University of BRISTOL

DEPARTMENT OF COMPUTER SCIENCE

The Dynamics of Dialects

Eddie Jones

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Bachelor of Science in the Faculty of Engineering.

Declaration

This dissertation is submitted to the University of Bristol in accordance with the requirements of the degree of BSc in the Faculty of Engineering. It has not been submitted for any other degree or diploma of any examining body. Except where specifically acknowledged, it is all the work of the Author.

Eddie Jones, 2019

Abstract

Every culture has coordinated a practical method of communication. The vast majority of these languages have existed since time immemorial, yet there are also examples emerging more recently in the absence of a universally adopted system. Within any given system, most users will be homogeneous due to regular contact. However, the world is increasingly shrinking and no longer home to pockets of isolated communities, yet it can support a huge diversity of language use. In this study, I will explore how the structure of a social network influences and is influenced by its language use, specifically focusing on the inter-language phenomena of dialect continua, language contact, and the formation of homophilic communities. Using a series of simulations on a community of autonomous agents I will investigate the conditions for emergence and stability of these phenomena.

By extending an existing model of vocabulary formation to support a general network structure, I have shown that diverse dialects can stably cohabitate the same social network without non-linguistic factors, and that network clustering plays a more significant role is this stability than the more intuitive path length. Additionally, there is some evidence that large social networks are inherently unstable and tend to decay into smaller sub-networks without convergence.

The model proved inaccurate in the context of language contact. Although its robustness could be limited, it is also possible that the mechanism of second-language acquisition (and consequently language contact) is fundamentally distinct, for which I have proposed an explanatory model and a hypothetical experiment.

Contents

Chapter 1

Introduction

Natural languages are highly structured ways to communicate that have consistently emerged from all cultures. Populations arrive at a consensus as to how to practically describe the potentially infinite set of (conceptual) objects that constitute their environment and culture using only a finite set of terms. It seems, however, to be rare for a population to be linguistically homogeneous. Instead, languages are often composed of dialects, both regional and social, as well as registers (a variety used only in a specific context) [1].

Language divergence has historically been attributed to a breakdown in communication that causes languages to fracture into several child languages, imitating the tree-like model of the evolution of species [2]. This hypothetical isolation, typically stemming from geographical separation or migration, allows the language of each population to evolve independently. However, newer theories don't presuppose total isolation and so are more compatible with the often blurry boundaries between languages [2]. For example, internet slang evolves quickly through social media [3]. This societal structure excludes parts of the population, but there is continued contact between the two groups. Consequently, only a register forms to accommodate the specific requirements of this context. Additionally, users retain the ability to readily realign their usage with the standard language, making it distinct from a dialect.

The converse situation also occurs, albeit less frequently, where two previously isolated languages gain contact [1, 2]. This process often results in the exchange of words and other constructions. In some instances, a new hybrid language arises. The official language of Haiti, for example, is Haitian Creole (kreyòl ayisyen), which inherits features from French and various West African languages. It evolved as a way for French colonists and African slaves to communicate [4].

This paper will explore how the structure of a social network influences, and is influenced by, the variety of language use on it, focusing specifically on the following three phenomena:

1.1 Dialect Continua

Leonard Bloomfield first described a dialect continuum (or area) as having the following property [5].

> *"The difference from place to place is small, but, as one travels in any one direction, the differences accumulate, until speakers, say from opposite ends of the country, cannot understand each other, although there is no sharp line of linguistic demarcation between the places where they live. Any such geographic area of gradual transitions is called a dialect area."*

These continua are not infrequent to both rural and more connected areas of the world [6]. They also serve as a motivating example for why the tree model of language families is overly-simplistic [2]. In the tree model of evolution, the fundamental unit (a leaf of the tree) is a species, which is concretely defined by the capacity to reproduce. Similarly, for a family tree of languages to be drawn, a formal definition is required. However, dialect continua clearly can't be assigned to discrete nodes. Hence the model falls short of accurately describing such relationships.

The question is, therefore, why is such a continuum stable? It is possible that the stability emanates from friction in the system; factors such as finite memories, noisy communication channels, and cultural/environmental differences may prevent total covnergence. On the other hand, the phenomena may be entirely reproducible from network structure alone.

1.2 Language Contact

Languages are not usually isolated; instead several are often used in the same context and consequently come into contact. Through this contact, the languages themselves converge, most commonly in the form of loanwords, calques (loan-phrases) and other minor influences. However, new languages can also form through the accumulation of these smaller changes [7]. When there is no one primary source from which a language inherits its structure, it is said to be either mixed, a pidgin or a creole. A mixed language arises from a bilingual population. A pidgin occurs when there is no such cross-fluency and a creole form when a pidgin becomes the native language of a population [8]. Interestingly, in many cases, the emergence of a pidgin or creole may be impressively quick taking only a generation or so [9]. Another common trait between hybrid languages is their simplicity in comparison to the languages they stem from [7, 8].

However, it is not common for the two languages that come into contact to be on an equal footing; the environments, cultures and prestige of each group may vary. This asymmetry invariably has a lasting impact on the resulting hybrid languages. For example, in Middle English words for farm animals (e.g. cow) originate in Old English, as the lower status farm worker were Anglo-Saxon. The terms used to describe meat (e.g. beef), however, came from Old French, the language of the aristocracy, who could afford to buy meat [10]. It is not clear whether the unique traits possessed by languages that arise from contact scenarios originate in the mere a need to communicate or from social structures.

1.3 Community Formation

Similar people naturally group themselves together (homophily), and equally, isolate themselves from dissimilar groups [11]. The resulting community often becomes more homogeneous as they reaffirm each other's characteristic. This phenomenon is extremely prevalent within social media where users have the option to filter content that opposes their world view.

This tendency to splinter into dialects (more homogeneous subgroups of the population) unites both languages and a vast array of more intricate social features, for example, opinion dynamics. Although not explicitly a linguistic structure, the two are intimately linked (as discussed further in the following section) and thus an explanatory model of the formation of dialects may also be used as a lens through which to study the dynamics of discrete cultural symbols and homophily.

As almost all languages consist of several dialects, despite the continual effort of official bodies, it is natural to ask why large linguistic communities are unstable.

As with all models, there are standard payoffs between accuracy and being analytically tractable, or in the case of computational models practical to simulate. The study of emergent properties of languages, however, poses some additional unique challenges.

First, natural languages can be viewed from two perspectives as proposed by Noam Chomsky, neither of which alone constitutes the whole: an internal formulation (the I-language) and an external one (the E-language) [12]. The former, as a subject of psychology, is extremely complex to model, and the latter consisting of an ever-changing series of interactions is equally intangible. Both arise from complex adaptive systems that feed into one another [9, 13]. The task of generating and relating properties of the latter to the former is therefore non-trivial.

Computational models are well suited to the study of complex systems and their emergent properties. While analytic models abstract individual interactions looking at overall trends, computational (stochastic) models can simulate each interaction and consequently maintain their non-intuitive synergistic effects that give rise to non-linear relationships between local and global properties [9].

Another complication when modelling the emergence of language is its time scale. Although continually changing [9], noticeable differences only arise after an unimaginable quantity of interactions that take place in a comparably large community. No tractable simulation would be able to model every individual agent and interaction in conjunction. As noted above, however, the simulation of each interaction is the key advantage of computational models, and thus the number of agents must be limited to an unrealistic quantity.

Although it is perhaps dubious whether category based semantics (where terms only describe a broad range of possible stimuli) are an accurate model of the I-language, it has been proposed that the specific categories that emerge are not arbitrary but rather the result of a socio-psychological optimisation process that hopes to achieve accuracy with minimal cognitive effort [14]. Formal analytic methods have demonstrated that the categorisation of colour hues is, in fact, optimal [15], and recent analysis has also pointed to the emergence of categories (and indeed the optimisation process) as fundamentally derived from the (constrained) communication [16]. Additionally, agent-based models have been successful in recreating this conclusion [17].

Colours have become the standard grounding for studying the formation of categorical vocabulary [18] for several reasons. Most importantly because of the relative ease of gathering data, but also as the space of colour hues is one dimensional and bounded; it is easier to model. However, any explicative mechanism, and indeed results, are equally applicable to any arbitrary ground set. In particular, there is no need to limit a categorical mapping of some perceptual space to semantic assignments. An alternative (and analogous study) is that of opinion dynamics.

Agent-based models of opinion dynamics have also been successful in demonstrating the impact of network structure on the ability of minority opinions to propagate [19]. In this somewhat typical model, the opinions were simple binary stances to which agents had a preferential attachment. Models of semantic alignment have been constructed similarly: each agent (with varying level of authority) asserts the truth-value of a "hypothesis" based on their interpretation of a sentence (the stimulus) [20]. In both models, agents adjust their stance through social interaction. There are clear parallels between these models that demonstrate the correspondence between opinion dynamics and the semantic alignment of a population.

Recently, the previously ignored topic of how interactions on the network and the network structure itself may co-evolve has been investigated [21]. If the network is rewired based on the success of each interaction, the tolerance of an agent (how likely it is to communicate with an agent of a different opinion) determines whether the population comes to global consensus or divides into distinct communities [21].

It is my hypothesis that agent-based models may also illustrate that communication alone is responsible for the inter-language relationships discussed above. Computation linguistics will, therefore, be the primary vehicle for this study using an existing model of vocabulary formation [17] where agents attempt to correctly distinguish, communicate and identify an object from a scene. The model of vocabulary is a simple symbol table that maps categorical concepts to an associated set of words. Additionally, the robustness of the model, in this extended context, will also be assessed.

Specifically, in this paper I will attempt to: (1) determine if a dialect continuum may form without non-linguistic factors, the conditions necessary and sensitivity to those conditions; (2) establish whether the model presents naturalistic features in the context of language contact, showing that the process of creolisation is either comparable or distinct from the emergence of a language in a population with no preexisting symbols; and finally (3) assess whether the convergent evolution of languages use significantly contributes to the formation of homophilic communities and dialects.

Chapter 2

Background

A simple categorical agent-based negotiation scheme has been shown sufficient to guarantee the emergence of a functional communication system that is homogeneous across a population [17]. Periodically a scene (collection of stimuli) was presented to pairs of these agents. One of the agents must then communicate to the other which of the stimuli is the (arbitrarily selected) "topic". If the second agent understands the first, and so can distinguish the topic from the scene, the communication is said to have succeeded. Here the perceptual space (set of stimuli) abstractly encodes colour hues [17].

In the original model, each agent may communicate uniformly with any other agent. Therefore, the social network on which they reside is complete and unweighted, where the edges encode the capacity for communication. This paper will add structure to this graph, to model more complicated social networks.

The core model uses several key metrics to track the evolution of the population:

Perceptual Categories. The perceptual channel of each agent consists of separate categories of stimuli (their internal representation or I-language). The average number of these categories per agent shows the precision of the agent's perception, that is, agents may only distinguish stimuli that they recognise as falling into distinct perceptual categories.

Linguistic Categories. When two or more adjacent perceptual categories have the same name, together they are referred to as a linguistic category. Hence, although an agent may be able to differentiate two stimuli, it may still use the same name for both of them. The number of the linguistic categories in an agent's perceptual channel corresponds to the number of distinct concepts which that agent understands, and so the average of the entire population is an approximate measure of the size of the E-language's lexicon.

Overlap. The overlap between any two agents is the range of perceptual space for which they share the same preferred name. When extended to the entire population by taking the average of all pairs of agents this is an effective method for measuring homogeneity that correlates to the success rate of interactions.

As seen in figure 2.1, the development of the population occurs in two distinct stages. The first stage is characterised by a low success rate and overlap between agents. Here agents divide their perceptual space to distinguish new stimuli, increasing the number of categories. At the same time agents are exposed to each other's varied vocabulary the number of words associated with these categories (synonymy) also rapidly increases. The dynamics abruptly changes in the second stage where the agents establish homogeneity with a high overlap and success rate. This also results in a fall in synonymy and the number of linguistic categories. Eventually, the dynamics stabilise; once a set of sufficiently precise linguistic categories are agreed upon there is little pressure to change [17].

