Master's Thesis in General Linguistics

Exploring the Effects of Population Size on Language Change: An Integrated Approach using Agent-Based Modeling and Statistical Analysis

Yuliya Mkhayan

Student number: 4080819

yuliya.mkhayan@student.uni-tuebingen.de

December 2023

First Examiner and Supervisor:
Dr. Christian Bentz

Second Examiner:
Prof. Dr. Gerhard Jäger

Antiplagiatserklärung

Hiermit versichere ich, dass ich die Arbeit selbständig verfasst habe, dass ich keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe, dass ich alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe, dass die Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist, dass ich die Arbeit weder vollständig noch in wesentlichen Teilen bereits veröffentlicht habe, und dass das in Dateiform eingereichte Exemplar mit dem eingereichten gebundenen Exemplar übereinstimmt.

Tübingen, Dezember 22, 2023

G. Mkhayan

Yuliya Mkhayan

Contents

Abstract

The effect of population size on language change has been the subject of a range of linguistic studies, with results differing in their evaluation of its significance. This thesis aims to contribute to this discussion by applying a hybrid approach that integrates agent-based modeling and statistical analysis. The implemented agent-based model is a multi-speaker Moran model. It is used to simulate language change under the action of neutral change, replicator selection, and interactor selection. The data generated by this model are then subjected to a quantitative analysis using multiple linear regression. The statistical analysis reveals that population size significantly affects language change when selection is at play, both in terms of replicator and interactor selection. Notably, the calculated effect sizes indicate a stronger relationship between population size and language change given the presence of replicator selection. Furthermore, the study shows that replicator selection is the only mechanism by which an innovation can completely replace a convention or become the predominant variant.

Acknowledgements

First and foremost, I would like to thank Christian Bentz for his excellent supervision and the countless hours he devoted to our extensive discussions on my Master's thesis. His guidance was a great support to me, and I am very grateful for his commitment.

I would also like to thank my husband, family and friends who have always supported me along the way. Their encouragement and belief in me have kept me moving forward.

List of Figures

List of Tables

1 Introduction

Language, as a complex adaptive system (Beckner et al., 2009), is a dynamic, constantly changing entity that evolves throughout the processes of human interaction by adapting to the communicative needs of its speakers. The structure and evolution of language is influenced by a variety of factors, one key among them being the size of the population. The effect of population size on language change has been explored in a number of scientific studies, with equivocal results. This thesis is therefore an attempt to contribute to the existing body of knowledge in this area and provide insights that can be applied to the study of language change in various contexts.

For this purpose, I restrict my attention to one step in the process of language change, i.e., propagation or, in other terms, diffusion, which represents the change in the proportions of linguistic variants in a community of speakers over time. I therefore assume the existence of two competing linguistic variants called innovation and convention, with the former being the minority variant and the latter the majority. My aim is to investigate how the proportion of innovation varies over time in populations of different sizes. Moreover, I assume that populations of speakers are organised into three different types of social networks: regular, small-world, and random. Previous research has shown that the structure of social networks plays a crucial role in shaping the dynamics and outcomes of language change. Thus, in this thesis, it serves as an additional factor influencing the process of innovation diffusion.

Furthermore, various constraints have been shown to affect the propagation of linguistic variants. These are expressed here through three mechanisms of language change identified in Baxter et al. (2009): neutral evolution, replicator and interactor selection, each thought to drive the diffusion process of innovation in a certain way. Neutral evolution assumes that the change in the proportion of innovation is driven solely by random fluctuations, i.e., without any selection processes involved. On the contrary, both replicator and interactor selections assume the presence of some kind of bias, the former being functional, related to the innovative variant itself, and the latter being social, related to the social status of the speaker who uses the innovative variant. This means that in the case of interactor and replicator selections, language change no longer occurs purely by chance but under the influence of selective pressures. For instance, in the context of replicator selection, innovation may have an easier pronunciation than convention, which increases the probability of its successful propagation. If, for instance, interactor selection is involved, innovation being used by a

speaker of higher social status may be more likely to propagate than convention being used by a speaker of lower social status.

This thesis thus investigates how population size affects, and whether it affects at all, the propagation of innovation depending on which mechanism of language change is involved. Moreover, the effects of three additional factors on the diffusion process are examined. These are: social network structure, the proportion of leaders, i.e., speakers with a high social status, and selection strength.

In order to explore the trends between language change and the factors of interest, I use an approach that combines computer simulations and statistical analysis. A similar approach in examining propagation has been employed in some other studies (e.g. Nettle, 1999a; Ke et al., 2008). However, the methodology employed in this thesis differs from that used in previous research. To simulate language change, I implement the multi-speaker generalization of the Moran model (Blythe, 2012), a version of the Utterance Selection Model (Baxter et al., 2006) whose mathematical definition is based on the evolutionary model of language change proposed by Croft (2000). The multi-speaker Moran model is a kind of an agent-based model where language users are agents endowed with certain attributes and behaviours, each agent performing a predefined range of actions when interacting with his or her interlocutor. Computer simulations are carried out using different values of the factors of interest. In this way, different scenarios of language change are modelled. The results of the simulations are analysed by means of multiple linear regression models, each defined for a specific mechanism of language change. This allows to identify general trends between the proportion of innovation and population size, as well as other factors such as social network type, selection strength, and the leader group size, in relation to each individual mechanism of language change. In addition, factors that significantly influence the propagation of innovation are identified in this way.

The rest of the thesis is organized as follows. In Section 2, I give a general introduction to the empirical and computational background surrounding the relation between language change and population size as its driving factor. Section 3 introduces the agent-based model of language change and describes its main components. In addition, the process of implementation of the baseline model is described. Mechanisms of language change and their integration into the agent-based model are the subject of Section 4. Section 5 is concerned with the parameter definitions for the agent-based model used in individual simulations. In addition, simulation results are presented and described for each of the mechanisms of language change. The output of statistical analysis for each mechanism is presented in Section

6. The effect of each individual factor on language change is discussed as well. Section 7 concludes the thesis by summarising the main results and outlines the future work.

2 Language Change and Population Size

The question of whether there is a relationship between language change and population size has been addressed in a number of empirical studies. However, these studies reveal contradictory results. Some suggest that the size of population has a significant effect on various aspects of linguistic structure as well as the rate by which language changes (e.g. Hay and Bauer, 2007; Fenk-Oczlon and Pilz, 2021; Lupyan and Dale, 2010; Bromham et al., 2015). On the other hand, the findings of other studies do not show significant relationship between population size and grammatical properties of language (e.g. Moran et al., 2012; Greenhill et al., 2018). In addition to the empirical studies, computer simulations have also been conducted to investigate the impact of population size on patterns and rates of language change. These simulations have yielded mixed results on the relation between population and language change as well (e.g. Nettle, 1999a; Wichmann and Holman, 2009). Therefore, the relationship between population size and grammatical properties of language, as well as the rate of language change, remains a subject of ongoing debate.

This section provides an overview of the empirical and computational background related to the relationship between language change and population size. It presents the evidence from both empirical and computational studies that explore the relationship between population size and the structural systems of language, such as morphology and phonology. In addition, the section reports the outcomes on the effect of population size on rates of language change, as well as on non-structural properties of language, such as vocabulary size.

2.1 Population Size and Language Complexity

In the study conducted by Hay and Bauer (2007), a positive correlation between population size and phoneme inventory size was discovered for a sample of 250 languages from diverse families. This correlation was observed for both vowel and consonant inventories. Atkinson (2011) replicated these findings using data from the database WALS (Word Atlas of Language Structures). In addition, he reported a negative correlation between the distance from Africa of a given language and phoneme inventory sizes. The results of Hay and Bauer (2007) and Atkinson (2011) were further confirmed by Wichmann et al. (2011) through analysis of a much larger sample of languages. Fenk-Oczlon and Pilz (2021) reproduced the results of

former studies that found a positive correlation between population size and phoneme inventory size for the sample of 61 languages from 18 families. They attributed this correlation to syllable complexity, which in turn correlates highly positive with phoneme inventory size. According to this, languages with larger speaker populations tend to have complex syllable structures and are therefore expected to have a large phoneme inventory size.

On the other hand, Moran et al. (2012) find little support for population size as an explanatory predictor of phoneme inventory size. They thus oppose the positive correlation between these two variables found in Hay and Bauer (2007) and see it as an artefact of the authors' statistical technique and biased data set. Other than Hay and Bauer (2007), Moran et al. (2012) use a much larger and more diverse sample of the world's languages and consider their genealogical relatedness. Donohue & Nichols (2011) also believe the correlation between population size and phoneme inventory size is artefactual and even doubt its existence. By examining many hundreds of languages with extensive genealogical and geographical coverage and distribution, the authors find that while there may be correlations within specific geographical areas, they cannot be observed globally.

Lupyan & Dale (2010) utilized the WALS dataset containing information on over 2,200 languages to explore the relationship between morphological complexity and the number of language users. Their findings suggested that there is a negative trend between population size and the level of morphological complexity: Languages spoken by larger groups demonstrated simpler inflectional morphology compared to languages spoken by smaller groups. In other words, languages with a larger number of speakers tend to be structurally simple, while languages spoken in smaller communities tend to develop greater structural complexity.

Trudgill (2011) also claimed that high levels of complexity, morphological as well, are likely to be observed in smaller populations. However, it is important to note that not only the small size of the population may account for high levels of linguistic complexity but also other factors such as high density of social networks, high isolation of speech communities and a high proportion of child learners (Trudgill, 2011).

That there may be other factors, besides population size, that affect the level of linguistic complexity was also demonstrated in computer simulation studies. The results of simulations conducted in Spike (2018) showed that larger populations are quite capable of maintaining linguistic complexity but under the condition that there is no strong preference for simplicity (p. 490). In addition, small populations have even been shown to lose complexity as a result of stochastic drift effects, unless there is a strong anti-simplicity bias.

Relatively recent computer simulation study in Frank & Smith (2020) has also shown that languages spoken by large populations tend to have simpler grammatical structure. However, population size was found to be not the only factor affecting the complexity level in a language – the proportion of younger learners appeared also to play a significant role in the process of language simplification (Frank & Smith, 2020). In other words, as population grows, the number of younger learners is also assumed to increase which in combination results in lower linguistic complexity (Frank & Smith, 2020).

2.2 Population Size and Vocabulary Size

Interestingly, compared to the negative trend observed in the relationship between population size and structural properties of language such as morphology, the opposite trend is seen for non-structural properties of language such as vocabulary size. Here, the relationship turns out to be positive, meaning that as population size increases, the level of complexity in vocabulary tends to increase.

Using computer simulations, Reali et al. (2014) examined the effect of the factor ease of learning on vocabulary size, or more precisely on the number of content words, in populations of various sizes. They demonstrated that when the population is small, hard-to-learn words represent a sizable proportion of the total linguistic inventory (Reali et al., 2014). In other terms, small populations tend to have more words that may be more complex in their structure, i.e. morphological or phonological, and as a consequence harder to learn. As population size increases, the number of easy-to-learn properties increases, whereas the frequency of those that are hard to learn decreases systematically (Reali et al., 2014). This in turn suggests that larger populations tend to have more words that are simpler in structure and therefore easier to learn. These results show that the level of linguistic complexity is not solely related to the size of population. More specifically, there may be other factors such as ease of learning that affect the richness of different structural or non-structural properties of language.

2.3 Population Size and Rates of Change

Bromham et al. (2015) tested the influence of population size on rates of language evolution by comparing rates of gain and loss of cognate words for basic vocabulary in Polynesian languages. Their results demonstrated that larger populations tend to have higher rates of gain of new words, whereas smaller populations tend to have higher rates of word loss. Based on population genetic theory, Bromham et al. (2015) linked higher rates of gain of new words in

larger populations to higher rates of adaptation. Thus, compared to smaller populations, larger populations are supposed to have a higher number of individuals that can produce a novel lexeme (Bromham et al., 2015, p. 2100). Furthermore, in larger populations there are fewer random fluctuations that could disrupt the successful fixation of new variants (Bromham et al., 2015, p. 2100). On the contrary, higher rates of word loss in smaller populations can be explained by the conservative character of the social networks in which they are organized, which can result in the enforcement of linguistic norms and resistance to change (Bromham et al., 2015, p. 2100). However, although these findings are strikingly consistent with general predictions of evolutionary models, as the authors themselves state, Greenhill et al. (2018) emphasized that this test in Bromham et al. (2015) was limited to 20 closely related languages from small Oceanic islands.

Greenhill et al. (2018) therefore examined comparative language data for 153 pairs of closely related sister languages from three of the world's largest language families: Austronesian, Indo-European, and Niger-Congo. They discovered some evidence that word loss rates are significantly higher in smaller languages for the Indo-European comparisons. However, they found no significant patterns in the other two language families. Based on these results, Greenhill et al. (2018) suggested that either the influence of population size on rates and patterns of language evolution is not universal, or in some cases it may be overshadowed by other influences.

The relationship between population size and rate of language change has also been the subject in computer simulations which has revealed different results as well. Nettle (1999b) presented a framework for simulating language change in the social networks derived from Social Impact Theory, as adapted by Nowak et al. (1990). In Nettle (1999a), the simulation model was extended to speech communities of different sizes which functioned as a predictor for the rate of language change. Simulation results demonstrated that there is a relationship between linguistic change and group size. It showed that the overall rate of change in a linguistic item tended to decrease as the size of the community increased. In short, change appeared to be faster in small communities.

Wichmann et al. (2007) replicated the study of Nettle (1999a) using a more advanced language model for simulations, which combined the Schulze model with a Barabasi-Albert network, and a more extensive empirical dataset. Interestingly, their findings indicated that language change rates can be either weakly or strongly affected by population sizes, depending on the parameter settings used. More specifically, it showed that when linguistic features diffuse globally, i.e. from anyone in the group, rather than locally, such as only from

nearby connections, and when the probability of diffusion is high, the rate of language change decreases drastically, given that the population size increases from 100 to ten million (Wichmann et al., 2007, p. 14). In other words, larger populations, combined with global diffusion and higher probabilities of propagation, are assumed to result in slower rates of language change.

Using an agent-based computer model, Ke et al. (2008) demonstrated that the rate of language change is affected not solely by population size but also by the type of social network wherein a population of speakers is organised. More specifically, they found that as the population size increases, so does the diffusion of a linguistic variant, which is a similar result to that presented in Nettle (1999a). However, it is important to note that this pattern was observed exclusively in social networks of regular structure. Compared to regular networks, in other types of networks, such as small-world, random and scale-free networks, diffusion has been found to increase only minimally (Ke et al., 2008, p. 945).

3 Agent-Based Modeling

3.1 Theoretical Foundations

As shown in the previous section, computer simulations can be utilized to explore the relationship between language change and various factors that may affect it. In general, the main advantage of a computer simulation is that it allows a range of potential scenarios to be modelled. In this way, the effects of many factors on a particular item or quantity can be explored, including the effects of factors for which data are not available or are difficult to collect. In terms of language change, computer simulations provide a way of observing how language changes through time and how the factors of interest, such as population size or network structure, influence these changes.

In this thesis, I use an agent-based modeling approach to simulate language change or, more specifically, the change in the proportion of a linguistic variant in populations of different sizes. In short, an agent-based model assumes the existence of some number of agents or, in terms of a model simulating language change, speakers and hearers who interact with each other and, in the course of such interactions, inevitably influence each other's behaviour. Such a model can also be seen as a representation of language as a complex adaptive system (Beckner et al., 2009), where the term complex implies many interconnected relationships between agents and a variety of messages that they send, and the term adaptive means that both agents' behaviours and language structures can change diachronically due to

their interdependence (Bentz, 2018, p. 9). Agent-based modeling, in its various definitions, has been applied in a number of research studies focusing on the relationship between language change and various internal and external factors potentially affecting it (e.g. Nettle, 1999b; Wichmann et al., 2007; Ke et al., 2008; Gong and Jäger, 2012; Spike, 2018).

In this study, a multi-speaker generalization of the Moran model (Moran, 1962) introduced in Blythe (2012) is used to implement an agent-based model of language change. The type of this model is a version of Utterance Selection Model (Baxter et al., 2006), a mathematical formulation of the evolutionary model for language change proposed by Croft (2000). The multi-speaker generalization of the Moran model is proposed in Blythe (2012) as "a null model of language change" (Blythe, 2012), since the only force pushing change in language is assumed to be neutral evolution, also known as genetic drift (Crow and Kimura, 1970) or random copying (Bentley et al., 2007).

