify that creolization as a conventionalization process and code-switching as flexible language choice can be optimal solutions for the same cognitive model in different social environments. [2]

[1]There are certainly different levels of explanations in creole studies (e.g., individual strategies, the role of the inductive bias in L1 acquisition, etc.). However, many studies conceptualize *creoles* (and other similar communicative systems, e.g., pidgins) as new languages/linguistic systems, which implies certain degree of conventionalization that is only meaningful at the population or subcommunity level.

[2]Our use of the term creolization here presupposes the answer to a theoretical question that is still under debate, in particular whether creolization is a distinct process or whether it is fundamentally similar to other forms of language change (Jourdan, 2021; Mufwene, 2004; McWhorter, 1998). By referring to a mix of constructions drawn from multiple languages as 'creolization', we follow recent work which eschews pidginization as a step in creolization. However, this debate between pidginization and creolization is beyond the scope of the current study: our model can be applied to other frameworks. The optimal codes we identify could be construed as operating during pidginization or any process where new signals are con-

Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 207–218
May 3, 2025

Through an agent-based cognitive model, we aim to provide a linkage between creolization as a population-level process and code-switching as an individual-level effect. Our model postulates a social structure of bilingual and monolingual populations, in which a set of agents engage in communication with each other. For each agent, their communicative strategy is shaped by multiple cognitive constraints. Specifically, each agent seeks to minimize their cognitive effort while maintaining the communicative success with their partner, a computational problem that we implement information-theoretically.

To preview the results, our model successfully captures both creolization and code-switching on the continuum of statistical regularization. Specifically, our model shows that when bilinguals put more cognitive effort in tracking their partner's identity and develop a partner-specific communicative strategy, they can rely on a simpler strategy per partner, leading to creolization as a conventionalization process. In contrast, when bilinguals put less effort into partner tracking, they will have to use a strategy that is more uncertain, leading to code-switching. Moreover, our model reveals a subtle dynamic between the social structure and the individual-level cognitive constraints: more intensive contact within the bilingual population is more likely to lead to creolization, but only when the bilingual agent puts enough effort into developing a partner-specific communicative strategy.

2 Background

A creole is a fully developed natural language with native speakers, often found to emerge in communities with intensive multilingual contact (e.g., colonies) as a strategy to address the need to communicate among speakers of different languages (Mufwene, 2004; Thomason and Kaufman, 1988). How creoles emerge is still an ongoing debate. McWhorter (1998) believes that creole is a special synchronically definable typological class that emerges from pidgins. However, opponents including Mufwene see creoles as more of a sociohistorical construct—they are just contact-heavy vernaculars of their lexifiers (i.e., the language from which the lexicon of a creole is primarily drawn) (Mufwene, 1996, 2004). Under this view, the constructions of a creole reflect features of the various source languages in contact as a result of a selection process. That is, from the combined "feature pool" of source languages, individuals select features for the creation of the new language. Such idiolectal selections then gradually converge through negotiation and compromise during communications, allowing the language to evolve into a new communal system that is unique enough from all source languages (Mufwene, 2002, 2004). In this paper, we adopt this "feature pool" viewpoint, and we model the bilingual strategy in language contact as a selection process from the pool of linguistic constructions from source languages.

Another contact-induced phenomenon in bilingual communication is code-switching, which unlike creolization, gives the speaker freedom to choose constructions from multiple languages during language production. According to Green and Abutalebi (2013), code-switching can happen in three interaction contexts: (1) single-language, where one language is used in one environment (e.g. at work) and the other in a second distinct environment (e.g. at home); (2) dual-language, where both languages are used but typically with different speakers; and 3) dense code-switching, where speakers switch languages within a single utterance, or even adapt words morphologically from one language in the context of the other. Later, Green and Wei (2014) proposed a control process model of code-switching, where the above interaction contexts determine whether the two languages are competitive (as in the single- and dual-language contexts) or cooperative (as in the dense code-switch context). If the two languages are competitive, cognitive efforts are needed to suppressed the non-target language, preventing it from entering into planning. If, however, the two languages are cooperative, neither needs to be suppressed. Speakers can choose whichever available construction that is most appropriate on the basis of semantic, syntactic and collocational considerations.

Conceptually, we view the key difference between creolization and code-switching as a matter of the degree of regularization with regard to the usage of linguistic constructions (Mufwene, 2020). At the one end is creolization, where linguistic constructions are used in a highly conventionalized way although the speakers are at the disposal of a repertoire of constructions from multiple source languages. That is, given a specific intended meaning, native speakers of a creole will agree on the

ventionalized. Meanwhile, the term code-switching might also have a different emphasis here than in many code-switching studies. We use this term to indicate the freedom to choose between languages when conveying a certain meaning.

construction to be used to express that meaning. At the other end is code-switching. Although bilinguals of the same languages may share some intuitions about where a switch can happen within a sentence, possibly due to some grammatical constraints (Poplack, 1980; Muysken, 2000), it will not be surprising at all if they choose different constructions and switching points to express the same intended meaning.

3 Modeling Framework

Through an agent-based model of bilingual strategies in a language contact scenario, we explore how creolization and code-switching are optimal solutions to a cognitive constraint problem within the bilingual population at two ends of a continuum which varies the regularization of linguistic constructions from multiple source languages.[3] Our model has two components. First, there is a **social structure** consisting of bilingual and monolingual populations, and a set of agents communicate with each other within this social structure. Second, at the individual level, each agent follows a communicative policy that is shaped by multiple cognitive constraints.