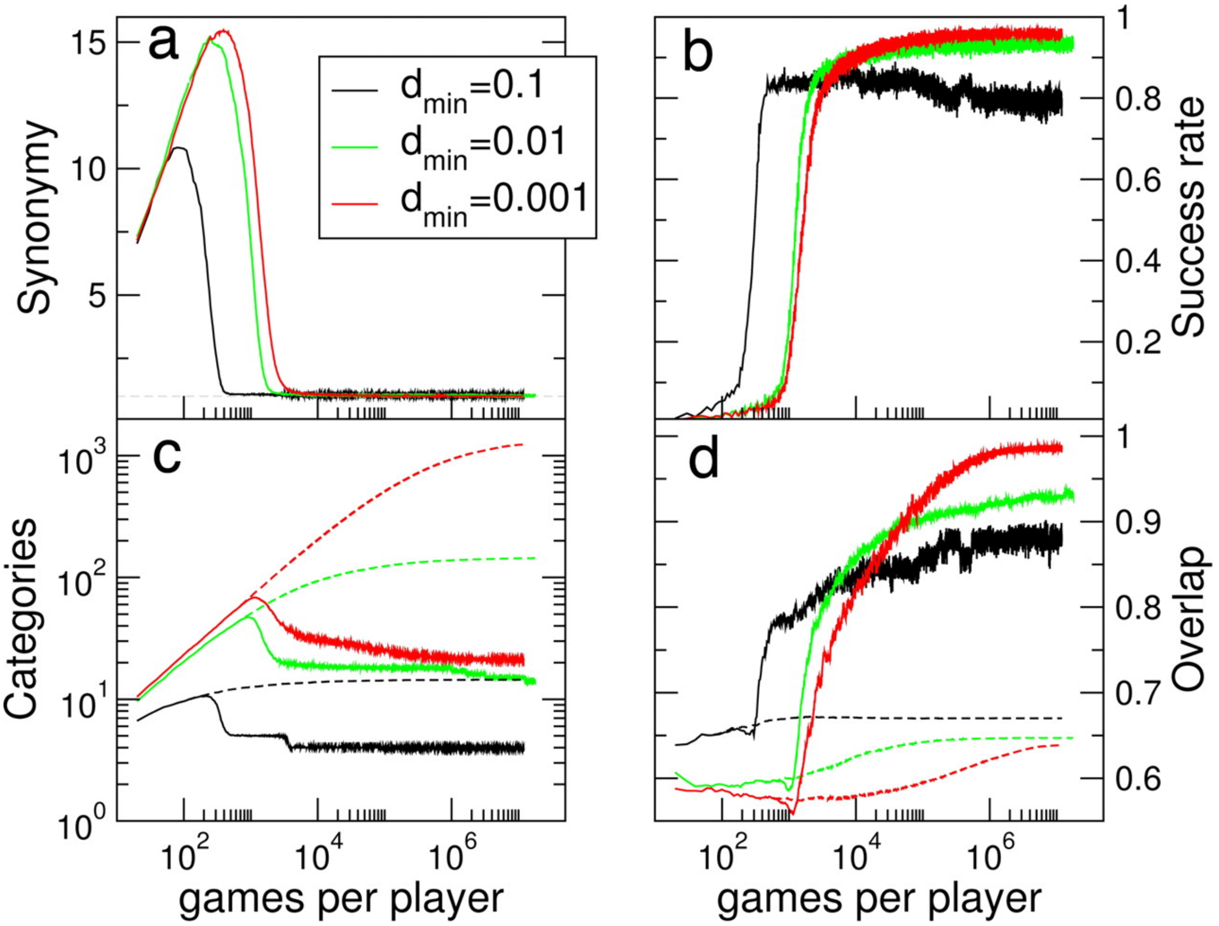


Figure 2.1: The dynamics of the simulations with $N = 100$, at various values of $d_{\min}$. Dotted and solid lines represent the average number of perceptual and linguistic categories respectively. This graphic is from the original paper [17].

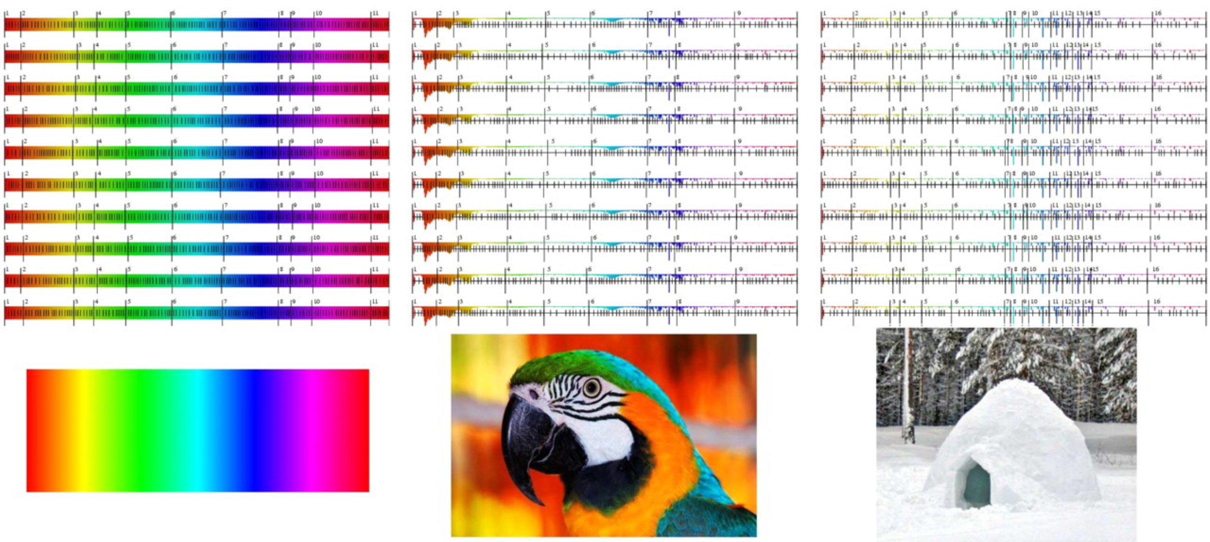


Figure 2.2: The impact of the environment. The perceptual channel and linguistic categories of ten agents, illustrated by vertical lines, above the corresponding environments from which the stimuli were sampled. This graphic is also from the original paper [17].

Figures 2.2 shows the vocabularies of ten agents. Although in the standard model stimuli are sampled uniformly, here they come from the associated images. The environment of the agents refers to this distribution (the hue histogram). In general, the boundaries of linguistic categories (denoted by the large vertical lines) are aligned between agent's vocabulary. Furthermore, there is a correlation between the frequency of stimuli and the precision of these category boundaries; agents have more opportunity to come to a precise consensus on stimuli to which they are more regularly exposed [17].

The original paper noted the evolution of two distinct hierarchies in the category structure. Precise perceptual categories develop in the first stage to allow agents to distinguish stimuli. Higher level linguistic categories evolve as words diffuse across perceptual boundaries that are misaligned between agents [17]. The core mechanic was later extended to allow for the emergence of further linguistic layers. In natural languages, there are usually an increasingly specific set of words for any given concept, for example, purple vs lilac. In the extension, the dictionary used by each agent was remodelled to encode this hierarchy [22].

The follow-up paper also added more structure to the stimuli presented. In the core model, the minimum distance between stimuli presented was constant across the perceptual space, which doesn't reflect the non-linearity of human vision. In the extended model, it instead varied to correspond to the JND (just-noticeable difference) of the human eye [22]. The JND is effectively the sensitivity, or more formally the minimum change required before there is an observable difference.

The order in which colour terms develop is relatively consistent in the evolution of natural languages [23]. For example, no languages have a word for blue without also having a word for red. The convergent evolution of different populations demonstrates that the semantics of colour terms aren't solely cultural, but are at least in part a byproduct of human physiology. In the version of the model extended to incorporate the human JND, category boundaries evolve according to this progression. The reflection of linguistic universals in the model provides more evidence for its realism [22].

Chapter 3

The Core Model

3.1 Formal Specification

The network consists of a collection of N agents. Each agent maintains its own partitioning of the perceptual space $[0, 1)$. The partitions are defined to be non-intersecting and contiguous so that each is a left-closed right-open interval referred to as "perceptual category". Perceptual categories are associated with a set of names, one of which is the agent's preferred name for that category; the rest are unordered. Initially, the agents have a trivial partition - a single category that spans the entire perceptual space with no associated names.

A linguistic category is a collection of adjacent perceptual categories, within an agent's partitioning of the perceptual space, that have a common preferred name.

On each iteration (or game), two agents are randomly selected to communicate: a speaker and a listener. They are both presented with a pair of stimuli (points in the perceptual space) that constitute the scene. Additionally, they are required to have a minimum Euclidean distance of $d_{\min}$. A specified distribution referred to as the environment is used to sample the stimuli.

Communication occurs in three stages:

Discrimination. First, the speaker learns to distinguish each element in the scene. If both stimuli are part of the same perceptual category, the agent creates a new division at the midpoint of the two stimuli. Each new category on either side inherits all the names of its parent category each with an additional new name. The new name is marked as the agent's preference and is guaranteed to be distinct from all previous names used by the population so far.

Reception. Once the speaker has learned to discriminate the scene, the listener becomes aware of the speaker's (new) preferred name for the topic. If and only if the agent recognises this name as referring to the topic, and not the other stimulus, then the communication is successful. Note this name need not be the listener's preference for it to recognise it.

Resolution. If the communication is successful, both speaker and listener remove competing words associated with the topic's category. On the other hand, if it has failed, then the listener inherits the speaker's preferred word but retains its original preference.

There are three primary metrics used to explore the core model's properties: perceptual categories, linguistic categories and overlap. The first two refer to the average number of each type of category per agent. The overlap measures the homogeneity of the population and is the average overlap between each pair of agents, which is the range of the perceptual space in which they share the same preferred name.

Although parts of the algorithm and metrics will be extended to fit the demands of each particular study, the discrimination, reception and resolution stages will remain unchanged.

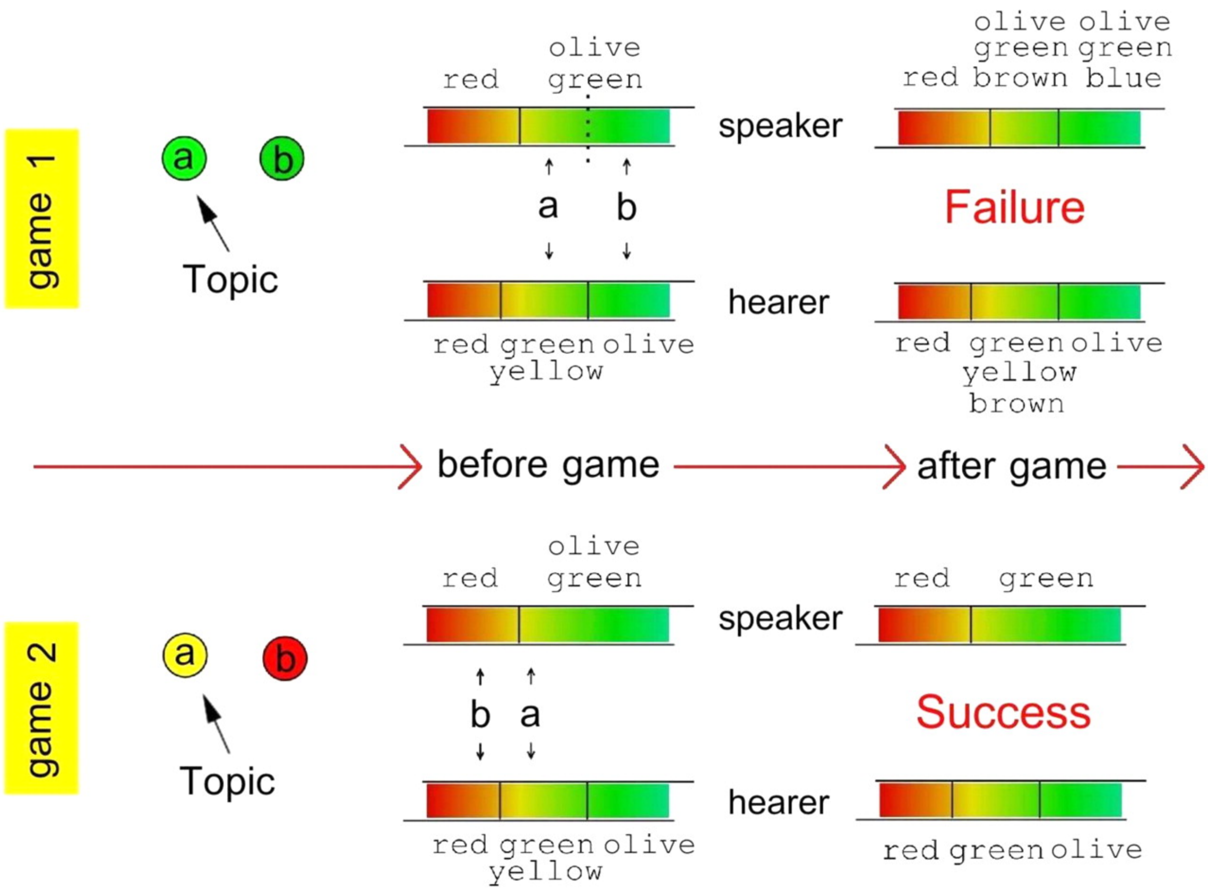


Figure 3.1: Two sample games. In the first, the speaker learns to distinguish the stimuli by splitting the category (illustrated by the dotted line) and names are newly generated and marked as the preference for these categories ("brown" and "blue" respectively). The communication has failed as the listener doesn't recognise the name "brown". Consequently, it adds this name to a's category, while maintaining a preference for "green". The second game is successful as the listener recognises the name "green"; this becomes the only name both agents associated with a's category. Again this graphic is from the original paper [17].

