Cooperation, Competition, and Linguistic Diversity*

Haiyun Chen† Leanna Mitchell‡

Version: August 20, 2018

Abstract

We develop a model that explains linguistic diversity in terms of the strategic incentives faced by linguistic groups. In our model, autonomous groups interact periodically in games that represent either cooperation, competition, or the lack of interaction. A language common to groups facilitates cooperation such as trade between them; whereas a language unique to one group affords that group an advantage in competitions such as warfare against other groups. The relative frequency of cooperation and conflict in a region provides incentives for each group to modify their own language, and therefore leads to changes in linguistic diversity over time. Our model predicts that higher frequency of cooperation relative to conflict reduces a region's linguistic diversity. Thus, a main contribution of our paper is to model strategic incentives as a cause of linguistic divergence.

*We thank Arthur Robson, Greg Dow, Tom Cornwall, Songzi Du, Dave Freeman, Curtis Eaton, Shih En Lu, Clyde Reed, Daniele Signori, and participants in SFU's ECON 900 seminars, Brown Bag Seminar, HESP Journal Club, Canadian Economics Association 2013 Meeting, the Summer School of Econometric Society in Seoul, and the 29th International Conference on Game Theory for helpful comments and suggestions. We are especially grateful to Arthur Robson and Greg Dow for their continual and excellent supervision. Mitchell acknowledges financial support from the Human Evolutionary Studies Program at SFU.

†Department of Economics, Grinnell College, Grinnell, IA, USA. Email: chenhaiy@grinnell.edu.

‡Department of Economics, Simon Fraser University, Burnaby, BC, Canada. Email: lmm7@sfu.ca.

1 Introduction

1.1 Motivation

The *Ethnologue*, a common reference for language classification, documents 7,105 human languages that are currently spoken worldwide (Lewis et al., 2013).[1] These languages are distributed very unevenly on earth.[2] Papua New Guinea, for example, accounts for only 0.3% of the world's land mass, yet it is home to 12% of all living languages. Australia, by contrast, accounts for 5.1% of the world's land area, but hosts merely 3% of its languages.[3] Further examples of regions with high linguistic diversity include sub-Saharan west Africa and south-central Mexico (Lewis, 2009).[4] Examples of areas with low linguistic diversity include northern Asia, Australia, and Brazil. In previous studies, scholars have documented relationships between geography and linguistic diversity. The most notable among those studies, from an economics perspective, is the one by Michalopoulos (2012), who finds that variance in land quality and in altitude are positively and significantly correlated with linguistic diversity.

Furthermore, human linguistic diversity changes over time. Pagel (2000) estimates that between 130,000 and 500,000 languages have been spoken on this planet, and that linguistic diversity peaked between 20,000 and 50,000 years ago. Despite this historical plurality, the number of languages is decreasing rapidly. Recent evidence suggests that we are on a trend towards the "hegemony" of a few dominant languages. It has been estimated that between 50% and 90% of the current languages will not survive to the next century (Hale et al., 1992; Austin and Sallabank, 2011), although there are also many examples of linguistic groups in the process of reviving their traditional languages (Bentahila and Davies, 1993).

A third puzzle is the diversity of trajectory between pairs of languages or dialects. A

[1]The *Ethnologue* uses the ISO 639-3 standard, which classifies languages using a three-letter coding system. The basic criteria that the *Ethnologue* uses to identify languages (as opposed to, for example, dialects) are (i) mutual intelligibility between speakers of variants of a language; and (ii) existence of a common literature or of a common ethnolinguistic identity. (Lewis et al., 2013) The criterion for including a language in the *Ethnologue* is that it must be "known to have living speakers who learned [it] by transmission from parent to child as the primary language of day-to-day communication". (Lewis, 2009)

[2]See Figure 6 in Appendix C for an illustration.

[3]These figures are the authors' calculation based on data from WolframAlpha (www.wolframalpha.com) and Lewis et al. (2013). Specifically, the world's land area is $1.4894 \times 10^8\,\text{km}^2$, that of Papua New Guinea is $462,840\,\text{km}^2$, and of Australia, $7.618 \times 10^6\,\text{km}^2$. The number of languages currently in use is 836 for Papua New Guinea and 245 for Australia.

[4]Following the precedent of Michalopoulos (2012), we use the term *linguistic diversity* to refer to the number of languages in a region. Other terms, such as "language diversity" (Nettle, 1998; Pagel, 2000), and "density of language groups" (Mace and Pagel, 1995), are also used to refer to similar though not necessarily identical concepts.

living language changes constantly, borrowing words from other languages, developing new ones, and abandoning others. Each language has many variants, or so-called dialects, over both space and time, such as Canadian English vs. Australian English, or Modern English vs. Middle English. Two languages can merge into one. When they merge within a generation or two the daughter language is referred to as a creole (Hall, 1966). A language may also split and diverge many times. All of the 1,200 modern Malayo-Polynesian languages, for example, are believed to have descended from a common linguistic ancestor (Gray et al., 2009). In the case of divergence, the rates at which pairs of languages diverge are not uniform, nor is the rate of divergence for a single pair over time. As illustrated by Figure 7, studies on the Indo-European and Malayo-Polynesian language families reveal that, while pairs of languages in the former family have diverged at a more or less common rate, languages in the latter family exhibit high variation in their rates of divergence, sometimes up to a threefold difference (Swadesh, 1952; Kruskal et al., 1971; Pagel, 2000).

1.2 Preview of the Results

In this paper, we show how, over time, cooperative and competitive incentives determine linguistic diversity. High frequency of cooperation, such as trade, has a homogenizing effect on language, thus leading to low linguistic diversity. High frequency of competition such as warfare, on the other hand, raises the value of the privacy of information, thus leading to the generation of new linguistic expressions and, over time, high linguistic diversity.

In situations involving potential benefit from cooperation between autonomous groups, the parties' ability to understand each other is undoubtedly crucial. Trade, exogamous marriage and political alliances all rely on clear communication. Perhaps less conventionally, communication is also critical in situations of conflict. Theoretical work on conflict suggests that information plays a vital role in determining the outcome of a hostile contention. In particular, there are "incentives to misrepresent information [...] specifically, each party would like to appear tougher than they really are" (Garfinkel and Skaperdas, 2007). Information regarding the strength of a group and any plans they have regarding attack or defense is particularly important. Success is more likely in these situations when a party can communicate such information within the group while keeping it indecipherable from their opponents. The invention of new terms that are understood only by the "insiders" of the group facilitates the control of information.[5]

[5]Group-specific vocabulary and other common elements of language (e.g. accents) may also serve the

In our model, groups interact periodically in pairs, in one of three types of activities: cooperation, competition, and non-interaction (hereafter referred to as a "null game"). Each group speaks a *dialect*, and two dialects are considered the same *language* if they are sufficiently similar.[6] In Appendix A, we show that the groups' expected payoffs in a game are a function of the relationship between their dialect and their opponent's dialect. In cooperation, modeled as a coordination game, group i's payoff is increasing in the commonality between i and j's dialects. This captures the intuition that coordination is more likely to be successful when the two groups can communicate in an accurate and inexpensive manner.

During conflict, modeled as a zero-sum game, i's expected payoff increases in how much it understands the opponent j's dialect, while the payoff decreases in the level of j's comprehension of i's dialect. Intuitively, when i and j are in an antagonistic relationship, the more information i has about j, the more advantaged i would be in that relationship. Since language is the most common medium of intra-group communication, a better understanding of j's dialect leads to a higher chance of obtaining j's private information. By the same token, the less j understands i's dialect, the better i would be able to protect its own information.[7]

Lastly, if two groups do not interact in a period, which we model by a "null game" where payoffs are constant in all outcomes, then the relationship between the groups' dialects has a neutral effect on linguistic change.

In the model, each group interacts with every other group exactly once per period, and the nature of each interaction (i.e. cooperative, competitive, or null) is determined by a random draw for every pair from among the three games. For a pair i, j, the probability of drawing a cooperative game is p_{ij}, the probability of drawing a competitive game is q_{ij}, and the probability of drawing a null game is $1 - p_{ij} - q_{ij}$. The null game is more likely to occur when groups are farther apart geographically.

We also assume that the groups are myopic, so that there is no need to consider

purpose of identifying insiders from "outsiders" of the group. Accurate identification of fellow members enables groups to communicate strategically sensitive information to its membership without accidentally revealing it to outsiders. Accents would be especially useful for the purpose of identifying members of groups that are closely related linguistically. We do not formally model differentiated language as enabling identification in this way, but this use of language is definitely consistent with our hypothesis that linguistic groups will choose to invent more new terms in high conflict settings. Language as an identifier is especially useful in groups large enough that not all members know each other personally.

[6] The definitions of *dialect* and *language* will be made more precise in Section 3.

[7] There is an additional, complementary benefit to a group of speaking a differentiated dialect, which we do not discuss at length or model explicitly. Differentiated language—including both vocabulary and accent—can help group members identify each other, which in turn facilitates the in-group distribution of sensitive information.

intertemporal tradeoffs. To maximize expected payoff, group i can make a costly effort to change its dialect at the beginning of a period by (i) learning parts of the dialect(s) spoken by some other group(s), and/or (ii) inventing novel linguistic expressions that are only understood by members of group i. The benefit of learning from an existing dialect is twofold: first, it boosts i's chance to succeed in a cooperative interaction with speakers of that dialect; second, it increases i's competitive edge in conflict against speakers of that dialect, because i would be more likely to decipher j's intra-group communication. Inventing new expressions for intra-group communication, on the other hand, improves i's odds of winning a conflict against all other groups, as the new expressions make it harder for all other groups to acquire information on i.

The main results we derive in this paper fall into three categories: (i) optimal linguistic change in a period game; (ii) the existence and characterization of a steady state in the infinitely repeated game, and convergence towards this state; and (iii) the steady state number of languages in a region and its comparative statics.

In any given period, it will always be optimal for a group to invent some new linguistic expressions, assuming that competitive games happen with positive probability. With regard to learning existing dialects, we show that if the probability of a non-trivial interaction (i.e. cooperation or competition) with a neighbor drops sufficiently fast with an increase in geographical distance, then it is optimal for a group to learn from the dialect spoken by its closest neighbor(s) first.

In the dynamic setting, where the period game is played repeatedly, we establish the existence and uniqueness of a steady state in which the set of groups speaking each language remains unchanged over time. We emphasize that while languages themselves continue to evolve in the steady state, *the set of groups* speaking any particular language does not. This in turn generates a unique steady state number of languages. If the initial linguistic composition is of a particular symmetric structure, we show that the number of languages in the region converges to that in the steady state. Lastly, we show that the steady state number of languages is weakly decreasing in the relative probability of cooperative vs. competitive interactions.

1.3 Main Contributions

Our paper makes three main contributions to the economics literature. First and foremost, we develop a formal economic model to explain how strategic incentives can induce language change and therefore linguistic diversity. Economic theorists have a long, albeit sporadic, interest in language. Marschak (1965) argued that the adaptability of languages to the environment in which they are used helps shape the features that they possess. Blume et al. (1993), Wärneryd (1993), and Robson (1990), among others, studied how messages in a language become associated with meaning in cooperative interactions with communication. Rubinstein (2000) characterized an "optimal" language based on criteria such as the ability of a language to identify objects and the ease by which the language is learned.[8] More recently, Blume and Board (2013b) explored the implications of language competence, and the knowledge thereof, on the efficiency of communication in common interest games. We contribute to this line of research by focusing on yet another aspect of language—the plurality of languages—and providing theoretical arguments for why such a phenomenon may be observed.

Second, our paper highlights the possibility that conflict generates a force that works against linguistic convergence. This is similar in spirit to the thesis of Blume and Board (2013a), that conflict of interest among speakers would lead people to favor a language that is vague even though more precise alternatives are available. Our model formalizes and extends the argument given by Pagel (2012), who hypothesizes that the multitude of languages exists to prevent people from understanding each other. In addition to conflict as a diverging influence on languages, our model also incorporates the traditional, more intuitive idea that cooperation fosters linguistic homogeneity. As a result, our model is more complete and produces a richer set of predictions regarding global linguistic diversity and linguistic change.

Third, this paper complements the existing literature on trade and conflict by explaining how trade and conflict may affect the diversity of languages and ethnicity. It is common in the economic literature to regard language as a facilitator of trade (Lazear, 1999) and as a cause of conflict (Esteban et al., 2012). Our theory, in contrast, suggests an alternate channel of causality, namely, that conflicts or the anticipation thereof could lead to more numerous languages.

The rest of the paper is organized as follows. Section 2 provides an overview of the stylized facts regarding linguistic diversity, existing explanations, and evidence supporting our hypothesis. While important in its own right, this section may appear digressive for some theoretically-minded readers due to its length. Such readers may skip directly to Section 3, which introduces our model in a formal setting, and refer back to the previous section for reality checks. In Section 4, we derive the optimal language change in a typical period. Section 5 introduces the dynamic setting, establishes the existence, uniqueness and characterization of the steady state, and convergence thereto. The comparative statics of the steady state are given in Section 6. Section 7 concludes.

[8] See Lipman (2003) for a good review.

2 Related Literature

2.1 Linguistic Diversity and Its Correlates

Scholars in several disciplines have documented empirical correlations between linguistic diversity and various ecological and geographic factors. Using historical data on native North American populations at the time of European contact, Mace and Pagel (1995) find a significant positive correlation between linguistic diversity and the diversity of mammal species, both of which, in turn, exhibit a pronounced negative latitudinal gradient. On a global scale, Harmon (1996) documents a positive correlation between linguistic and biological diversity. Also on a global scale, Nettle (1998) identifies climate as a key factor influencing global language distribution. In particular, he observes that areas with low rainfall and short growing season sustain fewer languages. Michalopoulos (2012) shows that contemporary linguistic diversity is related to geographic heterogeneity. He finds that variation in land quality (suitability for agriculture) and in elevation are both positively and significantly correlated with the number of languages in a particular region. Both Michalopoulos (2012) and Nettle (1998) find that average precipitation is significantly positively correlated with linguistic diversity.

There is a large and active literature in economics establishing a correlation between ethnolinguistic diversity and conflict. In this literature, measures of linguistic diversity or distance are used as proxies for differences in preferences over public goods (Esteban et al., 2012; Fearon, 2003; Desmet et al., 2012). The authors generally use linguistic data for the independent variable of interest because it is more available than data on genetic distance or on differences in preferences.

Additionally, there is a literature on the correlation between linguistic difference and trade or settlement patterns. Falck et al. (2012) find that similarity of dialect affects settlement decisions in Germany. Anderson and Van Wincoop (2004) review the literature showing that communication barriers increase the cost of trade between countries.

2.2 Existing Explanations of Linguistic Diversity

Several theories have been put forth to account for the empirical relationship between linguistic diversity and its environmental covariates. Common among these theories is a combination of isolation and drift. Linguistic divergence begins when populations of a linguistic community become isolated from one another. Drift in linguistics is a phenomenon analogous to mutation and drift in population genetics; it is random and unconscious change that can occur in language. When isolated, languages become dissimilar over time due to drift. Once the dissimilarity passes some threshold of mutual unintelligibility, those dialects would be considered different languages.

Nettle (1998) and Pagel (2000) argue that regions with favorable geographic conditions, e.g. those that are conducive to steady food supply, tend to sustain small, self-sufficient groups which seldom interact with each other. Consequently, populations in those regions are separated into various linguistic communities. In unproductive territories, on the other hand, survival demands that people cooperate on a large scale, and so language there will likely be uniform.