3.1 Communication within a social structure

There are two languages L_{A} and L_{B} in our toy scenario, and they share a discrete common meaning space $\mathcal{M}$. Each language consists of a disjoint set of $|\mathcal{M}|$ possible constructions (representing words, morphemes, syntactic patterns, and other strategies languages may use to convey meaning) for these meanings, $\mathcal{C}_{\mathrm{A}}$ and $\mathcal{C}_{\mathrm{B}}$. Therefore, there are $2|\mathcal{M}|$ constructions in total, forming a joint construction space for the entire society $\mathcal{C}_{\mathrm{A+B}} = \mathcal{C}_{\mathrm{A}} \cup \mathcal{C}_{\mathrm{B}}$.

In a communicative event, a speaker agent a_i, first comes up with an intended meaning $m \in \mathcal{M}$. Then, they select a construction $c \in \mathcal{C}$ to realize the intended meaning and send it to the listener agent a_j. The agent a_i's selection of construction c given meaning m follows a **communicative policy**, which is a probability distribution $p(c \mid m, a_j)$ to produce construction c given meaning m for listener a_j. In general, the goal of the sender in each communication is to choose a construction c in a way that helps the receiver a_j reconstruct the meaning m from it.

Figure 1 shows the social structure of a language

[3]Code for this model is available at `https://github.com/cj-torres/creolization-codeswitching`.

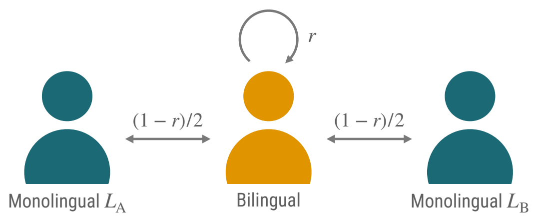

Figure 1: The social structure in a language contact scenario. A bilingual agent (in the middle) communicates both with monolingual agents (on the sides) and with other agents from the same bilingual population. The arrows represent the communication pathway along with the intensity of contact for each pathway.

contact scenario that defines the dynamics between agents. It consists of three agents, corresponding to three linguistic communities[4]: (1) two separate monolingual agents, one speaking L_{A} and the other L_{B}, and (2) a bilingual agent, which speaks both languages. For monolingual agents, given an intended meaning, they can only select a construction from their corresponding language, that is, either $c \in \mathcal{C}_{\mathrm{A}}$ or $c \in \mathcal{C}_{\mathrm{B}}$. For the bilingual agent, it can select constructions from the joint set of both languages $c \in \mathcal{C}_{\mathrm{A+B}}$.

The arrows in Figure 1 represent the dynamic of interaction among the three linguistic communities.[5] First, the bilingual speakers communicate among themselves, with the proportion of communicative events, or contact intensity, being r. Second, the bilingual agent communicates with each monolingual agent separately, with $(1 - r)/2$ representing the intensity of contact, that is, the proportion of communicative events that the bilingual agent has with the corresponding monolingual agent.[6] There is no direct communication between the two monolingual agents in this social structure.

[4]In the current version of our model, we assume the homogeneity within each community, in the sense that each individual within a community follows the same cognitive model and the same dynamic of interaction. Therefore, each community is represented as only one agent in our model.

[5]For simplicity, we assume that the communicative policy of monolingual groups remains stable and is not influenced by bilingual speakers, although this is not necessarily the case in a real language contact scenario (Thomason and Kaufman, 1988). Therefore, we only focused on the evolution of communicative policy within the bilingual group, without specifying any interaction within each monolingual group.

[6]Again, for simplicity, we assume that the intensity of contact with each monolingual group is equal, but our model is capable of accommodating unequal intensity as well.

3.2 Cognitive model for communicative policy

The cognitive model specifies at the individual level the communicative policy of the bilingual agent. This policy is subject to multiple competing constraints, and the goal of the agent is to search for an optimal policy striking a balance among these constraints. In this section, we first give a conceptual overview of these constraints and how they shape our individual-level cognitive model. We then turn to the mathematical formalization of our model.

3.2.1 Conceptual Overview

The fitness of the communicative policy is governed by two primary constraints. The first one is **communicative success**. That is, when encoding an intended meaning, the speaker should choose a construction such that their communicative partner can reconstruct the intended meaning based on the construction they receive. The second constraint is **cognitive effort**. A policy that involves more complicated decision-making processes may be more costly, inducing greater demand for working memory and cognitive control, and therefore will be less favored due to the limited cognitive resources that agent is equipped with.

The constraint of cognitive effort can be further decomposed into two components, namely, **partner tracking** and **construction selection**. The idea behind partner tracking is that the communicative policy needs to be partner-specific (Hawkins et al., 2023; Kleinschmidt and Jaeger, 2015). This means that the speaker should adopt a distinct policy based on the identity of the specific partner they interact with. Such a partner-specific policy is especially important for the bilingual population in a language contact scenario, since the policy effective for monolingual speakers of L_{A} does not work on monolingual speakers of L_{B}. In addition to keeping track of the partner's identity, the second part of the cognitive effort results from construction selection. Specifically, given a communicative partner, the agent needs to select a specific construction to encode the intended meaning. Higher uncertainty in selecting the appropriate construction increases the cognitive demand for decision-making (Hick, 1952; Kuperberg and Jaeger, 2016; Fan, 2014; Zénon et al., 2019; Zheng and Meister, 2025). If the agent can consistently rely on a specific construction to convey a particular meaning to a given partner, the selection process becomes more automatic, thus reducing cognitive effort.