Neutral evolution implies that change in language happens solely in consequence of replication processes without selection being involved (Blythe, 2012). This means that a linguistic variant is replicated at random. In other words, replication under neutral evolution is not conditioned on the functional or social status of the linguistic variant, such as occurs under selection. Such neutral model is considered a good starting point for the study on language change presented in this thesis, as it allows for the stepwise inclusion of additional factors that may have an influence on innovation propagation. In this way, a better understanding of the effects of individual factors on language change can be gained.

Figure 3.1: Memory of a speaker as a Moran-type population. The illustrated memory consists of 10 tokens and 2 different linguistic variants depicted by dark grey and light grey circles, respectively.
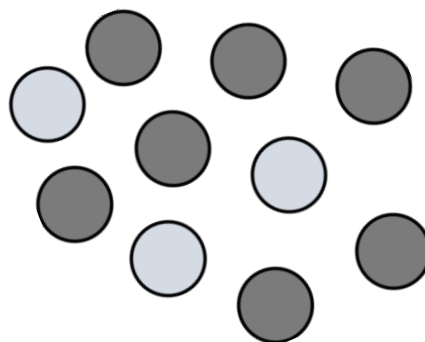
Note. Adapted from *Neutral evolution: A null model for language dynamics*, by R. A. Blythe, 2012, p. 3.

In the multi-speaker generalization of the Moran model, agents are speakers, each introduced as a separate Moran-type population, a set of tokens which represents a speaker's grammar,

also called memory or store (see Figure 3.1). In what follows, I will adhere to the term memory, as I think it best represents the essence of this entity.

To put it in more abstract terms, memory of a speaker displays his entire knowledge of the language. This knowledge is based on experience gained in past interactions with other agents. In more precise terms, however, memory represents a distribution of various linguistic variants. For the sake of simplicity, assuming the existence of just two linguistic variants, one variant may be represented by only a few tokens in the memory of a speaker, while the other by a much larger number of them. In this case, it means that the agent uses the majority variant more often than the minority when communicating with other agents.

In the multi-speaker generalization of the Moran model, linguistic variants are regarded as "different ways of saying the same thing" (Baxter & Blythe, 2006, p. 2). Thus, continuing to assume the existence of only two linguistic variants, in the real world these could be two lexemes expressing the same meaning or two different pronunciations of a specific vowel sound, or different ways of expressing a particular grammatical function like the future tense. Graphically, this idea is represented in Figure 3.1. As seen in the figure, there are two linguistic variants in the memory of an agent, with the variant depicted by the dark grey circles clearly predominating compared to the variant depicted by the light grey circles. What is seen in Figure 3.1 is the state of memory at a particular point in time implying that it does not remain stable, since it changes from interaction to interaction.

In general, the state of the memory of a speaker is assumed to be dependent on the frequencies he or she has heard particular linguistic variants within his or her speech community (Baxter & Blythe, 2006; Bybee, 2001). Therefore, referring again to the memory state of a speaker illustrated in Figure 3.1, it can be inferred that in the process of communication, he or she often encountered the variant indicated by dark grey circles.

In formal terms, the memory of an agent can be defined as a list consisting of N tokens, each conveying a particular meaning M and having a particular linguistic structure v used on n_v occasions (Blythe, 2012, p. 3). The number of occasions $n,$ where a token of a particular variant v is used, changes over time or, more specifically, in the course of interactions.

The multi-speaker Moran model assumes that, at any given interaction, two agents are randomly sampled from a population. When interacting with each other, both produce an utterance by randomly sampling a token from their memories. Subsequently, each speaker reinforces his own behaviour by copying the sampled token and replacing it in his own memory. Blythe (2012, p. 4) also describes this process as listening to himself. In addition, an agent can store a copy of the token sampled by the interlocutor in his own memory. This

allows the agent to better adapt his behaviour to that of the interlocutor (Blythe, 2012, p. 4). Both actions – listening to himself and copying from the interlocutor – are illustrated in Figure 3.2.

Figure 3.2: An interaction between two agents in the multi-speaker Moran model. Here, an interaction of two speakers, i and j, is depicted. Each of them randomly samples a token and copies it (a dashed circle) while deleting any other token (a crossed-out circle) in his own grammar store. This is the standard Moran update which is always performed for both interlocutors. Whether speaker i retains a copy of speaker's j sampled token and, conversely, whether speaker j retains a copy of speaker's i sampled token is defined by the probabilities p_{ij} and p_{ji}, respectively. Note that keeping a copy of the token sampled by the interlocutor in its own grammar also implies that any other token in this same grammar will be removed.
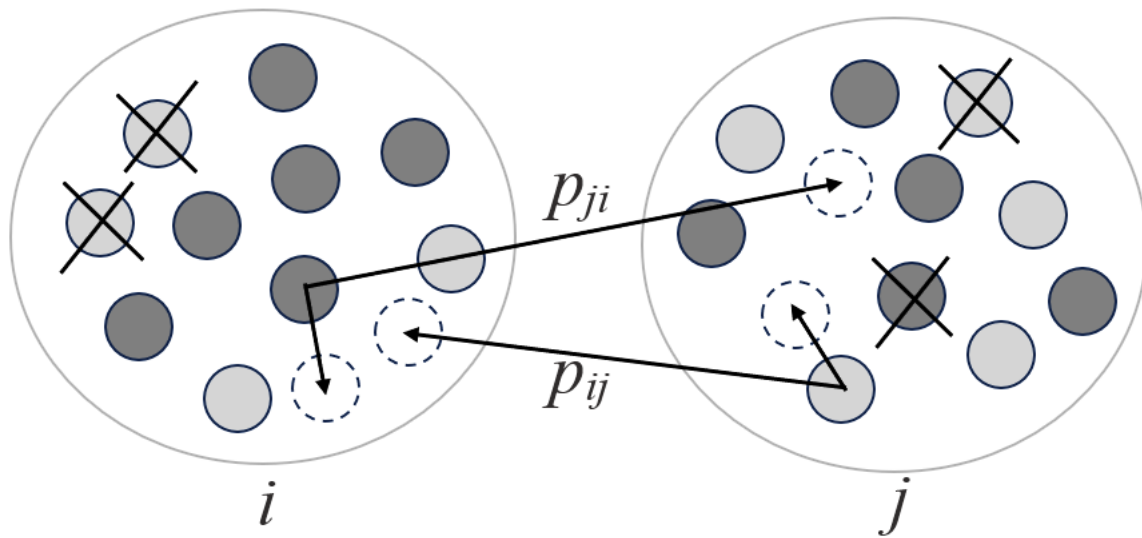
Note. Adapted from *Neutral evolution: A null model for language dynamics*, by R. A. Blythe, 2012, p. 4.

Besides this, different probabilities can be assigned to the events of copying tokens between memory stores – see Figure 3.2. In this way, the strength of the influence of one group of speakers on the other can be modelled (Blythe, 2012, p. 4). The variation of these probabilities between speakers is linked to identity-related phenomena in sociolinguistics (Blythe, 2012, p.4).

3.2 Implementation

Agent-based models have been shown to be a useful approach to investigate many factors and mechanisms involved in language change. However, the implementations of the agent-based models used so far in studies of language change are not publicly available. In addition, in most cases it has proven very difficult to understand how exactly individual components of such a model were implemented in the first place, as there is often a lack of information about the overall architecture of the model and detailed descriptions of the algorithms. Furthermore,

most descriptions of these implementations lack information regarding the programming language used or any specific framework.

To facilitate the from-scratch implementation of the multi-speaker Moran model, I decide to use an open-source library AgentPy (Foramitti, 2021) written in Python (Van Rossum & Drake, 1995). When selecting an appropriate framework for agent-based modeling, Python emerged as the preferred choice due to my familiarity with the language. The main alternative to AgentPy in Python is the agent-based modeling framework Mesa (Masad and Kazil, 2020). However, AgentPy provides some additional useful features not available in Mesa. For a detailed comparison of both frameworks, see the section[1].

To the best of my knowledge, no previous studies have employed the AgentPy framework for agent-based modeling with a focus on language change. Thus, this thesis represents the first attempt to utilise it in this context. It is important to note that AgentPy is a universal agent-based modeling framework which means that it is applicable to any field, be it ecology, cognitive sciences, sociology, economics, or linguistics. Some examples of agent-based models, such as wealth transfer, virus spread and flocking behaviour, implemented by using AgentPy are provided here[2]. The main advantage of AgentPy is that it allows to customize agents and the environments they exist in and to perform numeric experiments over multiple runs, making it suitable for scientific purposes.

Another distinguishing feature of AgentPy is that it allows for sample generation and different types of parameter ranges. Beyond this, AgentPy is designed to be compatible with established Python libraries such as NetworkX, NumPy, pandas, SciPy, and seaborn. Finally, the library provides methods and tools for generation and analysis of output data. An overview of the most important classes, methods and functions of AgentPy and how they are meant to be used can be found in this section.[3]

As illustrated in Figure 3.3, the structure of AgentPy incorporates various components, which build up one upon the other. The fundamental building block in the nested structure of AgentPy are agents, their attributes and actions. Customized agents can then be positioned within various environments with diverse topologies such as a network, a spatial grid or a continuous space. Next, both agents and environments are integrated into a model which can be run in an experiment over multiple iterations and with several parameter combinations. At the end of an experiment, the output data is saved and can be analysed and visualised.

---

[1] *Comparison – agentpy 0.1.6.dev0 documentation.* (n.d.). https://agentpy.readthedocs.io/en/latest/overview.html
[2] *Model Library – agentpy 0.1.6.dev0 documentation.* (n.d.). https://agentpy.readthedocs.io/en/latest/model_library.html
[3] *Overview — agentpy 0.1.6.dev0 documentation.* (n.d.). https://agentpy.readthedocs.io/en/latest/overview.html

Figure 3.3: Nested structure of the AgentPy framework.
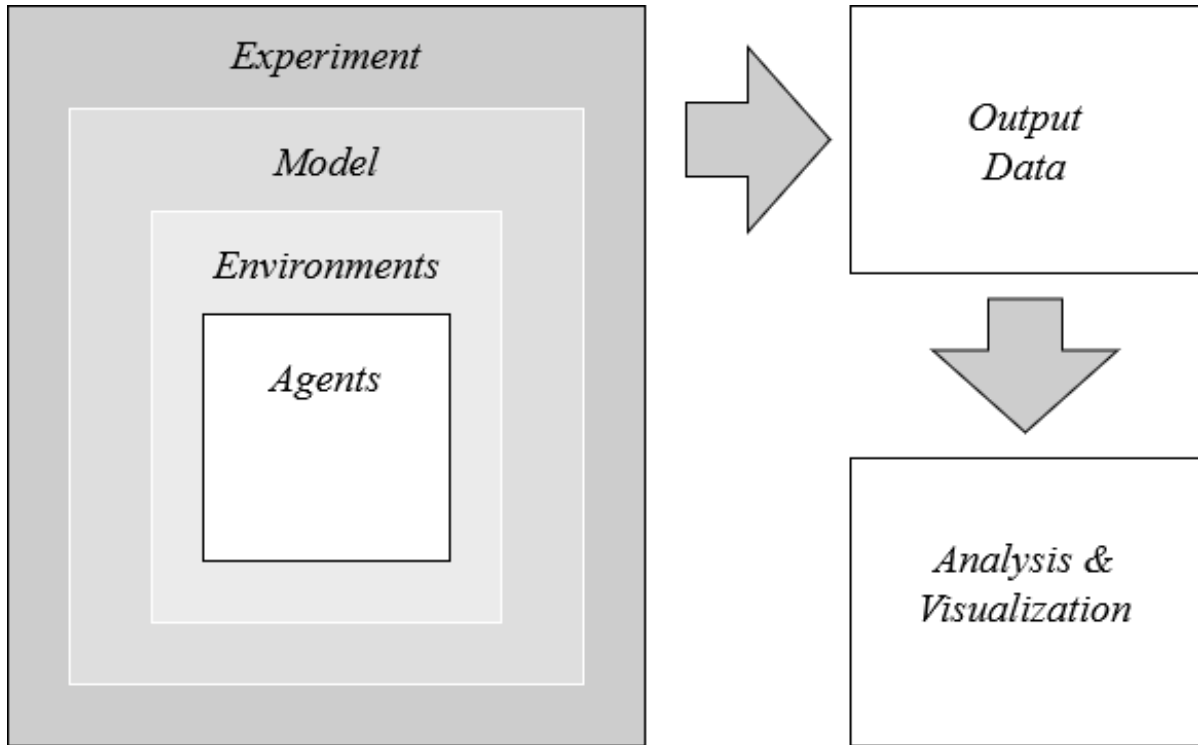
The architecture of the multi-speaker Moran model follows the nested structure of the framework AgentPy illustrated in Figure 3.3. This means that the basic components of the model are agents and an environment they exist in. Hereinafter, each component will be described in detail.

3.2.1 Agents and Their Attributes

To define an agent within the multi-speaker Moran model, a base class *Agent* of the AgentPy framework is used. It is designed to serve as a template for creating custom agent types. Each agent object in this model is conceptualised as a unique entity, characterised by a distinct set of attributes that allow it to be identified and distinguished from other agents within the same population. In total, each agent is endowed with three attributes: *id*, *memory*, and *probability*. These attributes are part of the *setup()* method detailed below.

The *setup()* method is used to define the main attributes of an agent in a given population. In the multi-speaker Moran model, each agent has a unique identifier denoted as *id*. This is a static identification attribute which represents each individual agent and allows him to be distinguished from and recognized by other agents. In formal terms, the set of all agent identifiers *ID* can be mathematically expressed as follows: $ID = \{id \mid id \in \mathbb{N}, 1 \leq id \leq N_A\}$, where N_A represents the overall number of agents or population size in the model.

Besides the static identification attribute *id*, each agent possesses two additional attributes *memory* and *probability*. The variable *memory* is implemented according to the theoretical foundations of the multi-speaker Moran model. It represents an agent's knowledge of language and is implemented as a list which consists of N_T tokens. For the sake of simplicity, the model assumes the existence of two competing linguistic variants, hereinafter denoted as A and B. Thus, going back to *memory*, of N_T tokens, the variant with the linguistic structure A occurs n_A times and the variant with the linguistic structure B occurs $N_T - n_A$ times. The variable *probability* provides an estimate of how likely it is that an agent will decide to choose the linguistic variant A when interacting with other agents. In other words, it is the relative frequency or fraction of the time an agent uses the variant A as opposed to the variant B. In the multi-speaker Moran model, it is calculated from the state of *memory*:

$$p_A = \frac{n_A}{N_T}, \; \{p_A \mid 0 \leq p_A \leq 1\}. \tag{1}$$

Accordingly, the probability that the agent decides to use the variant B when interacting with others is equal to:

$$p_B = \frac{N_T - n_A}{N_T}, \; \{p_B \mid 0 \leq p_B \leq 1\}. \tag{2}$$

Both variables *memory* and *probability* are dynamic as they vary over time due to interactions between agents. One interaction represents thus one point in time t. To give an example, imagine that *memory* of a randomly selected agent at some time t may be such that *probability* of the variant A is $p_A = 0.3$ and thus *probability* of the variant B is $p_B = 1 - p_A = 0.7$. This would imply that the linguistic variant B is perceived by this agent to be seven times more likely than the linguistic variant A.

It should be noted that as the states of individual memories and hence the probabilities of variants change, so does the whole language spoken in a given population. In the multi-speaker Moran model, the state of the language, denoted as L here, at a particular time t is defined as the sum of all individual probabilities p_A divided by the overall size of the population N_A:

$$L(t) = \frac{\sum_{i=1}^{N} p_i}{N}, \; \{L(t) \mid 0 \leq L(t) \leq 1\}. \tag{3}$$

That is, $L(t)$ represents the fraction of time agents use the linguistic variant A. Therefore, $1 - L(t)$ is the fraction of time agents use the linguistic variant B. At this point it should be noted that changes in L are the focus of this thesis. In other words, language change is studied at the

level of the whole population rather than the individual, which provides a broader view of how language changes over time.

3.2.2 Agents and Their Actions

The multi-speaker Moran model assumes a dynamic system in which agents are not just passive participants but social individuals who influence others and are influenced by others in the interaction process. In accordance with the theoretical foundations of the model, each agent object is endowed with a specific behaviour that is defined through three main actions implemented as the following methods of class *Agent*: *speak()*, *reinforce()*, and *listen()*.

The process of speaking is implemented as a method called *speak()*. When speaking, an agent produces an utterance by randomly sampling one token from his own *memory*. The sampled token is assumed to be one of the two competing variants, that is either the linguistic variant A or the linguistic variant B.