Boyd and Richerson (1988) point out another channel through which geography may increase isolation and therefore linguistic diversity. They maintain that high geographic heterogeneity increases migration costs, and thus facilitates isolation of different groups of population. Then, due to the force of linguistic drift, the languages of these geographically separate communities become dissimilar over time. Michalopoulos (2012) also theorizes that geographically heterogeneous regions foster location-specific human capital that is not easily transferable to a different environment. As a result, possession of location-specific human capital contributes to the immobility of language groups. Via a mechanism similar to the one in Boyd and Richerson (1988), then, this immobility leads to a higher linguistic diversity within the region. Michalopoulos (2012) also notes that, as geographically uniform territories are easier to conquer, and invasion has a homogenizing effect on language and culture, one would expect to see a positive correlation between linguistic diversity and geographic heterogeneity within a region.

Closest to our line of thinking is a short article by Pagel (2012). Pagel speculates that conflict over resources must be part of the reason why there are so many languages. He further mentions that humans are highly attuned to differences in speech, which helps them to identify the social group(s) to which individuals belong.

2.3 Strategic Incentives as a Determinant of Linguistic Diversity

We argue that changes in linguistic diversity are, at least partially, influenced by strategic incentives. Over many generations, strategically induced changes may accumulate and generate new languages. We propose a causal channel for such linguistic change: a region's geographic make-up (e.g. climate, terrain, soil, vegetation, natural resources, bodies of water, etc.) affects the relative probability of cooperative vs. competitive interaction. This relative probability in turn determines the strategic incentives that drive linguistic change. Our theory therefore differs from the ones reviewed in the previous subsection by highlighting strategic considerations as an important causal factor.[9]

The relationship between geography and conflict has received much attention from geographers and political scientists. Geographic variables that are found to affect regional levels of conflict include general resource abundance (Kratochwil, 1986; Fairhead, 2000; Le Billon, 2001; Renner, 2002; Smillie and Forskningsstiftelsen, 2002); high value resources such as gems, fuel, or narcotics (De Soysa, 2002; Fearon and Laitin, 2003; Buhaug and Lujala, 2005; Buhaug et al., 2009); presence of mountains (Fearon and Laitin, 2003; Buhaug and Rød, 2006); forest cover (de Rouen and Sobek, 2004); and whether there is a rainy season (Buhaug and Lujala, 2005).[10]

There are several ways in which regional geography affects potential gains from trade. Factors such as climate and soil differ across sites in a region, generating comparative advantages in the production of particular goods. The magnitudes of these comparative advantages, in turn, affect the potential for gains from trade. Secondly, geography affects the transaction costs of trading. The costs of travel, and therefore the costs of trade, depend on geographic variables including ruggedness of terrain, availability of freshwater, and the location of water bodies.

There are abundant examples of cooperative incentives resulting in linguistic change. During the colonial period, many linguistic groups sought to develop new trading relationships with each other. As a result, hundreds of pidgin languages emerged (Hall, 1966). Pidgins are informal linguistic hybrids that arise when two groups of formerly isolated peoples need to communicate extensively with one other. In subsequent generations, a pidgin may become a creole, which has a more formal set of vocabulary and grammatical rules and serves as a first language for many people. Some examples of creoles are the Chavacano language in the Philippines, Krio in Sierra Leone, and Tok Pisin in Papua New Guinea (Hall, 1966). Cooperative incentives can also cause linguistic change on a much smaller scale. The English words "ranch", "alligator", and "barbeque", for example, were borrowed from Spanish with only slight modification (Simpson et al., 1989).

There is a great deal of anecdotal evidence that conflict can lead to the invention of new words in the short run, and to linguistic divergence in the long run. As an example of the first case, Kulick (1992) reports that a group of Selepet speakers in Papua New

[9] While we agree that some linguistic change is random, i.e. that there is linguistic drift, we have not included this force in the current version of the model in the interest of simplicity. Degree of isolation, on the other hand, does play a role.

[10] To our knowledge, no empirical study has been done to examine the effect of local variation in territorial quality on local conflict. We expect that if such a study were done, it would find a positive relationship between these two variables. Such evidence would link our theory to the empirical findings of Michalopoulos (2012); i.e. that variance in land quality is an important determinant of linguistic diversity.

Guinea changed their word for "no" from *bia* to *bune* to be distinct from a neighboring group. He also observed another Papua New Guinean linguistic group switching all of their masculine and feminine words—the words for mother and father, for example, switched in meaning. Kulick states that "people everywhere use language to monitor who is a member of their tribe." It is very difficult to acquire the ability to speak a non-native language, λ, perfectly. If a person manages to do so, then she is very likely to have a close relationship with members of the group whose native language is λ.

Most social groups innovate linguistically, as exemplified by nicknames, slang, and "inside" jokes. Such terms can be useful in situations of inter-group conflict: for example, teenagers vs. parents, police vs. criminals, high school A vs. high school B, or government vs. insurgents. In these cases, a group can gain a strategic advantage by using newly invented expressions to identify group membership and communicate without revealing information to outsiders. A well documented and extensive case of this type of linguistic invention began in East London during the 1840s (Partridge et al., 2008). Speakers of the English dialect Cockney began generating "rhyming slang" which is relatively easy for insiders to learn but incomprehensible to outsiders. Various hypotheses have been proposed to explain the appearance of Cockney rhyming slang. These include linguistic accident, a form of amusement, to confuse the authorities, or, as Hotten (1859) believed, by "chaunters and patterers", i.e. street traders, possibly to assist with collusion. Developing some group-specific language naturally facilitates the control of information, and thereby generates a competitive advantage. In the case of Cockney rhyming slang, the police and customers were both engaged in competitive interactions with speakers of Cockney in the time period during which the slang emerged.

On a larger scale, there are many examples of distinct languages that have emerged in high conflict situations. "Thieves' Cant", a term that refers collectively to dozens of dialects of English which arose during the 17th century in Great Britain, was primarily spoken by criminal groups (Coleman, 2008). The Middle East (Goldschmidt and Davidson, 1991), the northern border of Italy (Kaplan, 2000), Nigeria (Osaghae and Suberu, 2005; Suberu, 2001) and Papua New Guinea (Johnson and Earle, 2000) are all examples of regions with both long histories of conflict and according to the *Ethnologue*, very high current linguistic diversity.

There are also many cases where groups in a competitive environment invest resources in maintaining a distinct language. An important category is the revival of indigenous languages, such as Halkomelem in Canada (Galloway, 2007), Welsh in Wales (Aitchison and Carter, 2000), Xibe in China (Jang et al., 2011), and Basque in Spain (Gardner et al., 2000). Minority groups will often fiercely protect their languages, as is the case for the

Quebecois in Canada (d'Anglejan, 1984) and the Catalan in Spain (Roller, 2002). When facing conflict with sub-groups, a national government will sometimes try to eliminate the sub-group's language or dialect. Some examples of such elimination include the mandatory residential school system for aboriginals in Canada in the 19$^{\text{th}}$ and 20$^{\text{th}}$ centuries (Milloy, 1999), and the establishment of the L'Académie française in 1635 with the goal of standardizing the French language (Rickard, 1989). Both policies arose during periods of internal conflict.

If drift is the only force affecting rates of linguistic divergence, these rates should be roughly constant over time. The two graphs in Figure 7 from Pagel (2000), show that rates of divergence are instead highly variable over time. Rates of divergence are particularly high initially, when a pair of groups undergoes a linguistic split, and then decrease quickly, finally flattening out. Note that the scale on the y axis is logarithmic, so the change over time in the rate of divergence is very pronounced. We hypothesize that a single language would split into two during a period of conflict, when each faction found it beneficial to differentiate their language. As the pair of languages became less mutually intelligible, the incentives to differentiate them further would diminish. Furthermore, any change in underlying circumstances that happened to increase the relative frequency of cooperation vs. conflict would further decrease incentives to differentiate. In both graphs, the rate of divergence between any pair of languages clearly starts high, decreases quickly initially and then tapers off.

Furthermore, rates of divergence should be roughly the same for pairs of groups that are completely isolated from one another. An interesting fact about Figure 7 is that there is a large range of divergence rates among Polynesian languages, and a somewhat lesser range of divergence rates among pairs of Indo-European languages. The higher variance in divergence rates among the Polynesian languages is consistent with strategic incentives affecting these rates. The Polynesian islands host a very large variety of political relationships, with some being intensely and chronically violent, and others almost perfectly peaceful (Younger, 2008).

Last but not least, as Chen (2017) reports, in a randomized and controlled experimental setting, linguistic diversification is observed when subjects interact in competitive zero-sum games and linguistic convergence emerges when subjects interact in coordination games. These findings provide strong support for the theoretical predictions made in this paper.

3 Model

In our model, groups residing in a region interact myopically in every period. There are three types of interactions—cooperative, competitive, and null—which we use to model cooperation, competition, and no interactions, respectively. The probability of each type of interaction occurring is determined by the geographic environment, which is captured by a fixed, exogenous, region-wide parameter, and the distance between the interacting groups. The payoffs of the interactions depend on the groups' understanding of each other's dialects. Therefore, at the beginning of a period, the groups, anticipating the type of interactions they may be involved in, choose to update their existing dialect by either learning words from other groups or inventing new linguistic expressions. These linguistic changes, as well as the region's linguistic composition, in turn determine the groups' payoffs in the periodic interactions.[11] The geographic parameters are constant over time, while the linguistic variables, to be introduced in Section 3.2, are in general time varying. We use a superscript t on the linguistic variables to index time. However, in this and the next section, where it is clear from the context that the variable refers to one in a particular period t, we may suppress this time index.

3.1 Geography

A region of size normalized to 1 is populated by groups from the set $\mathscr{G} = \{1, \ldots, G\}$, where $G = 2^n$, $n \in \mathbb{N} \cup \{0\}$. Each group $i \in \mathscr{G}$ controls a site in the region. To model distances between sites, we introduce a neighborhood structure analogous to a binary tree, as depicted in Figure 1. A degree 1 neighborhood is inhabited by two groups; they are each other's closest neighbors. A degree 2 neighborhood consists of two degree 1 neighborhoods, or four groups; and a degree k neighborhood consists of two degree $k-1$ neighborhoods. Hence, there are 2^k groups living in a degree k neighborhood.

The degrees of neighborhood between two groups can be interpreted as the number of natural geographic barriers between them. For example, we can think of Sub-region 1 in Figure 1 as a common meeting ground for groups 1 and 2 (living in Sites 1 and 2), and the Region is the common place for groups 1 or 2 to meet groups 3 or 4. To get to a Sub-region, a group needs to travel a short distance, getting across a river, for example. On the other hand, to get to the Region, a group has to travel a longer distance, for instance, overcoming a big mountain range. For any two groups i and j, let $d_{ij} \in \{0, 1, \ldots, D\}$

[11] A region's linguistic composition is a description of which group speaks which language, and how much each group understands of the other groups' dialects. The term is formally defined in Definition 5.

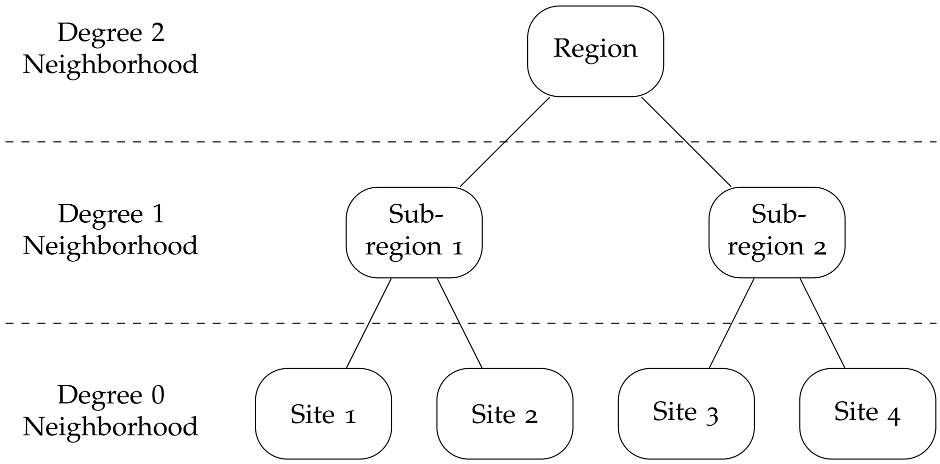


Figure 1: Neighborhood Illustration

denote the smallest degree neighborhood i and j share.[12] Then d_{ij} is a proxy for the traveling time from the site of group i to that of group j. A higher d_{ij} indicates a farther distance between groups i and j.[13] As will be clear in Section 3.3, the higher a pair's neighborhood degree, the less likely they will interact in a given period. D is the highest degree of neighborhood possible in the region (in Figure 1, for example, $D = 2$). Thus, $D = \log_2 G$, or equivalently, $G = 2^D$.

3.2 Language

For simplicity, we model a language as a list of sound-meaning pairs, which we call *linguistic elements*. Let $\mathscr{L}$ be the set of all possible linguistic elements.[14] For technical convenience, we assume that $\mathscr{L} = \mathbb{R}$.[15] Let $\mathcal{B}$ be the Borel σ-algebra on $\mathbb{R}$, and $|\cdot|$ be the Lebesgue measure on $(\mathscr{L}, \mathcal{B})$. Therefore, $(\mathscr{L}, \mathcal{B}, |\cdot|)$ forms a measure space.

Definition 1. A *dialect* spoken by group i at the beginning of period t is a non-empty, measurable (with respect to $|\cdot|$) subset $L_i^t \subset \mathscr{L}$ with a finite measure $|L_i^t| < \infty$. Let $L_0^t = \mathscr{L} \setminus \bigcup_{k \in \mathscr{G}} L_k^t$ denote the set of linguistic elements that are not used in any dialect.

[12] Trivially, $d_{ij} = d_{ji}$. Since any group is a degree 0 neighbor with itself, $d_{ij} = 0 \Leftrightarrow i = j$.

[13] The binary tree neighborhood structure is a highly stylized way to model distance between groups. We choose this approach mainly for its tractability, rather than its resemblance to any geographic structure in the real world. This binary tree neighborhood structure allows us to use a single parameter, d, to keep track of both the distance between any two groups and the number of groups within a particular degree of neighborhood (2^d). The latter feature is useful in proving Lemma 1, and subsequently establishing the existence of a steady state.

[14] That is, $\mathscr{L}$ contains all possible sound-meaning pairs. For example, "perro"-dog, "haha"-dog, "cat"-dog and "dog"-dog would be four of the elements of $\mathscr{L}$. Note also that $\mathscr{L}$ is not time dependent.

[15] We do not require a notion of closeness of any two linguistic elements, however. Using an alternative assumption—letting $\mathscr{L}$ be countable—will not change our results qualitatively. It is just awkward to work with a discrete set.

Group i can choose to enrich its dialect in two ways: (i) adopting linguistic elements from other groups' dialects, i.e. from $\{L_j^t \setminus L_i^t : j \neq i\}$, or (ii) acquiring elements from L_0^t, which are not in use by any group.

Denote by E_i^t the set of linguistic elements i learns from other groups' dialects. We write

$$E_i^t = \bigcup_{k \in \mathscr{G} \setminus \{i\}} E_{ik}^t$$

where $E_{ik}^t \subseteq L_k^t \setminus L_i^t$ is the (possibly empty) set of linguistic elements adopted from L_k^t by i. Assume that each E_{ik}^t is measurable. A non-empty E_i^t indicates that group i is adopting/learning part of other groups' vocabulary. This captures, for example, the adoption of English words by Spanish as discussed in Section 2.3.