Importantly, in policies that achieve a constant level of communicative success, partner tracking and construction selection compete with each other for limited cognitive resources during bilingual-to-bilingual communication, resulting in a trade-off between the two.[7]

For the communication with monolinguals, bilinguals always need to track the monolingual's identity to select an appropriate construction under the pressure for communicative success. Moreover, communication with different monolinguals results in exposure to a mixed selection of constructions from both languages, leading to a marginal distribution with an irreducible amount of uncertainty.

For the communication between bilinguals, since they know both languages, language choice does not affect communication success anymore, and a trade-off between partner tracking and construction selection now occurs. When agents attempt to minimize attention to partner identity they will revert to policies which reflect a mix of the languages they've been exposed to when they can do so without impacting communicative success. However, a mixture of L_{A} and L_{B} has a high uncertainty over construction selection since constructions from both languages are valid for each meaning. Attempting to reduce this uncertainty in all policies means that bilinguals must devise a third, separate, conventionalized communication system with one another. This creation of a conventional system means the bilinguals must be attentive of construction selection even among each other, thus increasing the load on partner tracking.

Therefore, partner tracking and construction selection are in a trade-off in the bilingual-to-bilingual policies: increasing effort in partner tracking reduces the effort needed to address the uncertainty in construction selection, and vice versa. Given this trade-off, no policy can simultaneously minimize both partner tracking and construction selection while maintaining communicative success. Instead, the agent needs to figure out an optimal balance.

To sum up, at the individual level, the communicative policy followed by an agent is under the constraints of communicative success and cognitive effort, with latter being further decomposed

[7]There is actually also a trade-off between cognitive effort and communicative success, such that a policy that achieves higher communicative success may require greater cognitive effort. However, this trade-off is not the main focus of the current investigation, and the constraint of communicative success, given our setup, primarily impacts how the bilingual agent speaks with monolinguals (see Appendix B).

into partner tracking and construction selection. To find an optimal policy, the agent seeks to minimize the cognitive effort by striking a balance between the effort for partner tracking and the effort for construction selection.

3.2.2 Mathematical Formalization

For each listener agent a and meaning m, a speaker agent has a policy $p(c \mid m, a)$. As mentioned above, we are interested in a tradeoff between partner tracking and construction selection while ensuring communicative success. We model this with the following objective function to be minimized for the communicative policy for speaker agent a^*, consisting of three terms reflecting partner tracking, construction selection, and communicative success:

$$\begin{aligned}\mathcal{J}_{a^*} = &\underbrace{\mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c \mid m,a\right)\log\frac{p\left(c \mid m,a\right)}{p\left(c \mid m\right)}\right]}_{\text{partner tracking; } I(a;c \mid m)} \\ &+ \alpha \underbrace{\mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c \mid m,a\right)\log\frac{1}{p\left(c \mid m,a\right)}\right]}_{\text{construction selection; } H(c \mid m,a)} \\ &+ \beta \underbrace{\mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c \mid m,a\right)\log\frac{1}{q_a\left(m \mid c,a^*\right)}\right]}_{\text{communicative success; } \mathcal{L}(p)},\end{aligned} \tag{1}$$

where q_a is the decoder policy for the listener agent a used with speaker agent a^* derived from its encoder via Bayes' rule.[8] The scalar α governs the tradeoff between the two components of cognitive effort: partner tracking and construction selection. The relative weight of cognitive effort compared to communicative success is governed by scalar parameter β. However, we are not primarily interested in the behavior of these interactions, so for our purposes β will remain fixed at a value ensuring that communication systems bilinguals share with monolinguals do not degrade. More information on this choice can be found in Appendix B.

Here in Equation 1 we have used information-theoretic quantities $I(a; c \mid m)$ and $H(c \mid m, a)$ to represent cognitive effort. This is in line with recent work in neuroscience, cognitive science, and psycholinguistics which has constrained the complexity of policies in this way (Tishby and Polani, 2011; van Dijk and Polani, 2013; Genewein et al., 2015; Zaslavsky et al., 2018; Gershman, 2020; Futrell, 2021; Lai and Gershman, 2021; Futrell, 2023).

The first term of Equation 1 is the mutual information between the variable a and the variable c given m or $I(a; c \mid m)$. It represents how much information an agent must use about its interlocutors in determining which constructions c to select given its policy. The second term in Equation 1 is the conditional entropy $H(c \mid m, a)$, a value which represents, among other things, the uncertainty inherent in the distribution $p(c \mid m, a)$. The final term $\mathcal{L}(p)$ represents the communicative success expected given policy p.

What we will seek to answer in this investigation is how this policy varies with respect to α between 0 and 1. The behavior on the extremes are relatively easy to predict. For $\alpha = 0$, the objective reduces to

$$\mathcal{J}_{a^*} = I(a; c \mid m) + \beta\mathcal{L}(p), \tag{2}$$

and we would predict speakers will attempt to choose policies that are close to the marginal distribution $p(c \mid m)$, that is, a policy which does minimal partner tracking to achieve the desired level of communicative success. For $\alpha = 1$ we instead get

$$\begin{aligned}\mathcal{J}_{a^*} &= I(a; c \mid m) + H(c \mid m, a) + \beta\mathcal{L}(p) \\ &= H(c \mid m) + \beta\mathcal{L}(p),\end{aligned} \tag{3}$$

which means that speakers choose as deterministic a policy as possible when $\mathcal{L}(p)$ is not impacted (see Appendix A for the full derivation of Equation 3). At these extreme ends we will see that social demography does not matter. However, we will also see that for intermediate values of α, the speaker's policy is heavily mediated by the social structure of their environment.