Next, the action of reinforcing follows. It is implemented as the method called *reinforce()*. When this method is called for, the agent is animated to reinforce his own behaviour, firstly by copying the token he produced when speaking, and secondly inserting it in place of another randomly selected token in his *memory*. As a consequence of this process, the state of the agent's dynamic attributes, i.e., *memory* and *probability*, may change.

The final action, the process of listening, follows the similar pattern. Here, on the other hand, the agent copies the token produced not by himself but by his interlocutor into his own *memory*. The action of listening can also lead to changes in *memory* and thus in *probability*. The implemented method of the process is called *listen*(*neighbour*) where the variable *neighbour* represents an agent's interlocutor from the neighbourhood – the set of all the connections the agent has to others (Kauhanen, 2016, p. 333).

In addition, each of the N_A agents has the method *update()*. It is called by the model at the end of the interaction, i.e. after the agent has performed all three actions described above. The method updates the value of *probability* based on the *memory* state that is expected to change due to the agent's actions of reinforcing and listening.

3.2.3 Environment

A population of agents is assumed to exist in a certain kind of the environment characterised by a particular structure of social connections. Several agent-based simulations show that the way agents are connected to each other affect how change in language propagates through population over time. The links between agents are typically determined by the structure of

the social network. Some of the studies have shown that there are significant differences between diffusion processes depending on which types of social networks they occur in (e.g. Ke et al., 2008; Gong et al., 2012, Spike, 2018). In general, it has been shown that changes propagate faster in the networks characterized by a high number of short distance connections, that is in the networks having many clusters of closely linked agents (Ke et al., 2008; Gong et al., 2012).

Table 3.1: Network characteristics. The values were computed based on 50 randomly sampled population sizes ranging from 10 to 10,000 agents. The clustering coefficient and shortest path length values reported here for each network result from the sum of individual averages, divided by 50 – the total number of sampled population sizes. Both metrics are rounded.

Network	p	$C(p)$	$L(p)$
Regular	0	0.5	626
Small-world	0.01	0.49	39.99
Random	1	0.01	6.34

In this study, I use three different types of social networks which are typically used in computational models simulating language change (e.g. Ke et al., 2008; Gong et al., 2012). These are: regular, small-world and random networks. Table 3.1 reports the two main characteristics of these networks: the average clustering coefficient C and the average shortest path length L. The clustering coefficient C measures how likely it is that two nodes connected to the same node are also connected to each other. In other words, it quantifies the likelihood that a certain set of nodes form a clique within the network. For this reason, Watts and Strogatz (1998) also refer to the clustering coefficient C as a measure of cliquishness (p. 440). The path length L measures the average shortest distance between two nodes in the graph.

It should be noted that the network structure is controlled by the parameter rewiring probability p, i.e. the probability that an edge is reconnected to a node chosen uniformly at random (Watts and Strogatz, 1998, p. 441). As p increases, both average clustering coefficient C and average shortest path length L decrease (see Figure 3.5). From Figure 3.4, it is obvious that at $p = 0$, the network is a highly clustered regular lattice ($C = 0.5$) with high average shortest path length ($L = 626$). On the contrary, the random network at $p = 1$ is a poorly clustered network ($C = 0.01$) with low average shortest path length ($L = 6.34$). Small-world networks at $p = 0.01$ are highly clustered networks ($C = 0.49$), like regular networks, yet with low average shortest path length ($L = 39.99$), like random networks. Note that p is set to 0.01 for the small-world network as it is within the range where the small-world properties are best represented (Watts and Strogats, 1998, p. 441). The same value of p for the generation of a small-world network was also used by Ke (et al., 2008).

Figure 3.4: The rewiring procedure. The parameters used here are the following: n = 20 and k = 4. In regular network, if p = 0, each node has exactly four connections. As p increases, the degree of disorder within the graph increases. That is, in small-world network, not each node has exactly four connections: some nodes have exactly four, some more, and some fewer. As p reaches 1, a random graph emerges which exhibits a highly chaotic behaviour of connections.

Regular   Small-world   Random

$p = 0$ $p = 1$

Increasing randomness
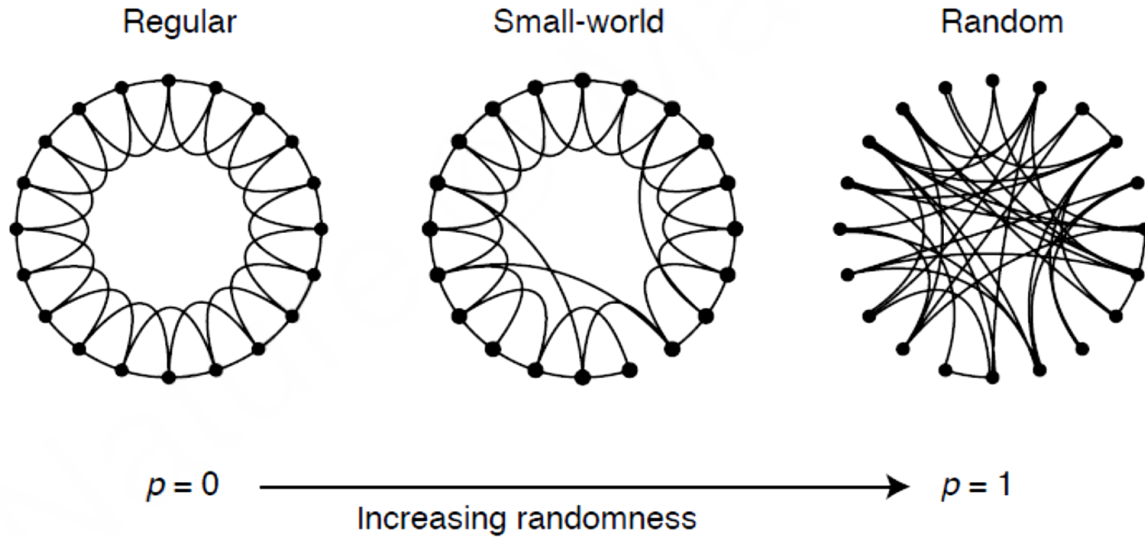
Note. From *Collective dynamics of 'small-world' networks*, by Duncan J. Watts & Steven H. Strogatz, 1998, p. 441.

The number of nearest neighbours k for each agent in all three types of networks was set to 4. This configuration assumes that, on average, each agent is connected to four other agents. The choice of $k = 4$ is considered an appropriate representation of real social connections, given the assumption that only a small proportion of individuals in the population has a large number of connections, while the majority has a small number of connections.

It should be kept in mind that $k = 4$ describes the average number of connections per agent in a population. Thus, this does not mean that each agent has exactly four connections in the networks: Their number varies depending on p, and therefore some agents may have more and some even less than four connections with others in the network. The only exception here is the regular network, in which the number of connections remains the same for each of the agents.

In the implementation of the multi-speaker Moran model, regular, small-world and random networks represent three different types of the environments that agents are placed in. Each is integrated into the model by the means of the class *Network* which is part of the AgentPy framework. To implement the different network structures, the Python package called *NetworkX* (Hagberg et al., 2008) is utilized. The generated networks are symmetric and are not multiplex and the connections are binary. In other words, if an agent i is connected to an agent j, then the agent j will also be connected to the agent i. Only one connection is allowed between any two agents and each pair of agents is either connected or not connected.

Furthermore, the networks remain static over time, which means that on the one hand agents are neither removed from nor added to the networks, and on the other hand, the connections between agents remain unchanged. It should be noted here that I recognize that social networks are dynamic systems in which ties between individuals change throughout time, which is also emphasized in research on language change (e.g. Kauhanen, 2016). Nevertheless, for the sake of simplicity as well as for ease of interpretation of the results of subsequent simulations, the agent environments in this thesis are realized as static social networks.

Figure 3.5: The relation between clustering coefficient C and shortest path length L. The data shown in the figure are average values of clustering coefficient *C* and shortest path length *L* given in Table 3.1 for regular, small-world, and random networks. For better comparability, the values have been normalised to a range between 0 and 1. While regular and random networks demonstrate high and low values of *C* and *L*, respectively, the small-world network is marked by relatively low *L* and high *C*, which is consistent with the observations reported in Watts & Strogatz (1998).
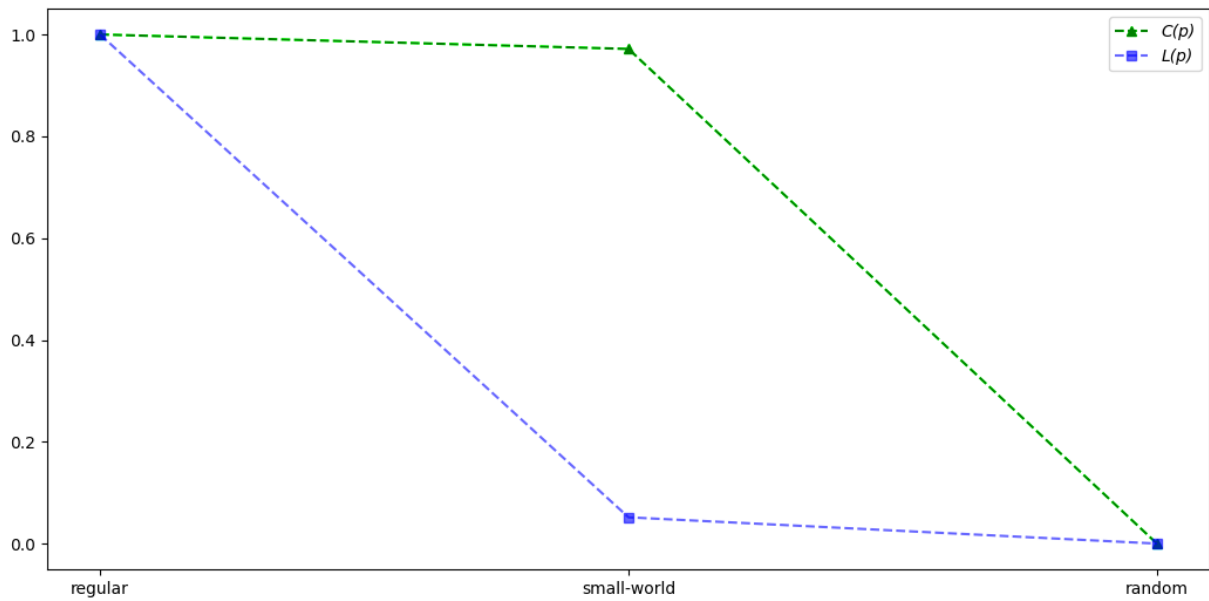
Note. Adapted from *Collective dynamics of 'small-world' networks*, by D. J. Watts & S. H. Strogatz, 1988, p. 441.

3.2.4 Model Implementation Architecture

The multi-speaker Moran model is defined as a class called *Model*. In the AgentPy framework, it functions as a template for custom model classes. The model incorporates all main components introduced and described in the previous chapters. The architecture of the model is illustrated in Figure 3.6. The GitHub repository link for the model implementation, which includes the source code and instructions for use, can be found in Appendix A.

The model integrates two main components, i.e. the class *Agent* and the class *Network*, which have been described in the previous chapters. Both are initialized before the first simulation step with the *setup()* method. In this way, both the population of agents and the type of network wherein they are organised and interact with each other are determined.

The model's method *step()* specifies the actions to be carried out during each simulation step, except for the very first step, where $t = 0$. One single simulation step corresponds to one single interaction between two randomly selected agents. Thus, at any given simulation step, agents i and j are selected at random to interact with each other. More specifically, the agent i is chosen from the entire population and his interlocutor, the agent j, is chosen from his neighbourhood, i.e., from those agents he has connections to. After selecting two agents for an interaction, for each of them all four methods of the *Agent* class, i.e. *speak()*, *reinforce()*, *listen()*, and *update()*, are executed.

Figure 3.6: Simplified model implementation architecture. In the diagram presented, the individual boxes constitute three different classes named *Model*, *Agent*, and *Network*. The sections under the class names *Model* and *Agent* show the public operations of these classes. The *Network* class and the *Agent* class are directly related to each other in the sense that the former consists of one or more agents. This relationship is indicated by a bold line with an arrow. The *Model* class calls both *Agent* and *Network* classes within its structure, as indicated by the two dashed lines with an arrowhead.
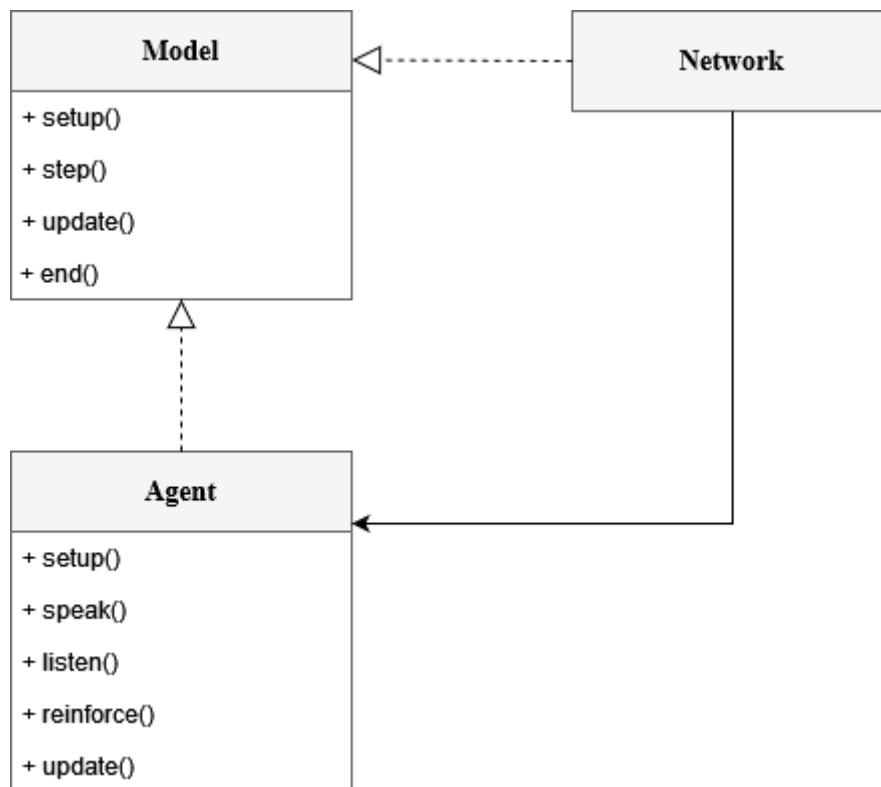
Another method of the model called *update()* determines the actions of the model after each simulation step. Compared to the *step()* method, the initial step, i.e. $t = 0$, is included here. In this way, the initial state of the model is determined. The method is used to update the state of

the entire language *L* based on the agents' states of *memory* stores at each simulation step *t*. The calculation of the state of the entire language *L* is conducted according to the equation given in (3). The output of the calculation is recorded with the model's method *record()* which is integrated into the model's method *update()*.

Finally, the *end()* method determines the actions of the model to be executed after the last simulation step. Specifically it saves the result of the model, that is the final state of the entire language *L*, through the method called *report()*. This means that the output of a single model run that involves a predetermined number of individual simulation steps is the most recent record of *L* for a particular population.

4 Language Change and its Mechanisms: Theory and Implementation

Language change is believed to occur through the process of replication, which involves the repeated use and transmission of linguistic structures in utterances (Blythe & Croft, 2012, p. 271). Thus, language change is not just a one-time event but rather a continuous and iterative process. In this context a speaker is assumed to reproduce a certain linguistic form based on the acquired knowledge about his language, which includes previous utterances produced either by him or by his interlocutors (Croft, 2000). In this thesis, language change is seen as a two-step process, as suggested by Blythe & Croft (2012, p. 271). The first step involves the generation of variation in the replication process which is associated with an innovation entering the community of language users. The second step comprises the process of selection[4], also called propagation or diffusion. This process implies the replication of linguistic variants by language users.

In this thesis, I follow the approach of Blythe & Croft (2012) in that I restrict my attention exclusively to one step of the language change process – propagation. Therefore, I assume the existence of two competing linguistic variants, hereafter referred to as A and B, the first representing the innovation that has already entered the community of language users but remains the minority variant, and the second representing the convention used by the majority.

---

[4] The term selection may be misleading here, since in the context of language change it usually implies a social or functional bias associated with a particular linguistic variant. However, it should be noted that the use of the term in this particular context rather refers to the differential replication of linguistic variants, which can also occur as a result of stochastic drift, i.e. when no bias is attached to linguistic variants.