Let $N_i^t \subseteq L_0^t$ denote the set of linguistic elements i acquires from L_0^t. Again, each N_i^t is measurable.[16] We further assume that groups acquire elements from L_0^t in an uncoordinated fashion, and therefore that N_i^t and N_j^t are disjoint (up to a subset of measure zero) for any i and j.[17] The latter assumption is plausible, in that L_0^t is of an infinite measure while each N_i^t has a finite measure, and so it is unlikely that any two groups would adopt the same subset of elements from L_0^t when their actions are uncoordinated. This type of learning can be thought of as people in group i inventing new expressions for their own dialect. Thus, a non-empty N_i^t corresponds to the examples of "code" words developed by groups such as criminal organizations, the police, the Selepet, etc. when facing conflict. The emergence of Cockney rhyming slang is an example of very large N_i^t.

We consider several dialects to be variants of the same language if they are sufficiently similar, as follows:

Definition 2. A subset $\widetilde{\mathscr{G}}_{\lambda^t} \subseteq \mathscr{G}$ of groups speak the *same language* λ^t at the end of period t if $L_i^t \cup E_i^t = L_j^t \cup E_j^t$ up to a subset of measure zero for all $i, j \in \widetilde{\mathscr{G}}^t$. The language λ^t is defined as the set $\lambda^t = \bigcup_{k \in \widetilde{\mathscr{G}}_{\lambda^t}} (L_k^t \cup E_k^t \cup N_k^t)$.

Thus, two dialects are deemed the same language if they consist of the same set of linguistic elements, acquired either through inheritance from a previous generation (i.e. elements in the set L_k^t) or learning (i.e. elements in the set E_k^t). Some readers may wonder why we choose $L_i^t \cup E_i^t = L_j^t \cup E_j^t$, as opposed to $L_i^t = L_j^t$ (or $L_i^t \cup E_i^t \cup N_i^t = L_j^t \cup E_j^t \cup N_j^t$),

[16] In the remainder of the paper, whenever we refer to a subset of $\mathscr{L}$, we assume that it is measurable, unless otherwise noted.

[17] Note also that L_i^t, E_i^t and N_i^t are disjoint pairwise. The names of the sets are chosen to signify their properties: E stands for "existing", so elements in E_i^t are chosen from existing dialect (other than L_i^t); and N stands for "non-existing", so that it contains elements that are not from an extant dialect.

as the criterion of similarity between i and j's dialects. Our choice is based on two reasons. First, languages evolve continually, and so it is unlikely for two dialects—in our model as well as in reality—to be identical at any point in time. Therefore, $L_i^t = L_j^t$ as a criterion of similarity, which requires i and j's dialects to be identical up to a subset of measure zero, would be too stringent. Second, in our model, especially the dynamic part in Section 5, groups learn and invent linguistic elements in every period. In this setting, $L_i^t \cup E_i^t = L_j^t \cup E_j^t$ is the most similar two dialects can get *at the end of a given period*. Hence Definition 2 is already using a very strict criterion, given the setup of our model.

Definition 2 is also consistent with the two criteria of language classification in the *Ethnologue* (see footnote 1): common ethnolinguistic source and mutual intelligibility. Every language has many variants, and these variants contain features specific to a particular linguistic group. The degree of similarity of pairs of variants is of a continuous nature. Above some arbitrary degree of dissimilarity, variants are referred to as distinct dialects, and above a second but higher arbitrary degree of dissimilarity they are counted as distinct languages.[18] The elements in N_i^t and N_j^t represent exactly this aspect of dialects; their presence does not disqualify dialects from being counted as a single language.

Elements in N are linguistic innovations, as discussed in Section 3.3, and often considered to be new "slang". By counting English as a single language we declare Canadian English and American English, for example, to be the same language. These two variants of English, however, are by no means identical. What distinguishes them, differences in spelling, accent, language use, etc., exist partially due to the diverging force of competition. Hence, we consider it appropriate to ignore elements in N_i^t and N_j^t when judging whether groups i and j speak the same language.

Defining a language in this way allows us to determine the speakership of each language, and then count the number of languages in the region. This can be done for any period, and becomes particularly important in Section 5, when the model is extended to a dynamic setting where speakerships may change from period to period.

The cost of acquiring new linguistic elements within a period depends on the sizes of the sets acquired, E_i^t and N_i^t,[19] and is assumed to have the following (time independent) functional form:

$$C\left(|E_i^t|, |N_i^t|\right) = c(|E_i^t|) + c(|N_i^t|)$$

[18]In defining language and dialect, the *Ethnologue* makes the following comment: "Every language is characterized by variation within the speech community that uses it. Those varieties, in turn, are more or less divergent from one another. These divergent varieties are often referred to as dialects. They may be distinct enough to be considered separate languages or sufficiently similar to be considered merely characteristic of a particular geographic region or social grouping within the speech community." (Lewis et al., 2013)

[19]In this paper, we use the words "size" and "measure" interchangeably.

where $c : [0, \infty] \to \mathbb{R} \cup \{\infty\}$ is a strictly increasing and strictly convex function with $c(0) = 0$.[20] Here we assume that there is no complementarity between the costs of the two modes of acquisition. Learning another language and inventing new expressions occur in very different contexts. Learning happens in an inter-group environment, in which multiple groups must spend time together; whereas inventing occurs in an intra-group context, where members of a single group spend time together inventing novel linguistic expressions. We assume the social and mental resources required for engaging in these activities are different enough to justify a zero cross-partial derivative.[21]

To make our analysis more tractable, we assume that c is quadratic:

$$c(|\cdot|) = \frac{1}{2}|\cdot|^2, \tag{1}$$

where $|\cdot|$ is the measure of a set. Convexity represents the increasing marginal costs that are likely to occur with such activities.

3.3 Strategic Interactions

Our model is set in the context of small scale farming or foraging economies. Pairs of groups in the region interact periodically in one of three activities: cooperation, competition, or non-interaction.[22]

A common example of a cooperative interaction is a situation in which there are potential gains from trade. The coastal First Nations of British Columbia, for example, often traded oolichan oil as far as 300 miles into the interior of the continent in exchange for commodities such as copper, flint, dried meat, and furs (Phinney et al., 2009). Accurate communication is very helpful in generating surplus in such transactions. Larger surpluses are possible if traders can easily and accurately communicate such information as when and where they will meet, and what the demand and supply are likely to be for each good.

Formally, we model a cooperative interaction as a pair of groups trying to coordinate on accomplishing some task using language. The expected payoff from a cooperation is increasing in the measure of the intersection $(L_i \cup E_{ij}) \cap (L_j \cup E_{ji})$.[23] The intuition is

[20] Since the Lebesgue measure maps $\mathcal{B}$ onto $[0, \infty]$, it follows that the image of $|\cdot|$ is the entire non-negative half of the extended real line.

[21] We do not believe that relaxing the additive separability assumption would change our results in a qualitative way. At the very least, we conjecture that our conclusions would still hold if we allow for a sufficiently small cross partial between the costs in the two modes of acquiring new linguistic elements.

[22] This subsection deals with what happens in a typical period t, so we suppress the superscript t on the linguistic variables.

[23] See Appendix A for an explanation.

that, since i and j use language to coordinate their actions, the more linguistic elements they have in common, the more accurate their communication will be, and consequently the better the two groups will perform in the cooperative interaction. As we show in Appendix A, i's expected payoff from the cooperative interaction has the following reduced form:

$$u_i(E_i, E_j) = |(L_i \cup E_{ij}) \cap (L_j \cup E_{ji})|. \tag{2}$$

Observe that N_i and N_j do not affect u_i. It would thus be equivalent to define u_i as $|(L_i \cup E_{ij} \cup N_i) \cap (L_j \cup E_{ji} \cup N_j)|$, since $N_i \cap (L_j \cup E_{ji} \cup N_j) = \varnothing$.

A competitive game represents a situation where the two groups are antagonistic towards each other, for example when they are competing over the use of resources.[24] We formally model the competitive game as a form of zero-sum game. The role of language in this context is to communicate *within* a group itself, for example for group i to organize a show of strength, an attack, or a plan for defense. With positive probability, i's within-group communication may be intercepted/overheard by someone from group j. If members of group j know a large portion of the elements in i's dialect, there is a high probability that an intercepted communication will be understood and used to group i's disadvantage. We show in Appendix A that the expected payoff from the competitive game has the following reduced form:

$$v_i(E_i, N_i, E_j, N_j) = \beta \left[|(L_i \cup E_i \cup N_i) \setminus (L_j \cup E_j)| - |(L_j \cup E_j \cup N_j) \setminus (L_i \cup E_i)| \right]. \tag{3}$$

This is the difference between how much of i's dialect is private from j, and how much of j's dialect is private from i, weighted by β. The first term in the square brackets captures i's ability to conceal information from j, and the second term reflects j's ability to conceal information from i. Therefore, i's expected payoff from the competitive game is increasing in the former and decreasing in the latter.[25] The parameter β represents the relative magnitude of competitive payoffs vs. cooperative payoffs. Note also that $v_i(\cdot) = -v_j(\cdot)$.

The null game can be interpreted as a scenario in which i and j do not interact. When a pair of groups play a null game, each gets a payoff of zero with certainty.

For any matched pair i and j, the probability of a cooperative interaction occurring is p_{ij}, a competitive game q_{ij}, and a null game $1 - p_{ij} - q_{ij}$. These probabilities are related to the distance between i and j, d_{ij}, and the geography of the region through a parameter

[24] See, for example, Dawson (1891). The real world counterparts to this game, however, need not involve actual violence.

[25] Notice that the functionality distinction is only relevant in a cooperative game. Thus, even if both i and j learn the same (measurable) subset of elements from a third group k, knowledge of that subset of elements does not affect the payoffs of a competitive game between i and j.

r which summarizes the region's geographic conditions.[26] Let the relative probability of the cooperative vs. competitive game in a given period be

$$\frac{p_{ij}}{q_{ij}} = r, \tag{4}$$

where r is exogenous and constant throughout the region. A high r could mean high complementarity of between the resources in the region, so that groups need to cooperate in order to produce final goods. The fraction p_{ij}/q_{ij} is therefore the same for all pairs $i, j \in \mathscr{G}$. Let $\pi : \{0, \ldots, D\} \to [0, 1]$ be a function that relates the geographic distance between i and j, d_{ij}, to the probability that these two groups will interact either cooperatively or competitively, i.e. $p_{ij} + q_{ij}$. Then, the probability of i and j playing a null game, i.e. they do not interact in a period, is

$$1 - p_{ij} - q_{ij} = 1 - \pi(d_{ij}). \tag{5}$$

We assume that $\pi(\cdot)$ satisfies the following two properties:[27]

$$\frac{\pi(d)}{\pi(d+1)} \geq 2^d, \qquad \forall d \in \{0, \ldots, D\} \tag{6}$$

$$\pi(0) = 1. \tag{7}$$

Property (6) implies that $\pi(\cdot)$ is a decreasing function. Thus, the farther apart two groups are geographically (i.e. the larger d_{ij} is), the less likely they will interact with each other. Furthermore, property (6) requires that the probability of interaction between any two groups drops sufficiently fast as the distance between them increases. Specifically, we assume that the probability of a group interacting with one degree d neighbor is greater than the sum of the probabilities of interacting with all of its degree $d+1$ neighbors.[28] This assumption is necessary for the proof of Lemma 1, which shows that i always prefers to learn, when possible, from a closer neighbor's dialect than from a farther neighbor's.[29]

[26] Our theory does not depend critically on the relationship between r and any specific geographic variables, as long as it is determined jointly by a set of relevant geographic factors, and is roughly constant throughout the region.

[27] An example would be $\pi(d) = \gamma 2^{-d(d-1)/2}$, where $\gamma = \pi(1) \leq 1$.

[28] Note that property (6) can be written as $\pi(d) \geq 2^d \pi(d+1)$, where 2^d is the number of degree $d+1$ neighbors of a group.

[29] As will be clear in the next section, the marginal benefit of learning from a dialect is directly related to the probability of interacting with the group that speaks it. Property (6) basically ensures that learning a word only known by one degree d neighbor is more useful than learning a word known by 2^d degree $d+1$ neighbors. If property (6) does not hold, then it becomes extremely difficult to characterize an equilibrium in this problem, as equilibria would then depend on the initial composition of dialects—the pattern of their

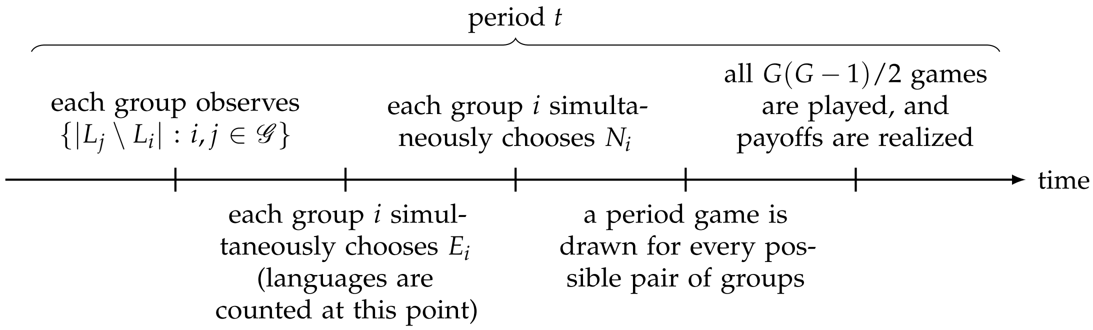


Figure 2: Timing of events in a typical period t

Property (7) can be loosely interpreted as that a group always "interacts with itself". This is a requirement mainly for technical purposes.

Together, (4) and (5) imply that

$$p_{ij} = \frac{r}{1+r}\pi(d_{ij}) \qquad \text{and} \qquad q_{ij} = \frac{1}{1+r}\pi(d_{ij}). \tag{8}$$

The parameter r and the set of pairwise distance measures, $\{d_{ij} : i,j \in \mathscr{G}\}$ are exogenous and constant over time. They are also common knowledge. Within a given period, each group plays a total of $G-1$ games, one with every other group in the region. In every period, therefore, a total of $G(G-1)/2$ games are played.

The timing of events within a period, illustrated by Figure 2, is as follows:

i) At the beginning of a period, each group observes $\{|L_j \setminus L_i| : i,j \in \mathscr{G}\}$. In other words, each group knows the measure of the set that exists to be learned from every other group's dialect.
ii) $\{E_i\}_i$ are chosen simultaneously.
iii) $\{N_i\}_i$ are chosen simultaneously.
iv) A period game is then drawn for every possible pair of groups, according to the probabilities described in (5) and (8).
v) Lastly, all $G(G-1)/2$ games are played, and payoffs are realized.

In Section 4, we derive a group's optimal decisions, N_i^* and E_i^*, for a typical period. In Section 5, we examine the long run implications of the short run results.

pairwise intersections.

4 Short Run Results

Our first result establishes the optimal size of N_i for every group i. We show that each N_i^*, the optimal subset of linguistic elements acquired by group i from L_0, has the same measure.