4 Procedure

With the agent communication structures and objective functions defined we are able to calculate optimal policies using gradient descent. We investigate whether the resulting policies that bilinguals share with each other entrench constructions from one or the other language, or whether the policies show freedom to choose between constructions from L_{A} and L_{B}. The former resembles creolization in the feature pool model, where bilinguals select mixes of features from both languages (Mufwene, 2004). The latter resembles code-switching, with agents

[8] For monolinguals the encoder is fixed as language L_{A} or L_{B} and does not update (see Figure 2 for a depiction of these policies).

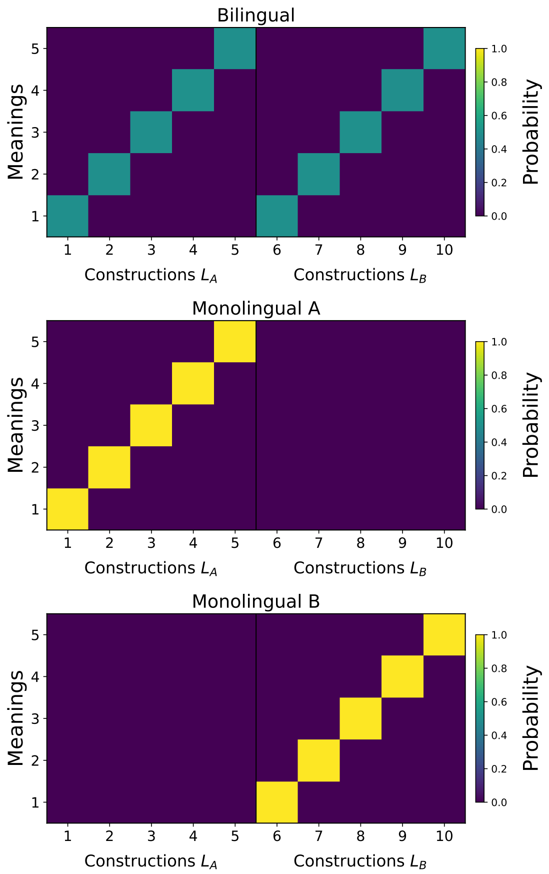

Figure 2: Model initialization: all bilingual agents are initialized to perform code-switching, where the agent has equal probability to choose constructions from L_A or L_B given a meaning m. Monolingual agents are initialized to use only one language.

being able to freely use constructions from either language. We assume for sake of simplicity that the monolingual populations are so large that they resist changing in response to contact, but we leave the bilingual population policies free to update.

Model initialization. We set the number of meanings $|\mathcal{M}|$ to 5. Three agents are initialized: the monolingual L_A speakers, the monolingual L_B speakers, and the bilingual speakers. Monolingual speakers were initialized with language policies reflecting their language, either L_A or L_B, which were disjoint mappings from meanings to constructions. Bilingual speakers were initialized with mappings reflecting a uniform mixure of L_A and L_B. We show initializations for bilingual and monolingual agents in Figure 2.

Model training. During training, the bilingual agent's language policies were updated using gradient descent. Training was performed for five different values of $\alpha \in \{0.0, 0.25, 0.5, 0.75, 1.0\}$ and for five different values of $r \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. β was kept constant at a value of 10.0. Training was performed until successive improvements of the loss due to gradient descent were less than 10^{-4}.

5 Result

Our model successfully captures both creolization and code-switching, as we explain below. Figure 3 shows the optimal bilingual-to-bilingual communicative policies with varying α and bilingual-to-bilingual contact intensity r.[9] Figure 4 shows the overall cognitive effort in communicative policies, again with varying α and r. We now highlight three critical findings in our model result.

Effect of uncertainty in construction selection. First, the model result shows that as the bilingual agents prioritize partner tracking (i.e., lowering α), they become more likely to adopt code-switching as their optimal communicative policy, such that the constructions from both source languages L_A and L_B are kept active for each meaning (as seen in panels from the top rows in Figure 3). In contrast, when the bilingual agents prioritize minimizing construction selection costs (i.e., increasing α), they are more likely to develop creolization within the bilingual population: for each meaning, only one construction from either L_A or L_B is selected deterministically, resulting in a highly conventionalized use of linguistic constructions (as seen in panels from the bottom rows in Figure 3).

Effect of contact intensity. We also observed that higher contact intensity within the bilingual population leads to increased creolization. As shown in Figure 3, when r increases (indicating stronger contact among bilinguals), the optimal policy usually assigns the full—or nearly full—probability mass to a single construction. In other words, for each meaning a convention is established using source constructions from either L_A or L_B. To illustrate why this is the case we analyzed the effects of choosing constructions from L_A over L_B in conveying a single meaning m_i for varying levels of r and α. The results can be seen in Figure 4. The effect of variable r on policy optima is clearer in these charts: as r increases, the optimal policy with respect to cognitive effort gradually moves

[9] As mentioned above, β was set such that bilingual-to-monolingual communication was always accurate. However, to see how these policies appear after training, and to see the effect of varying β on them, see Appendix B.