3.2 Implementation and Performance

The primary issue with the model was the performance sensaitive nature of data gathering. Although each iteration is simple to run, the efficiency is crucial due to the large quantity necessary ($\sim 10^9$). The first naive prototype simulation was extremely slow to run. It was implemented in Python (chosen for its powerful numerical and plotting libraries), using dynamic arrays to represent agents, categories and names.

The simulation was translated into C to improve performance while recording data to later be processed by Python scripts. However, this was still too slow to be practical. The set o name associated with each category was remodelled as a linked list with a known last element to allow for constant time insertion at both the beginning and end of the list. Dynamic arrays only provide amortised (average) constant time for inserting at the end and linear time for insert at the beginning. The beginning of the list is used to mark the preferred name.

The category structure within an agent, however, was adapted to fit a binary tree. This structure is more reflective of its growth via the splitting of existing categories, which is done in logarithmic time. A binary tree also takes logarithmic time to search for an element, or in this case, find the category in which a topic falls. The naive performance of these task in a dynamic array is linear with its length. There are typically around 20 perceptual categories per agent, and 100 agents.

Non-leaf nodes are only references to historical categories, so aren't relevant to the calculation of most metrics. Additional pointers are kept during the splitting process to allow for leaf-traversal without unnecessary computation, which results in a significant speedup due to the size of the trees.

The naive Python implementation took an average of 29.274s, over ten runs, to complete 100×10^3 iterations with $N = 100$ and $d_{\min} = 0.05$, this is significantly slower than optimised/refactored C version, which took an average of 0.063s under the same conditions. Although the iterations are inter-dependent and so cannot run in parallel, several simulations (with varying parameters) are launched simultaneously on different nodes.

3.3 Dynamics of the Core Model

Figure 3.2 shows the evolution of the core model (which is on a complete uniform network) running with a population of size $N = 100$ and with precision of $d_{\min} = 0.01$. Unless specified otherwise all data points are an average of 10 runs and these parameters will be used for the rest of the project.

Perhaps surprisingly for a stochastic model, the behaviour is remarkably consistent between runs, as illustrated by the small standard deviation plotted on the graph as error bars. As noted in the introduction, natural languages are often remarkably similar in structure (or at least share strong universal traits) despite having little to no contact. The convergent evolution of simulations is, therefore, a promising result.

As with the original paper, there are two distinct stages of evolution.

In the first phase, the overlap between agents is very low as agents haven't sufficiently divided the perceptual space. Consequently, the agents split their perceptual space further, as seen in the increasing perceptual and linguistic curves.

There is a rapid transition into the second phase where the agents begin to agree on a common vocabulary. The overlap (or equivalently the success rate) drastically increases, stabilising at around 0.97 ± 0.01. Although the agents have perfect memory and each communication is noiseless, the small number of linguistic terms limits this figure. The number of perceptual categories is still increasing which shows that the language has become homogeneous before it can adequately describe the perceptual space. It eventually surpasses the optimal division of $1/d_{\min} = 100$ categories; there is very little growth beyond this point. Often perceptual categories

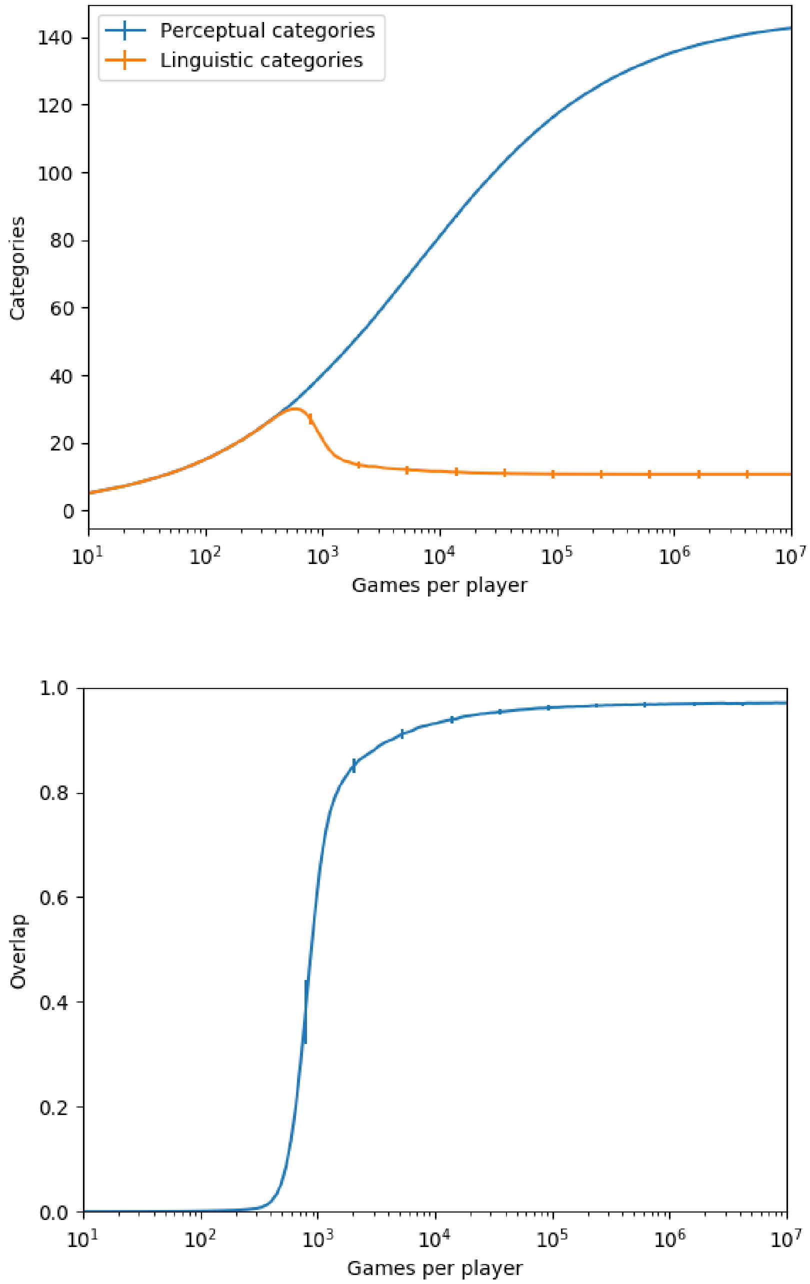


Figure 3.2: The dynamics of the core model. Above: the average number of perceptual and linguistic categories, and below: the overlap, both with respect to iterations per agent.

are misaligned between agents and agents cannot adjust their boundaries once they have formed. For communication to be successful, therefore, increasingly broad linguistic categories must develop with a single name, and consequently, their number falls. If there are too few, however, the communications will fail as the listener will not be able to distinguish the two objects by name. These two opposing pressures cause the average number of perceptual categories to plateau to around 11 ± 2. There is no intuitive reason for this figure. Although there is some variation in this value between simulations, in any given simulation it is effectively constant.

Having a significantly smaller number of terms than the capacity of the resolution power is a very naturalistic property of the model [17]. For example, many different shades of red are easily distinguished despite all being referred to as red (this applies to more specific terms as well).

Both the number of perceptual and linguistic categories follow qualitatively similar curves to the original paper. By visual inspection (only plots and not the raw data are present in their report), the number of perceptual categories is also quantitively the same. Initially, the number of linguistic categories closely replicates the original. However, it diverges from the number of perceptual categories slightly earlier in the replicated version, at around $(5.6 \pm 0.9) \times 10^2$, compared to $\sim 10^3$ in the original. Another difference is that the number of linguistic categories in the original data doesn't fully plateau. The last value is ~ 16 which is notably higher than in the replication. The cause of these differences is not clear. However, as there is no change in the essential phenomenological properties of these curves, the differences are in some sense trivial. Therefore, the validity of the replication, as a model of the emergence of vocabulary, still stands.

The dynamics of the overlap also differs from the original paper, despite tending towards the same value. The results presented here show a trivially small overlap at the start of the recorded window, whereas it was approximately ~ 0.6 in the original paper. Thus the overlap also increased considerably faster in the replication. Again, it is unclear what has caused these difference. Unlike with the categories, however, these difference are more significant. A slower increase in overlap is intuitively more probable, but an initially high value seems unlikely; as with any model, it is hard to asses which is more accurate. As an increasing success rate directly causes an increase in overlap (by unifying preferred names), one would expect them to be closely correlated. Interestingly, the replicated success rate and overlap curves follow almost identical trends, and so in some sense is more coherent. The follow-up paper, with a slightly modified mechanism, also recorded an initially small overlap [22].

Chapter 4

Dialect Continuum

The model of a dialect continuum will deviate from the core model in several ways. The first and simplest of which is to replace the complete network with a chain; agent i may now only communicate with agents $i \pm 1$ (without the edges wrapping). This network structure is a hypothetical (over-)idealisation of a dialect continuum.

This study will also introduce a new metric: the local overlap, which only averages the overlap between agents that share an edge. Of course, on a complete network, this is indistinguishable from the previously defined global overlap. The ratio between these two concepts (the GLOR for short) is key to assessing if the defining characteristics of a dialect continuum are present; a population with low GLOR will be more homogeneous on the local than the global scale and vice versa.

An additional variation will use multiple environments in the same network with the hypothesis that this will exacerbate these properties. For the sake of simplicity, I will only use two symmetrical environments on the network. These will be piecewise uniform distributions, with a region of high- and low-frequency stimuli. The two parameters h and t determine the height and transition point of the first distribution respectively, and the values $h' = {}^{(1-ht)}/_{1-t}$ and $t' = 1 - t$ determine the corresponding second distribution (a mirror image of the first around the midpoint $t = {}^{1}/_{2}$). Equation 4.1 shows the probability density function of a stimulus x. For the case when the two agents communicating reside in different environments, the speaker's distribution is used to generate the scene.

$$\operatorname{pdf}(x) = \begin{cases} h & x \leq t \\ \frac{1-ht}{1-t} & x > t \end{cases} \tag{4.1}$$

As seen in figure 2.2 the variance of category boundaries in regions of lower frequency is significantly higher. Measuring the homogeneity should take this into account, especially when the population spans multiple environments. The formula for the environmentally weighted overlap between two agents i and j can be seen in equation 4.2, where pdf_i is the probability density function of agent i.

$$\int_0^1 \frac{\operatorname{pdf}_i(x) + \operatorname{pdf}_j(x)}{2} \delta\left(\operatorname{pref}_i(x), \operatorname{pref}_j(x)\right) dx \tag{4.2}$$

After investigating the dynamics of the chain, the simulation will be rerun on a ring lattice where each agent may communicate with its ${}^{K}/_{2}$ neighbours on either side. As the now connected endpoints of the chain still maintain distinct environments, in some sense there is more friction along that edge. Therefore, in theory, the ring lattice should behave similarly to the chain. Although the ring doesn't directly correspond to the topography of dialect continua in real-world examples, it has two motivating advantages over a chain. First, it removes the edge effect of agents with a lower degree than the others. Second, it forms a natural base case for the construction of small-world networks.

There are two defining characteristics of small-world networks that often contradict one another. Despite a low average degree, there is a short path between most nodes; this trait is stereotypical of traditional random networks [24]. Additionally, they have a high level of clustering which is predominantly found regular lattices [24]. Clustering refers to the percentage of triples that are closed, a triple (being a set of three nodes with at least two edges) is closed if it has exactly three edges. The balance of these two properties makes this type of network particularly efficient in many contexts. Typically it is found in self-organising natural systems, especially social networks [24].

The Watts-Strogatz model is the classic method for randomly generating graphs with small-world properties [24]. The algorithm starts with a ring lattice of N node. Every node (or agent in this case) initially has $K/2$ edges to either side, connecting it to its nearest neighbours. Therefore, the degree of each agent is K. For each node, in turn, the $K/2$ edges to the right are rewired with probability β. Rewiring consists of uniformly selecting one of the other nodes (with the additional requirement that the graph remains simple) and replacing the current edge with a new edge connecting to the selected node. If K is too small, rewiring may isolate some nodes and cause the small-world properties to break down. This can be avoided by restricting K to be sufficiently larger than $\ln N \approx 4.6$ [24]. At $\beta = 0$, this produces a standard ring lattice, for $0 < \beta \ll 1$ small-world proprieties are observed, and for $\beta \approx 1$ the network is effectively random [24].

4.1 Chain Network

Figure 4.1 shows how the simulation evolves on a chain where every agent is exposed to the default uniform environment.

The dynamics of the number of perceptual categories is identical to that of the core model. The development of linguistic categories, on the other hand, shares many of the same features such as its divergence from the perceptual category curve and tend towards a lower value, but it quantitatively differs. Specifically, it diverges from the number of perceptual categories earlier, at around $(1.2 \pm 0.1) \times 10^3$. Paradoxically the period over which the number of linguistic categories is decreasing after peaking is significantly longer, and it is not clear if it eventually stabilises. However, as the same previously discussed pressures apply to this network, one would expect a final stable value to be reached.

Unlike in the core model, the local overlap between agents is increasing from the start of the simulation. It eventually reaches a plateau similar to the global overlap of the core model: 0.95 ± 0.01. The global overlap here, however, takes longer to develop and doesn't reach the same level.