Proposition 1 (Optimal size of N_i)**.** *For $i \in \mathcal{G}$, we have $|N_i^*| = |N^*|$, where*

$$|N^*| = \frac{\beta}{1+r} \sum_{k=1}^{D} 2^{k-1} \pi(k). \tag{9}$$

Proof. From (2) and (3), the *ex ante* expected payoff of group i is

$$\begin{aligned} U_i(\cdot) = &\sum_{j \neq i} p_{ij} \left[|(L_i \cup E_{ij}) \cap (L_j \cup E_{ji})| \right] - c(|E_i|) \\ &+ \sum_{j \neq i} \beta q_{ij} \left[|(L_i \cup E_i \cup N_i) \setminus (L_j \cup E_j)| - |(L_j \cup E_j \cup N_j) \setminus (L_i \cup E_i)| \right] - c(|N_i|). \end{aligned} \tag{10}$$

Observe that only the second line involves N_i. Thus group i's maximization problem is

$$\max_{|N_i|} \sum_{j \neq i} \beta q_{ij} \left[|(L_i \cup E_i \cup N_i) \setminus (L_j \cup E_j)| - |(L_j \cup E_j \cup N_j) \setminus (L_i \cup E_i)| \right] - c(|N_i|).$$

Since benefit is linear and cost is strictly convex in $|N_i|$, and $c'(0) = 0$, there exists a unique maximum. The maximum is given by the first order condition where marginal cost is equal to marginal benefit.[30] Notice that marginal benefit of $|N_i|$ is constant, described by

$$\beta \sum_{j \neq i} q_{ij} = \beta \sum_{j \neq i} \frac{1}{1+r} \pi(d_{ij}) = \frac{\beta}{1+r} \sum_{k=1}^{D} 2^{k-1} \pi(k),$$

where the first equality follows from equation (8) and the last equality follows from the fact i has 2^{k-1} degree k neighbors. The marginal cost of acquiring elements is just $|N^*|$, according to (1). Therefore condition (9) is precisely the first order condition. Since the problem is symmetric for every i, it follows that (9) holds for all $i \in \mathcal{G}$. □

The intuition for this result is the standard marginal analysis in economics, as depicted in Figure 3. Acquiring linguistic elements from L_0 means inventing new expressions that no other groups but i can understand. Doing so enhances i's ability to keep secrets from all other groups, which in turn raises i's expected payoff in all of its competitive games. The

[30] Linear benefit and strictly convex cost ensures that the second order condition holds as well.

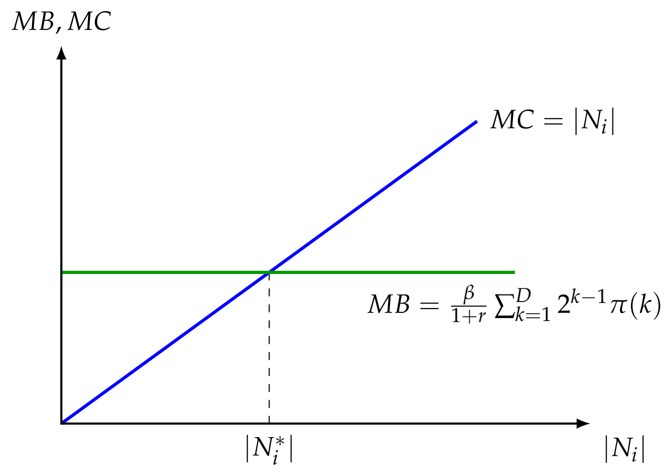


Figure 3: Optimal $|N_i^*|$

marginal benefit is therefore the sum of the probabilities of playing a competitive game with each other group in the region, weighted by the relative magnitude of competitive payoffs to cooperative ones, β. At the (interior) optimum, the marginal benefit must equal marginal cost of inventing new linguistic expressions. Observe that $|N^*|$ depends only on r, β and D, which are all exogenous parameters and constant over time. Therefore, $|N^*|$ is constant over time as well.

To obtain a similar characterization for $|E_i^*|$, we need to put some structure on the initial set of dialects, $\{L_i\}_{i\in\mathscr{G}}$. Specifically, we restrict attention to sets of dialects that are localized, as defined below.

Definition 3 (Localization)**.** The set of dialects $\{L_i\}_{i\in\mathscr{G}}$ is *localized* if the following condition is satisfied:

$$\ell \in L_i \setminus L_j \quad \Rightarrow \quad \ell \notin L_k, \qquad \forall i,j,k \in \mathscr{G} \text{ such that } d_{ij} < d_{ik}. \tag{11}$$

Localization means that if there is an element that a group j can learn from L_i, then that element cannot also be in the dialect of any of i's and j's mutually farther neighbors.[31] In other words, localization requires that whenever L_i and a distant neighbor's dialect L_k have a common element, then all the dialects of i's closer neighbors must also have that element.

Localization captures the intuition that it is very improbable for two distant languages to independently develop the same word that has the same meaning. Evidence from

[31] It is worth noting that both i and j are equally distant from k when $d_{ij} < d_{ik}$. In Figure 1, suppose i, j, k are Sites 1, 2, and 3, respectively. Sites 1 and 2 are degree 1 neighbors with $d_{12} = d_{21} = 1$, and both are degree 2 neighbors of Site 3 with $d_{13} = d_{23} = 2$. Thus, $d_{12} < d_{13}$ implies that $d_{21} < d_{23}$.

historical linguistics suggests that the elements of pre-colonial languages at least roughly satisfy the localization property. When a linguistic group splits into two, the speakers of the two new dialects naturally tend to live close together. Studies of the Indo-European languages, for example, found that languages with a more recent common linguistic ancestor—e.g. Spanish and Portuguese, both of which share many common features of Latin, their "parent language"—are also geographically close to each other (Finegan, 2008).

According to Pagel et al. (2007), "[l]anguages, like species, evolve by a process of descent with modification". Linguistic divergences are also commonly represented as trees, similar to those relating biological species (Kruskal et al., 1971; Pagel, 2000; Esteban et al., 2012). Suppose that i and j have a more recent common linguistic ancestor than i, j, and k. An element that satisfies any one of the three following conditions satisfies the localization property: i) all three retain the element from a common ancestor, ii) an ancestor of both i and j split from k and then acquired it before i and j split from each other, or iii) any element acquired by a single group after all three had split.

There is also a theoretically appealing reason for our focus on symmetric and localized sets of dialects. As we show later in Lemma 2, if the region begins with a set of symmetric and localized dialects, then the set of dialects will always be symmetric and localized. If all groups in a neighborhood of some degree $d \in \{0, 1, \ldots, D\}$ speak the same language, then the set of dialects is localized. Localization, however, holds for a much more general set of dialects.

Assuming localization enables us to establish the order in which a group i learns from existing dialects. This order is closely connected to the magnitude of the marginal benefit of learning a subset of another dialect. Let $MB_i(d_{ij}, s) = 2^{d_{ij}-s}(p_{ij} + \beta q_{ij})$ denote the marginal benefit that i receives from learning a subset of elements that is shared by $2^{d_{ij}-s}$ neighbors of degree d_{ij}, where $s \in \{1, \ldots, d_{ij}\}$.[32]

Lemma 1 (Order of Learning). *Suppose the set of dialects $\{L_i\}_i$ is localized. Then, $MB_i(d_{ij}, s)$ is decreasing in d_{ij}. As a result, it is always optimal for i to learn all of $\bigcup_j (L_j \setminus L_i)$ before learning anything from $\bigcup_k (L_k \setminus L_i)$, where j, k are such that $d_{ij} < d_{ik}$.*

[32] Localization, together with symmetry of dialects (see Definition 4), imply that within each degree d_{ij} neighborhood, $MB_i(d_{ij}, s)$ has at most d_{ij} values, and s indicates the s^{th} highest value. Take $d_{ij} = 3$ for example. Group i has four neighbors of degree 3. Symmetry and localization require that any linguistic element i learns must be either (i) commonly shared by all four neighbors, (ii) shared by two groups (who are degree 1 neighbors with each other), or (iii) unique to one of the four groups. Hence, the three possible values of $MB_i(3, s)$ would be, from the highest to lowest, $MB_i(3, 1) = 4(p_{ij} + \beta q_{ij})$, $MB_i(3, 2) = 2(p_{ij} + \beta q_{ij})$, and $MB_i(3, 3) = (p_{ij} + \beta q_{ij})$. Bear in mind that some of these steps need not exist. For example, when $L_i = L_j$ (up to a subset of measure zero) for all $i, j \in \mathcal{G}$, $MB_i(d_{ij}, s) = 0$ for all d_{ij}. If not all dialects in the region are identical, however, symmetry and localization implies that the lowest step always exists.

Proof. From Definition 3 and equation (2), it follows that after learning all the elements of all the languages where $d < d_{ij}$, learning a subset of elements $E_{ij} \subseteq L_j \setminus L_i$ is only going to affect i's payoff when it is interacting with neighbor(s) of degree d_{ij}. According to (10) and (8), therefore, given an arbitrary $d_{ij} \in \{1, \ldots, D\}$, $MB_i(d_{ij}, s)$ is at least

$$p_{ij} + \beta q_{ij} = \frac{r+\beta}{1+r}\pi(d_{ij}), \tag{12}$$

as is the case when $s = d_{ij}$, and at most

$$2^{d_{ij}-1}(p_{ij} + \beta q_{ij}) = 2^{d_{ij}-1}\left(\frac{r+\beta}{1+r}\pi(d_{ij})\right), \tag{13}$$

as is the case when $s = 1$. $s = 1$ is the case where all of i's degree d_{ij} neighbors know the element, so it will be useful in games with any of them. Observe that (12) and (13) are the same when $d_{ij} = 1$, because i only has one degree 1 neighbor. Observe also that the lowest marginal benefit of learning from a degree d_{ij} neighbor, i.e. $MB_i(d_{ij}, d_{ij})$, is higher than the highest marginal benefit of learning from a degree $d_{ij}+1$ neighbor, i.e. $MB_i(d_{ij}+1, 1)$:

$$\frac{r+\beta}{1+r}\pi(d_{ij}) - 2^{(d_{ij}+1)-1}\left(\frac{r+\beta}{1+r}\pi(d_{ij}+1)\right) \geq 0, \qquad \forall d_{ij} \in \{0, \ldots, D\}.$$

The inequality follows from (6). This completes the proof. □

Lemma 1 allows us to partition the set of i's learnable elements, $\bigcup_k (L_k \setminus L_i)$, according to the marginal benefit they confer, and hence the neighborhood degrees. Let

$$P_i(d) = \begin{cases} \{\bigcup_k \left(L_k \setminus \left(\bigcup_{k'} L_{k'}\right)\right) : d_{ik} = d \text{ and } d_{ik'} < d\} & \text{if } d \in \{1, \ldots, D\} \\ \varnothing & \text{if } d = 0 \end{cases}$$

describe the elements of this partition.

$P_i(d)$ is the set of learnable linguistic elements in the dialects of i's degree d neighbors, excluding the elements that are in the dialects of i's closer neighbors, of degrees less than d. Note that elements of i's own language are never learnable to i; since i is a degree 0 neighbor with itself, $P_i(d) \cap L_i = \varnothing$ for all d. Moreover, for any $d \neq d'$, $P_i(d)$ and $P_i(d')$ are disjoint.

It follows, therefore, that $|\bigcup_d P_i(d)| = \sum_d |P_i(d)|$. By Lemma 1, it also follows that, for any $A \subseteq P_i(d)$ and $A' \subseteq P_i(d')$ where $|A| = |A'|$ and $d < d'$, the marginal benefit of learning $|A|$ is strictly greater than that of learning $|A'|$. Moreover, each $P_i(d)$ can be further partitioned into d cells, based on how many neighbors of degree d share those elements.

By the definition of $MB_i(d_{ij,}s)$, the more neighbors sharing an element, the higher the marginal benefit that element confers. Therefore, assuming localization, we can order i's set of learnable elements by their associated marginal benefits: first by neighborhood degree d, and then by s within each $P_i(d)$. Such an ordering, together with Lemma 1, implies that as $|E_i|$ increases, marginal benefit decreases in a step-wise fashion.

Proposition 2 (Optimal size of E_i^*)**.** *Let the set of dialects $\{L_i\}_i$ be localized. The optimal size of E_i^* is contained within the following interval:*

$$|E_i^*| \in \left[\sum_{k=0}^{d^*} |P_i(k)|, \quad \sum_{k=0}^{d^*+1} |P_i(k)| \right),$$

where $d^* \in \{0, \dots, D\}$ is determined by

$$\frac{r+\beta}{1+r}\pi(d^*) \geq \sum_{k=0}^{d^*} |P_i(k)| \qquad \textit{and} \qquad \frac{r+\beta}{1+r}\pi(d^*+1) < \sum_{k=0}^{d^*+1} |P_i(k)|. \tag{14}$$

Our dynamic results in Section 5 do not require that we know the exact value of $|E_i^*|$, only the degree of neighborhood d^* within which i chooses to learn all the remaining elements of its neighbors dialects.

Proof. First, observe that a unique $|E_i^*|$ exists. The size of acquired elements from existing dialects, $|E_i|$, takes a value from a compact set $\left[0, \sum_{k=0}^{D} |P_i(k)|\right]$, on which the objective function (10) is continuous. Hence a maximum exists. Since marginal benefit is weakly decreasing in $|E_i|$ while marginal cost is strictly increasing, uniqueness of $|E_i^*|$ is ensured. Figure 4 provides an illustration.

We can verify that $|E_i^*|$ is indeed bounded by the proposed interval through the first order condition. Notice that the benefit function is an increasing, piece-wise linear function that is not differentiable at a finite number of points, namely the points at which marginal benefit makes a discrete jump downwards. Nevertheless, for each value of $|E_i|$, there exists a set of (super-)derivatives for the benefit function, bounded by the left- and right-derivatives at $|E_i|$, and the set is a non-singleton at the points where the benefit function has a kink. The first order condition requires that at $|E_i^*|$, there exists a (super-)derivative of benefit function that is equal to the derivative of the cost function (the latter of which is uniquely defined at every $|E_i|$). Moreover, it has to be true that (i) at $|E_i| = \sum_{k=0}^{d^*} |P_i(k)|$, marginal benefit is weakly higher than marginal cost; and (ii) at $|E_i| = \sum_{k=0}^{d^*+1} |P_i(k)|$ marginal cost exceeds marginal benefit. But according to (14), d^* is

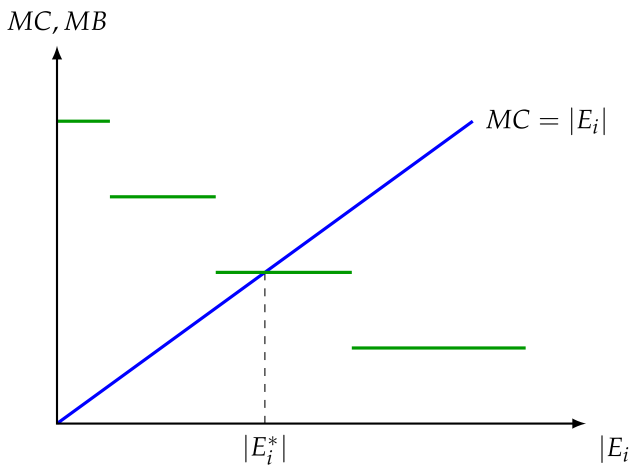


Figure 4: Optimal $|E_i^*|$

chosen such that these two conditions are simultaneously satisfied. The marginal benefit of learning the last element in $P_i(d^*)$ is described by (12); for otherwise i would have learned it sooner. The marginal cost of learning this last element is $\sum_{k=0}^{d^*} |P_i(k)|$. Therefore, equation (14) ensures that d^* is chosen such that the marginal benefit of learning the last element in $P_i(d^*)$ is higher than the marginal cost. However, the marginal cost exceeds the marginal benefit if $P_i(d^*+1)$ is fully acquired.

We know $|E_i| = \sum_{k=0}^{d^*} |P_i(k)|$ when everything up to and including $P_i(d^*)$ is fully learned and $|E_i| = \sum_{k=0}^{d^*+1} |P_i(k)|$ when everything up to and including $P_i(d^*+1)$ is fully learned. Therefore, $|E_i^*|$ must lie between these two values. □

Corollary 1. *At the optimum, group i will acquire (i) all the learnable elements from its neighbors with degree smaller than or equal to d^*; (ii) a proper subset of learnable elements from its degree d^*+1 neighbors; and (iii) no elements from its neighbors with degree greater than d^*+1.*

Proof. This follows directly from Lemma 1 and Proposition 2. □

Furthermore, if the region's dialects satisfy a symmetry condition, then Proposition 2 immediately implies that all groups learn the same measure of elements in a typical period.