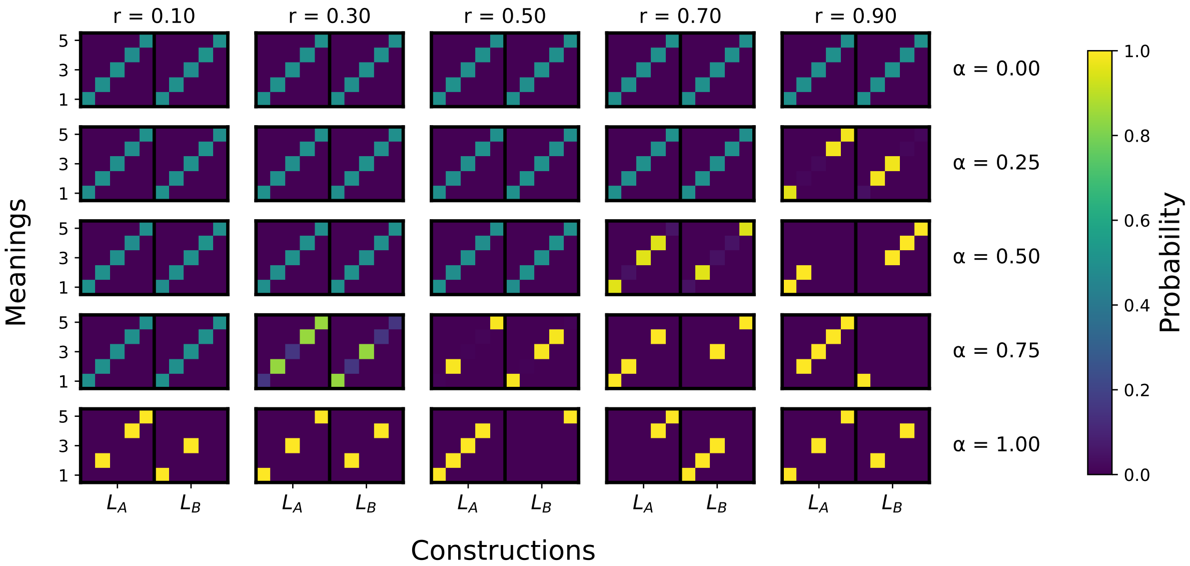

Figure 3: Optimal communicative policies $p(c \mid m)$ for bilingual agents speaking to bilingual agents, as a function of (1) the penalty α applied to the uncertainty of construction selection, and (2) the contact intensity r within the bilingual population (as opposed to bilingual-to-monolingual contact). Code-switching corresponds to a communicative policy that probabilistically alternates between the languages L_{A} and L_{B}, as found at $\alpha = 0.0$ (top row). Creolization corresponds to a policy that combines constructions from the two languages, but where the outputs are deterministic, as found at $\alpha = 1.0$ (bottom row). For intermediate values of the uncertainty penalty α, the outcome (code-switching vs. creolization) is determined by the contact intensity r.

away from an equal probability of using either language towards the highly conventionalized use of one language for a given meaning (i.e., moving away from the middle point to the sides).

For all values of α we see another general effect: higher r, which means less monolingual contact, leads to lower possible cognitive costs. Such a result can be explained by Green and Wei (2014)'s model: when the contact with monolingual population is weak, bilinguals are less likely to encounter single- and/or dual-language contexts, where cognitive control is needed to suppress the non-target language.

When $\alpha = 0$ or $\alpha = 1$, no effect of contact intensity. Our result also reveals an interaction between the bilingual-to-bilingual contact intensity and α, a model parameter that represents how the two sub-components of the cognitive constraint are prioritized. As shown in Figure 4, the contact intensity only influences optima when $0 < \alpha < 1$. When $\alpha = 0$, the bilingual population always adopts code-switching as their optimal policy regardless of their internal contact intensity. On the other hand, when $\alpha = 1$, the optimal policy is always creolization.[10] Counterintuitive as it may seem at first glance, this dynamic suggests that, on the one hand, in order for creolization to emerge, the bilingual agent must put enough effort to develop a somewhat partner-specific policy (i.e., $\alpha > 0$), regardless of the bilingual-to-monolingual contact intensity. On the other hand, in order to derive code-switching, that is, to keep the representation from both languages active for bilingual communication, the bilingual agent must to some extent ignore partner identity and put at least some effort to address a more uncertain selection over constructions (i.e., $\alpha < 1$), regardless of the bilingual-to-monolingual contact intensity.

6 Discussion

This paper presents an agent-based cognitive model aiming to capture two common phenomena in bilingual language use, namely creolization and code-switching, as two ends of a continuum of varying regularization. We set up the stage for our modeling in a language contact scenario with bilingual agents communicating among themselves and with mono-

[10]Technically, the value of α can go beyond 1. However, the result of $\alpha > 1$ will be qualitatively the same as when $\alpha = 1$, with creolization being the optimal strategy.