The generally lower level of homogeneity compared to the complete network may be responsible for the slower stabilisation of the number of linguistic categories.

Towards the end of the simulation, both global and local overlap were extremely stable with a derivative of 1.83×10^{-10} and 4.21×10^{-12} respectively. Because of this level of stability, the network can be said to form a dialect continuum. There is very little change from agent to agent, but these accumulate to a lower level of global homogeneity. The GLOR (global-local overlap ratio) at the end of the simulation was 0.77 ± 0.07. In comparison, the complete network had a GLOR of 1.00.

The model was then extended further to investigate the impact of exposing agents to two different environments (distribution of stimuli). The first environment covers agents 0 to $N/2$, and the second covers the remaining agents.

The heat map, figure 4.2, shows how the environment affects the average GLOR (weighted by the environment) at the end of the simulation. The range of valid distributions is bounded

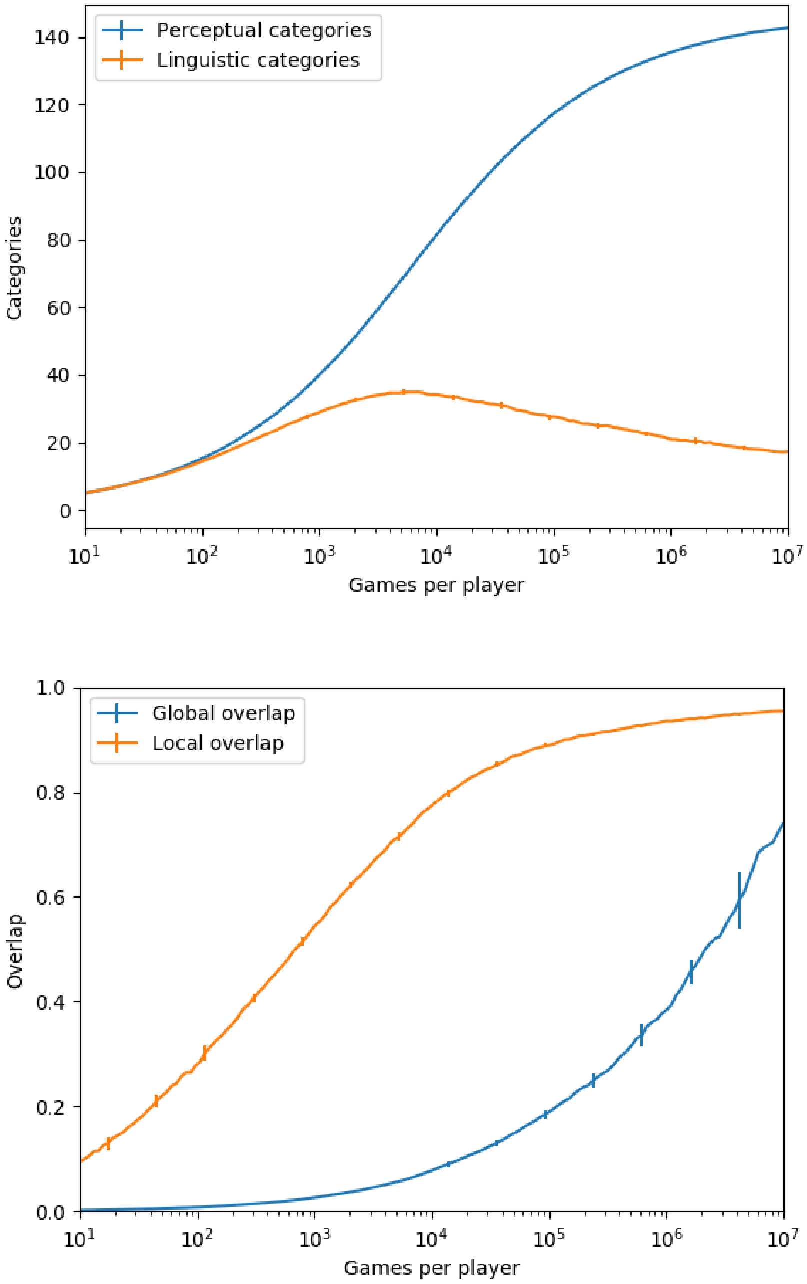


Figure 4.1: The dynamics of a chain. Above: the average number of perceptual and linguistic categories, and below: the local and global overlap, both with respect to iterations per agent.

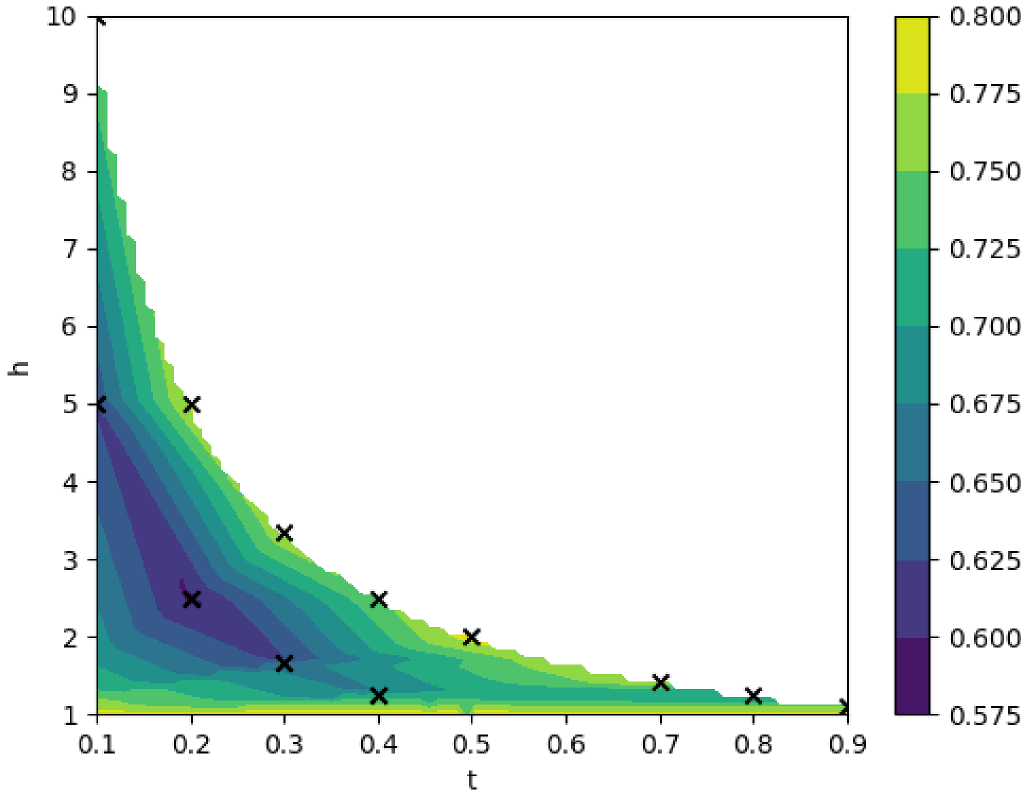


Figure 4.2: The effect of the environment on a chain. The GLOR at the end of the simulation with varying values of t and h that determine the environmental distribution.

by $ht \leq 1$, and as the distributions are symmetric (see above) only the range $h \geq 1$ needs to be investigated. Specifically, linear interpolation of a series of sample points (also marked on the plot) was used to generate the plot. The GLOR appears to converge towards local minima at the centre of the region (furthest from any boundary), GLOR $= 0.59$ at $h = 2.5$, $t = 0.2$. Unless specified otherwise these parameters will be used as they provide the most contrast for further investigations.

The next section will investigate how this model and indeed dialect continua respond to the path length and the clustering of a network by running the simulation on a modified ring lattice. Although this contradicts the natural topology of a dialect continuum, distinct environments on either end of the network can be used to maintain some sense of "distance" despite becoming connected.

4.2 Small-World Network

Before investigating the effect of small-world properties, it is necessary to check that a stable dialect continuum can also form on a ring lattice. Figure 4.3 shows the evolutionary dynamics (GLOR) of the simulation running on a ring lattice with varying K values (the degree of each agent). Initially, the value of K has a small impact on the GLOR. Figures 4.4 and 4.5 demonstrate that this difference originates in the local overlap, whereas the global overlap is effectively constant between values of K. When K is smaller agents have fewer neighbours, and so coming to a consensus (in that neighbourhood) is naturally easier, that is, takes fewer iterations. On the global scale, there hasn't been enough time for any substantial agreement regardless of the size of the neighbourhoods. However, as time goes on the reverse effect is observed; having larger neighbourhoods enables long-distance interactions that speed up the global convergence. For all values tested the GLOR reached a plateau that was less than that of a complete network, although this value varied with K.

In summary, although a ring lattice doesn't directly map to the topology of a dialect continuum, with the addition of multiple environments, the differing levels of homogeneity on the local and global scale can be preserved.

Using the Watts-Strogatz model the ring lattice (with $K = 6 \gg \ln 100$) was randomly rewired with varying β values. The two principal characteristics of the network that will change across this range are the average path length and the clustering. The average path length is $N/2K$ for a ring lattice and falls rapidly towards $\ln N / \ln K$ for a random network. The clustering coefficient also decreases with β. However, it does so much slower with the expected value being proportional to $(1-\beta)^3$. For many complex systems, these contrasting rates of change produce unusual dynamics that are not present at either extremum [24].

In this scenario, one would expect both the factors to have a synergistic effect on the GLOR. The decreasing average path length increases the global overlap. Intuitively clustering would have a strong positive feedback effect on the local overlap as agents that are part of strongly connected cliques have more affirmation of their vocabulary use. Both of which would, in theory, contribute to a decreasing GLOR.

Figure 4.6 shows how the path length, clustering coefficient and GLOR change with β. Surprisingly, for the first half of the graph ($\beta < 10^{-2}$) there is minimal change in the GLOR despite a changing path length so there is little dependence between the two. In the second half, the clustering coefficient begins to fall as the GLOR consistently increases demonstrating their correlation. Counter-intuitively the degree of separation (the average path length between agents) doesn't have as significant an impact on the divergence of dialects as the extent to which each dialect group is tightly connected.

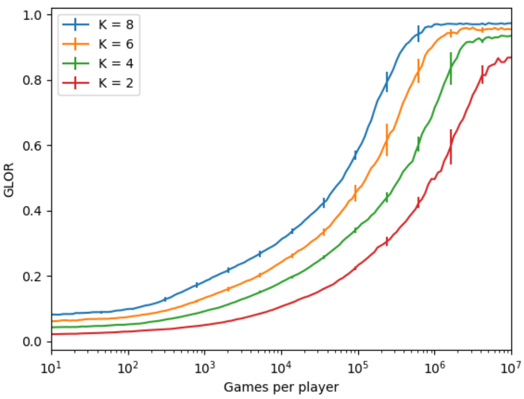


Figure 4.3: The dynamics of the GLOR with respect to iterations per agent on a ring lattice for differing values of K.

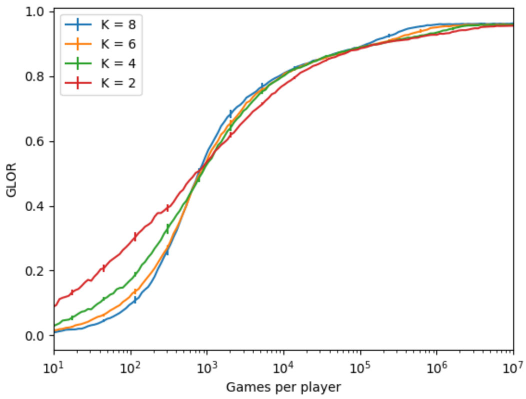


Figure 4.4: The dynamics of the local overlap with respect to iterations per agent on a ring lattice for differing values of K.

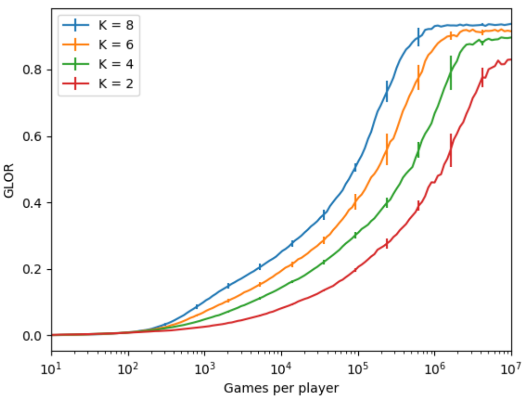


Figure 4.5: The dynamics of the global overlap with respect to iterations per agent on a ring lattice for differing values of K.

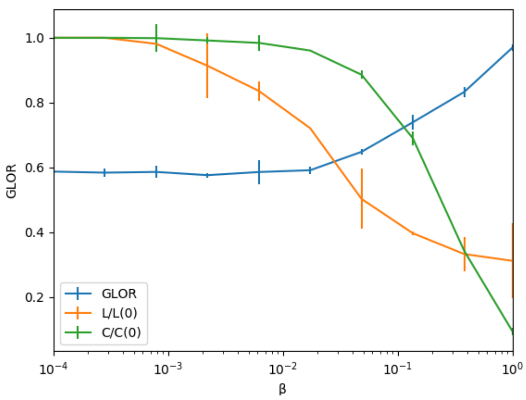


Figure 4.6: The GLOR at the end of the stimulation on small-world networks generated from the Watts-Strogatz model ($K = 6$) with varying values of β.