Definition 4 (Symmetry)**.** The set of dialects $\{L_i^t\}_{i\in\mathscr{G}}$ is *symmetric* at t if the following two conditions are satisfied:

$$|L_i^t| = |L_j^t|, \qquad \forall i, j \in \mathscr{G} \tag{15}$$

$$|L_i^t \cap L_j^t| = |L_i^t \cap L_k^t|, \qquad \forall i, j, k \in \mathscr{G} \text{ such that } d_{ij} = d_{ik}. \tag{16}$$

Condition (15) requires that the measure of each dialect is the same. Condition (16) requires that the intersection of any dialect L_i^t with that of any equally distant neighbors has the same measure.

Corollary 2. *If the set of dialects $\{L_i\}_i$ is symmetric and localized, then the optimal size of E_i^* is the same for all groups in any given period, i.e. $|E_i^*| = |E^*|$ for all $i \in \mathscr{G}$.*

Proof. This follows directly from the properties of symmetry and localization of the set of dialects, and the symmetry of each group's decision. □

While Proposition 2 makes no requirement about which subset of the degree $d^* + 1$ neighbors' dialects a group should learn, we make two assumptions about how groups learn from their $d^* + 1$ neighbors' dialects. The first assumption ensures that the set of dialects stay symmetric and localized in the subsequent period, so that results derived in this section can be applied to analyze long-run properties of the region's languages in a dynamic context. The second assumption enables us to use the definition of same language (i.e. Definition 2) consistently in the dynamic setting.[33]

Assumption 1. When a group is indifferent between learning from several dialects, it will learn the same measure of elements from each of them; the elements learned are chosen randomly from the set of learnable elements in those dialects.

Assumption 2. All groups within a degree d^* neighborhood learn the same subset of elements from their degree $d^* + 1$ neighbors.

Notice that the characteristics of the set of learnable elements may vary from period to period. Consequently, unlike $|N^*|$, $|E^*|$ need not be the same across time.

5 Long Run

In this section, we examine the long run implications of the model. Suppose that a period represents a human generation of approximately twenty years. Each new generation of group members inherits the dialect of the previous generation, and then may choose to learn from the neighbors' dialects as well as to invent novel expressions.

We assume that groups are myopic, so that each generation is only interested in payoffs in the current period. While this assumption frees us from considering repeated game

[33] Without Assumption 2, we would have to modify the criterion for same language to a slightly more complicated version: $L_i^t \cup \left(\bigcup_{k=1}^{d^{t*}} E_{ik}\right) = L_j^t \cup \left(\bigcup_{k'=1}^{d^{t*}} E_{jk'}\right)$. The other aspects of the model are unaffected by this assumption.

effects, the groups' myopia over payoffs is not implausible. Group membership in small scale societies is fluid across generations. Most groups are patrilocal or matrilocal, so adults expect that approximately half of their offspring, at maturity, will leave for another group. This consideration reduces the incentives of the current generation to consider the costs and benefits of their choices on future generations of the group. Myopic play implies that each group's decision in any given period is characterized in Section 4. This allows us to analyze the trajectory of a region's linguistic composition over generations.

Definition 5. The *linguistic composition* in the region at time t is the set

$$\Lambda^t = \left\{ (\lambda, \widetilde{\mathscr{G}}_\lambda) : \lambda = \bigcup_{k \in \widetilde{\mathscr{G}}_\lambda} (L_k^t \cup E_k^t \cup N_k^t) \text{ where } \widetilde{\mathscr{G}}_\lambda \text{ is the set of groups speaking the same language } \lambda \right\}.$$

Recall that the criterion for "same language" is given in Definition 2. We are interested in the steady state of a region's linguistic make-up; that is, a linguistic composition that is stable in the long run. We denote such a composition $\overline{\Lambda}$, wherein each language $\lambda \in \overline{\Lambda}$ is spoken by the same subset of groups $\widetilde{\mathscr{G}}_\lambda \subseteq \mathscr{G}$ in every subsequent period.[34] Henceforth, we use a superscript t on the variables to index time. In accordance with Definition 2, we formally define a steady state as follows:

Definition 6. The region's linguistic composition is in a steady state if for all t,

$$L_i^t \cup E_i^t = L_j^t \cup E_j^t \quad \Rightarrow \quad L_i^{t+1} \cup E_i^{t+1} = L_j^{t+1} \cup E_j^{t+1} \tag{17}$$

and

$$L_i^t \cup E_i^t \neq L_j^t \cup E_j^t \quad \Rightarrow \quad L_i^{t+1} \cup E_i^{t+1} \neq L_j^{t+1} \cup E_j^{t+1} \tag{18}$$

for all $i, j \in \mathscr{G}$.[35]

A linguistic composition is in a steady state if the following two conditions hold: (i) every pair of groups i, j that speak the same language in any period t will continue to speak the same language in $t+1$; and (ii) every pair i, j that do not speak the same language in t will not speak the same language in $t+1$. Therefore, a steady state of the

[34] Note however that the languages themselves will not be the same; they will grow in size over time, according to the results in Section 4.

[35] In our model, the sizes of the languages are growing over time. Hence, the steady state of a linguistic composition is steady in the sense that each language in the composition has a stable speakership.

linguistic composition is a set $\overline{\Lambda}$ of languages, each of which is spoken by the same subset $\widetilde{\mathscr{G}}_\lambda$ of groups over time.

Recall from Section 3.3 that the order of events in a typical period t is as follows:

i) $\{|L_j^t \setminus L_i^t| : i, j \in \mathscr{G}\}$ is observed by all groups.
ii) $\{E_i^t\}_i$ are chosen simultaneously.
   - The number of languages for period t, $\#(\Lambda^t)$, is counted at this point.[36]

iii) $\{N_i^t\}_i$ are chosen simultaneously.
   - The set of dialects at the beginning of period $t+1$, $\{L_i^{t+1}\}_i = \{L_i^t \cup E_i^t \cup N_i^t\}_i$, is determined at this point.

iv) A period game is drawn for every possible pair of groups according to (5) and (8).
v) The period games are played, and payoffs are determined based on $\{L_i^t, E_i^t, N_i^t\}_i$.

For the results in Section 4 to apply in every period, it must be the case that the set of dialects is symmetric and localized at the beginning of every period. The following lemma shows that if $\{L_i^t\}_i$ satisfies symmetry and localization in the initial period, it will continue to do so in subsequent periods.

Lemma 2. *Suppose $\{L_i^t\}_i$ is symmetric and localized. Then $\{L_i^{t+1}\}_i$ is also symmetric and localized.*

The proof of this lemma can be found in Appendix B.1.

Before we state our main dynamic results, it helps to define an important value, $\overline{d}$. As we will state formally in Proposition 4 and Proposition 5, $\overline{d}$ is the threshold value of neighborhood distance. Groups with $d_{ij} \leq \overline{d}$ will speak the same language in the steady state, and groups with $d_{ij} > \overline{d}$ will speak different languages.

Definition 7. Define $\overline{d}$ as an element of $\{0, \ldots, D\}$ that satisfies the following condition:

$$\frac{r+\beta}{1+r}\pi(\overline{d}) \geq \sum_{k=0}^{\overline{d}} |P_i^t(k)| \qquad \text{and} \qquad \frac{r+\beta}{1+r}\pi(\overline{d}+1) < \sum_{k=0}^{\overline{d}+1} |P_i^t(k)|, \tag{19}$$

where

$$|P_i^t(k)| = \begin{cases} 2^{k-1}|N^*| & \text{if } k \in \{1, \ldots, \overline{d}\} \\ 0 & \text{if } k = 0 \end{cases}$$

and $|P_i^t(k)| \geq 2^{k-1}|N^*|$ if $k \geq \overline{d}+1$.

[36] $\#(\Lambda)$ is the number of elements in Λ.

The particular neighborhood distance $\overline{d}$ has some special properties. The next proposition, for example, establishes that if $d_{ij} > \overline{d}$ for a pair of groups i, j, then their languages will, in some sense, become less similar over time. This is because the size of the set that j invents every period, $|N_j|$, will always be larger than the size of the set that i learns from j, $|E^*_{ij}|$. With each passing period, the set of elements that i speaks but j doesn't will become larger. Due to symmetry, the set of elements that j speaks but i doesn't will also become larger.

Proposition 3. *If $d_{ij} > \overline{d}$ then $|E^*_{ij}| < |N_j|$. Thus, $|L_j \setminus L_i|$ increases every period.*

The proof can be found in Appendix B.2. Proposition 3 is a result regarding the trajectory of a pair of languages. From (5), $\pi(d)$ is the probability of a non-null interaction occurring between the groups. According to (6), $\pi(\cdot)$ is a decreasing function of d, so the further apart i and j are geographically, the less they interact. Proposition 3 establishes that if meaningful interactions between i, j are sufficiently infrequent, that is, if d is high enough, their dialects will become less and less similar over time. This process might look empirically similar to linguistic drift, as both types of divergence theoretically decrease with frequency of interaction.

The next proposition builds on the results of Proposition 3, establishing that if $d_{ij} \leq \overline{d}$, and i,j begin by speaking the same language, then their languages will continue to be the same. That is, in every period i will learn all of j's newly invented elements, and vice versa.

Proposition 4 (Existence and characterization of a steady state)**.** *There exists a steady state, $\overline{\Lambda}$, which is characterized by the following condition:*

$$L_i^t \cup E_i^t = L_j^t \cup E_j^t \quad \Leftrightarrow \quad d_{ij} \leq \overline{d}, \qquad \forall i, j \in \mathscr{G}, \tag{20}$$

where $\overline{d} \in \{0, \ldots, D\}$ is determined by (19).

The proof of this proposition can be found in Appendix B.3. The proposition establishes that there is a steady state of the region's linguistic composition which can be fully described by a particular degree of neighborhood, $\overline{d}$. In this steady state, all groups that are members of the same $\overline{d}$ neighborhood speak the same language, and furthermore only groups which are members of the same $\overline{d}$ neighborhood speak the same language. In other words, if $d_{ij} \leq \overline{d}$ then $L_i \cup E_i = L_j \cup E_j$, and if $d_{ij} > \overline{d}$ then $L_i \cup E_i \neq L_j \cup E_j$. Note that $\overline{d}$ depends only the exogenous parameters of the model ($|N^*|$ is also a function of those parameters). Thus, $\overline{d}$ is constant over time.

Corollary 3 (Steady state number of languages)**.** *The number of languages in the steady state is* $\#(\overline{\Lambda}) = G/2^{\overline{d}}$.

Proof. This follows directly from condition (20). Each degree $\overline{d}$ neighborhood is inhabited by $2^{\overline{d}}$ groups, and these groups share a common language in the steady state. Since there are G groups, the number of languages in the steady state is therefore $G/2^{\overline{d}}$. □

Next we turn our attention to convergence, showing that for a general set of initial conditions, the linguistic composition converges towards the $\overline{\Lambda}$ described in Proposition 4.

Proposition 5. *If* $\{L_i^0\}_{i\in\mathcal{G}}$ *is symmetric and localized, then after a finite number of periods, the steady state* $\overline{\Lambda}$ *is reached.*

The proof of this proposition is presented in Appendix B.4. This proposition says that, after finitely many periods, i will learn the entire set of learnable elements from its degree $\overline{d}$ neighbors and will therefore, by definition, speak the same language as each of them. It is worth noting the two special initial linguistic compositions where all groups have identical initial dialects or completely distinct ones both satisfy the symmetry and localization conditions.

Every period, i learns a larger set of the languages of its $d \leq \overline{d}$ neighbors than those groups collectively invent. In other words, $|E_i^*| > 2^{\overline{d}-1}|N^*|$. Suppose i does not initially speak the same language as its $d = 1$ neighbor, and $\overline{d} \geq 1$. In the first period, i begins "catching up" with its $d = 1$ neighbor, then later with its $d = 2$ neighbors, and so forth until it has learned the entire set of learnable elements from its neighbors of degree $d \leq \overline{d}$. At this point, $L_i \cup E_i = L_j \cup E_j$ holds for all of i's $d \leq \overline{d}$ neighbors.

From this point onwards, i learns at least $2^{\overline{d}-1}|N^*|$ elements every period, including all those invented in the last period by its neighbors of degree $d \leq \overline{d}$. i may learn some elements from its degree $\overline{d}+1$ neighbors, but as was shown in Proposition 3, not as many as this set of neighbors invents. For this reason, i never catches up with these neighbors.[37] From Corollary 3, $\#(\overline{\Lambda}) = G/2^{\overline{d}}$, therefore $G/2^{\overline{d}}$ is the steady state number of languages. Last but not least, since the steady state is uniquely determined by $\overline{d}$, we conclude that it is the unique steady state for the set of initial linguistic compositions that we consider.

6 Comparative Statics

Now we return to our original purpose: to show the effect of cooperative and competitive incentives on the number of languages in a region. In this section, we establish the key

[37] The divergence of i's language from those of its higher degree neighbors actually occurs immediately, as a result of its choice of E_i in the first period.

comparative static result for the number of languages in the steady state: $\#(\overline{\Lambda})$ is weakly decreasing in r, the ratio of the probabilities of cooperative and competitive interactions.

Furthermore, we examine the results of exogenous changes in the geographic locations of groups on the relationships between particular dialects. Specifically, we determine under what circumstances such a change would alter the steady state identity of the dialects of an arbitrary pair of groups.

Proposition 6. *The steady state number of languages, $\#(\overline{\Lambda})$, is weakly decreasing in r.*

In other words, the higher the regional ratio of the probability of cooperation vs. competition, the fewer languages there will be in the region in steady state.

Proof. According to Corollary 3, $\#(\overline{\Lambda}) = G/2^{\overline{d}}$. We show that $\#(\overline{\Lambda})$ is weakly decreasing in r by proving that $\overline{d}$ is weakly increasing in r. Recall that $\overline{d}$ is determined by (19). In particular, it must be the case that, for any $\overline{d} \in \{1, \ldots, D\}$, the first inequality of (19) can be rearranged as follows:

$$\begin{aligned}
\frac{r+\beta}{1+r}\pi(\overline{d}) &\geq (2^{\overline{d}}-1)|N^*| \\
\frac{r+\beta}{1+r}\pi(\overline{d}) &\geq (2^{\overline{d}}-1)\frac{\beta}{1+r}\sum_{k=1}^{D} 2^{k-1}\pi(k) && \text{by (9)} \\
(r+\beta) &\geq \frac{2^{\overline{d}}-1}{\pi(\overline{d})}\left[\beta\sum_{k=1}^{D} 2^{k-1}\pi(k)\right]. && (21)
\end{aligned}$$

Similarly, for any $\overline{d} \in \{0, \ldots, D-1\}$, the second inequality of (19) can be written as

$$(r+\beta) < \frac{2^{\overline{d}+1}-1}{\pi(\overline{d}+1)}\left[\beta\sum_{k=1}^{D} 2^{k-1}\pi(k)\right]. \tag{22}$$

(21) and (22) together imply that

$$\frac{2^{\overline{d}}-1}{\pi(\overline{d})}\left[\beta\sum_{k=1}^{D} 2^{k-1}\pi(k)\right] \leq r+\beta < \frac{2^{\overline{d}+1}-1}{\pi(\overline{d})}\left[\beta\sum_{k=1}^{D} 2^{k-1}\pi(k)\right], \tag{23}$$

where the terms in the square brackets are independent of either r or $\overline{d}$. Note that the fraction $(2^k-1)/\pi(k)$ is strictly increasing in k, because $\pi(\cdot)$ is decreasing. Thus, when the increase in r is sufficiently large, $\overline{d}$ would also have to increase in order for (23) to hold. Likewise, when the decrease in r is large enough, $\overline{d}$ needs to decrease as well to satisfy (23). This completes the proof. □

Proposition 7. *The steady state number of languages, $\#(\overline{\Lambda})$, is not affected by multiplying the probability of a non-null interaction, $\pi(d)$, by any positive constant, $\phi > 0$.*

Proof. The statement follows trivially from (23): note that a constant ϕ multiplying $\pi(\cdot)$ gets canceled out immediately. □

This result shows that an exogenous change in the frequency of meaningful interactions in the region, as long as it does not affect either the relative frequency of cooperative vs. competitive interactions, r, nor the set of ratios $\pi(d)/\pi(d+1)$, $d \in \{1, \ldots, D-1\}$, will have no effect on the steady state number of languages in the region.