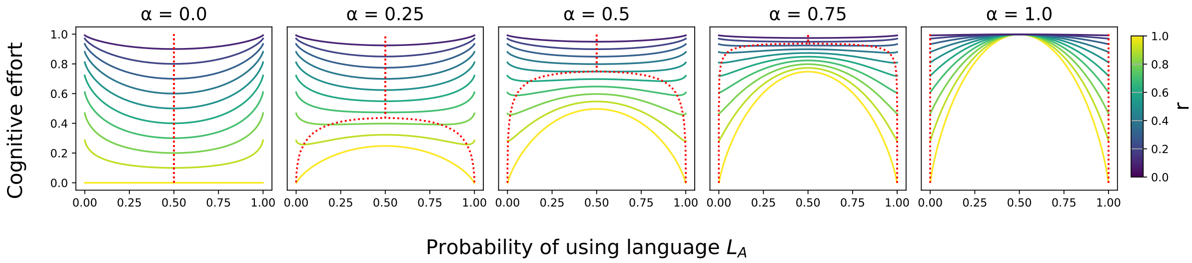

Figure 4: Cognitive effort in communicative policies for a random meaning and its two candidate constructions, with varying penalty α applied to the uncertainty of construction selection. Cognitive effort is calculated as the sum of partner tracking and the uncertainty in construction selection (i.e., the first two terms in the objective function given by Equation 1). Each line within each panel shows how the choice to use constructions from L_{A} versus L_{B} (x axis) affects the overall optimality of the cognitive effort (y axis), given the contact intensity r within the bilingual population (line color). Optimal solutions with minimal cognitive effort for each r are shown as red dotted lines. These optima were calculated for 1000 values of $r \in [0.0, 1.0)$, with only $r \in \{0.0, 0.1, ..., 0.8, 0.9\}$ shown here for ease of visualization.

lingual agents of two different languages. Within the model, the bilingual agent's communicative policy is constrained by communicative success and cognitive effort, with the latter pressure further breaking down into partner tracking and construction selection subcomponents. The bilingual agent is trained with the goal of striking a balance among these constraints by updating their language policies using gradient descent. We find that the optimal solution—creolization versus code-switching—varies with the contact intensity bilinguals have with monolinguals, but the effect varies by relative importance of the cognitive effort subcomponents.

The reason for the transition in bilingual-to-bilingual communication policies from code-switching to creolization as a function of contact intensity can be seen in Figure 4, which shows how the choice of meaning–construction mapping affects the cognitive effort (the sum of partner tracking and construction selection in Equation 1). At extremum $\alpha = 0.0$, the optima (indicated by a red line) all lie at the .5 mark, meaning that the optimal strategy to choose a construction given meaning is always an even probability between L_{A} and L_{B}, that is, a code-switching strategy. However, as α increases, a bifurcation in optima appears for critical bilingual population values. This bifurcation corresponds to the bilingual-to-bilingual contact intensity at which bilingual speakers can reduce cognitive effort by entrenching the use of L_{A} or L_{B} at the exclusion of the other. As α rises, so does this critical contact intensity value until $\alpha = 1.0$, where code-switching is never cognitively preferred. In reality, of course, both conventionalization and flexible use of constructions are observed in bilingual communities, and the critical question is more about which strategy is preferred under which condition. Therefore, we expect the actual value of α, if we are able to fit our model on some form of empirical data, would lie between these two extreme values.

From the perspective of code-switching, our model of bilingual strategy shares similarities with some other models previously proposed in the literature, especially the one in Green and Abutalebi (2013). Specifically, the model in their work tracks how different social contexts change the mode of control bilinguals may apply. Such a context-dependent control echos the partner-tracking component in our model, in the sense that the identity of communicative partner forms one aspect of the contextual information which the bilingual speaker depends on to select their optimal policy. Despite this similarity, our model differs from many others in that we aim to capture the phenomenon at a different timescale. That is, for our model, it characterizes the optimal strategy of code-switching at the population level. In contrast, many studies in the existing code-switching literature focus on the communicative strategy within each individual communicative event (e.g., the processing of a single sentence), and they often look into what specific conditions within that communicative event (e.g., grammatical configurations, processing mode, etc.) trigger or allow code-switching. This is an important question that the current study has left unaddressed, and is worth investigating in future extensions of our model.

From the perspective of creolization, unlike code-switching, the phenomenon is typically approached with a highlight on the special role of inductive bias during the language acquisition in children. The learning process in adults, in contrast, was not at the center of theoretical focus in the literature, or at least has been viewed as only playing a secondary role. Certainly, our model by no means aims to downplay the significant role of children's language acquisition in creole genesis. However, our result also suggests that, at least under certain cognitive pressures and social circumstances, creolization seems to be inevitable even though the model does not specify any acquisition process for inter-generational transmission, which is a setup typically considered pidginization. In fact, more and more psycholinguistics research have proposed that the learning mechanism in children may not be fundamentally different from the one in adults (Chang et al., 2006, 2000). It is therefore crucial to ask to what extent the effect of children's acquisition bias on creolization is fundamentally different from the effect of adults' imperfect learning, and to what extent the empirical difference between the two observed in the literature is a qualitative versus a quantitative one.

One thing which remains unaddressed is the possible typological differences creoles display (McWhorter, 1998). This is outside the scope of this model due to the model's very simple nature. Our own model assumes no preference between constructions, such that constructions from either language are equally likely to be selected by bilinguals in creolization. In other words, the model incorporates no inductive biases with respect to constructions. Explaining the common traits of creoles—for example the loss of gender and case—may require resorting to a more detailed characterization of the learnability and inductive bias of certain linguistic features, which is more sophisticated than our model can provide. To what degree creoles are typologically unique is still a matter of heated debate. We do not intend this work to stake any claim on this matter.