In the basic model of language change, a linguistic variant, such as A or B, represents a token of a specific linguistic structure and is referred to as the replicator[5] (Blythe & Croft, 2012). The speaker who determines which replicator, or linguistic variant, is copied or not is called the interactor[6]. When producing an utterance, an interactor decides which of available variants to replicate. Such decision process is assumed to be driven by mechanisms of linguistic selection that "lead to the differential replication, that is, propagation, of some variants at the expense of others" (Blythe & Croft, 2012, p. 271). Differential replication can occur due to social or functional biases associated with linguistic variants or purely by chance. It has been suggested that language users do not simply copy the most frequent variants they hear; they rather adapt their language to the social environment in which they move or of which they wish to become a part. Thus, in terms of social bias, the differential weighting of linguistic variants may be related to the varying types of social valuation of different language users (Labov, 2001): Some groups of speakers may have more prestige than others and thus the linguistic forms they use may have a higher diffusion probability. Functional bias[7], on the other hand, is associated with internal properties of linguistic variants. Thus, for example, some replicators may be easier to articulate or to learn than others (Nettle, 1999b, p. 100). Consequently, these linguistic variants might be replicated by language users with a higher probability.

However, the replication of linguistic variants can also occur purely by chance, without any social or functional bias being at play. Such absence of any bias corresponds to what is called neutral evolution which "is a process that one is ineluctably drawn to whenever the changes occur due to replication acting alone (i.e., without selection) are considered". (Blythe, 2011, p. 2). In the case of neutral evolution, change is driven by random fluctuations in frequencies of a specific linguistic variant: These fluctuations can either decrease to 0%, leading to extinction of the linguistic variant, or increase to 100%, causing the linguistic variant to become fixed (Blythe & Croft, 2012). A significant property of the model with a neutral evolution mechanism is that the probability of a linguistic variant becoming the majority variant in a population is a function of its frequency (Baxter et al., 2009, p. 270). In

---

[5] The term replicator originates from the generalized models of evolutionary change such as that proposed by Hull (1988), which use it to denote the entity that is replicated, for example, the gene. Croft (2000) subsequently applies this term in a linguistic context in defining an evolutionary model of language change.
[6] Blythe & Croft (2012) borrow the term interactor from the realm of biological evolution, referring to Hull's definition, which characterizes it as an entity that interacts with its environment and thereby causes differential replication. (p. 271).
[7] The term functional is adopted from the term functional selection described in Nettle (1999b). Here it is used to refer to different types of biases, i.e. phonetic or structural biases that are directed towards specific linguistic variants rather than towards people, as in the case with social bias.

other words, the frequency of a linguistic variant is the primary factor that determines the probability of its successful propagation in a population.

The neutral evolution, social and functional biases thus introduce the different mechanisms of language change or propagation of linguistic variants, the concepts of which are offered by Baxter et al. (2009) and more precisely defined in Blythe & Croft (2012). However, these concepts are integrated into the multi-speaker Moran model implemented here in a slightly modified form: More precisely, there are some differences in the definitions and implementations of the mechanisms. The primary reason for these modifications is related to the structure of the base model, i.e. the multi-speaker Moran model, which is different from the base used by Blythe and Croft (2012). The main properties of these three mechanisms and their implementation within the multi-speaker Moran model are described in detail in the following chapters.

4.1 Neutral Change Mechanism

In the context of language change, neutral evolution is characterised by properties such as symmetric interactor and replicator behaviours and equal interaction frequency (Blythe & Croft, p. 276). In other words, interactors have equal positions in a population by speaking with the same number of other speakers and each interaction occurs with equal frequency: Thus, speaker i interacts equally often with speaker j and with speaker z. Replicators are equally weighted: There are no social or functional biases associated with the linguistic variants.

The mechanism of neutral evolution is set in relation to another mechanism called neutral interactor selection which has the same properties, namely the symmetry of interactors and replicators (Baxter et al., 2009; Blythe & Croft, 2012). The only difference between the two mechanisms is in the interaction frequency between speakers. That means, compared to neutral evolution, neutral interactor selection allows the frequency of interaction between speaker i and his interlocutors to vary (Blythe & Croft, 2012, p. 276). In other words, speaker i may interact more with speaker j than with speaker z. This unequal interaction frequency is considered a type of selection as it is assumed to cause differential replication of linguistic variants (Baxter et al., 2009, p. 271).

In this thesis, I decide to implement the neutral interactor selection mechanism as a mechanism of neutral change into the multi-speaker Moran model. There are two reasons for this. First, agents are determined to exist and interact with each other in a specific kind of social network. That means that some agents might have more connections than others. In

other words, agent *i* may interact with agent *j*, but not with agent *z*, because the link between *i* and *z* might not exist. Second, according to the theorical foundations of the multi-speaker Moran model, the selection of an agent and his interlocutor is required to be random. Hence, if agent *i* has two neighbors *j* and *z*, he may interact with one of them more often than with the other. Moreover, I believe that communities in which each speaker has the same number of connections and the same frequency of interaction with each of them are rather rare, and thus, compared to the neutral evolution mechanism, the mechanism of neutral interactor selection better reflects real-world social relationships. Note however, that I am intentionally avoiding the term selection in the name of the mechanism: It can be misleading, since selection usually implies some sort of functional or social bias assigned to a linguistic variant. Furthermore, I also avoid the term neutral evolution, since according to the definition in Baxter et al. (2009), it would imply the equal interaction frequency between individual agents, which is not the case in the implementation presented here. So in what follows I will use the following term: mechanism of neutral change.

At this point it should be noted that the mechanism of neutral change is in the nature of the multi-speaker Moran model, since it assumes the symmetry of replicators and interlocutors. In more formal terms, both interactors *i* and *j* are always equally weighted. That means, both probabilities p_{ij} and p_{ji} are set to 1: In any given interaction, agent *i* will always retain the token produced by agent *j* and vice versa. Moreover, both replicators A and B are also equally weighted. This means that in any single interaction, when the action of listening is performed, agent *i* will always keep the token sampled by speaker *j* and vice versa. Therefore, it does not matter who produces a variant, as well as which variant, A or B, is produced. Similarly, when an agent reinforces his own behaviour, he always retains the sampled token, regardless of whether it is a variant A or a variant B. Thus, changes in agents' memories and hence in probabilities of using the innovative variant A are all due to drift or, in relation to the definition proposed in Baxter et al. (2009), due to the frequency of interaction with the interactor. That said, I tend to attribute the change in frequency of a linguistic variant to the term drift, since the interactions between agents are purely random and not controlled in any way.

4.2 Replicator Selection Mechanism

In comparison to neutral change, replicator selection is characterised by the presence of some sort of bias attached to linguistic variants. This can be social in that different social values can be associated with different linguistic variants, that is, replicators (Baxter et al., 2009, p. 269).

In other words, the innovation A may have a higher social value than the convention B due to the higher prestige of the group that uses it, and as a result it may be more frequently replicated. Furthermore, the bias associated with the variant A can also be functional which means that it can be articulated more easily than the variant B. The replicator selection mechanism therefore assumes that linguistic variants are not symmetric, i.e., that one variant is weighted differently over another (Blythe and Croft, 2012, p. 276). This asymmetry in valuations of different variants is the essential property of the model which includes the replicator selection mechanism. Furthermore, the results of the modeling performed by Blythe & Croft (2012) show that replicator selection is an essential mechanism for producing an S-curve, the trajectory of change, where a conventional variant is replaced by a competing variant (p. 294). In the implementation of the replicator selection mechanism, I follow the approach of Blythe & Croft (2012) in the sense that I do not focus on a specific type of selection and implement a general replicator selection mechanism which potentially could represent social or functional biases assigned to the innovation A.

The replicator selection mechanism is integrated into the multi-speaker Moran model as a parameter named *selection strength*, hereinafter referred to as *s*. The parameter regulates the degree of preference to the innovative variant A. It represents the probability with which an agent, in the case of hearing the innovation A from his interlocutor, will retain it in his memory store. The range of the parameter is as follows: $s \in (0, 1]$. The value of $s = 0$ is excluded as this would neutralize the presence of selection in the model and return it to a neutral state when changes occur by stochastic drift. The value of $s = 0.1$ represents the weakest selection strength, while the value of $s = 1$ the strongest.

To give an example, if an agent as a user of the convention B hears the innovative variant A from his interlocutor, he will adopt it into his memory with the probability given by *s*. That means, a higher selection strength means a greater chance for the agent to retain the variant A: For instance, if the selection strength *s* is set to 0.8, there is an 80% chance that the agent will store the innovative variant A in his memory. If, on the other hand, *s* is equal to 0.1, the probability that the agent will retain the variant A is significantly lower.

4.3 Interactor Selection Mechanism

The mechanism of interactor selection implemented in this thesis corresponds to what Baxter et al. (2009) call the mechanism of weighted interactor selection.[8] This is a type of selection where interactors are non-symmetric or, in other words, differently weighted: They are "preferred or dispreferred by a speaker no matter how frequently or infrequently she interacts with them, and their linguistic replications (utterances) are weighted accordingly" (Baxter et al., 2009, p. 271). Thus, the variant uttered by an interactor who has a more influential role in the community is likely to be replicated more often, regardless of how frequently one interacts with him. It is important to note, however, that in the case of interactor selection, the replicators remain symmetric. Therefore, in comparison to replicator selection, the bias in interactor selection is attached to the interactor but not to the replicator.

The implementation of the mechanism of interactor selection is carried out following the implementation proposed by Blythe & Croft (2012, p. 287), where a population of speakers is divided into two groups called leaders and followers. The partition of a population into two contrasting clusters derives from Roger's model in which a community of speakers is divided into five adopter categories (Rogers, 1995). From these five categories, Blythe & Croft (2012) select two – leaders and followers - following Labov (2001) and Katz and Lazarsfeld (1955).

Leaders represent early adopters and followers represent later adopters since the former adopt an innovation earlier than the latter. In addition, leaders or early adopters have a greater social weight than the followers in the community and thus receive more respect (Blythe & Croft, 2012, p. 275). That is, later adopters or followers follow early adopters or leaders: They put more weight on the utterances of early adopters than on the utterances of their own group. Early adopters however do not follow later adopters and, as a consequence they do not put more weight on the behaviour of the followers than that of their own group.

The mechanism of weighted interactor selection is integrated into the multi-speaker Moran model as follows. The network of agents is divided into two groups which correspond to the group of leaders and the group of followers. From the population of N_A agents, the first n of them represent leaders, and the remaining $N_A - n$ are followers. Leaders use the innovation A with the initial probability 1, that is $p_A = 1$ for $i \leq n$, and followers use the convention B, that is $p_A = 0$ for $i > n$. Thus, in the multi-speaker Moran model, n represents the number of leaders in the population in relation to the total population size. It is defined as

---

[8] Note that Baxter et al. (2009) use this term as a contrast to neutral interactor selection. Since I have chosen a different terminology for this thesis, this contrast no longer seems relevant here and in the following I simply use the term interactor selection.

follows: $n \in [0.1, 0.2]$. Therefore, a population of agents can have either 10% or 20% of leaders in the model.[9]

In addition, token copying events between memory stores are assigned different probabilities depending on the social status of the agents in the model. At this point, it should be noted that only leaders can influence the memory states of followers: In other words, followers listen exclusively to leaders and leaders do not listen to anyone. Therefore, an agent as a follower retains a copy of the token produced by his interlocutor only in the case if the latter holds a leader role. Otherwise, the copy operation is not executed. In formal terms, this yields the following: $p_{ji} = 1$ for $i > n$ and $j \leq n$, where p_{ji} is the probability that agent i keeps a copy of the token produced by agent j in his memory if agent i is a follower and speaker j is a leader. Conversely, if the probability p_{ij} occurs, the roles are reversed, and agent j retains a copy of speaker i's selected token: $p_{ij} = 1$ for $j > n$ and $i \leq n$.

Furthermore, I decide to integrate into the mechanism of interactor selection an additional parameter, i.e., selection strength s, which is assumed to control for the degree of the influence of leaders on followers. In this case, selection strength s represents the probability that a follower will retain a copy of the token produced by a leader, that is: $s \in (0, 1]$. Again, as with the replicator selection mechanism, $s = 0$ is not included in the range as this would neutralize the presence of selection strength. The strongest influence of leaders on followers is thus ensured with $s = 1$ and the weakest with $s = 0.1$. To illustrate the process, here is an example: When $s = 1$ there is a 100% chance that follower i will copy the token produced by leader j into his memory. On the other hand, when $s = 0.1$, the chance that follower i will retain the token from leader j is only 10%.

4.4 Mechanisms in Comparison

Characteristics for each mechanism are given in Table 4.1. The main difference between the mechanisms is whether or not there is asymmetry in replicators or interactors. The main property of the model including the replicator selection are nonsymmetric replicators. As Table 4.1 shows, there are no differential values associated with interactors in this kind of model. Furthermore, whether a replicator will propagate or not does not depend on the frequency of interactions: The weight attached to the replicator alone is a crucial factor which determines whether its propagation is likely to be successful or not.

---

[9] The values of n are set so low due to the assumption that groups of people with a higher social weight are expected to be quite small in a real-world situation.

Table 4.1: Mechanisms of propagation within the multi-speaker Moran model. Dashes indicate the irrelevance of the respective characteristics in terms of the spread of a linguistic variant.

Mechanism	Interaction frequency	Interactor behaviour	Replicator behaviour
Neutral change	unequal	symmetric	symmetric
Replicator selection	-	-	nonsymmetric
Interactor selection	-	nonsymmetric	symmetric

Note. Adapted from *S-curves and the mechanisms of propagation in language change*, by R. A. Blythe and W. Croft, 2012, p. 276.

In interactor selection, the replicator behaviour is symmetric, but the interactors are nonsymmetric. The interaction frequency also does not matter as the successful propagation of a replicator depends solely on the weight of an interactor.

In neutral change, both interactors and replicators are symmetric, and the force affecting the spread of a replicator in the community is the unequal interaction frequency. Therefore, to successfully propagate, a replicator needs to be very frequently uttered: The more often a specific variant is used, the higher the probability that it will propagate in the community.

5 Simulations

5.1 Parameters and Experiments

The multi-speaker Moran model has a range of parameters. Some of them remain constant over the whole simulation process and some are varied. The parameter settings for each of the three mechanisms of language change are presented in Table 5.1. It should be noted that each of the mechanisms is integrated into the model as a boolean value. Therefore, in order for a specific mechanism to be activated, the value True must be passed to it.

Constant parameters are a single lingueme[10] including a specified number of V variant forms, memory size consisting of N_T tokens, initial probability of the variant A denoted as p_A, the average number of neighbours k per agent and, finally, the number of interactions, which indicates time t in the model. The model assumes the existence of two competing variant forms: the innovative variant A and the conventional variant B, i.e. $V = 2$. The memory size of each agent is an integer set at $N_T = 10$ tokens, of which 20% are initially the tokens of the variant A and 80% are therefore the tokens of the variant B: $p_A (t = 0) = 0.2$ and $p_B (t = 0) = 1$

---

[10] The term lingueme is used in the context of the evolutionary model of language change to represent variants of a word or a set words (Blythe and McKane, 2007, p. 13). In short, a lingueme is a representation of the language spoken by the agents in an agent-based model.

– p_A = 0.8. That is, in the initial state of the system, agents are assumed to use the convention B eight times more likely than the innovation A on average. In defining such a distribution of both variants, I follow the approach of Blythe and Croft (2012, p. 285) by claiming that the innovation A has reached a small frequency at $t = 0$, but do not explicitly model the process by which this happens. Note that the initial probability of the innovation p_A is expected to change during interactions between agents. Its values can vary between zero (0) and one (1), with zero (0) indicating the extinction of the innovation and one (1) its successful propagation across the entire population. The change in p_A under the influence of neutral change, replicator and interactor selection mechanisms over time is what the model is designed to investigate.

The remaining parameters, that is population size N_A, rewiring probability p, selection strength s, and proportion of leaders n_L, are all varied in separate simulations. The parameter of population size N_A represents the overall number of agents. It is defined as a list containing 50 different values of N_A, which range from 10 to 10,000 agents, and functioning as a parameter of one single experiment. The values of rewiring probability p are set for each mechanism to 0, 0.01 and 1, which represent regular, small-world and random networks, respectively.

Table 5.1: Parameters in the multi-speaker Moran model for neutral change, replicator selection and interactor selection.