Chapter 5

Language Contact

In this study and model of language contact, two isolated populations (each of size $N = 50$) will first evolve to homogeneity on separate complete networks, before they are fully connected. It is possible that both languages survive, but the emergence of a new hybrid language would be more typical. In the latter case, if the model is accurate, the resulting language should be simpler, as indicated by fewer linguistic categories. Additionally, the emergence of a hybrid language should take place in a short period.

A snapshot of the two populations (and their respective languages) before contact will be saved and compared, using the overlap metric, with the new combined population and language. It is also possible to measure how the overlap varies across the perceptual space using a sliding window, equation 5.1 shows the formula for this.

$$\frac{\int_{x-\frac{w}{2}}^{x+\frac{w}{2}} \left(\mathrm{pdf}_i\left(y\right) + \mathrm{pdf}_j\left(y\right)\right) \delta\left(\mathrm{pref}_i\left(y\right), \mathrm{pref}_j\left(y\right)\right) dy}{\int_{x-\frac{w}{2}}^{x+\frac{w}{2}} \left(\mathrm{pdf}_i\left(y\right) + \mathrm{pdf}_j\left(y\right)\right) dy} \tag{5.1}$$

Each population will use a distinct set of names, as to trace the origin of names in use. Any new names generated after contact will again be marked. For each parent population, I will consider the average range of the perpetual space with preferred names that originate in that population, indicating what percentage of the new hybrid language originated in the lexicon of the language corresponding to that population. This metric scales each name by the range of stimuli that it covers. It also disregards any change in meaning as it is perfectly natural for semantics to morph over time, and especially with a significant cultural shift such as the introduction of a new group. It is also worth noting, however, that the overlap metric will incorporate any semantic changes.

The environment in which the two parent populations and the combined population develop will vary across the different simulations to explore the interplay of the environment and the resulting language's vocabulary. This will provide additional evidence for the model's accuracy. Specifically, I will contrast the following three scenarios: all three environments are uniform, the two parent populations develop in distinct environments but meet in a uniform environment, and arbitrarily they meet in the first environment. In this final scenario, the associated language should be a distinctly more competitive system of communication and consequently will form a large percentage of the emergent language's lexicon.

5.1 Uniform Environment

In the first simulation, the two parent populations (A and B) evolved under the same conditions as the combined population (C), that is, in a uniform environment.

Before contact the number of linguistic categories used by each population behaves much like the core model; initially, it increases to account for the rapid exposure to new stimuli, followed by a sharp fall as the population become more homogeneous and settles upon a smaller set of names. As with the core model, the local overlap before contact is initially low but rapidly increases and stabilises as the number of linguistic categories starts to fall and stabilise.

After contact, the average number of names used by an agent, surprisingly, more than doubles as the two groups are exposed to both vocabulary sets. The process of coming into contact with a mutually exclusive language appears to cause widespread destabilisation where names become more specialised as there are more competitors, linguistic categories narrow, and the number of disjoint concepts increase (although it eventually stabilises as the population re-establishes homogeneity). At the end of the simulation, there were 21.60 ± 3.40 linguistic categories, which is notably more than with the core model. As a corollary, the local overlap (figure 5.3) post-contact stabilises in a similar time frame, but not to the same extent, as pre-contact. The existence of two disjoint languages in the daughter language's genealogy results in more variation in language use, unlike in natural languages.

The post-contact overlap, however, is high hence the two languages have fully creolised, as opposed to the merely exchanging loanwords or the formation of a pidgin, where both populations retain their original language in conjunction with an inter-language. The convergence pre- and post-contact occur in a comparable period.

New names are generated as the perceptual space divides, not from miscommunication. Each agent's perceptual space becomes substantially divided before contact, and so there is no need for the creation of new names after this point. Hence the range for which the preferred name originating in population A and B respectively, cover the entire perceptual space - they are in complementary distribution. Thus the analysis may focus solely on the former (the range for which the preferred name originates in population A) which I shall refer to as the name ratio. The mean name ratio was 0.51, which is insignificantly higher than 0.5 (p-value = 0.55) its variance was also extremely low: 0.0035. In general, around half the vocabulary of the child language comes from population A and B respectively. Intuitively this is quite naturalistic as an overly-dominating language seems unlikely without any additional social pressures. It is hard to qualify this judgement with data as no such examples of uniform language contact exist.

Figure 5.3 shows how the overlap, between the new mixed population C and the snapshots of pre-contact populations A and B respectively, as it varies across the perceptual space. This is calculated using a window width of $w = \frac{1}{20} = 0.05$ (approximately the inverse of the number of linguistic categories). It also includes the total; as linguistic categories may have become misaligned, this isn't necessarily 1. For both population A and B, the overlap fluctuates around 0.5. Interestingly and unrealistically, the value seems to oscillate. The Pearson Correlation Coefficient between each data point and the average of adjacent data points was -0.49; there is a relatively strong negative correlation that proved significant (p-value = 0.00145). The origin of names used by the population C appears to consist of a regular alternating pattern. Natural languages don't show this behaviour; adjacent terms tend to have similar origins. For example, "university" and "college" are both from Latin via Old French. On the other hand, this could be due to non-linguistic factors such as prestige or emphasis on organised education (i.e. the environment). Additionally, it hard to quantify adjacency of real stimuli to formalise this argument.

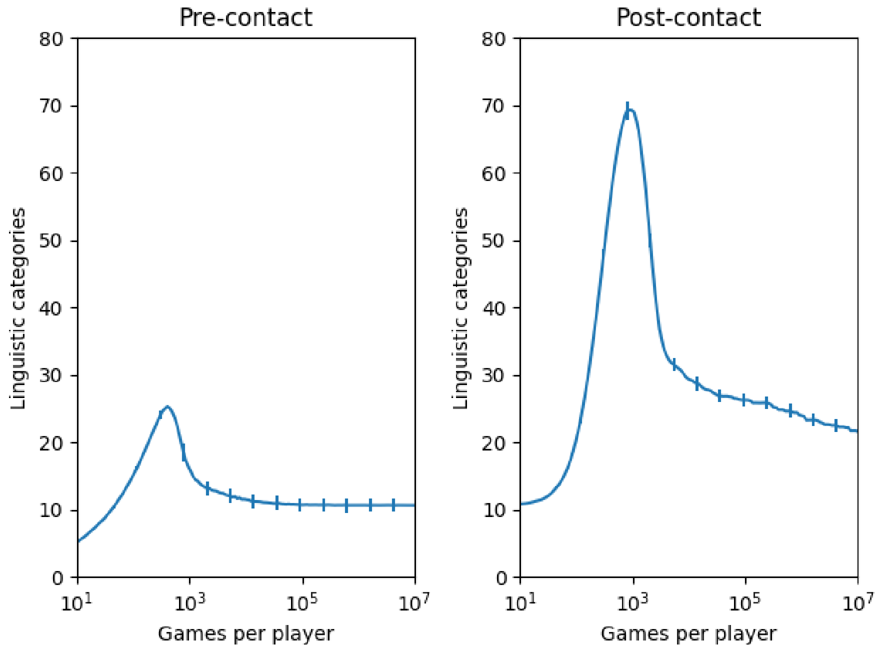


Figure 5.1: The linguistic categories with respect to the iterations per agent before and after contact where all populations are exposed to a uniform environment.

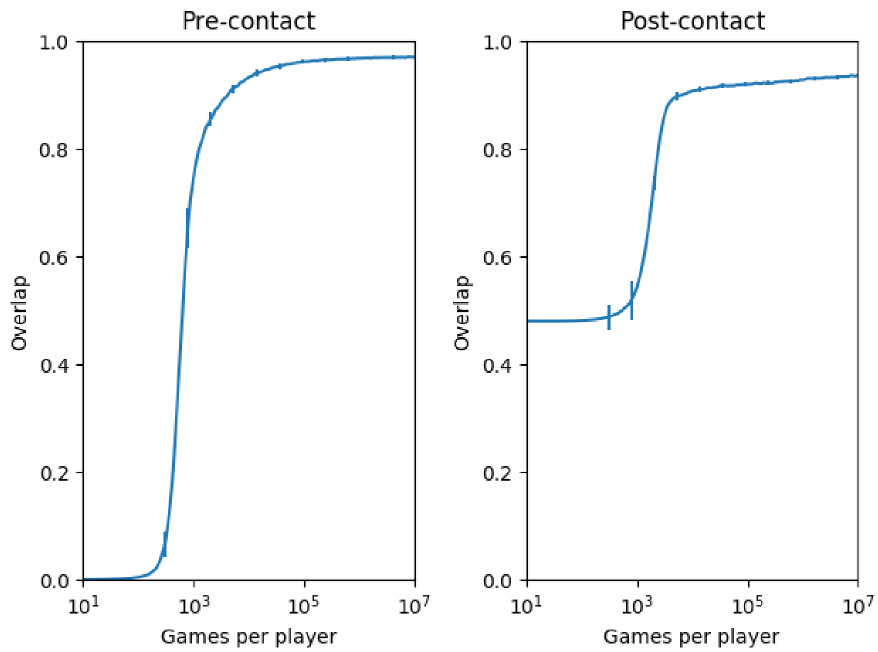


Figure 5.2: The local overlap with respect to the iterations per agent before and after contact where all populations are exposed to a uniform environment.

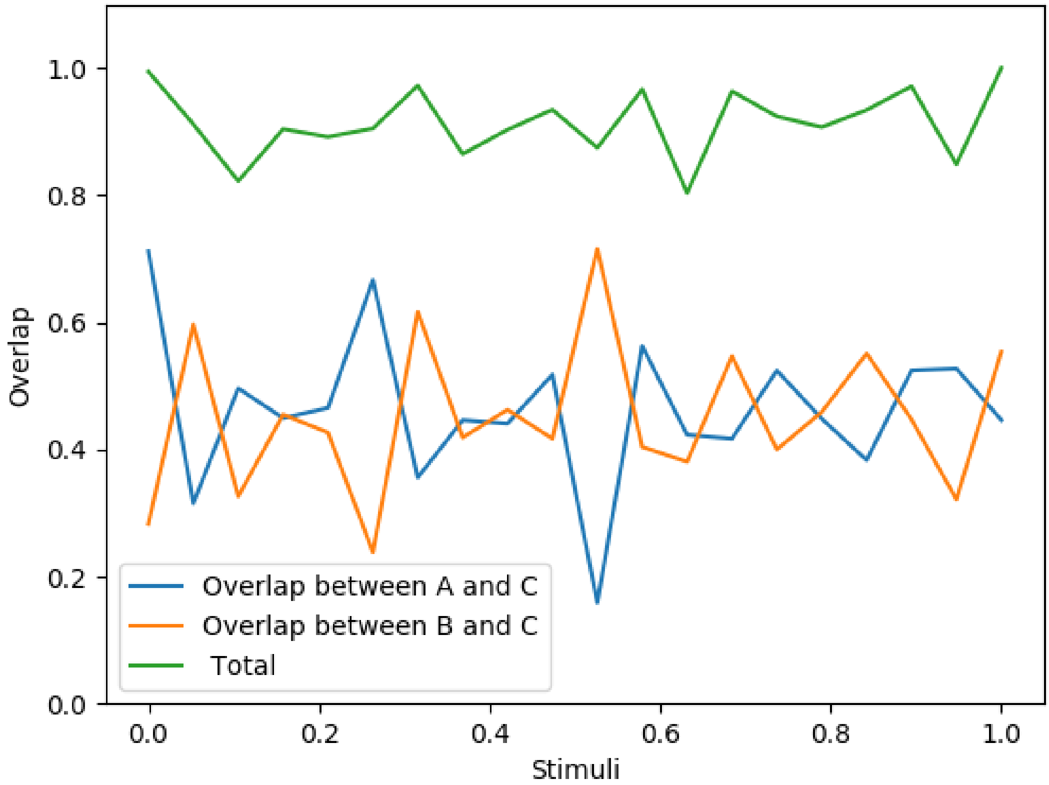


Figure 5.3: The overlap between population A and C, B and C, and the total, across the perceptual space (calculated using a window width of $w = 0.05$.