7 Discussion

As a means of transmitting information between human beings, nothing can be more fundamental than language. Yet languages themselves are complex and dynamic. Understanding the mechanism behind linguistic change, therefore, is an interesting and worthwhile undertaking. In this paper, we propose that the convergence and differentiation of languages can be partly explained in terms of variations in the strategic nature of the environment in which linguistic groups reside. In our model, languages homogenize when cooperation prevails in a region, and that languages become more numerous as more conflicts arise.

Most noteworthy in our model is that divergence of languages is endogenously generated. This explanation contrasts with the conventional explanation of linguistic differentiation as a result of isolation and drift. Both theories predict that the languages of a pair of geographically distant groups will become less similar over time. The distinction is that with drift, the change is due to random mutation, whereas in our model, each group chooses to differentiate their language from all neighbors. Since $d_{ij} > \overline{d}$, i and j do not interact frequently enough to make it worthwhile to keep up with the changes in each other's language via learning.

The context of our model—small scale farming or foraging economies—may cause some readers to overlook its relevance to contemporary linguistic diversity. It is important, however, to note that linguistic change is usually slow (Pagel et al., 2007). Sometimes a generation may not even be conscious about the changes occurring in their language. Anecdotal evidence of the English language suggests that it was already close to its present form before the advent of modern transportation and communication technology. An average English speaker today, for example, has little trouble understanding Jonathan Swift's *Gulliver's Travels*, written in 1726, decades before the industrial revolution. This time lag in the determination of linguistic diversity suggests that we should not expect much contemporaneous effect of current prevalence of trade and/or conflict on current linguistic diversity. Despite our focus on small-scale societies, we hypothesize that cooperative and competitive incentives are important causes of linguistic change in the modern world. The factors affecting the type and frequency of interaction, however, are complex and change quickly.

Appendix

A Justification for the Reduced Form Expected Payoffs

In this section we provide a justification for the two reduced form expected payoffs, (2) and (3), described in Section 3.3. In particular, we show how communication in games of cooperative and competitive natures can lead to the specific forms of these payoff functions.

Our argument is developed in a framework of two-player extensive form games, whose payoff structure is of either cooperative or competitive nature. The game begins with nature choosing a state and one randomly selected player to observe it. This player, referred to as the sender, then decides whether to transmit a message regarding the state. If a message is sent, then the other group, called the receiver, has an opportunity to choose an action based on the sender's message.

Each group is endowed with a set of messages that refer to a subset of the states. Elements in the intersection of the two spaces are understood by both groups, while those unique to one group's message space are understood only by that group. We derive a group's expected payoffs in the subgame perfect equilibria, and show that they are related to the sizes of intersection and relative complement of the two message spaces in the way specified by (2) and (3).[38]

A.1 General Environment

Suppose at the beginning of each period t, a two dimensional state (τ, a) is drawn to determine a game $\Gamma^t(\tau, a)$ played by groups i and j. The parameter $\tau \in \{\text{coop}, \text{comp}\}$ determines the strategic nature of the interaction, and $a \in A = [0, 1]$ determines the optimal action(s) for each group. The two parameters are drawn independently of each other: $\tau = \text{coop}$ with probability $\frac{p}{p+q}$ and $\tau = \text{comp}$ with probability $\frac{q}{p+q}$; a is drawn uniformly from A. Let σ_k, $k = i, j$, denote group k's action in $\Gamma^t(\tau, a)$. The payoffs in

[38] While we interpret a group's message space to be its dialect, the framework proposed in this section does not integrate perfectly with the model presented in the paper. Specifically, in this section, the measure of the intersection of the message spaces is assumed to be smaller than the measure of the action space. For this to hold every period in the main sections of the paper, the action space would have to grow exogenously every period at a rate no smaller than the growth rate of the intersections.

$\Gamma^t(\text{coop}, a)$ are

$$\pi_i(\sigma_i, \sigma_j) = \pi_j(\sigma_i, \sigma_j) = \begin{cases} 1 & \text{if } \sigma_i = \sigma_j = a \\ 0 & \text{otherwise,} \end{cases} \qquad \forall a \in A; \tag{24}$$

and the payoffs in $\Gamma^t(\text{comp}, a)$ are

$$\pi_i(\sigma_i, \sigma_j) = \begin{cases} 2 & \text{if } \sigma_i = a \neq \sigma_j \\ -2 & \text{if } \sigma_i = \sigma_j = a \end{cases} \quad \text{and} \quad \pi_j(\sigma_i, \sigma_j) = -\pi_i(\sigma_i, \sigma_j), \qquad \forall a \in A. \tag{25}$$

Let $M_k = [x_k, y_k]$, where $0 \leq x_k < y_k \leq 1$ and $k = i, j$, be the set of messages that group k can use to refer to actions in A. Note that $M_i \cap M_j$ need not be empty, and that M_i and M_j need not have the same size. Each message $m \in M_k$ refers to one and only one action $a \in A$. More precisely, let $\mu_k : M_k \to A$ be a measure-preserving injection that maps elements in M_k to those in A. Thus, the meaning (or referent) of a message $m \in M_k$ is given by $\mu_k(m)$. Further, let μ_i, μ_j satisfy the property that $\mu_i(m) = \mu_j(m)$ for all $m \in M_i \cap M_j$, so that messages in the intersection of the two message spaces refer to the same actions. Denote by m_k^a a message sent by k that refers to action a; thus $m_k^a = \mu_k^{-1}(a)$. We say that group j *understands* a message m_i sent by group i if and only if $m_i \in M_i \cap M_j$.[39]

Let $\widetilde{\Gamma}^t$ be an extensive form game based on $\Gamma^t(\tau, a)$:

i) With equal probability, nature chooses either i or j to be the sender.
ii) The sender, say i, observes the states (τ, a).
    - If the state a is such that there exists no $m_i \in M_i$ with $\mu_i(m_i) = a$, the game ends and both i and j get zero payoff.
    - If there exists $m_i \in M_i$ with $\mu_i(m_i) = a$, then i decides whether to engage the receiver, j, in $\Gamma^t(\tau, a)$.
        - If i chooses not to engage, then the game ends with both groups getting zero payoff.
        - If i chooses to engage, then it sends m_i^a and chooses $\sigma_i = a$ in $\Gamma^t(\tau, a)$. Both τ and m_i^a are then observable to j.
iii) Based on τ and m_i^a, j chooses an action $\sigma_j \in A$.
iv) The payoffs are determined by σ_i and σ_j according to (24) and (25).

[39]We assume that messages in $M_i \cap M_j$ is common knowledge.

The game $\widetilde{\Gamma}^t$ can be interpreted as follows. The leader of group i discovers an interaction opportunity characterized by (τ, a) that would involve another group j. Group i's optimal payoff is a function of a. If group i does not have the requisite language (i.e. m_i^a) to communicate a among its members, then it cannot take advantage of this opportunity. Hence there is no change in either group's payoff. Suppose i has a message m_i^a that refers to a. The the leader of i can still decide whether to take advantage of the opportunity based on the anticipated response of j. If she decides not to, then there is no change in either group's payoff. If she decides to take the opportunity, then she sends m_i^a, which serves as a coordination signal for members of group i. Group j can also observes m_i^a with some probability, and for simplicity, we assume this probability to be 1. Based on whether j understands m_i^a and the observed nature of the interaction, j responds (optimally).

In the next two subsections, we focus on the case where the sender has a message for the drawn state a, and analyze a group's expected payoff in a subgame perfect equilibrium (SPE).

A.2 Cooperative Interaction

Suppose $\tau = \text{coop}$, and i is the sender. The following is an SPE of $\widetilde{\Gamma}^t$:

- i always engages, sends m_i^a and chooses $\sigma_i = a$;[40]
- j chooses $\sigma_j = a$ if $m_i^a \in M_i \cap M_j$, and chooses uniformly from A if $m_i^a \in M_i \setminus M_j$.

In this SPE, the payoff of both groups are positive whenever a is such that $m_i^a \in M_i \cap M_j$. This event occurs with probability $|M_i \cap M_j|$, since $|A| = 1$. Therefore, the expected payoff of i is

$$|M_i \cap M_j|.$$

In fact, it is easy to verify that all SPEs (except a subset of measure zero of them) have the same derived expected payoffs. If we interpret $M_i = L_i \cup E_{ij}$, then (2) is derived.

A.3 Competitive Interaction

Suppose $\tau = \text{comp}$, and i is the sender. The following is an SPE of $\widetilde{\Gamma}^t$:

- i engages, sends m_i^a, and chooses $\sigma_i = a$ if and only if a is such that $m_i^a \in M_i \setminus M_j$;
- j chooses $\sigma_j = a$ if $m_i^a \in M_i \cap M_j$, and chooses uniformly from A if $m_i^a \in M_i \setminus M_j$.

[40] Note that this is a weakly dominant strategy for i.

In this SPE, the sender guarantees a non-negative payoff. The payoff is positive whenever a is such that $m_i^a \in M_i \setminus M_j$. As a result, the *ex ante* (before the identity of the sender is determined) expected payoff of group i is

$$\frac{1}{2}\Big[\Pr\left(\{a : m_i^a \in M_i \setminus M_j\}\right)(2) + \Pr\left(\{a : m_j^a \in M_j \setminus M_i\}\right)(-2)\Big]$$
$$= |M_i \setminus M_j| - |M_j \setminus M_i|.$$

Again, all SPEs (except a subset of measure zero of them) have the same derived expected payoffs. If we interpret $M_k = L_k \cup E_k \cup N_k$, $k = i, j$, then (3) is derived.

B Proofs

B.1 Proof of Lemma 2

Proof. Let $\{L_i^t\}_i$ be symmetric and localized. We want to show that $\{L_i^{t+1}\}_i$ is also symmetric and localized, namely, it satisfies conditions (15), (16), and (11).

It is obvious that $\{L_i^{t+1}\}_i$ satisfies condition (15) by Proposition 1 and Corollary 2.

Next, to see that $\{L_i^{t+1}\}_i$ satisfies condition (16), consider i, j, k such that $d_{ij} = d_{ik}$. If $d_{ij} = d_{ik} \leq d^{t*}$, where d^{t*} is determined by (14) for every $t \geq 1$, then we know by Corollary 1 that $L_i^t \cup E_i^t = L_j^t \cup E_j^t = L_k^t \cup E_k^t$. From the definition of L^{t+1}, it follows that $|L_i^{t+1} \cap L_j^{t+1}| = |L_i^{t+1} \cap L_k^{t+1}|$. Suppose $d_{ij} = d_{ik} > d^{t*}$. There are two cases to consider. First, if $d_{ij} = d_{ik} > d^{t*} + 1$, then i will learn nothing from either j or k (nor would the latter two learn anything from i). So $|L_i^{t+1} \cap L_j^{t+1}| = |L_i^{t+1} \cap L_k^{t+1}| = |L_i^t \cap L_j^t| = |L_i^t \cap L_k^t|$. Second, suppose $d_{ij} = d_{ik} = d^{t*} + 1$. Observe that

$$|L_i^{t+1} \cap L_j^{t+1}| = |L_i^t \cap L_j^t| + |E_{ij}^t| + |E_{ji}^t| + |E_{i-j}^t \cap E_{j-i}^t|$$
$$|L_i^{t+1} \cap L_k^{t+1}| = |L_i^t \cap L_k^t| + |E_{ik}^t| + |E_{ki}^t| + |E_{i-k}^t \cap E_{k-i}^t|,$$

where $E_{m-n} = E_m \setminus E_{mn}$. By Assumption 1, $|E_{ij}^t| = |E_{ik}^t|$. Since the decision problem is symmetric, it must be that $|E_{ji}^t| = |E_{ki}^t|$ as well. Moreover, that both j and k are i's degree $d^* + 1$ neighbors implies that j and k are within the same d^* neighborhood. By Assumption 2, all groups within a d^* neighborhood learn the same subset of elements from their degree $d^* + 1$ neighbors. It follows that $E_{j-i}^t = E_{k-i}^t$ up to a subset of measure zero. Therefore, $|E_{i-j}^t \cap E_{j-i}^t| = |E_{i-k}^t \cap E_{k-i}^t|$.[41] As a result, condition (16) holds in $t + 1$.

[41]We do not actually need Assumption 2 to prove that $|E_{i-j}^t \cap E_{j-i}^t| = |E_{i-k}^t \cap E_{k-i}^t|$; Assumption 1 alone

Lastly, condition (11) says that if an element ℓ is learnable by i from a degree d neighbor, then it cannot be in the dialect of a neighbor of degree greater than d. Suppose, for contradiction, that $\{L_i^{t+1}\}_i$ violates condition (11), i.e. there exists an element $\widehat{\ell}$ such that

$$\widehat{\ell} \in L_j^{t+1} \setminus L_i^{t+1} \quad \wedge \quad \widehat{\ell} \in L_k^{t+1}$$

for some i, j, k such that $d_{ij} < d_{ik}$. Since $\{L_i^t\}_i$ is localized and $L_k^{t+1} = L_k^t \cup E_k^t \cup N_k^t$, it must be the case that either $\widehat{\ell} \in E_{jk}^t \setminus E_{ik}^t$ or $\widehat{\ell} \in E_k^t$. $\widehat{\ell} \in E_{jk}^t \setminus E_{ik}^t$ contradicts the assumption that when $d_{ik} = d_{jk}$, i and j learn the same set from k.

$\widehat{\ell} \in E_k^t$ and $\widehat{\ell} \in L_j^{t+1} \setminus L_i^{t+1}$ imply that

$$\widehat{\ell} \in (E_k^t \cap (L_j^t \cup E_j^t)) \quad \wedge \quad \widehat{\ell} \notin (E_i^t \cap (L_j^t \cup E_j^t)).$$

But this is inconsistent with Lemma 1 and Proposition 2. If $d_{ij} \leq d^{t*}$, then there is no element in $L_j^t \cup E_j^t$ that i does not know, hence contradicting the second conjunct. If $d_{ij} > d^{t*}$, k will not choose to learn anything from j. This is true because when $d_{ij} < d_{ik}$, we have $d_{ki} = d_{kj}$. By $d_{ij} < d_{ik}$ and $d_{ij} \leq d^{t*}$, it must be the case that $d_{ik} > d^{t*} + 1$. Combining this fact with Lemma 1, k will not learn anything from i, and so neither will k learn from j. Hence we have a contradiction with the first conjunct. Therefore, condition (11) must hold for $\{L_i^{t+1}\}_i$.