Symbol	Meaning	Neutral change	Replicator selection	Interactor selection
V	number of variants	2	2	2
N_A	population size	[10, 10000]	[10, 10000]	[10, 10000]
N_T	number of tokens in memory	10	10	10
$p_A\,(t=0)$	initial probability of the variant A	0.2	0.2	0.2
k	number of neighbours	4	4	4
p	rewiring probability	{0, 0.1, 1}	{0, 0.1, 1}	{0, 0.1, 1}
s	selection strength	-	{0.1, 0.2, ..., 1.0}	{0.1, 0.2, ..., 1.0}
n_L	proportion of leaders	-	-	{0.1, 0.2}
t	number of interactions	100,000	100,000	100,000

The values of selection pressure s range from 0.1 to 1 with the step of 0.1. The value 0 is not included in the range, as this would mean that the selection pressure is absent. It should be kept in mind that the parameter s determines the probability an agent will decide to retain the innovative variant A in his memory store. Therefore, the value 0.1 represents a very weak

pressure towards the innovation A, i.e. only 10% chance of being retained by a speaker. Conversely, the value 1 indicates a very strong pressure towards A, i.e. 100% chance that it will be retained by a speaker. Furthermore, it should be considered that the *s*-parameter is used only in the simulations where replicator selection or interactor selection are involved.

The proportion of leaders *n* is varied in the simulations where interactor selection is at play. This parameter includes only two values, 0.1 and 0.2, indicating that 10% and 20% respectively of speakers in a given population are leaders. These two values are selected based on the assumption that the proportion of leaders in a population is expected to be generally low.

Figure 5.1: Architecture of one single experiment.
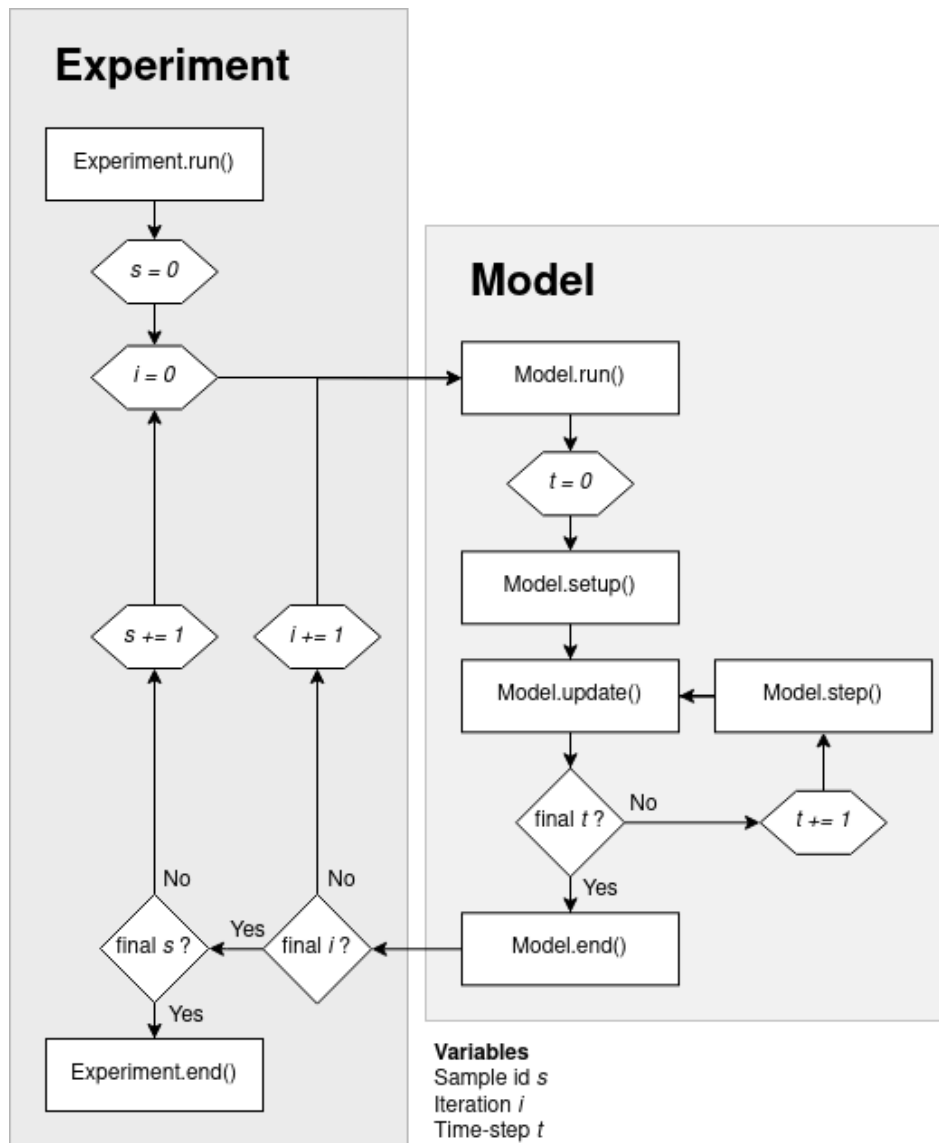
Note. From *Overview. Multi-run experiments – agentpy 0.1.6.dev0 documentation*, n.d. (https://agentpy.readthedocs.io/en/latest/overview.html#multi-run-experiments).

The simulations are performed for each mechanism of language change as part of multi-run experiments with a range of parameter combinations. The class *Experiment* from the framework AgentPy was used to run the multi-speaker Moran model multiple times. As shown in Figure 5.1, an experiment starts with the first parameter combination in the sample denoted as $s = 0$. This parameter combination is repeatedly used in the simulation for the defined number of iterations, denoted as i. After the final iteration i, the next parameter combination s is taken from the sample for which the same cycle of steps is performed.

For each experiment, 50 different values of population size N_A ranging from 10 to 10,000 agents are sampled. The sampling is performed by means of class *Sample* of the AgentPy framework by using the default method *linspace* which allows for sampling the specified number of evenly spaced values for a given range. At this point, it should be noted that through sampling, each individual population size N_A is combined with parameters and parameter ranges from Table 5.1, according to the mechanism of language change involved. Each parameter combination s from the generated sample is set to be repeated over $i = 3$ iterations in each simulation run. A single iteration would show only a fraction of the whole story and might even lead to wrong conclusions due to potential outliers. Therefore, multiple iterations are essential as they allow to get a more realistic picture of how specific parameter combinations tend to affect the propagation of the linguistic variant of interest. Indeed, it would be better to use a larger number of iterations for a single parameter combination. On the other hand, this would drastically slow down the simulation process. That is why, within the scope of this simulation study, three iterations are considered sufficient to identify the general trends in the propagation process.

At the end of each experiment, the output of the model represents a list of the language state values L_i at the final time step $t = 100,000$, where i stands for a single model run. In total, 150 experiment runs are conducted with the neutral change mechanism, 1500 experiment runs with the replicator selection mechanism and, finally, 3000 experiment runs with the interactor selection mechanism. The link to the simulated data can be found in Appendix A.

5.2 Results

I first simulate the propagation process of the innovative variant A under the mechanism of neutral change. That is, I use the base multi-speaker Moran model without any selection processes being involved. The output of the model is a range of language state values L, each corresponding to a particular population size from the sample. To demonstrate the simulation results, the average of L values for each population size is taken, which is shown in Figure

5.2. From the figure, it is obvious that in small populations drift processes generally tend to lead either to the increase or decrease in the proportion of the innovative variant A. On the other hand, larger populations appear to be less prone to change: After 100,000 interactions, agents of these populations keep favouring the conventional variant B.

Notably, this trend is evident in all three types of social networks. Some differences in the probabilities of the variant A are seen only between social networks with very small populations: The probability of A tends to increase over time in the regular and small-world networks, while it tends to decrease in the random network. Noteworthy is that the innovation A does not demonstrate a complete fixation, i.e. propagation to 100%, in any of the networks, independently of whether the population size is small or large.

Figure 5.2: Simulation results for the neutral change mechanism. Each subplot represents the results for a specific type of social network. The population size values N lie on the x-axis and the corresponding mean values of the language state L lie on the y-axis. The set of parameters for each network type and population size used: $V = 2$, $N_T = 10$, $p_A (t = 0) = 0.2$, $k = 4$, and $t = 100{,}000$.
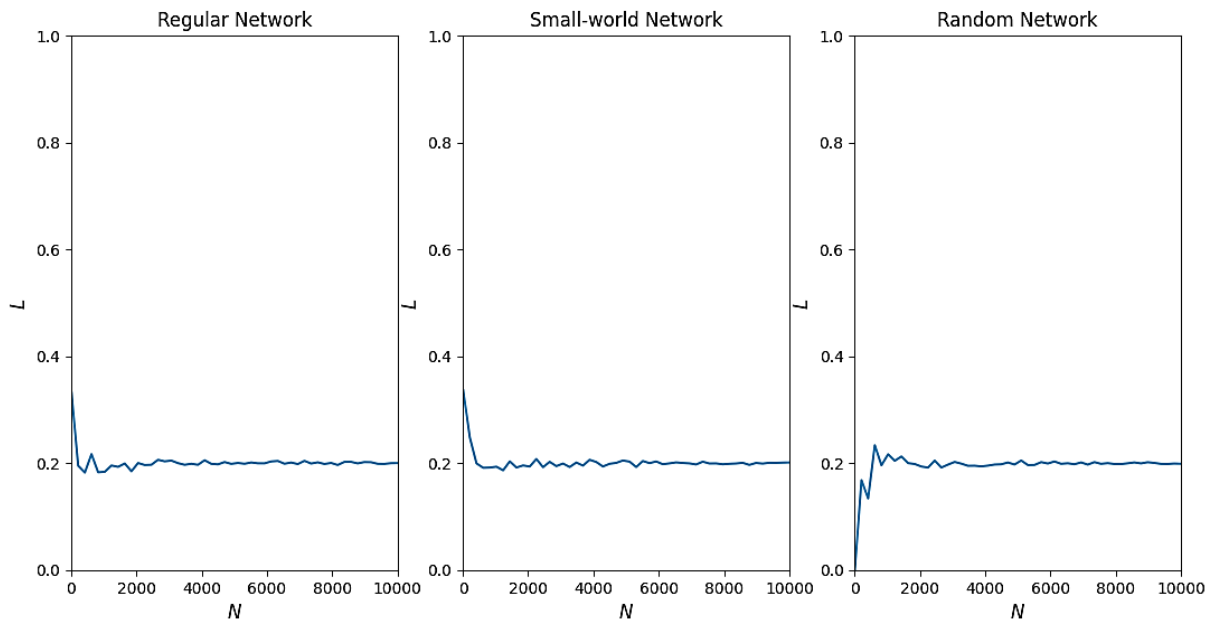
When the mechanism of replicator selection enters the process of the innovation propagation, the picture observed from the results of the null model changes. This is illustrated in Figure 5.3.

It is obvious that on average the probability of the innovative variant A tends to increase with increasing selection pressure s towards it. What is remarkable is that this pattern can be observed in all three network types, without any striking differences between them. When the selection strength is very low, i.e. $s = 0.1$, the innovation A appears to diffuse to a larger number of agents in small populations and even to become the majority variant, i.e., $L \geq 0.6$. On the other hand, not much change in the proportion of A seems to occur in large

populations. At this point however, it should be noted that compared to the mechanism of neutral change, the changes in the proportion of the variant A in larger populations are more obvious when replicator selection is involved. Nevertheless, the innovation A does not succeed in overcoming the threshold at which it would displace the variant B and thus become a new established convention. Only if the selection strength is strong enough, i.e. $s \in \{0.9, 1.0\}$, do agents tend to use the innovative variant A rather than the conventional variant B. As can be seen from Figure 5.3, given a very strong selection strength value $s = 1$ and the large size of the population $N = 10,000$, the innovation A reaches a probability of about 0.6. However, in spite of a very strong selection strength being at play, it should be noted that the variant A never succeeds in spreading to 100%. By contrast, in very small populations such as $N = 10$, the variant A tends to always spread to 100%, independently of the strength of replicator selection, and thus displaces the conventional variant B.

Figure 5.3: Simulation results for the replicator selection mechanism. Each subplot represents the results for a specific type of social network. The population size values N lie on the x-axis and the corresponding mean values of the language state L lie on the y-axis. The set of parameters for each network type and population size us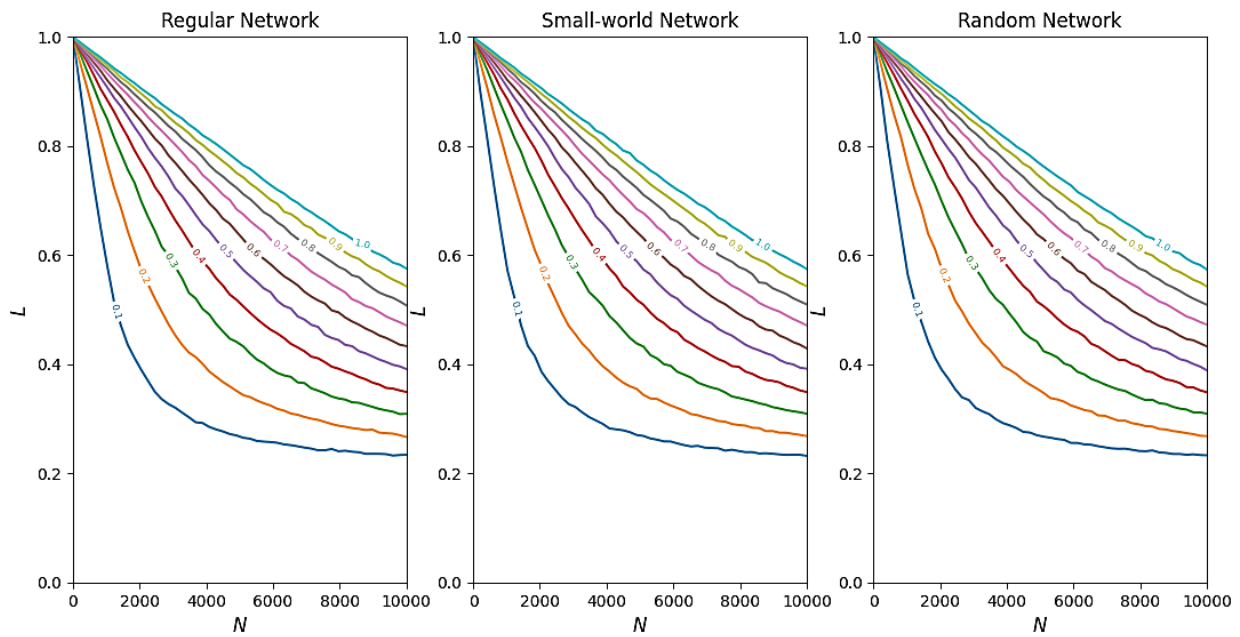ed: $V = 2$, $N_T = 10$, p_A ($t = 0$) $= 0.2$, $k = 4$, $s \in \{0.1, 0.2, ..., 1.0\}$, and $t = 100,000$. Note that each individual curve in each of the three subplots stands for a particular selection strength value s, which is marked with the corresponding floating point number.

When interactor selection is involved into the innovation propagation process, the proportion of the variant A tends to change in very small populations, that is in the populations of size $N = 10$, while remaining relative stable in the populations of larger size. These trends are illustrated in Figure 5.4, where the proportion of leaders accounts for 10% of the respective population, and in Figure 5.5, where it accounts for 20%. It is noticeable that when the proportion of leaders $n = 0.1$ and the population size is very small, the innovation A fails to

propagate and therefore goes extinct in the small-world and random networks. The exception is the regular network, where it sometimes reaches probability values greater than the initial state of 0.1. In the larger populations on the other hand, the change in the language state L is hardly visible, which means that the innovation A continues to be used with a probability of about 0.1.

A rather unexpected result is that the parameter selection strength s does not seem to exert any considerable influence on the spread of the innovative variant A. This becomes obvious when looking at the individual curves, each representing a specific value of s, in the separate subplots in Figure 5.4: They are all very close together.