7 Conclusion

We successfully provide a cognitive model which when paired with the correct social environment can explain both code-switching and creolization as behavioral optima among bilinguals for certain parameterizations. How to fit such parameterizations to real bilingual behavior, or whether a single parameterization exists remain open questions. Is there a single value for α or is it dependent on other factors? This remains to be seen.

Limitations

Our model makes a number of assumptions that might seem unrealistic. We aim to address some of these in future work, but some are inherent to attempting to explain such a complex phenomenon using such a simple model.

Among the inherent assumptions are that both languages share a discrete common meaning space $\mathcal{M}$, and a discrete message space. Constructions are also considered independent of one another. Neither of these are realistic assumptions, but addressing them directly would introduce substantial complexity to the model for unclear gain and so is not yet a priority.

However, we seek to address some limitations in the future. One is that we model communities as homogeneous. In reality bilingual communities exhibit heterogeneity, with different bilinguals having different levels of proficiency in each language and maybe different cognitive resources and therefore different preferred strategies (α). Modeling this is a priority for future work and will involve changes to the social network.

The second major limitation we hope to address with future work is the communicative policy of monolingual groups, which is currently assumed to remain stable. This is also not realistic. Language contact, even of a more moderate kind, can lead to the creation of *sprachbunds* with large regions of shared linguistic features. This may be addressable in future work with a more detailed social network, as with the case of heterogeneity. However, without any source of innovation, we should expect that a community where all policies update should eventually converge to a common policy.

References

Joan Bybee. 2006. From usage to grammar: The mind's response to repetition. *Language*, pages 711–733.

Franklin Chang, Gary S Dell, and Kathryn Bock. 2006. Becoming syntactic. *Psychological Review*, 113(2):234.

Franklin Chang, Gary S Dell, Kathryn Bock, and Zenzi M Griffin. 2000. Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29:217–230.

Jin Fan. 2014. An information theory account of cognitive control. *Frontiers in Human Neuroscience*, 8:680–696.

Richard Futrell. 2021. An information-theoretic account of semantic interference in word production. *Frontiers in Psychology*, 12:672408.

Richard Futrell. 2023. Information-theoretic principles in incremental language production. *Proceedings of the National Academy of Sciences*, 120(39):e2220593120.

Tim Genewein, Felix Leibfried, Jordi Grau-Moya, and Daniel Alexander Braun. 2015. Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI*, 2:27.

Samuel J Gershman. 2020. Origin of perseveration in the trade-off between reward and complexity. *Cognition*, 204:104394.

David W Green and Jubin Abutalebi. 2013. Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25(5):515–530.

David W Green and Li Wei. 2014. A control process model of code-switching. *Language, Cognition and Neuroscience*, 29(4):499–511.

Robert D Hawkins, Michael Franke, Michael C Frank, Adele E Goldberg, Kenny Smith, Thomas L Griffiths, and Noah D Goodman. 2023. From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review*, 130(4):977.

W. E. Hick. 1952. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4(1):11–26.

Christine Jourdan. 2021. Pidgins and creoles: Debates and issues. *Annual Review of Anthropology*, 50(Volume 50, 2021):363–378.

Dave F Kleinschmidt and T Florian Jaeger. 2015. Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2):148.

Gina R Kuperberg and T Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1):32–59.

Lucy Lai and Samuel J Gershman. 2021. Policy compression: An information bottleneck in action selection. In *Psychology of Learning and Motivation*, volume 74, pages 195–232. Elsevier.

John H McWhorter. 1998. Identifying the creole prototype: Vindicating a typological class. *Language*, pages 788–818.

Salikoko S Mufwene. 1996. The founder principle in creole genesis. *Diachronica*, 13(1):83–134.

Salikoko S Mufwene. 2002. Competition and selection in language evolution. *Selection*, 3(1):45–56.

Salikoko S Mufwene. 2004. *The ecology of language evolution*. Cambridge University Press.

Salikoko S Mufwene. 2011. Creolization is a social, not a structural, process. In *Degrees of Restructuring in Creole Languages*, pages 65–84. John Benjamins Publishing Company.

Salikoko S Mufwene. 2020. Language shift. *The International Encyclopedia of Linguistic Anthropology. Edited by James Stanlaw. JohnWiley & Sons, Inc. DOI*, 10:9781118786093.

Pieter Muysken. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge University Press.

Pieter Muysken. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition*, 16(4):709–730.

Carol Myers-Scotton. 1997. *Duelling languages: Grammatical structure in codeswitching*. Oxford University Press.

Shana Poplack. 1980. Sometimes I'll start a sentence in spanish y termino en espanol: toward a typology of code-switching. *Linguistics*, 18:581–618.

Florencia Reali and Thomas L Griffiths. 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3):317–328.

Sarah Grey Thomason and Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Univ of California Press.

Naftali Tishby and Daniel Polani. 2011. Information theory of decisions and actions. In *Perception–Action Cycle*, pages 601–636. Springer.

Sander G. van Dijk and Daniel Polani. 2013. Informational constraints-driven organization in goal-directed behavior. *Advances in Complex Systems*, 16(2–3):1350016.

Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. 2018. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942.

Alexandre Zénon, Oleg Solopchuk, and Giovanni Pezzulo. 2019. An information-theoretic perspective on the costs of cognition. *Neuropsychologia*, 123:5–18.

Jieyu Zheng and Markus Meister. 2025. The unbearable slowness of being: Why do we live at 10 bits/s? *Neuron*, 113(2):192–204.