This completes the proof. □

B.2 Proof of Proposition 3

To prove this proposition, it is helpful to first introduce the following lemma:

Lemma 3. *Let $\{L_i^t\}_{i\in\mathscr{G}}$ be symmetric and localized. Then, $d^{t*} \leq \overline{d}$, where d^{t*} is determined by (14) for every $t \geq 1$.*

Proof. First, observe that, for all $t \geq 1$, all $i \in \mathscr{G}$,

$$\sum_{k=1}^{d} |P_i^t(k)| \geq (2^d - 1)|N^*|, \qquad \forall d \in \{1, \ldots, D\}. \tag{26}$$

This is true because, from Proposition 1, we know that each group invents a measure of $|N^*|$ new linguistic elements in every $t-1$ ($t = 1, 2, \ldots$). Hence in t, the set of learnable elements within a degree d neighborhood for group i, $\bigcup_{k=1}^{d} P_i^t(k)$, must contain at least, and potentially more than, a measure of $(2^d - 1)|N^*|$ elements. Recall that d^{t*} and $\overline{d}$ are

is enough, although the exposition will be more complicated.

determined by conditions (14) and (19), respectively. According to (19), $|P_i^t(d)| = 2^{d-1}|N^*|$ for $d \leq \overline{d}$, and so (26) holds with equality for $d \leq \overline{d}$. Therefore, conditions (14) and (19) imply that $d^{t*} \leq \overline{d}$ for all $t \geq 1$. □

This lemma says that regardless of initial linguistic composition, no group will ever learn the all the learnable elements of all of its $\overline{d} + 1$ degree neighbors.

Now we are ready to prove Proposition 3 itself.

Proof. First, consider i, j such that $d_{ij} > \overline{d} + 1$. From Lemma 3, $d^{t*} \leq \overline{d}$. Trivially, $d^{t*} + 1 \leq \overline{d} + 1$. Therefore, if $d_{ij} > \overline{d} + 1$, then $d_{ij} > d^* + 1$. According to Corollary 1, i will never learn any elements of j's language. Every period, $|L_j \setminus L_i|$ increases by exactly $|N^*|$.

Next, consider i, j such that $d_{ij} = \overline{d} + 1$ and $d^* < \overline{d}$. Since $d^* + 1 < d_{ij}$, by Corollary 1 i will not learn any element of j's language.

Next consider i, j such that $d_{ij} = \overline{d} + 1$ and $d^* = \overline{d}$. From 9, $|P_i(d)| \geq 2^{d-1}|N^*|$ for all d. From Corollary 1, i learns $|E_i^*| - \sum_{k=0}^{d^*} |P_i(k)|$ from its $d^* + 1$ neighbors. The amount that i learns from its degree $\overline{d} + 1$ neighbors, $|E_i^*| - \sum_{k=0}^{\overline{d}} |P_i(k)|$, is maximized when $|P_i(d)| = 2^{d-1}|N^*|$ for all $d \leq \overline{d}$. Since $d^* = \overline{d}$ it must be true by Proposition 2 that $|E_i^*| < \sum_{k=0}^{\overline{d}+1} 2^{k-1}|N^*|$. According to Assumption 1, i learns an equal measure of each of its degree $\overline{d} + 1$ neighbor's languages, so$i$ will not learn more than $|N^*|$ of any single one of its degree $\overline{d} + 1$ neighbors' languages. Therefore $|E_{ij}^*| < |N_j|$ and $|L_j \setminus L_i|$ increases every period. □

B.3 Proof of Proposition 4

Proof. According to Definition 6, we need to show that if condition (20) holds for t then it also holds for $t + 1$.

Suppose (20) is true for t. We know from Proposition 1 that between t and $t + 1$, each group invents a set of new elements of measure $|N^*|$. Therefore, within a degree $\overline{d}$ neighborhood, a measure of $2^{\overline{d}}|N^*|$ new elements will have been invented by the end of period t. For $\overline{d} = 0$, it is trivially true that (20) holds for $t + 1$, because a dialect is always the same as itself. Consider $\overline{d} \geq 1$. At $t + 1$, dialects within a $\overline{d}$ neighborhood will stay as one language—i.e. $L_i^{t+1} \cup E_i^{t+1} = L_j^{t+1} \cup E_j^{t+1}$ for all i, j such that $d_{ij} \leq \overline{d}$—if and only if the marginal benefit at $|E^*| = (2^{\overline{d}} - 1)|N^*|$ outweighs the marginal cost. But this is true by how $\overline{d}$ is determined in (19). Therefore, condition (20) holds for $t + 1$. □

B.4 Proof of Proposition 5

One additional lemma will be useful in proving Proposition 5.

Lemma 4. *For any $t \geq 1$ and $\overline{d} \geq 1$, we have*

$$d^{t*} < \overline{d} \quad \Rightarrow \quad |E^{t*}| > (2^{\overline{d}} - 1)|N^*|,$$

and

$$d^{t*} = \overline{d} \quad \Rightarrow \quad |E^{t*}| \geq (2^{\overline{d}} - 1)|N^*|.$$

Proof. Let $|\overline{E}|$ denote the measure of elements learned by a group in the steady state $\overline{\Lambda}$. From Proposition 4, we know that $|\overline{E}|$ is at least $(2^{\overline{d}} - 1)|N^*|$. Consider the first implication. Suppose $d^{t*} < \overline{d}$, so that the region's linguistic composition is *out of* the steady state. Then the marginal benefit of learning elements on the interval $\left[(2^{\overline{d}} - 1)|N^*|, |\overline{E}|\right]$ is at least $\frac{r+\beta}{1+r}\pi(d^{t*})$, which is strictly higher than the marginal benefit of the elements on the same interval if a group were learning *in* the steady state. This is illustrated in Figure 5. In the top panel, the linguistic composition if out of steady state ($d^* = 2$); in the bottom panel, the linguistic composition is in steady state ($\overline{d} = 3$).[42] On the interval $\left[(2^{\overline{d}} - 1)|N^*|, |\overline{E}|\right]$, the marginal benefit in the top panel is higher than that in the bottom panel. Since the marginal cost of learning is the same both in and out of the steady state, it follows that a group must learn a strictly larger measure of elements out of steady state. As a consequence, $|E^{t*}| > |\overline{E}| \geq (2^{\overline{d}} - 1)|N^*|$ and the first implication is established. The second implication follows trivially from the fact that when $d^{t*} = \overline{d}$, the lower bound in Proposition 2 applies to both $|E^{t*}|$ and $|\overline{E}|$. But since the latter is bounded below by $(2^{\overline{d}} - 1)|N^*|$ according to Proposition 4, so must $|E^{t*}|$. $\square$

Now we are ready to prove Proposition 5 itself.

Proof of Proposition 5. If $\overline{d} = 0$, then by (14) and (19) we have $d^{t*} = 0$ for all $t \geq 1$. Consequently, the steady state $\overline{\Lambda}$ is achieved at the end of the first period at the latest. Suppose $\{L_i^0\}_i$ is such that $L_i^0 = L_j^0$ for all $i, j \in \widetilde{\mathscr{G}} \subseteq \mathscr{G}$. Then, at the beginning of $t = 1$, each L_k will differ by at least a measure of $|N^*|$ elements. Since $d^{t*} = 0$ for all t, we must have $|E_i^{1*}| < |N^*|$. Thus, for all $t \geq 2$, it must be the case that $L_i^t \neq L_j^t$ for all $i, j \in \mathscr{G}$. If, on the other hand, not all of the initial set of dialects are identical and $d^{t*} = 0$, then *a fortiori*, $L_i^t \neq L_j^t$ for any $i, j \in \mathscr{G}$ and any $t \geq 1$.

[42] To simplify the drawing, we assumed that the marginal benefit of learning has only one step for each d_{ij}. See footnote 32 for more detail.

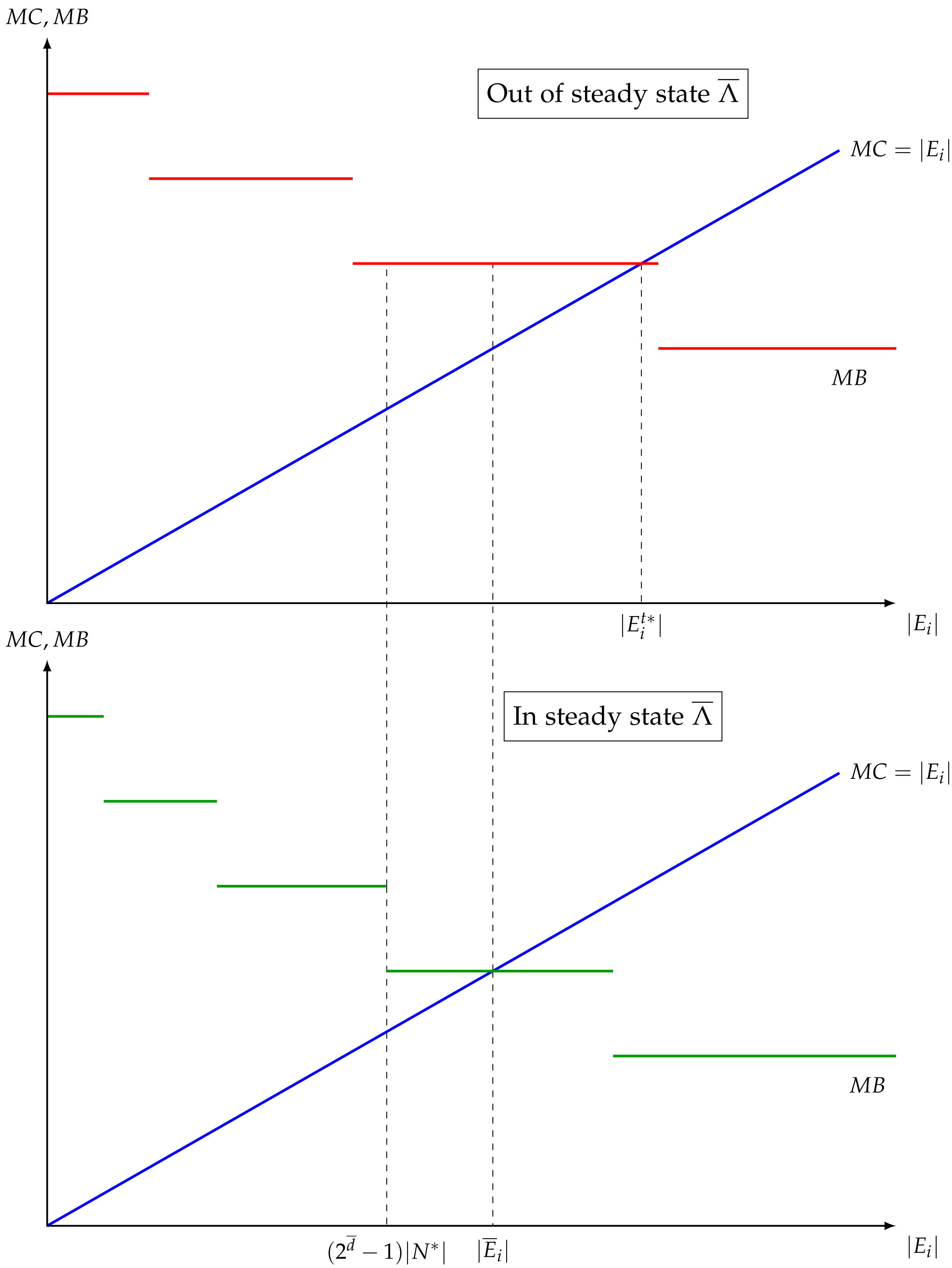


Figure 5: Marginal Benefit of Learning in and out of Steady State

For the remainder of the proof, we assume $\overline{d} \geq 1$, and the proof proceeds as follows: First we show that after the initial period, all pairs of groups i and j with $d_{ij} > \overline{d}$ will never speak the same language. Then, we show that all groups within a degree $\overline{d}$ neighborhood will eventually share a common language as defined in Definition 2.

Let $\{L_i^0\}_i$ be symmetric and localized, and consider a pair of groups i, j with $d_{ij} > \overline{d}$. By Lemma 3, we know that $d_{ij} > d^{t*}$. This implies that, for each t, the measure of i's set of learnable elements is strictly greater than the measure of the set that i actually learns: $\sum_{k=0}^{d_{ij}} |P_i^t(k)| > |E_i^{t*}|$. According to Lemma 1, therefore, it is never optimal for i to learn the entire set of $L_j^t \setminus L_i^t$ for every $t \geq 0$. As a consequence, we have

$$d_{ij} > \overline{d} \quad \Rightarrow \quad L_i^t \cup E_i^t \neq L_j^t \cup E_j^t, \qquad \forall t \geq 1.$$

Next, consider a pair of groups i, j for which $d_{ij} \leq \overline{d}$. In each period t, i's closest $2^{\overline{d}} - 1$ neighbors invent a measure of $(2^{\overline{d}} - 1)|N^*|$ new linguistic elements. Lemma 4 shows that when d^{t*} is less than (or equal to) $\overline{d}$, the size of the set of elements learned by i is strictly (or weakly) greater than the number of elements newly invented by the closest $2^{\overline{d}} - 1$ neighbors. Since each L_i is of a finite measure, and the rate of linguistic convergence (i.e. size of acquired elements from existing dialects) is higher than the rate of linguistic divergence (i.e. size of invented elements) within a degree $\overline{d}$ neighborhood, dialects within such a neighborhood will attain $L_i \cup E_i = L_j \cup E_j$ in finitely many periods and will remain that way thereafter. As a result, there exists a $T < \infty$ such that for all $t \geq T$, we have $L_i^t \cup E_i^t = L_j^t \cup E_j^t$ for all i, j with $d_{ij} \leq \overline{d}$. This completes the proof. □