Figure 5.4: Simulation results for the interactor selection mechanism with the proportion of leaders of 10% (n = 0.1) in three types of social networks. The population size values N lie on the x-axis and the corresponding mean values of the language state L lie on the y-axis. The set of parameters for each network type and population size used: V = 2, N_T = 10, p_A (t = 0) = 0.1, k = 4, s $\in$ {0.1, 0.2, ..., 1.0}, and t = 100,000. Each individual curve in each separate subplot stands for a particular selection strength value s, each represented in the legend box.
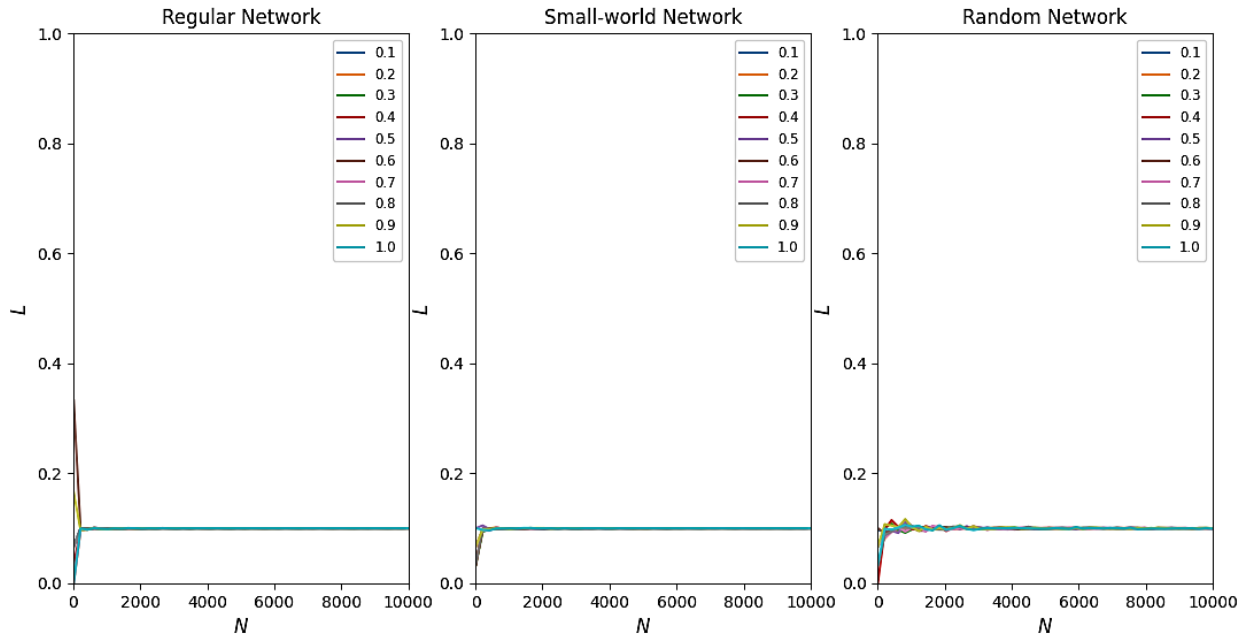

Similar results are observed when the proportion of leaders is increased to 20%, i.e. $n = 0.2$. One difference is that the innovation A sometimes succeeds in spreading in very small populations in all three networks: Figure 5.5 shows that although there are some unsuccessful propagations of the variant A, there are also some successful ones. As for larger populations, there are no considerable changes in the proportion of the innovative variant here: Its usage probability remains at around 0.2, which means that an agent would choose the innovative variant A in only two out of ten interactions on average.

Figure 5.5: Simulation results for the interactor selection mechanism with the proportion of leaders of 20% (n = 0.2) in three types of social networks. The population size values N lie on the x-axis and the corresponding mean values of the language state L lie on the y-axis. The set of parameters for each network type and population size used: $V = 2$, $N_T = 10$, $p_A (t = 0) = 0.2$, $k = 4$, $s \in \{0.1, 0.2, ..., 1.0\}$, and $t = 100,000$. Each individual curve in each separate subplot stands for a particular selection strength value s, each represented in the legend box.
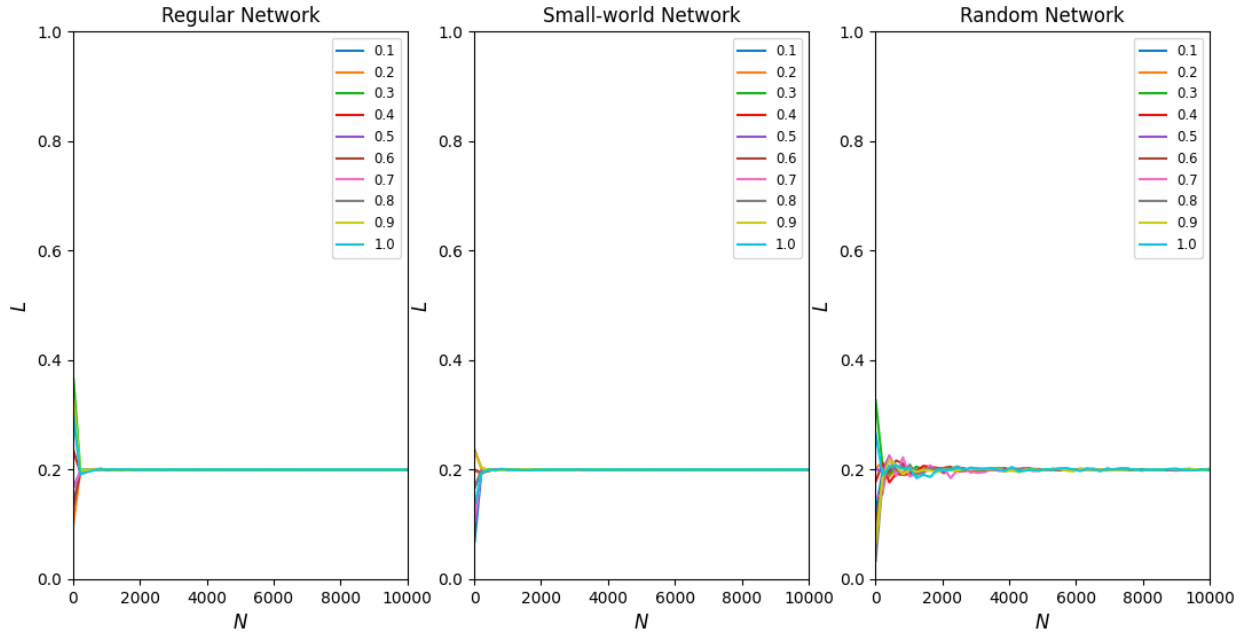
Looking at the results of the model with the interactor selection mechanism in general, the innovation A does not seem to succeed in becoming the majority variant under any of the specified conditions: In most cases, its probability tends to remain close to the initial state of the system.

6 Statistical Modeling

Simulation results show that neutral change, replicator and interactor selection differently affect the change in the proportions of the innovative variant A over time. Replicator selection turns out to be the only mechanism under which, if the selection pressure is strong enough, the innovative variant A manages to diffuse to a large number of agents, thereby becoming the majority variant. Here though the question arises as to what role different factors, such as population size, social network structure, selection strength and the number of leaders in the population, play in the diffusion of the innovation under each of the three mechanisms. Is population size, for example, a significant factor for the innovation diffusion within each mechanism, or does it only have a significant effect when some sort of selection is at play?

To quantitatively identify the trends between the innovation propagation and the named factors as well as their effects on it, I apply multiple linear regression on the data simulated under each individual mechanism of language change. Multiple linear regression is

considered to be a suitable approach to the relations to be investigated between the variables mentioned. First, it allows to include multiple predictors that influence the innovation diffusion and, second, to measure the effect of one predictor while holding all other predictors constant. In this way, the direct effects of specific predictors can be disentangled.

To examine the effects of population size N_A, social network type p, selection strength s and the proportion of leaders n_L on the diffusion of the innovative linguistic variant A, I define three types of the multiple linear regression model, hereafter referred to as the neutral change model, the replicator selection model, and the interactor selection model. All three models are fitted to the simulated data using the R package *stats* (R Core Team, 2022).

Multiple regression models rely on several assumptions, such as linearity, normality of the residuals, homoscedasticity, and the absence of multicollinearity. In this context, it should be noted that the validation of the models identifies violations of these assumptions, with the multicollinearity assumption being the only one that is met. More specifically, diagnostic plots provided in Appendix B clearly demonstrate violations of linearity, normality, and homoscedasticity for each of the models. The model outputs should therefore be viewed with caution, as they might be inaccurate due the violations of the assumptions detected. On the other hand, these violations are in some way unavoidable given the structure of the simulated data. Multiple regression thus turns out to be not the best model for the kind of data being analysed here. It cannot be ruled out that other modelling approaches such as nonlinear regression could provide a better fit to the data. However, this approach could also have disadvantages, as there could be a risk of under- or overfitting the observed data. Therefore, as this thesis is primarily concerned with examining the relationship between the predictors and the response variable rather than making exact predictions, multiple linear regression provides a good basis for exploring general trends that may be useful for further research. In the following, the formal definition for each model and the outputs are presented in detail. The link to the R script for the statistical analysis can be found in Appendix A.

6.1 Model Definitions and Outputs

6.1.1 Neutral Change Model
The neutral change model is applied to the data generated by the multi-speaker Moran model, which assumes the mechanism of neutral change as the sole driver of language change. The neutral change model is defined in the way that it predicts the language state L depending on population size N_A and the type of social network p. It is specified as follows:

$$L_i = \beta_0 + \beta_1 \log_{10}(x_{1i}) + \beta_2 x_{2i} + \varepsilon_i,$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2), \tag{4}$$

where each ith language state value is predicted by the corresponding $\log_{10}$-transformed population size ($\log_{10}(x_{1i})$)[11], and the three-level categorical factor social network type (x_{2i}). The state of language L is predicted by the intercept β_0, plus the coefficients β_1, and β_2, multiplied by the respective values of the independent variables. Furthermore, the model assumes the normal distribution of the residual error ε with mean 0 and variance σ_ε^2.

The numerical output of the neutral change model is given in Table 6.1. The estimated coefficient of $\log_{10}$-transformed population size is positive and not significant ($p = 0.77$, 95% CI [-0-01, 0.013]). Moreover, as indicated by Cohen's f2 value of 0.000191, the population size predictor has a remarkably tiny effect size.[12]

Table 6.1: Output of the neutral change model. Estimated regression coefficients, standard errors, t-values and p-values for the neutral change model presented in eqn 6.1.

Predictor	Coefficient	Estimate	SE	t-value	p-value
Intercept	β_0	0.1956818	0.021108	9.119	<2e-16 ***
$\log_{10}$(Population Size)	β_1	0.0017091	0.005866	0.291	0.771
Small-world Network	β_2	0.0002105	0.0075096	0.028	0.978
Random Network	β_2	-0.0093549	0.0075096	-1.246	0.214
					R² = 0.004912

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

The intercept estimate represents the average usage probability of the innovative variant A in the regular network which functions in the model as a reference level. From the intercept value it can be inferred that on average the value of language state L increases by 0.000211 in small-world networks compared to regular networks. However, the coefficient implies that the difference in L values in these networks is not statistically significant ($p = 0.98$, 95% CI [-0.015, 0.015]). In comparison to regular networks, random networks are associated with an

---

[11] The $\log_{10}$ function is used instead of the log (natural log) function to transform the population size data in the current and two subsequent models due to its simpler interpretation. That is, if one considers a $\log_{10}$-transformed population size value of 3.5, then he knows that the original number is somewhere between a number that has 3 zeros (3000 agents) and 4 zeros (4000 agents). Performing such mental calculations is not as straightforward for e.

[12] Cohen's f² (Cohen, 1988) is an effect size measure that is used in the context of multiple regression. It allows to determine individual effects sizes of categorical and continuous independent variables (Selya et al., 2012). According to Cohen's (1988) guidelines, f² ≥ 0.02 represents a small effect size, f² ≥ 0.15 a moderate effect size, and f² ≥ 0.35 a large effect size. The instructions for the exact calculation of the measure can be found in Selya et al. (2012). My R implementation of the computation can be found in the GitHub repository of the master's thesis project, which is linked in Appendix A.

average decrease of -0.0094 in the response variable. The estimated coefficient is also not significant ($p = 0.21$, 95% CI [-0.024, 0.005]) and thus the difference in mean values of language state L in these networks. In this context, the predictor social network type has a noticeably small effect size (Cohen's $f^2 = 0.004743$).

Finally, R^2 value of the neutral change model equals 0.004912, which means that only about 0.49% of the variance in the response variable is explained by the predictors in the model. Conversely, 99.51% of the variation in language state L is due to chance, or due to factors not included in the model.

6.1.2 Replicator Selection Model

The replicator selection model is applied to the data simulated under the influence of the replicator selection mechanism. The state of language L is modelled here as a function of population size N_A, social network type p, and selection strength s. The replicator selection model is specified as follows:

$$L_i = \beta_0 + \beta_1 \log_{10}(x_{1i}) + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i,$$
$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2), \tag{5}$$

where each ith language state value is predicted by $\log_{10}$-transformed population size ($\log_{10}(x_{1i})$), the three-level categorical factor social network type (x_{2i}), and selection strength (x_{3i}). The latter is specified as a ten-level ordered factor, with 0.1 representing the weakest selection pressure as well as the reference level and with 1 indicating the strongest selection pressure. Therefore, the language state L is predicted by the intercept β_0, plus the coefficients β_1, β_2, and β_3 multiplied by the respective values of the independent variables. Again, the model assumes that the residual error ε is normally distributed with mean 0 and variance σ_ε^2.

The numerical output of the replicator selection model is given in Table 6.2. The estimated coefficient of $\log_{10}$-transformed population size is negative and highly significant ($p < 0.001$, 95% CI [-0.193, -0.18]). This means that an increase in $\log_{10}$-transformed population size by one unit corresponds to an estimated decrease in the language state value by 0.19 units. In addition, the $\log_{10}$-transformed population size turns out to have a large effect on the language state L (Cohen's $f^2 = 0.72$).

The estimated regression slopes of the categorical predictor social network structure show two differences to one shared reference level, the regular network. Therefore, the p-values correspond here to the null-hypothesis that the difference between regular network and small-world network on the one hand and the difference between regular network and random

network on the other hand is equal to 0. That is, the first slope, small-world network, represents the difference between the values of L in regular and small-world networks. The coefficient is positive ($\beta = 0.000121$) and not significant ($p = 0.98$, 95% CI [-0.008, 0.008]), indicating that there is no significant difference between language states in these networks. Equivalently, the difference between language states in regular and random networks is showed by the negative estimated coefficient ($\beta = -0.00175$), which is also not significant ($p = 0.71$, 95% CI [-0.01, 0.006]). This also indicates that there is no significant difference between values of L in these networks. Additionally, the Cohen's f^2 value of 0 indicates that social network type has no effect on the response variable.

Table 6.2: Output of the replicator selection model. Estimated regression coefficients, standard errors, t-values and p-values for the replicator selection model presented in eqn 6.2.

Predictor	Coefficient	Estimate	SE	t-value	p-value
Intercept	β_0	1.1144809	0.0130546	85.371	< 2e-16 ***
$\log_{10}$(Population Size)	β_1	-0.1862372	0.0032835	-50.246	< 2e-16 ***
Small-world Network	β_2	0.0001209	0.0041989	0.029	0.977
Random Network	β_2	-0.0017466	0.0041989	-0.416	0.677
Selection Strength 0.2	β_3	0.1520276	0.0076662	19.831	< 2e-16 ***
Selection Strength 0.3	β_3	0.2616672	0.0076662	34.133	< 2e-16 ***
Selection Strength 0.4	β_3	0.3423204	0.0076662	44.653	< 2e-16 ***
Selection Strength 0.5	β_3	0.4009770	0.0076662	52.305	< 2e-16 ***
Selection Strength 0.6	β_3	0.4432964	0.0076662	57.825	< 2e-16 ***
Selection Strength 0.7	β_3	0.4735893	0.0076662	61.777	< 2e-16 ***
Selection Strength 0.8	β_3	0.4952613	0.0076662	64.603	< 2e-16 ***
Selection Strength 0.9	β_3	0.5103107	0.0076662	66.567	< 2e-16 ***
Selection Strength 1	β_3	0.5208770	0.0076662	67.945	< 2e-16 ***
					$R^2 = 0.735$

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

The regression output for the predictor selection strength s shows nine slopes, namely, nine differences to one shared reference level. In the case of the present dataset, the reference level is the weakest selection strength value, i.e. $s = 0.1$. The estimated coefficients indicate that there is a significant difference in the estimated mean response between each of the selection pressure levels presented in Table 6.2 and the reference level of 0.1.

For example, the estimated coefficient for selection strength of 0.8 ($\beta = 0.5$) is highly significant ($p < 0.001$, 95% CI [0.48, 0.51]), suggesting that there is a statistical evidence of a difference in mean response between the given value of s and the reference level. In other words, the probability that the innovative variant is used increases by 0.5 units when selection

strength changes from 0.1 to 0.8. Furthermore, the effect of selection strength on the response variable is substantial (Cohen's f^2 = 1.95). Therefore it can be concluded that selection strength has a significant positive effect on the propagation of innovation: When selection strength increases, the value of language state L also tends to increase.

Finally, the R^2 value of 0.735 indicates that the replicator selection model describes about 73,5% of the variance in language state L.