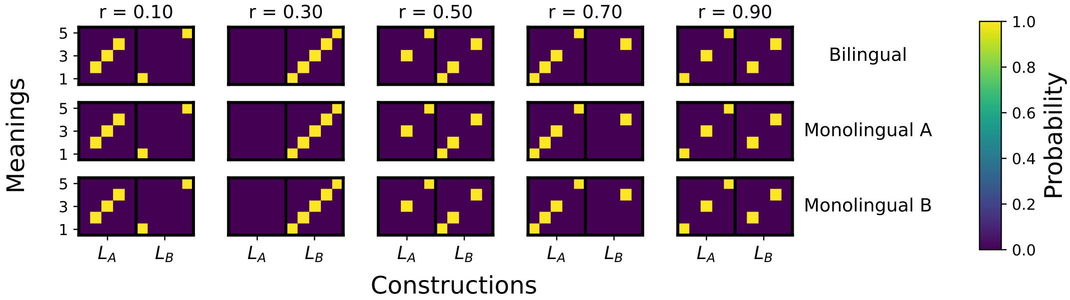

Figure 5: Optimal policies ($\beta = 0.01$) of the bilingual agent when communicating with different types of partners.

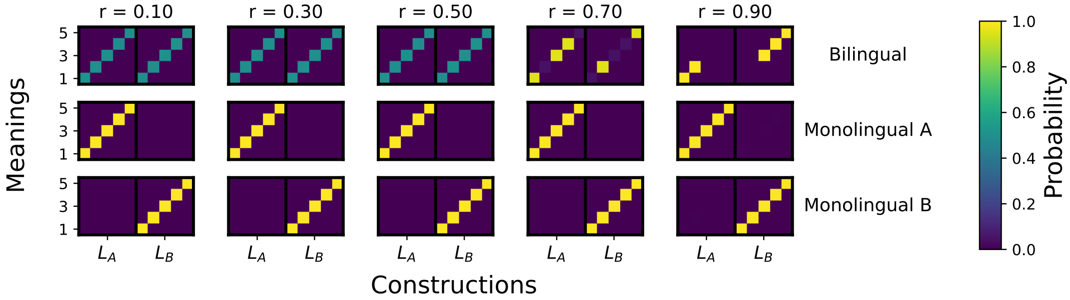

Figure 6: Optimal policies ($\beta = 10$) of the bilingual agent when communicating with different types of partners.

A Deriving Equation 3

Unlike in the $\alpha = 0$ case it might not be immediately obvious how setting $\alpha = 1$ results in Equation 1 becoming Equation 3 and so we figured it would be reasonable to devote a little space explaining why this happens. In fact, it is easiest to see why this is the case by merging the first two terms in Equation 1. If we do that, we get the condensed version

$$\mathcal{J}_{a^*} = \mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c|m,a\right)\log\frac{p\left(c|m,a\right)^{1-\alpha}}{p\left(c|m\right)}\right] + \beta \mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c|m,a\right)\log\frac{1}{q_a\left(m|c,a^*\right)}\right], \tag{4}$$

which we get by moving the α parameter inside of the log terms. Doing this makes it obvious that when $\alpha = 1$ the objective function is given as:

$$\mathcal{J}_{a^*} = \mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c|m,a\right)\log\frac{1}{p\left(c|m\right)}\right] + \beta \mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c|m,a\right)\log\frac{1}{q_a\left(m|c,a^*\right)}\right], \tag{5}$$

with the first term now becoming the expected value of a cross-entropy term. In fact, this can be reduced further. If we take the expectation of the first term with respect to a we get

$$\mathcal{J}_{a^*} = \underbrace{\mathop{\mathbb{E}}_{m}\left[\sum_c p\left(c|m\right)\log\frac{1}{p\left(c|m\right)}\right]}_{H(c|m)} + \beta \mathop{\mathbb{E}}_{a,m}\left[\sum_c p\left(c|m,a\right)\log\frac{1}{q_a\left(m|c,a^*\right)}\right], \tag{6}$$

since the various $p(a)$ will sum to 1 and the terms inside the first log do not depend on a. This means the first term is the conditional entropy $H\left(c \mid m\right)$.

B The effects of β

We stated above that the trade-off between cognitive effort and communicative success (manipulated through β in Equation 1) was not of primary interest in the current study. This is because, although including a pressure for communicative success is necessary to get the results we do, the manipulation of the term β, given our setup, primarily impacts how the bilingual agent speaks with monolinguals. For low β their accuracy with monolinguals does not matter, and with any $\alpha \neq 0$ the pressure to decrease entropy results in bilingual agent communication policies drifting off L_{A} and

L_B with their monolingual interlocutors the majority of the time. In other words, the consequences of entropy-reduction (creolization) become the only observed effect.

A demonstration of this phenomenon can be seen in Figure 5 and Figure 6, which show the optimal policies of bilingual speakers when interacting with different types of communicative partners. For both figures, we fix $\alpha = 0.5$, with r varying in the columns. We set $\beta = 0.01$ in Figure 5 and $\beta = 10$ in Figure 6. When β is low ($\beta = 0.01$), all policies used by the bilingual agent collapse into a single one regardless of the type of their communicative partner. When β is high ($\beta = 10$), the bilingual agent use separate policies corresponding to different types agent populations, and only innovate the linguistic system when interacting with another bilingual. Monolingual policies remain accurate only when β is high enough, and when β is set in this manner the monolingual policies remain accurate for all values of α.