C Figures

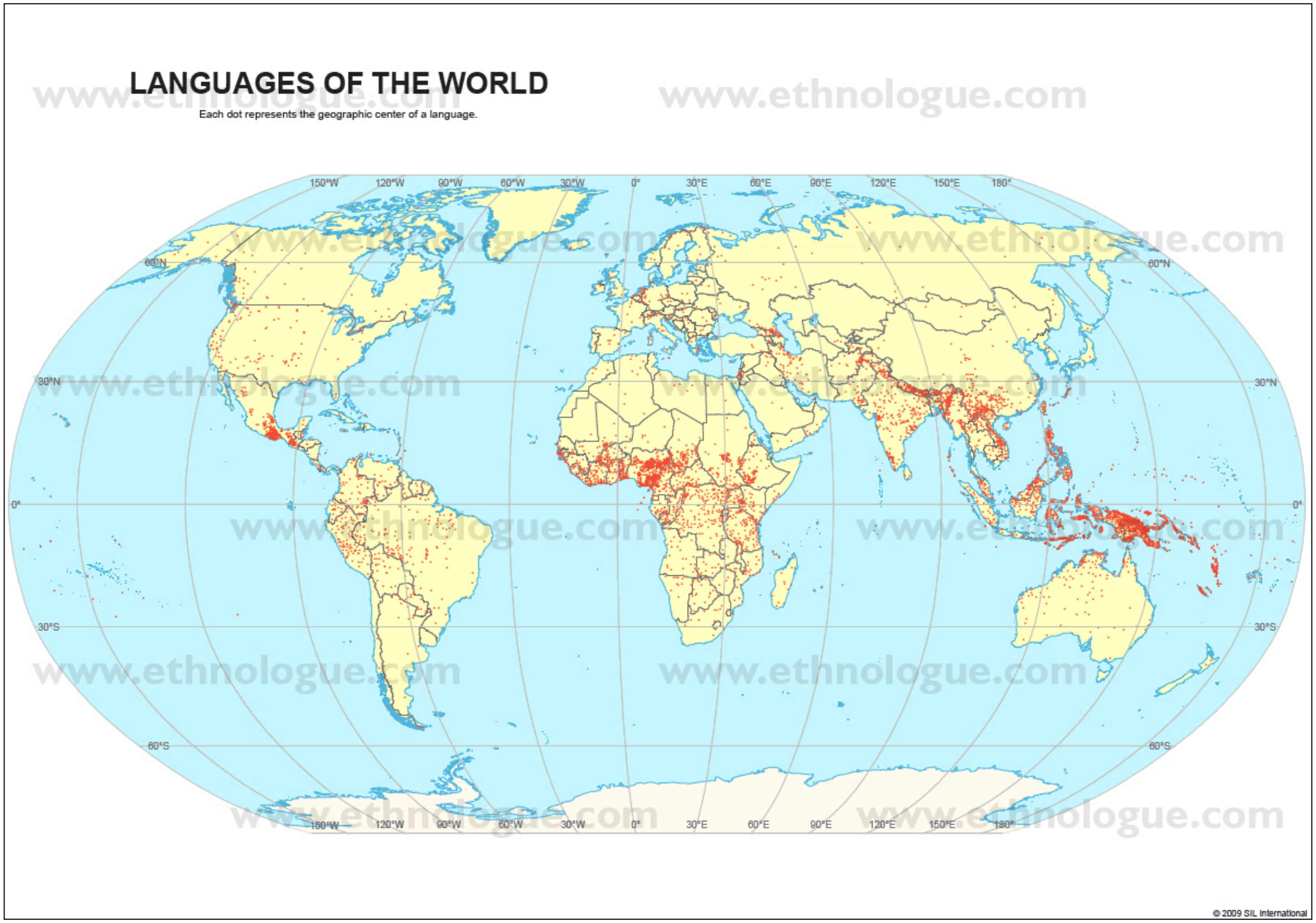
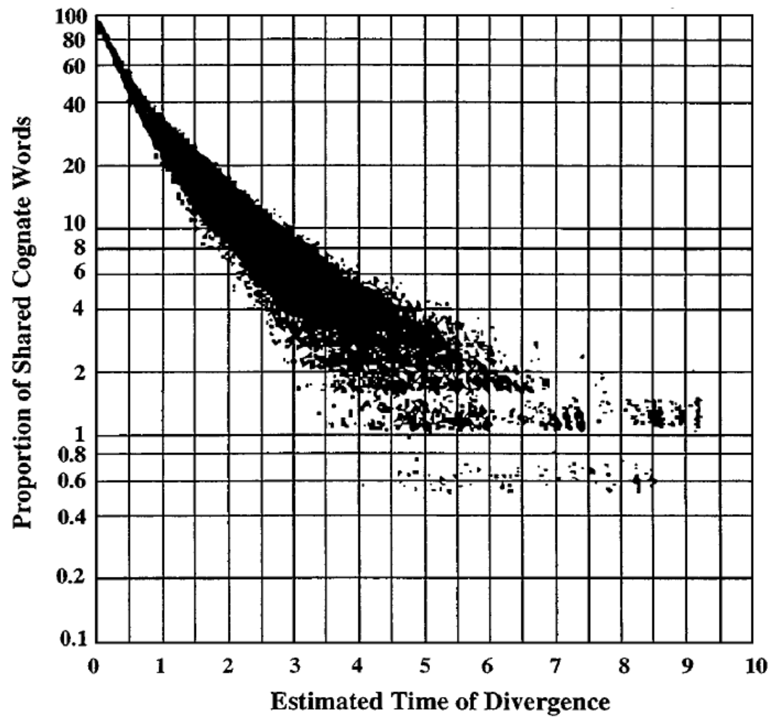
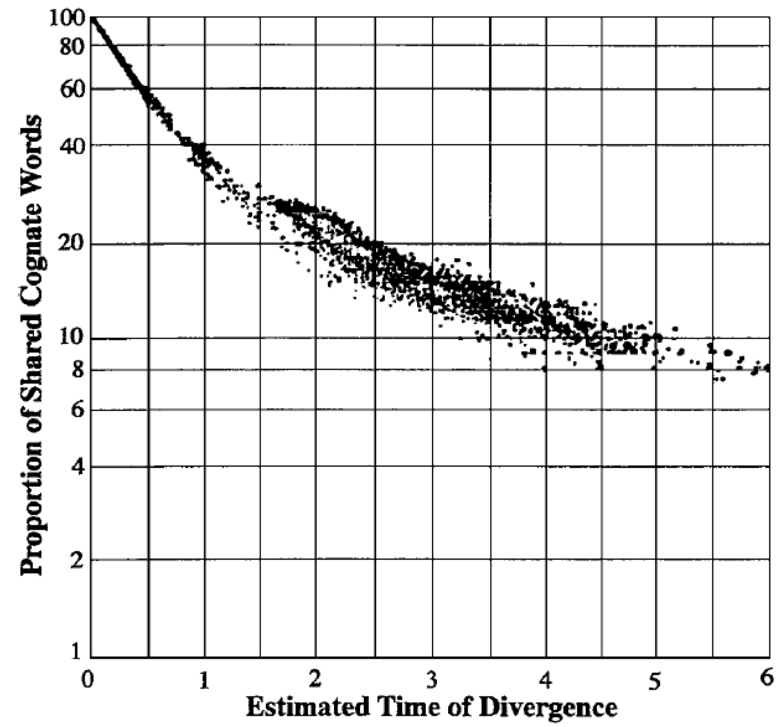


Figure 6: Distribution of Languages in the World
Each red dot indicates the geographic center of a language. Source: Lewis (2009, http://archive.ethnologue.com/16/show_map.asp?name=World&seq=10).



(a) Divergence of Malayo-Polynesian Languages



(b) Divergence of Indo-European Languages

Figure 7: Estimated Linguistic Divergence in Malayo-Polynesian and Indo-European Language Families
Each dot is a pair of groups. Note that the vertical axis is in log-scale. On both graphs, the rate of divergence is very high immediately after a split: pairs of languages seem to explode apart, and later settle down to a much low rate of divergence. Source: Pagel (2000, p.399–400).

References

Aitchison, J. and H. Carter (2000). *Language, economy, and society: The changing fortunes of the Welsh language in the twentieth century*. University of Wales Press. [p. 10]

Anderson, J. E. and E. Van Wincoop (2004). Trade costs. Technical report, National Bureau of Economic Research. [p. 7]

Austin, P. K. and J. Sallabank (2011). *The Cambridge handbook of endangered languages*. Cambridge: Cambridge University Press. [p. 2]

Bentahila, A. and E. Davies (1993). Language revival: restoration or transformation? *Journal of Multilingual & Multicultural Development 14*(5), 355–374. [p. 2]

Blume, A. and O. Board (2013a). Intentional vagueness. *Erkenntnis*, 1–45. [p. 6]

Blume, A. and O. Board (2013b). Language barriers. *Econometrica 81*(2), 781–812. [p. 6]

Blume, A., Y.-G. Kim, and J. Sobel (1993). Evolutionary stability in games of communication. *Games and Economic Behavior 5*, 547–575. [p. 6]

Boyd, R. and P. J. Richerson (1988). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press. [p. 8]

Buhaug, H., S. Gates, and P. Lujala (2009). Geography, rebel capability, and the duration of civil conflict. *Journal of Conflict Resolution 53*(4), 544–569. [p. 9]

Buhaug, H. and P. Lujala (2005). Accounting for scale: Measuring geography in quantitative studies of civil war. *Political Geography 24*(4), 399–418. [p. 9]

Buhaug, H. and J. Rød (2006). Local determinants of african civil wars, 1970–2001. *Political Geography 25*(3), 315–335. [p. 9]

Chen, H. (2017). *Essays on the Economics of Linguistic Diversity and Preference for Surprise*. Ph. D. thesis, Simon Fraser University, Burnaby, B.C., Canada. [p. 11]

Coleman, J. (2008). *A History of Cant and Slang Dictionaries: Volume III: 1859-1936: Volume III: 1859-1936*, Volume 3. OUP Oxford. [p. 10]

d'Anglejan, A. (1984). Language planning in quebec: An historical overview and future trends. *Conflict and Language Planning in Quebec 5*, 29. [p. 11]

Dawson, G. (1891). *Notes on the Shuswap people of British Columbia*. Dawson Brothers. [p. 17]

de Rouen, K. R. and D. Sobek (2004). The dynamics of civil war duration and outcome. *Journal of Peace Research 41*(3), 303–320. [p. 9]

De Soysa, I. (2002). Paradise is a bazaar? Greed, creed, and governance in civil war, 1989-99. *Journal of Peace Research 39*(4), 395–416. [p. 9]

Desmet, K., I. Ortuño-Ortín, and R. Wacziarg (2012). The political economy of linguistic cleavages. *Journal of Development Economics 97*(2), 322–338. [p. 7]

Esteban, J., L. Mayoral, and D. Ray (2012, September). Ethnicity and conflict: An empirical study. *American Economic Review 102*(4), 1310–42. [p. 6, 7, 22]

Fairhead, J. (2000). The conflict over natural and environmental resources. *War, hunger, and displacement. The origins of humanitarian emergencies 1*, 147–178. [p. 9]

Falck, O., S. Heblich, A. Lameli, and J. Südekum (2012). Dialects, cultural identity, and economic exchange. *Journal of Urban Economics 72*(2), 225–239. [p. 7]

Fearon, J. and D. Laitin (2003). Ethnicity, insurgency, and civil war. *American Political Science Review 97*(1), 75–90. [p. 9]

Fearon, J. D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth 8*(2), 195–222. [p. 7]

Finegan, E. (2008). *Language: Its Stucture and Use* (5 ed.). Thomson Wadsworth. [p. 22]

Galloway, B. (2007). Language revival programs of the Nooksack Tribe and the Stó:lō Nation. *Be of Good Mind: Essays on the Coast Salish*, 212. [p. 10]

Gardner, N., M. Serralvo, and C. Williams (2000). Language revitalization in comparative context: Ireland, the Basque Country and Catalonia. In C. Williams (Ed.), *Language Revitalization: Policy and Planning in Wales*, pp. 311–355. Cardiff: University of Wales Press. [p. 10]

Garfinkel, M. R. and S. Skaperdas (2007). Economics of conflict: An overview. Volume 2, pp. 649–709. Elsevier. [p. 3]

Goldschmidt, A. and L. Davidson (1991). *A concise history of the Middle East*. Westview Press. [p. 10]

Gray, R. D., A. J. Drummond, and S. J. Greenhill (2009). Language phylogenies reveal expansion pulses and pauses in pacific settlement. *Science 323*(5913), 479–483. [p. 3]

Hale, K., M. Krauss, L. J. Watahomigie, A. Y. Yamamoto, C. Craig, L. M. Jeanne, and N. C. England (1992). Endangered languages. *Language 68*(1), 1–42. [p. 2]

Hall, R. A. (1966). *Pidgin and creole languages*, Volume 7. Ithaca: Cornell University Press. [p. 3, 9]

Harmon, D. (1996). Losing species, losing languages: Connections between biological and linguistic diversity. *Southwest Journal of Linguistics 15*, 89–108. [p. 7]

Hotten, J. C. (1859). *A Dictionary of Modern Slang, Cant, and Vulgar Words*. Hotten. [p. 10]

Jang, T., K. Lim, and T. Payne (2011). Xibe language revitalization and documentation project. [p. 10]

Johnson, A. and T. Earle (2000). *The evolution of human societies: from foraging group to agrarian state*. Stanford University Press. [p. 10]

Kaplan, D. (2000). Conflict and compromise among borderland identities in northern italy. *Tijdschrift voor economische en sociale geografie 91*(1), 44–60. [p. 10]

Kratochwil, F. (1986). Of systems, boundaries, and territoriality: An inquiry into the formation of the state system. *World Politics 39*(1), 27–52. [p. 9]

Kruskal, J. B., I. Dyen, and P. Black (1971). The vocabulary method of reconstructing language trees: Innovations and large-scale applications. In F. R. Hodson, D. G. Kendall, and P. Tautu (Eds.), *Mathematics in the Archeological and Historical Sciences*, pp. 361–380. Edinburgh: Edinburgh University Press. [p. 3, 22]

Kulick, D. (1992). Language shift and cultural reproduction: socialization, self, and syncretism in a papua new guinean village. [p. 9]

Lazear, E. P. (1999). Culture and language. *Journal of Political Economy 107*(S6), pp. S95–S126. [p. 6]

Le Billon, P. (2001). Angola's political economy of war: The role of oil and diamonds, 1975–2000. *African Affairs 100*(398), 55–80. [p. 9]

Lewis, M. P. (Ed.) (2009). *Ethnologue: Languages of the World* (16th ed.). Dallas, Texas: SIL International. [p. 2, 43]

Lewis, M. P., G. F. Simons, and C. D. Fennig (Eds.) (2013). *Ethnologue: Languages of the World* (17th ed.). Dallas, Texas: SIL International. [p. 2, 15]

Lipman, B. L. (2003). Language and economics. In N. Dimitri, M. Basili, and I. Gilboa (Eds.), *Cognitive Processes and Economic Behavior*, Chapter 7, pp. 75–93. London: Routledge. [p. 6]

Mace, R. and M. Pagel (1995). A latitudinal gradient in the density of human languages in north america. *Proceedings: Biological Sciences 261*(1360), 117–121. [p. 2, 7]

Marschak, J. (1965). Economics of language. *Behavioral Science 10*(2), 135–140. [p. 5]

Michalopoulos, S. (2012). The origins of ethnolinguistic diversity. *American Economic Review 102*(4), 1508–39. [p. 2, 7, 8, 9]

Milloy, J. (1999). When a language dies. *Index on Censorship 28*(4), 54–64. [p. 11]

Nettle, D. (1998). Explaining global patterns of language diversity. *Journal of Anthropological Archaeology 17*(4), 354–374. [p. 2, 7, 8]

Osaghae, E. E. and R. T. Suberu (2005). *A History of Identities, Violence and Stability in Nigeria*. Centre for Research on Inequality, Human Security and Ethnicity, University of Oxford. [p. 10]

Pagel, M. (2000). The history, rate and pattern of world linguistic evolution. In C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Eds.), *The Evolutionary Emergence of Language*, pp. 391–416. Cambridge: Cambridge University Press. [p. 2, 3, 8, 11, 22, 44]

Pagel, M. (2012). War of words: the language paradox explained. *New Scientist 216*(2894), 38–41. [p. 6, 8]

Pagel, M., Q. D. Atkinson, and A. Meade (2007). Frequency of word-use predicts rates of lexical evolution throughout indo-european history. *Nature 449*(7163), 717–720. [p. 22, 32]

Partridge, E., T. Dalzell, and T. Victor (2008). *The Concise New Partridge Dictionary of Slang and Unconventional English*. Psychology Press. [p. 10]

Phinney, S., J. Wortman, and D. Bibus (2009). Oolichan grease: A unique marine lipid and dietary staple of the north pacific coast. *Lipids 44*(1), 47–51. [p. 16]

Renner, M. (2002). The anatomy of resource wars. *World Watch Paper 162*. [p. 9]

Rickard, P. (1989). *A history of the French language*. Routledge. [p. 11]

Robson, A. J. (1990). Efficiency in evolutionary games: Darwin, Nash and the secret handshake. *Journal of Theoretical Biology 144*(3), 379–396. [p. 6]

Roller, E. (2002). When does language become exclusivist? linguistic politics in catalonia. *National Identities 4*(3), 273–289. [p. 11]

Rubinstein, A. (2000). *Economics and Language: Five Essays*. Cambridge: Cambridge University Press. [p. 6]

Simpson, J. A., E. S. Weiner, et al. (1989). *The Oxford English Dictionary*, Volume 2. Clarendon Press Oxford. [p. 9]

Smillie, I. and F. Forskningsstiftelsen (2002). *Dirty Diamonds: Armed Conflict and the Trade in Rough Diamonds*. Fafo Institute for Applied Social Science. [p. 9]

Suberu, R. T. (2001). *Federalism and ethnic conflict in Nigeria*. United States Inst of Peace Press. [p. 10]

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society 96*(4), 452–463. [p. 3]

Wärneryd, K. (1993). Cheap talk, coordination, and evolutionary stability. *Games and Economic Behavior 5*, 532–546. [p. 6]

Younger, S. (2008). Conditions and mechanisms for peace in precontact polynesia. *Current Anthropology 49*(5), 927–934. [p. 11]