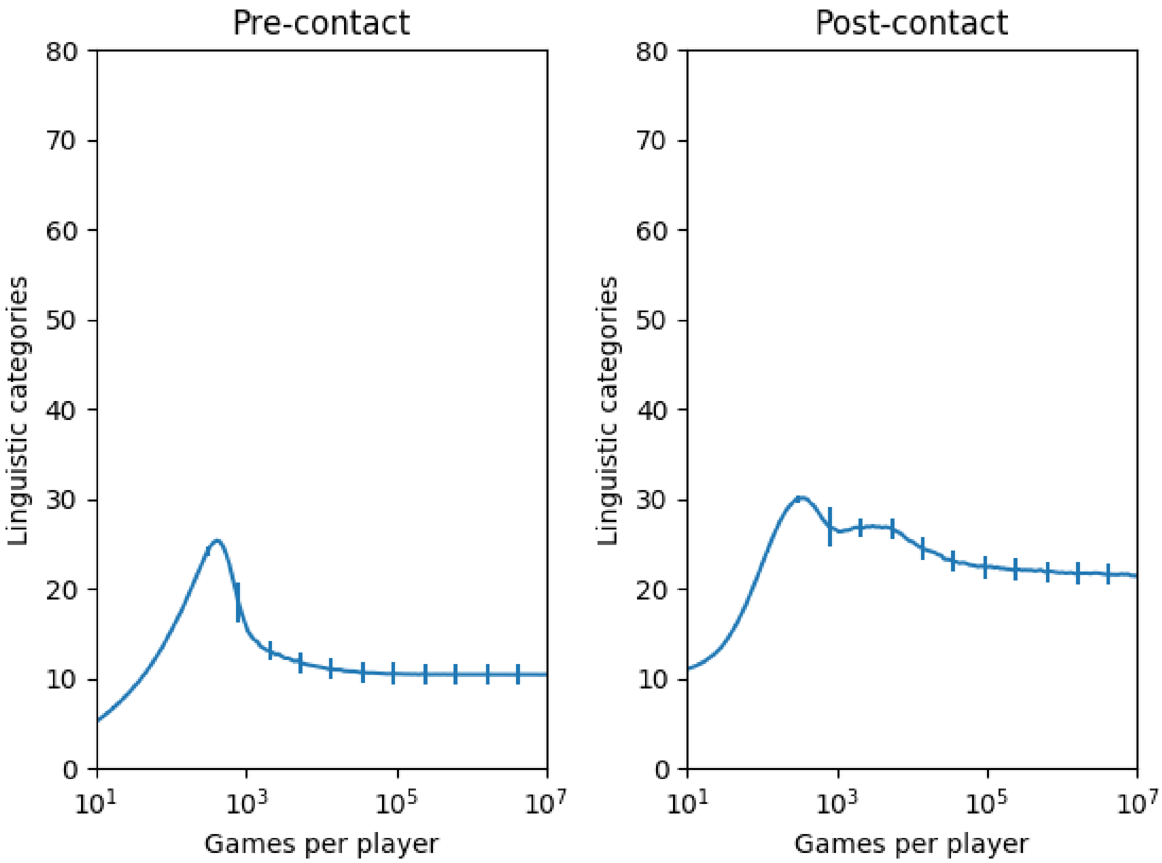


Figure 5.4: The number of linguistic categories in the total population before and after contact. Here populations A and B are exposed to distinct environments, and population C is exposed to a uniform one.

5.2 Non-Uniform Environment with Neutral Contact

In the next simulation the two parent populations, A and B, were exposed to different environments pre-contact. The two environments used were the ones found to produce the most significant separation on a chain network: $h = 4$, $t = 0.2$ and the mirror distribution respectively (see chapter 4 for details). However, when the two populations will come into contact, a uniform environment is still used.

In this scenario the number of linguistic categories post-contact (shown figure 5.6) didn't reach such an extreme value. The environments have mutually exclusive regions of higher frequency, so there is now a more obvious choice as to which name will propagate into the child language for any particular (category of) stimuli. This leads to certain terms dominating others, and the vocabulary overall being more stable. Interestingly, there was also an unexplained region of local stability at around 10^3 games per player. The local overlap had similar dynamics to the first scenario.

As with the first scenario, the name ratio was approximately 0.5. When considering how the overlap changes across the perceptual space, shown in figure 5.5, there is again an alternating effect in the centre of the perceptual space where both populations experience stimuli at equal frequencies. However, outside this range, as one would expect, it is significantly higher with one of the populations. The resulting mixed population, C, used names borrowed from A's vocabulary in regions that the population A was frequency exposed to and vice versa.

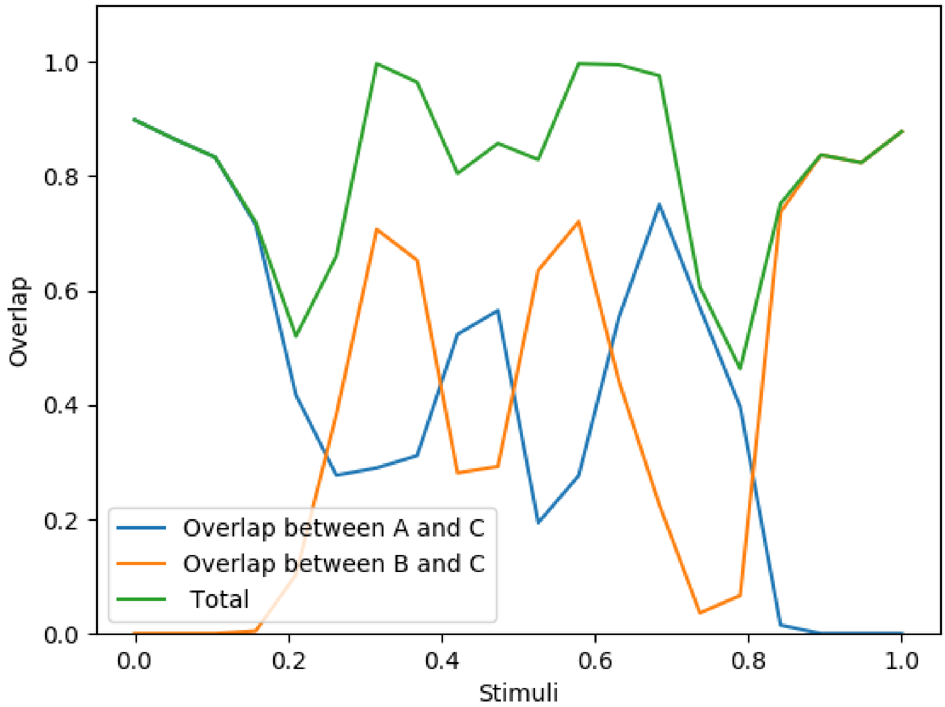


Figure 5.5: The overlap across the perceptual space when C is exposed to a uniform environment. Here populations A and B are exposed to distinct environments, and population C is exposed to a uniform one.

At the transition between these regions, at 0.2 and 0.8 respectively where the probability density function is discontinuous, the total overlap falls significantly. It is not clear why this is the case as the frequency of stimuli is still high for one of the environments.

5.3 Non-Uniform Environment with Bias Contact

When the two populations come into contact in one of the existing environments, without loss of generality A's environment was used, the number of linguistic categories increases slower after contact. There was also a second peak before finally stabilising.

Surprisingly the name ratio remained around 0.5. Although one of the parent languages was more suited to describing the environment, it still presented a region of instability that B's vocabulary could fill. However, the total overlap reflected the environment in which the combined population developed. There is some evidence of the oscillation in the centre of the perceptual space; it is however masked by other factors. The total overlap again falls at 0.8 but not at 0.2; as the environment in which the combined population develops has a region of high-frequency stimuli between 0.0 and 0.2, there is time for homogeneity to re-establish.

In summary, under this model when two developed communities were combined, a new hybrid language consistently forms. The model, however, failed to imitate the trend of simplification found in natural languages, and instead, the hybrid languages were more varied, often inheriting and maintaining near synonyms from both languages. Additionally, the period over which the hybrid language emerges is no faster from having existing language in use.

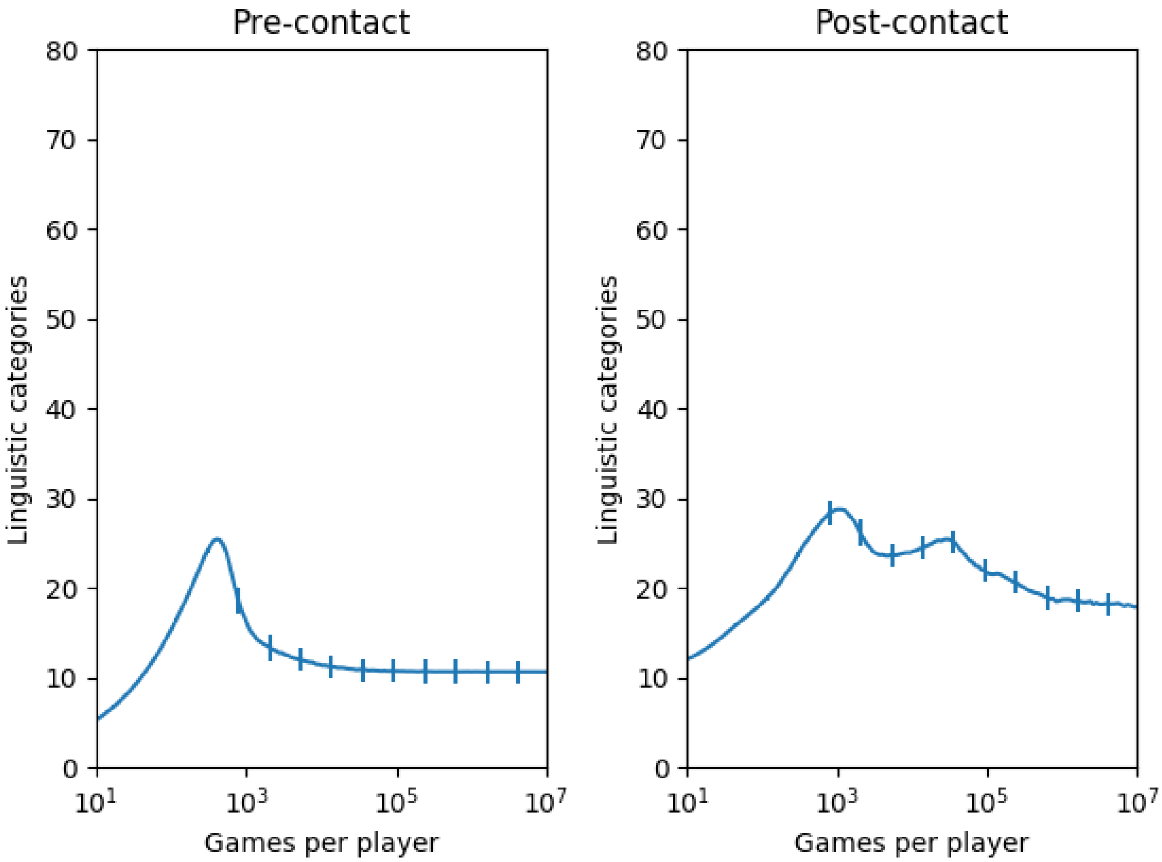


Figure 5.6: The number of linguistic categories in the total population before and after contact. Here populations A and B are exposed to distinct environments, and population A is exposed to a A's environment.

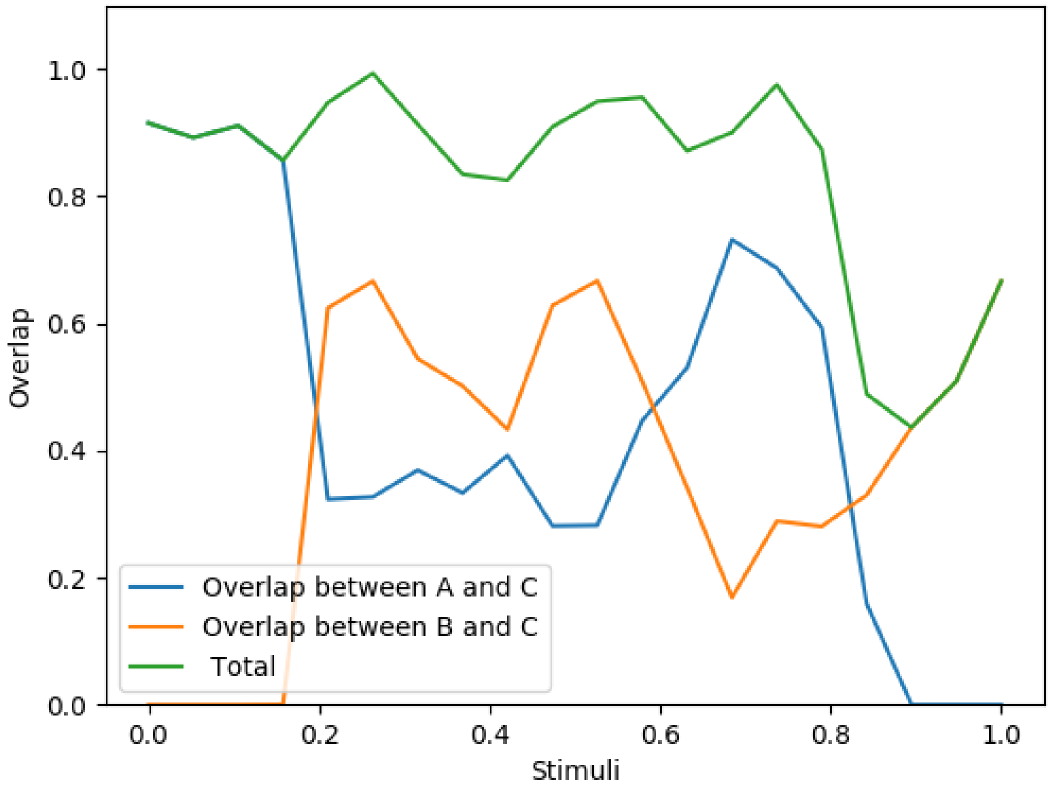


Figure 5.7: The overlap across the perceptual space. Here populations A and B are exposed to distinct environments, and population C is exposed to A's environment.