6.1.3 Interactor Selection Model

The interactor selection model is applied to the data simulated by the multi-speaker Moran model with the mechanism of interactor selection. The language state L in this model is defined as a function of population size N_A, social network type p, selection strength s, and the size of the leader group n_L. The model specification is as follows:

$$L_i = \beta_0 + \beta_1 \log_{10}(x_{1i}) + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i,$$

$$\varepsilon_i \sim N(0,\ \sigma_\varepsilon^2), \tag{6}$$

where each ith language state value is predicted by $\log_{10}$-transformed population size ($\log_{10}(x_{1i})$), the three-level categorical factor network type (x_{2i}), the ten-level ordered categorical factor selection pressure (x_{2i}), and the two-level ordered categorical factor proportion of leaders. The latter factor consists of two values, i.e. 0.1 and 0.2, which means that the proportion of leaders is either 10% or 20%. Therefore, the language state L_i is predicted by the intercept β_0, plus the coefficients β_1, β_2, β_3, and β_4 multiplied by the respective values of the predictor variables. As in the neutral change and replicator selection models, the residual error ε is assumed to be normally distributed with mean 0 and variance σ_ε^2. The numerical output of the interactor selection model is given in Table 6.3.

The estimated coefficient of $\log_{10}$-transformed population size is significant (p < 0.001, 95% CI [0.005, 0.006]) and positive (β = 0.0053). This means that there is a positive relationship between $\log_{10}$-transformed population size and language state L. In other words, as $\log_{10}$-transformed population size increases by one unit, the value of L tends to increase by 0.0053 units. On the other hand, despite the high significance of the estimated coefficient, population size has a very small effect on the response (Cohen's f^2 = 0.0199).

The estimated regression slopes of the categorical predictor network structure show two differences to regular network. The first slope of β = -0.00075 shows the difference between the mean values of language state L in regular and small-world networks. The estimated

coefficient is not significant (p = 0.15, 95% CI [-0.002, 0]), meaning that the difference between mean response values is not significantly different from 0 in these two networks. Equivalently, the second estimate of β = -0.0009 demonstrates the difference between means of L in regular and random networks. Again, the estimated coefficient is not significant (p = 0.09, 95% CI [-0.002, 0]), indicating that there is no significant difference between values of L in these networks. The effect of the predictor social network type on the response is very small (Cohen's f^2 = 0.00074).

Table 6.3: Output of the interactor selection model. Estimated regression coefficients, standard errors, t-values and p-values for the interactor selection model presented in eqn 6.3.

Predictor	Coefficient	Estimate	SE	t-value	p-value
Intercept	β_0	0.0799924	0.0016137	49.570	< 2e-16 ***
$\log_{10}$(Population Size)	β_1	0.0053440	0.0004024	13.279	< 2e-16 ***
Small-world Network	β_2	-0.0007454	0.0005146	-1.448	0.1475
Random Network	β_2	-0.0008791	0.0005146	-1.708	0.0876 .
Selection Strength 0.2	β_3	0.0003178	0.0009396	0.338	0.7352
Selection Strength 0.3	β_3	0.0016716	0.0009396	1.779	0.0753 .
Selection Strength 0.4	β_3	0.0006846	0.0009396	0.729	0.4662
Selection Strength 0.5	β_3	0.0007974	0.0009396	0.849	0.3961
Selection Strength 0.6	β_3	0.0013957	0.0009396	1.485	0.1375
Selection Strength 0.7	β_3	-0.0001904	0.0009396	-0.203	0.8394
Selection Strength 0.8	β_3	0.0004206	0.0009396	0.448	0.6545
Selection Strength 0.9	β_3	0.0014195	0.0009396	1.511	0.1309
Selection Strength 1	β_3	0.0009590	0.0009396	1.021	0.3074
Proportion of Leaders 0.2	β_4	1.002580	0.0004202	238.599	< 2e-16 ***
					R^2 = 0.8641

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; . $p < 0.1$

The estimated differences between individual selection strength values presented in Table 6.3 and the reference level of 0.1 are not significant. This differs from the output of the replicator selection model where these differences are significant. Thus, the selection strength does not appear to affect the innovation diffusion if the mechanism of language change is interactor selection, which is also indicated by its very small size of effect (Cohen's f^2 = 0.0015).

From the numerical output in Table 6.3 it can be further seen that the difference in mean response between the proportion of leaders of 20% and the reference level, i.e. the proportion of leaders of 10%, is highly significant ($p < 0.001$, 95% CI [0.099, 0.101]). Looking at the estimated coefficient for this difference (β = 1.0026), it can be further concluded that the

value of language state L is higher in populations with 20% of leaders. In addition, the effect of leader group size is very large (Cohen's $f^2 = 6.34$).

Finally, the R^2 value of the interactor selection model equals 0.864, meaning that all the predictors together explain 86,4% of the variance in the response variable.

6.2 Effects on Language Change

The aims of the current study were twofold. First, I implemented an agent-based model to simulate language change as a process of innovation diffusion under three different mechanisms of propagation: neutral change, replicator selection, and interactor selection. Second, I used simulated data to explore the effects of population size, network type, selection strength, and leader group size on the diffusion of the innovative variant A. This was achieved through statistical analysis involving the multiple regression model that was specified for each mechanism independently. Hereafter, the effects of each predictor variable on the response are compared to each other, with respect to the three main mechanisms of language change: neutral change, replicator and interactor selection. All effect displays are created using the package *effects* (Fox and Weisberg, 2019) in R (R Core Team, 2022).
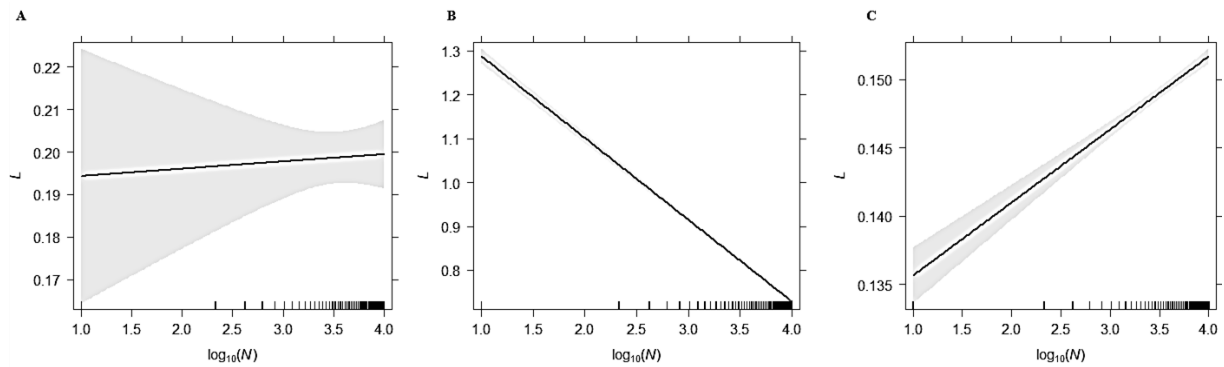
6.2.1 Population size
The outputs of the multiple regression models demonstrate that population size takes effect on the spread of the innovative variant only if the mechanism of language change is replicator or interactor selection. It is important to note however, that the trends in the relationship between population size and language change differ under these two mechanisms. It becomes obvious from Figure 6.1 that, if the mechanism of replicator selection is at play, the innovative variant A propagates more easily in smaller populations. On the other hand, in larger populations, innovation diffusion tends to slow down. In contrast, if the mechanism of interactor selection is involved, the trend in the relationship between the predictor and the response variable is positive. This is an interesting result as it indicates that in larger populations the chance that the innovation propagates is greater than in smaller populations. However, despite the estimated significance of the independent variable population size in the interactor selection model, it should be noted that its effect size was found out to be very small in comparison to the effect size of the proportion of leaders in the model. This leads to the assumption that the population size is not very important in predicting the probability of the innovative variant and that the predictor such as the number of leaders is more strongly associated with the response variable.

A positive relationship between the response and population size is also observed under the mechanism of neutral change, although the impact of population size becomes insignificant in terms of the spread of the innovation under this condition.

Moreover, in the context of the present results, Nettle's (1999a) claim that the diffusion of a linguistic variant is faster in smaller communities appears to be true only for the case where the replicator selection mechanism is involved. This suggests that the rapid spread of an innovation depends not only on the small population size, but also on what kind of mechanism is involved in the process of language change.

Figure 6.1: Effect of population size on innovation diffusion under the mechanism of neutral change (A), replicator selection (B), and interactor selection (C). The shaded region in each of the subplots represents the pointwise confidence intervals based on the standard errors computed from the covariance matrix of the fitted regression coefficients. The rug plot at the bottom of each subplot shows the location of the $\log_{10}$-transformed population size values.

Moreover, it is noticeable that the innovation reaches a high probability, either completely replacing the conventional variant in very small populations or becoming the majority variant in larger populations, if selection is associated with the innovative replicator. For the remaining two mechanisms, the innovative variant does not achieve probability values that are significantly higher than the initial state. These observations suggest that the innovative variant has a chance of successful spread in a population, i.e. of becoming a newly established convention, only if it is linked to some kind of bias. Going back to the theory, such biases can imply, to give some examples, an easier pronunciation, a simpler morphological structure, or an easier memorisation.

6.2.2 Social Network Structure

All three models show that the social networks do not differ significantly from each other in terms of the average probability of the innovative variant, whatever mechanism of language change comes into play. From Figure 6.2 (A), it is obvious that when change takes place at random, the innovation gets fixed at a value around 0.2, i.e. the initial probability value set at

the start of each simulation, in all three networks. A similar situation is observed in each network type in two other scenarios: When the change is due to replicator selection, the average probability of the innovative variant lies around 0.8 (Figure 6.2, B), and when the mechanism of interactor selection is involved, this value falls to around 0.15 (Figure 6.2, C).

At this point it should be noted that the non-significant differences in average usage probabilities of the innovation between the regular, small-world and random networks are surprising results, given the computational models that argue in favour of network structure effects (Ke et al., 2008; Gong et al., 2011). On the other hand, there are studies (Spike, 2018; Raviv et al., 2020) which show that network structure has no obvious impact on language change. However, I believe that it is more likely that the way the simulation experiments are defined in the current study might not express the differences between the network types well enough. One explanation might be that the distribution of population sizes is not normal: It is skewed towards the small population size values. In other words, population sizes are mostly clustered near the value of $\log_{10} = 4$ meaning that lower values of population size are underrepresented. By pointing this out, I would like to say that the effect of network structure on language change might be more visible in smaller populations and rather unremarkable in larger populations. Expressing it in other words, the non-significant effect of network structure may be biased towards larger population sizes.

Figure 6.2: Effect of social network type on innovation diffusion under the mechanism of neutral change (A), replicator selection (B), and interactor selection (C). The values on the x-axis, denoted as p, stand for a regular network, a small-world network, and a random network, respectively. The grey vertical lines in the subplots display the confidence intervals.
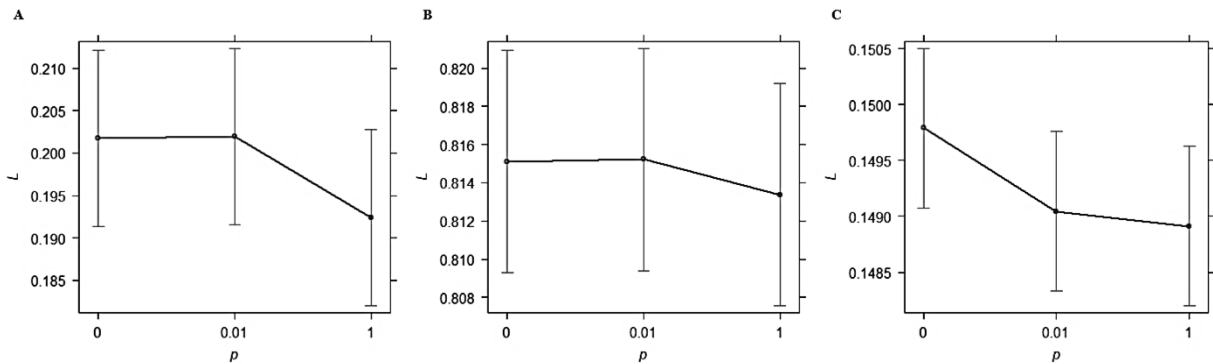
In addition, it can be assumed that network structure might interact in some way with population size that cannot be captured by the multiple linear regression models defined. More specifically, the influence of network structure on the innovation diffusion may change depending on the value of population size: The effect of network structure on language change might not be the same for all population sizes. To investigate this further, a larger sample of population size values as well as an inclusion of the multiplicative effect of both

predictors, i.e. network structure and population size, on the response variable into the multiple linear regression model, may be useful.

The observed non-significant effect of network type could also be the result of the selected parameters. For example, the values of the rewiring probability p for the regular and small-world network are very close to each other: A slightly higher p for the small-world network could create an even stronger contrast between the two network types, which could lead to different results. The same applies to the parameter k, i.e. the average number of neighbours per agent, which remains constant for any population size, whether it is small or large. Here, however, it could be assumed that the number of such connections depends on the population size and thus would tend to increase with increasing population size. It cannot be ruled out that such variation in the k parameter could yield different results. But here the question arises as to whether the increase in population size is actually associated with the increase in the number of connections of an individual in the real-world situation.

Despite the non-significant differences between the networks under each mechanism of language change, some tendencies can still be identified. It is evident from Figure 6.2 that the change in the probability of the innovative variant tends to be slower in the random network than in the regular and small-world networks regardless of which mechanism of language change is activated. This observation can be attributed to the fact that the structure of the random network is characterised by a relatively small number of clusters ($C(p = 1) = 0.01$) which implies many long-distance connections between agents. This means that the innovative variant needs to take a relatively large number of steps until reaching each individual agent in the social network, which ultimately slows down its propagation. By contrast, the structure of the regular and small-world networks exhibits a large number of clusters ($C(p = 0) = 0.5$; $C(p = 0.01) = 0.49$) and thus many local connections, which tends to accelerate the diffusion process.

Looking at the individual mechanisms of language change, it is interesting to note that the average values of language state L are almost identical when replicator selection is involved (Figure 6.2, B). A similar situation is observed when the change occurs due to random processes (Figure 6.2, A). Although it is worth noting here that, when neutral change is involved, the contrast between the regular and small-world networks on one side and the random network on the other is slightly higher than in the case with the replicator selection mechanism. On the other hand, when the interactor selection mechanism is activated, the contrast between the three network types becomes even stronger compared to the other two mechanisms (Figure 6.2, C). These observations may suggest that when the mechanism of

language change is replicator selection, the way agents are linked becomes a less significant factor in the spread of a linguistic variant. A similar assumption can be made about the mechanism of neutral change. By contrast, when the mechanism of language change is interactor selection, the character of the connections between agents appears to play a greater role in the diffusion process. This makes sense given that a population has a relatively small proportion of speakers with a leader role. Thus, in the regular network, where each agent has an equal number of ties, the chance of an agent interacting with a leader and as a result memorising the innovative variant produced by him is presumably higher than in the random network, where the number of connections an agent has can vary greatly. This could lead to a lower probability of such an interaction with a leader and thus to a slowdown in innovation propagation.

6.2.3 Selection Strength

The strength of selection is found to have a significant impact on language change when replicator selection is involved. On the other hand, when selection is associated with an interactor, the effect of the selection strength is not significant.

Figure 6.3: Effect of selection strength on innovation diffusion under the mechanism of replicator selection (A), and interactor selection (B). The grey vertical lines in the subplots display the confidence intervals.
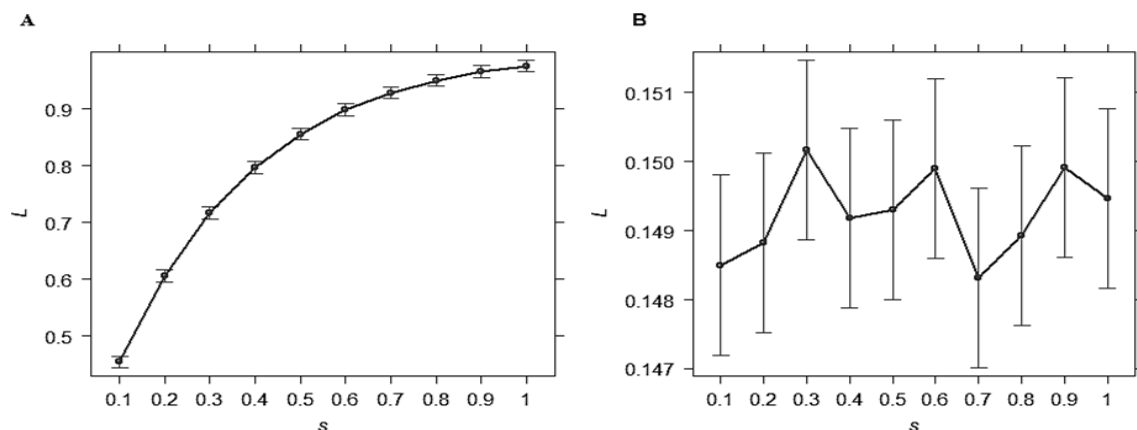

By comparing the two subplots shown in Figure 6.3, it becomes clear that the effects of selection strength differ considerably. At this point it must be said that the non-significant effect of selection strength that occurs when the mechanism of interactor selection is operating is a rather unexpected result. Selection strength is expected to boost the probability of the innovative variant, as evident in the case of language change under the replicator selection mechanism. More specifically, as seen in Figure 6.3 (A), as the strength of selection increases, the proportion of the innovation increases at a decreasing rate. In the case of interactor selection, however, selection strength loses its role as the key force for the

innovation propagation: There is no clear pattern in the relationship between language change and the strength of selection (see Figure 6.3, B). This may be due either to the definition of the interactor selection mechanism within the agent-based model or because other factors such as population size and the number of leaders in the population have a more significant impact on language change.