Chapter 6

Community Formation

The community structure (or modularity) of a given network quantifies how divided it is. A complete network for example, although maximally clustered, would have low modularity as it cannot be divided into communities with a higher degree of internal (or intra-community) connections than external (or inter-community) connection. Formally the modularity of a given partition is:

$$Q = \frac{1}{2m} \sum_{i,j=0}^{N} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta \left(c_i, c_j \right) \tag{6.1}$$

Where A is the adjacency matrix, k_i is the sum of the weights of edges emanating from node i, m is the sum of all weights in the matrix, and c_i is the proposed community of node i.

The task of finding the optimal partition is analogues to edge detection, and it is non-trivial. The Louvain heuristic algorithm is one of the fastest methods (although greedy and therefore not optimal) for finding such a partition [25]. Initially, each node is assigned to a unique community, it then proceeds in two stages [25]:

- The optimal community for node i is searched for by simply iterating through all possibilities that fix the remaining nodes. This process is done for each node in turn until there is no increase in modularity.
- The communities are now combined. That is a new (non-simple) graph is formed where each node represents a community from the previous step. Edges in the original network are mapped to the new network; if (i, j) is an edge between agents i and j, then there is an edge (c_i, c_j) between their prospective communities (this may result in self-loops and multi-edges). The algorithm is then repeated on this new network.

This paper will use an existing implementation of the algorithm that builds on the Python library NetworkX.

To simulate the formation of communities, the network in which the agents reside must be dynamic and updated based on the agent's preferences for with whom they interact (elective communication). These preferences evolve based on the success or failure of each interaction. However, the core unweighted structure would behave erratically under this model and be unlikely to converge. Instead, weights will be added to each edge of the network. These weights will encode the relative probability of those two agents communicating. The procedure for selecting agents on each iteration is, therefore, to first select the speaker s uniformly, and then to select the listener l with probability $\frac{w(s,l)}{\sum_{i=1}^{N} w(s,i)}$, i.e. the weight of the edge (s, l) encodes the unnormalised probability of l being selected given that s has been selected.

Sigmoidal functions have been successful in modelling natural adaptive processes. They also push values towards extrema and hence are an intuitive choice for updating weights. Traditional sigmoid functions, defined over the real numbers, will always have an image that is small than itself. In this application, however, we need to be able to repeatedly apply the function if there is a series of successful interactions and the combined effect should be monotonic. Hence, a bounded sigmoid function is needed which is defined on the, arbitrarily selected, unit interval (i.e., $[0, 1]$). In particular a smoothstep function (equations 6.2) fits these requirements. First used in computer graphics, it has a first and second derivative that is zero at the bounds making it particularly smooth. Note the implementation of this function is slightly adapted to avoid floating-point rounding issues.

$$\text{smoothstep}(x) = \begin{cases} 0 & x \leq 0 \\ 6x^5 - 15x^4 + 10x^3 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases} \tag{6.2}$$

The increasing and decreasing functions are $x \mapsto \text{smoothstep}(x + \lambda_1)$, and $x \mapsto \text{smoothstep}(x - \lambda_2)$ respectively where $\lambda_1, \lambda_2 \in [0.15, \infty)$. I shall refer to these two parameters as the inclusion and exclusion factor. Note, this range ensures repeated application of the function is monotonic, although in practice only small values of λ produce naturalistic results. The following series of experiments will test the values: 0.15, 0.30, and 0.45.

To assess the effect of convergent evolution on the formation of communities I will contrast the model with dynamic weights and elective communication (the live model) with a null model. The null model will differ in that no genuine communication occurs - each iteration succeeds or fails randomly with an even chance.

6.1 Results

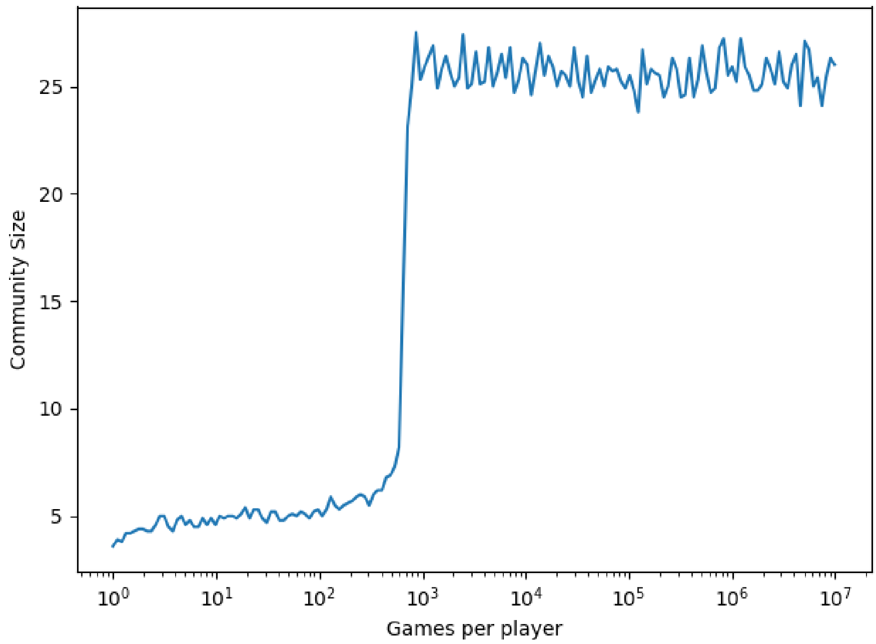
Figure 6.2 on the following page shows the evolutionary modularity of the null model with varying values of λ_1 and λ_2. For larger exclusion factors ($\lambda_2 > 0.15$), the community structure appears very rapidly between 10^2 and 10^3 games per agent (the exact value varies between parameters). Once the community structure has formed, it is very stable. These dynamics are very similar to that of the overlap in the core model in that they both exhibit a particularly steep, yet smooth, transition from disorder to order. In the context of the emergence of communication systems, this phenomena has been studied [26]. This analogous effect indicates a broader trend of systems rapidly transitioning.

When the exclusion factor was small, however, the dynamics were notably different. Although for all values of λ_1 there was a rise in modularity, it was only a nominal quantity and was not stable.

Surprisingly the modularity of the live model (figure 6.3) behaved almost identically. As the community structure appears firsts, the homogeneity must develop to fit its mould, oppose to inducing new communities. Of course, this is a result of the population initially being in heterogeneous; a population that is already developed may behave differently. Both the size and stability of the communities are comparable between the live model as seen in figure 6.1. Note this figure actually shows the number of communities which of course is inversely proportional to their average size.

Interestingly, these results show that a community structure can form without convergence evolution. Additionally, as the unhindered community structure (of the null model) emerges slightly faster than the language of the live model, the live model is at a disadvantage. Initially, it has a lower success rate which postpones community formation.

Intuitively both parameters would have symmetric effects on the dynamics. However, in both the null and live model only the exclusion factor made a qualitative difference. This parameter is responsible for the decrease in weight and determines to what extent will agents exclude others



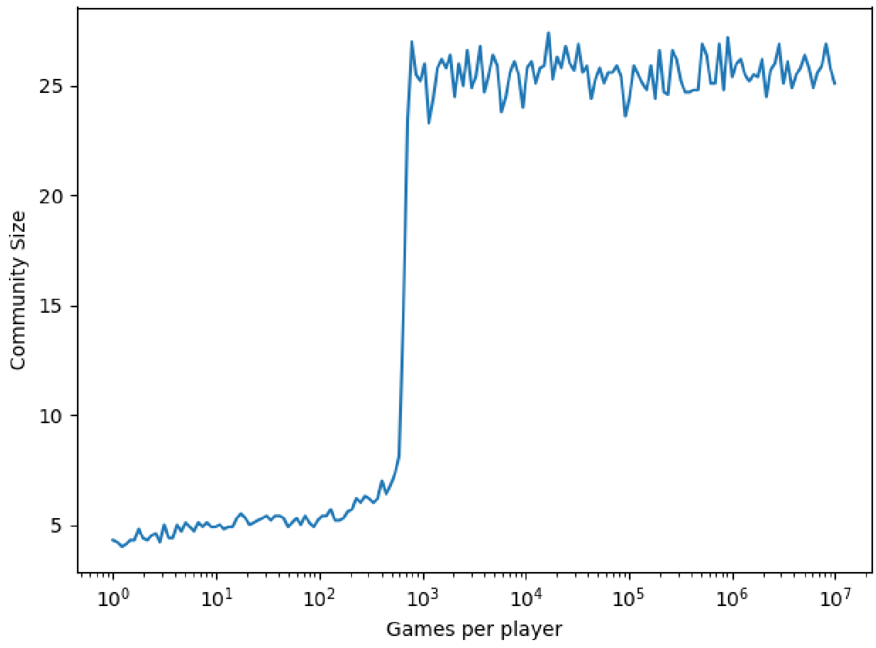


Figure 6.1: The average number of communities over time with $\lambda_1 = \lambda_2 = 3.0$, above: the null model, and below: the live model. For all parameters that induce community structure, the dynamics are comparable.

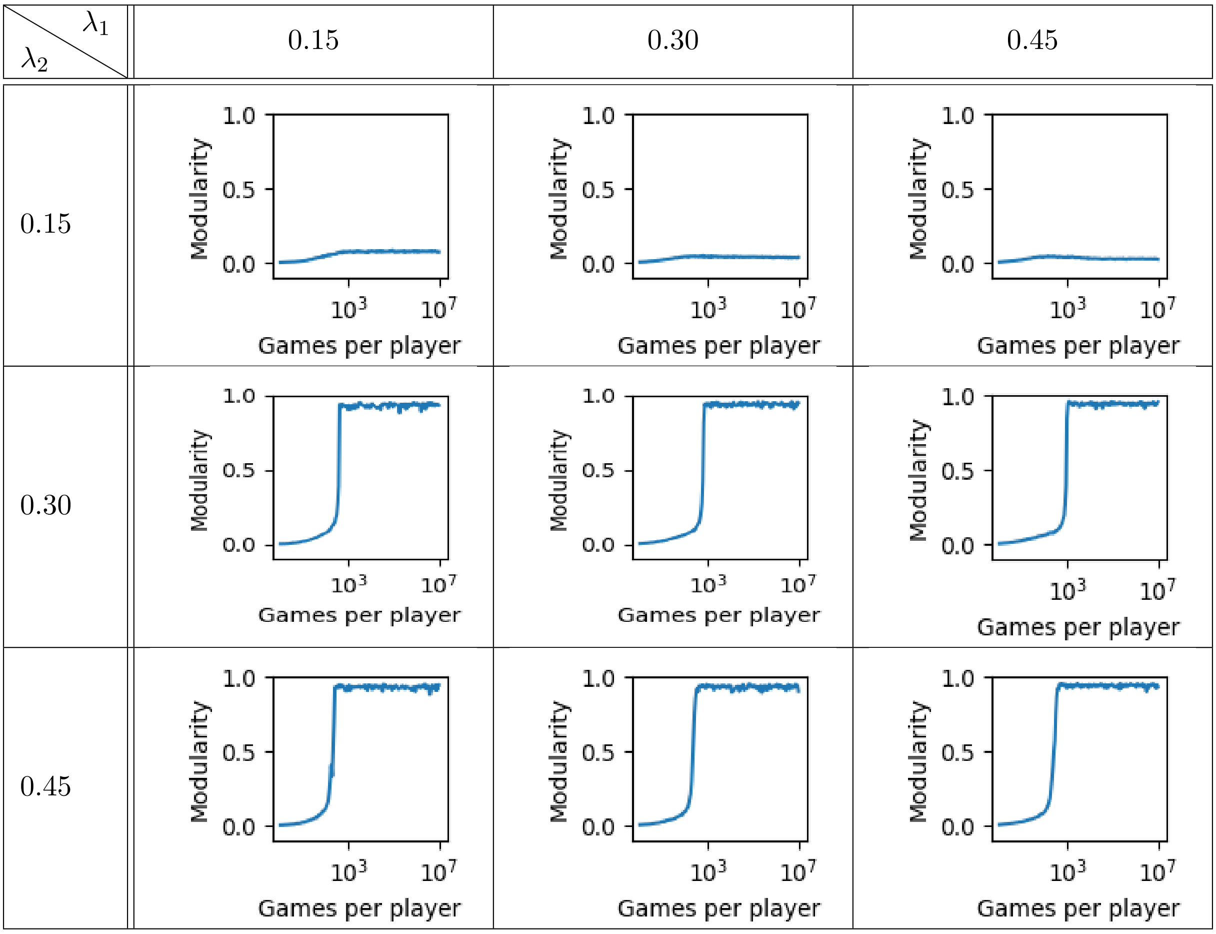


Figure 6.2: The modularity (community structure) of the null mode over time with varying inclusion and exclusion factors.

following unsuccessful interactions. In this sense, community formation is a fundamentally negative process.

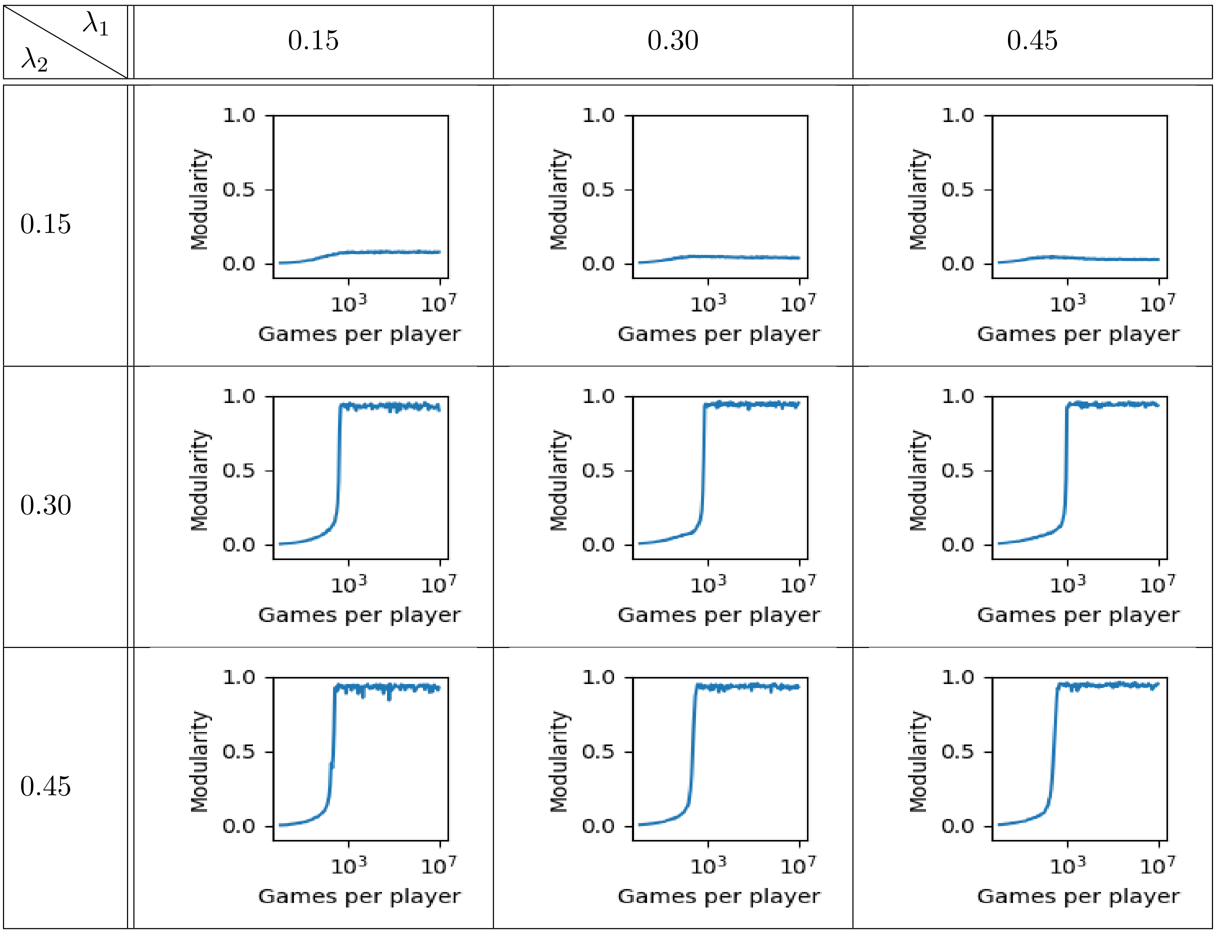


Figure 6.3: The modularity (community structure) of the live model over time with varying inclusion and exclusion factors.

Chapter 7

Conclusion and Evaluation

The core model failed to completely replicate the dynamics presented in the original paper. However, it did produce the same high-level naturalistic traits, namely a coordinated series of linguistic terms that functionally describe a continuous perceptual space. Therefore, it is still a valid building block for the extensions presented in this paper.

The first study successfully demonstrated that simple agents in a uniform environment are capable of establishing a dialect continuum on a chain network via communication alone - the primary objective. These agents have no cultural values or asymmetric ties to particular agents and have an unbounded memory. It is surprising, therefore, that this dialect structure is stable as intuitively the system would tend towards homogeneity, even if the longer average path length slows this process down. Although, as expected, the addition of multiple environments to the network caused a more distinct dialect continuum to form.

In addition to showing this is possible, I also had some success in uncovering some of the explicit requirements of the network, although this needs to be formally tested.

The same effect of local homogeneity without global homogeneity is achievable in a ring-like structure that agents wrap around, here there are several, depending on the specific parameter, long-distance connections that intuitively should breakdown the continuum. This is the first piece of evidence that dialect continua aren't a product of separation (a long average path length).

By exploring a series of networks with small-world like properties (that is with short average path length and high clustering) it appeared to be the case that there was no correlation between path length and the formation of a dialect continuum. Instead, the tendency for agents to cluster together determined the stability of the difference between local and global homogeneity. However, as a by-product of the method used to generate these networks, the Watts-Strogatz model, both path length and clustering are monotonic. As a result, without assuming the dependence on these measures is linear, is it is impossible to separate and statistically test any correlation. Possible future work may involve formally exploring this conclusion by examining the dynamics of a more comprehensive range of networks. On the other hand, this may diverge further from the standard topology of social networks in general, consequently bring into question whether any discovered mechanism is responsible for dialect continua in practice.

In the second study, the model presented many unnaturalistic properties. Although a hybrid language formed, it was considerably more complex than its parent's language, contradicting the general pattern of pidgins and creoles being relatively simple. On the other hand, it has been suggested that the transition from Old English to Middle English was, in fact, a creolisation with Old French, specifically one that resulted in a more complex language. Both these languages were comparatively very similar (both in internal structure and their environment) when compared to most other examples of creoles. It is possible, therefore, that the simplification process only occurs in languages that surpass a certain threshold of dissimilarity. Although the data presented in this paper supports this hypothesis; the number of linguistic categories that formed when the populations had developed in different environments was significantly smaller, it is fairly inconclusive.

Homogeneity between the mixed population and its two parents was equally inconsistent with natural languages. Although in ranges of the perceptual space where one language dominated the behaviour was as predicted. In other ranges (and in transitions between them) it was unrealistic: specifically, the origin of vocabulary oscillated. A more monotonic trend with smoother transitions between regions of vocabulary would be more consistent with natural languages.

In summary, this is a poor model of language contact, demonstrating that the emergence of hybrid languages is different from the process by which linguistic categories form in an initially language-less community, or at least is it inseparable from non-linguistic factors not present in the model. More naturalistic models, way include language users translate terms into their pre-existing language matrix, by implicitly aligning it to their linguistic boundaries. As a possible extension and hypothesis, a model which replaced linguistic boundaries with perceptual boundaries before the two groups come into contact might better replicate this phenomenon, and specifically induce the simplification process.

The study of homophilic community formation is usually approached from the perceptive of opinion dynamics. However, due to its similarities with the emergence of dialects, it is also a socio-linguistic phenomenon. I constructed a model that was capable of developing homogeneous groups from an initially unstructured population. However, I also have shown that this process can occur without convergent evolution where success is deemed arbitrarily. This observation would lead us to conclude that, the splintering of unstable large networks, later induces pockets of homogeneous sub-networks, instead of being caused by a feedback effect from the convergence. However, it is important to remember that this only explores initially unstructured networks with no preexisting language. It may also be the case that this abstraction doesn't preserve the timescale of the two processes (convergence of language and community formation) and so any interaction between them may have different dynamics in practice.

Additionally, the primary factor which determines the formation of these communities appears to be negative, that is an effect that weakens the weights between edges has more impact than the dual effect which strengthens them.

Overall, some of these inter-language relationships have been demonstrated to be intimately linked to network structure as hypothesised. For these case, loose explanatory models have been proposed, although more rigorous testing of the conclusion is required if the mechanism behind their emergence is to be completely isolated. On the other hand, some of the core traits of these phenomena weren't successfully modelled (e.g. in the second study), and so it is hard to make inferences about the traits origin. However, as discussed above, I believe this points to a different mechanism at play rather than the model proving ineffective in these extended contexts. Additionally, I think there is clear ground for further research, specifically into the broad tendency of large (social) networks to be unstable and sharply transition into stability. Often this is viewed as a side effect of the processes that occur on the network, but in fact, may be more of a driving force than previously thought.

Bibliography

[1] R. Wardhaugh. *An Introduction to Sociolinguistics*. 5th ed. Blackwell, 2006. ISBN: 9781405135597.

[2] A. François et al. *The Routledge Handbook of Historical Linguistics*. Ed. by C. Bowern and B. Evans. 1st ed. Routledge, 2014. ISBN: 9780415527897.

[3] D. Crystal. *Language and the Internet*. Ed. by K. Taylor. 2nd ed. Cambridge University Press, 2009. ISBN: 9780511487002.

[4] J. L. Bonenfant. "History of Haitian-Creole: From Pidgin to Lingua Franca and English Influence on the Language". In: *The Review of Higher Education* 3 (2011).

[5] L. Bloomfield. *Language*. 7th ed. Motilal Banarsidass, 2012. ISBN: 9788120811966.

[6] J. K. Chambers and P. Trudgill. *Dialectology*. 2nd ed. Cambridge University Press, 1998. ISBN: 9780521593786.

[7] R. Hickey. *The Handbook of Language Contact*. 1st ed. Wiley-Blackwell, 2010. ISBN: 9781405175807.

[8] J. Arends et al. *Pidgins and Creoles: An introduction*. Ed. by J. Arends, P. Muysken, and N. Smith. John Benjamins, 1994. ISBN: 9789027252364.

[9] S. Kirby. "Natural Language From Artificial Life". In: *Artificial Life* 8 (2002).

[10] S. Horobin and J. Smith. *An Introduction to Middle English*. Edinburgh University Press, 2002. ISBN: 9780748614813.

[11] M. McPherson, L. Smith-Lovin, and J. M. Cook. "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* (2001).

[12] N. Chonksy. *Knowledge of Language: Its Nature, Origins, and Use (Convergence)*. Praeger, 1986. ISBN: 9780275917616.

[13] J. R. Hurford. "The Evolution of Language and Languages". In: *The Evolution of Culture* (1999). Ed. by R. Dunbar, C. Knight, and C. Power.

[14] E. Rosch. *Cognition and Categorization*. 1st ed. Lawrence Erlbaum, 1978. ISBN: 9780470263778.

[15] T. Regier, P. Kay, and N. Khetarpal. "Color naming reflects optimal partitions of color space". In: *PNAS* 104 (2007).

[16] T. Belpaeme and J. Bleys. "Explaining Universal Color Categories Through a Constrained Acquisition Process". In: *Adaptive Behaviour* 13 (2005).

[17] A. Puglisi, A. Baronchelli, and V. Loreto. "Cultural route to the emergence of linguistic categories". In: *PNAS* 105 (2008).

[18] L. Steels and T. Belpaeme. "Coordinating perceptually grounded categories through language: A case study for colour". In: *Behavioural and Brain Sciences* 28 (2005).

[19] J. Alvarez-Galvez. "Network Models of Minority Opinion Spreading: Using Agent-Based Modeling to Study Possible Scenarios of Social Contagion". In: *Social Science Computer Review* 34 (201).

[20] D. Kalociński, M. Mostowski, and N. Gierasimczuk. “Interactive Semantic Alignment Model: Social Influence and Local Transmission Bottleneck”. In: *Journal of Logic, Language and Information* 27 (2018).

[21] B. Kozma and A. Barrat. “Consensus formation on adaptive networks”. In: *Journal of Physics A: Mathematical and Theoretical* (2008).

[22] V. Loreto, A. Mukherjee, and F. Tria. “On the origin of the hierarchy of color names”. In: *PNAS* 109 (2012).

[23] P. Kay et al. “The World Color Survey”. In: *Center for the Study of Language and Information* (2009).

[24] D. J. Watts and S. H. Strogatz. “Collective dynamics of “small-world” networks”. In: *Nature* 393 (1998).

[25] V. D. Blondel et al. “Fast Unfolding of Communities in Large Networks”. In: *Journal of Statistical Mechanics Theory and Experiment* (2008).

[26] A. Baronchelli et al. “Sharp transition towards shared vocabularies in multi-agent systems”. In: *Journal of Statistical Mechanics: Theory and Experiment* (2006).