6.2.4 Leader Group Size

The differences between the average probability of the innovation in the population with 10% of leaders and that in the population with 20% of leaders are found to be significant. Here, however, it is clear that the average probability values of both groups remain identical to the initial states, which is evident from Figure 6.4.

Figure 6.4: Effect of the proportion of leaders on innovation diffusion under the mechanism of interactor selection. The grey vertical lines in the plot display the confidence intervals.
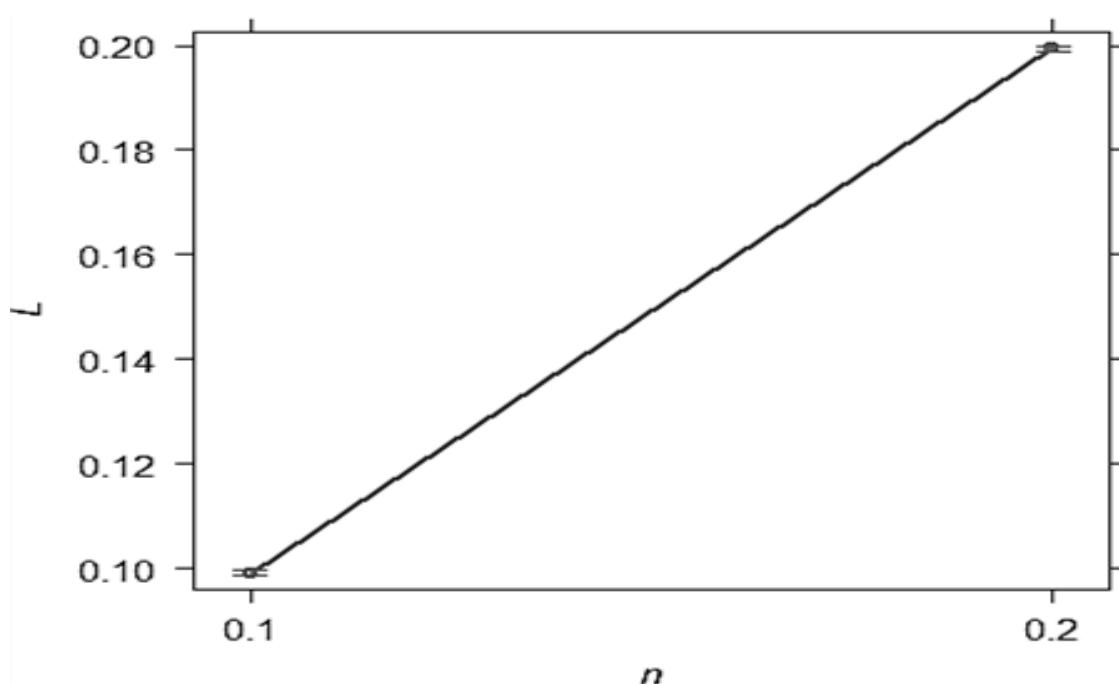
This observation suggests that the usage probability of the innovative variant is on average not higher than the proportion of leaders in the population. This means that the innovative variant remains to be used predominantly by leaders. The reason for this could lie in the definition of the interactor selection mechanism within the multi-speaker Moran model. It should be kept in mind that agents in the model are defined in such a way that they not only can listen to their counterpart but also reinforce their own behaviour. It can therefore be assumed that the self-reinforcement of agents may slow down the innovation propagation driven by the leaders.

7 Conclusions and Outlook

It is a matter of the continuing debate whether change in structural properties of language is affected by population size or not. This thesis has been an attempt to explore the effect of population size on language change in the different scenarios, using an approach that integrates agent-based modeling and statistical analysis. The multi-speaker Moran model simulates language change as a process of innovation diffusion driven either by neutral change or by replicator or interactor selection processes. The results of statistical analysis suggest that population size significantly effects the innovation propagation only if some kind of selection, be it replicator or interactor selection, is involved. In addition, it is to be noted that the trends in the relationship between population size and language change appeared to be different depending on the type of selection mechanism. Namely, it was shown that if replicator selection is at play, the relationship between population size and language change is negative, which is an indication that smaller populations are more susceptible to change than larger ones. If, on the other hand, the mechanism is interactor selection, this relationship is found to be positive, which suggests the opposite pattern: Change is more likely to spread in larger than in smaller populations. However, the interactor selection model shows a small effect size for the independent variable population size, indicating a weak relationship between this variable and language change.

The strength of selection was shown to have a significant impact on innovation diffusion only if replicator selection is present. In small populations, the innovative variant was shown to diffuse to a linguistic majority variant, independent of the strength level of selection. In larger populations, successful diffusion of the innovation to the majority of language users was observed after the selection strength threshold of 0.5 was exceeded. On the other hand, when interactor selection came into play, no significant differences were observed between the language states with respect to the various levels of selection strength. This led to the assumption that other factors such as population size and the proportion of leaders in a population might have a more dominant role in language change in this case.

Furthermore, the results of the statistical analysis showed no effect of network structure on the language change process, with regular, small-world, and random networks all demonstrating similar patterns of the propagation of the innovative variant, independent of the mechanism involved. However, these findings should be viewed with caution, as they could be biased due to the left-skewed nature of the sampled population size data or due to the lack of well recognizable differences between the individual network types. Therefore, in order to

further explore the effect of network type within the defined agent-based model and to confirm or refute its influence on language change, some aspects of social networks may be modified. Potential adjustments in this context could be the following: First, an average number of nearest neighbours per agent could be varied according to population size. Second, since social networks are dynamic systems in which the connections between agents change over time, a generation turnover could be added that would allow individual agents to be removed from or added to the network, thus mirroring the events of death and birth in the real world. Moreover, an interaction between population size and social network structure could be incorporated into the definition of the multiple regression models. This adjustment could enhance the models' fit to the data used in this study. In addition, further insights into the relationship between population size and language change could be gained separately for regular, small-world and random networks.

Based on the results obtained in this thesis it can be summarised that the significance of the effect of population size on language change is accompanied by selective pressures associated either with the innovative variant itself or with an interactor possessing a prestigious social status and using it. Nonetheless, in the context of replicator selection population size exhibits a more pronounced association with language change than in the context of interactor selection.

8 Appendix A

Source Code and Data

The source code of the implementation of the Moran model with multiple speakers, instructions for use, practical examples and the script for statistical analysis can be found here: https://github.com/yuliyamkh/ABM_AgentPy.git

The data generated by the multi-speaker Moran model for each mechanism of language change can be found under the following link:

https://github.com/yuliyamkh/thesis_project/tree/master/sim_data

9 Appendix B

Multiple Regression Assumptions

Linearity

The multiple linear regression runs here assume that there is a linear relationship between the dependent variable and the predictor variables. At this point, it should be noted that the linearity assumption is checked only for the continuous predictor, that is population size. For multicategory predictors, such as in this case social network structure, selection strength, proportion of leaders, this assumption is always met.

Figure 9.1: Linearity assumption. Relationship between proportion of the innovation and population size under the mechanism of neutral change (A), replicator selection (B), and interactor selection (C). The blue line is the line of best fit and the red line is the loess line representing the local regression smoother. Gray areas show 95% confidence intervals.
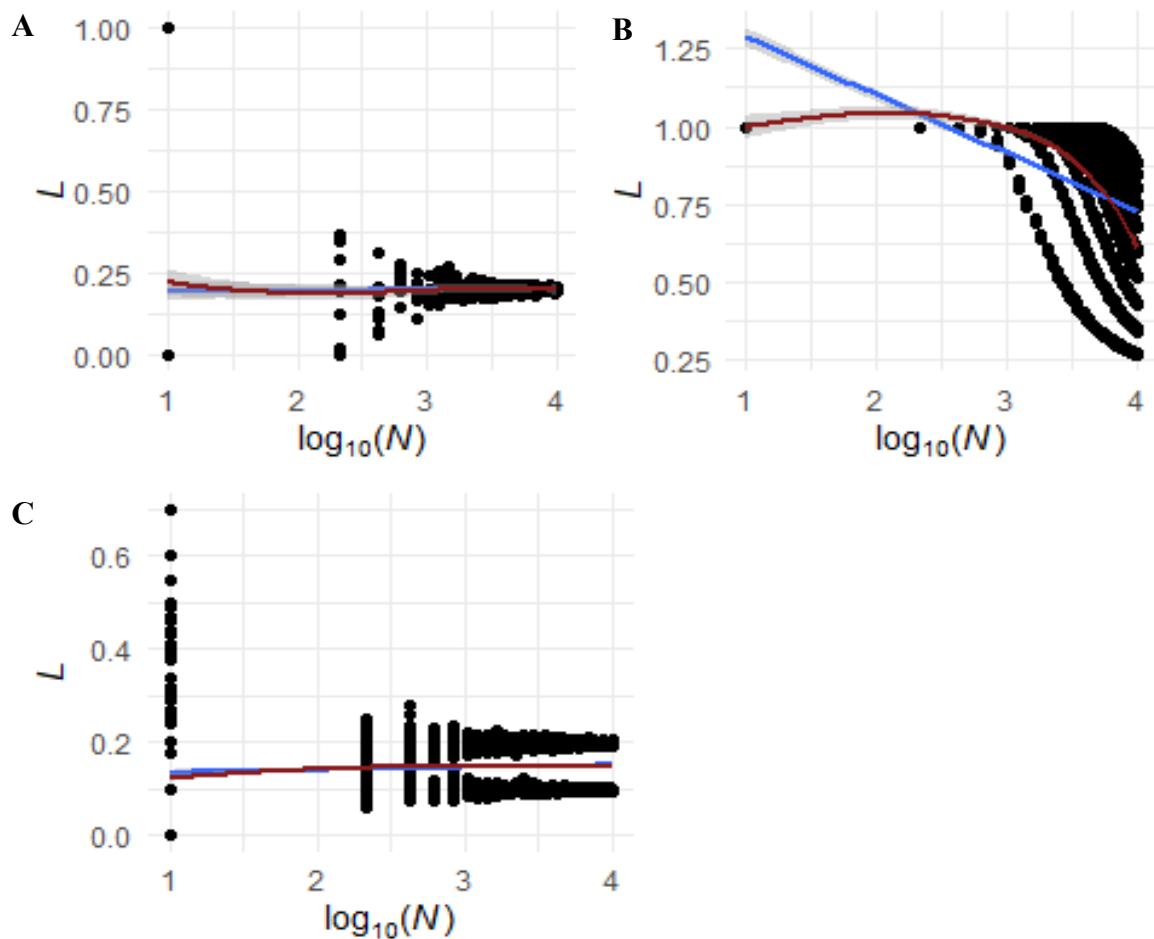

Here, the loess line is added to illustrate the shape of the trajectory when the linearity assumption is relaxed. The confidence intervals of the two lines, blue and red, should always overlap. The violation of the linearity assumption can be clearly observed in the relationship

between *L* and population size in the model with the mechanism of replicator selection (Figure 9.1, B). The line of best fit and the line of the local regression smoother diverge significantly from each other. For the two other models, the violation is not as severe as for the replicator selection model.

Normality

The normality assumption assumes that the residuals are normally distributed. In all three cases, this assumption is not met. As seen in Figure 9.2, the two curves, the empirical density curve and the theoretical density curve, are far apart in each subplot. The density plot in A and C reveals positive skew, and the density plot in B negative skew.

Figure 9.2: Normality assumption. The blue solid line provides a smoothed estimate of the distribution of the observed residuals. The dashed line is the best fit normal curve.
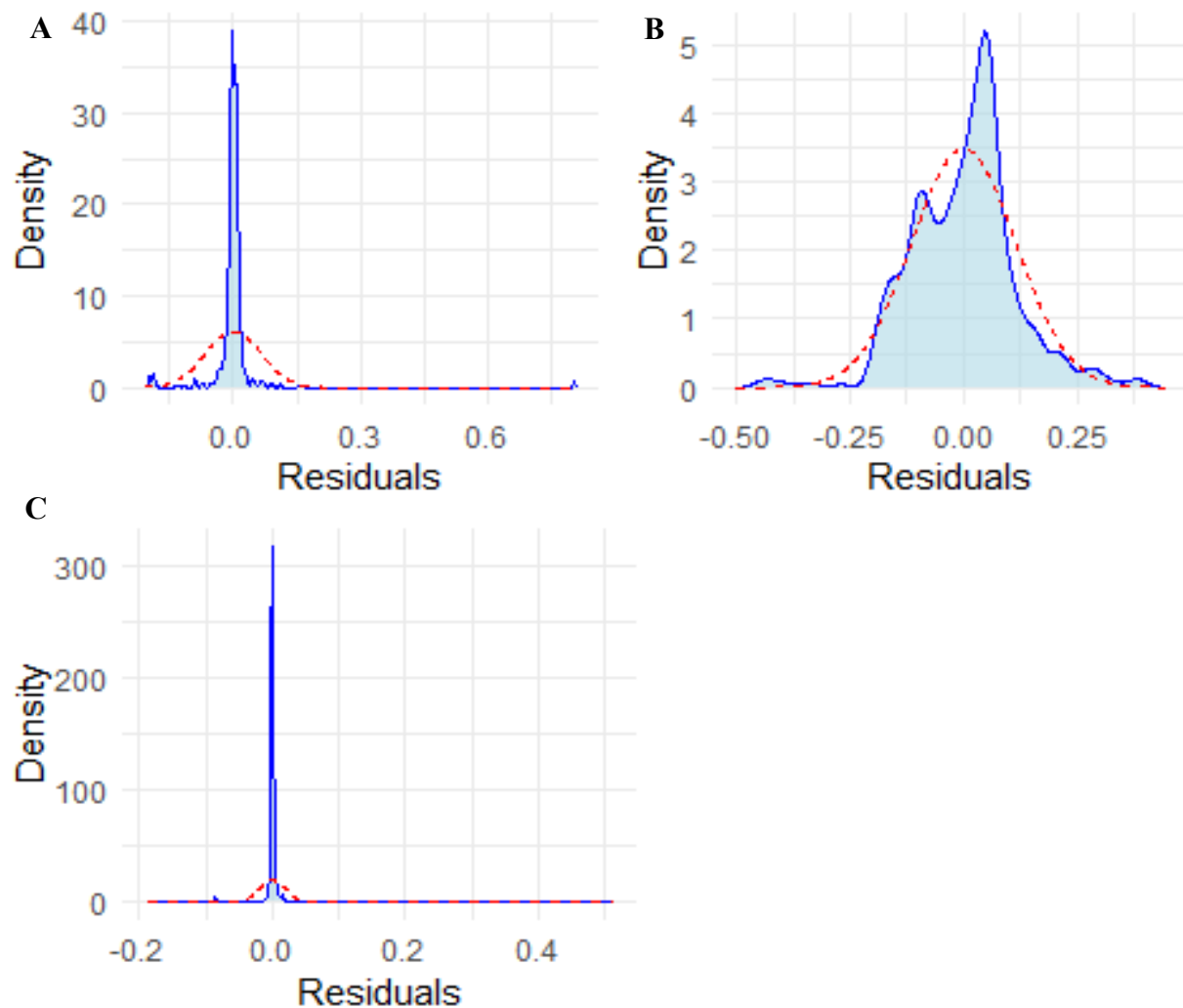
Homoscedasticity

The assumption of homoscedasticity implies that the spread of residuals is nearly identical along the regression line. A clear violation of the assumption is observed in each model. An unequal variance can be seen in each subplot in Figure 9.3. The following pattern can be observed in all three models: As the fitted values increase, the variance of the residuals decreases.

Figure 9.3: Homoscedasticity assumption.
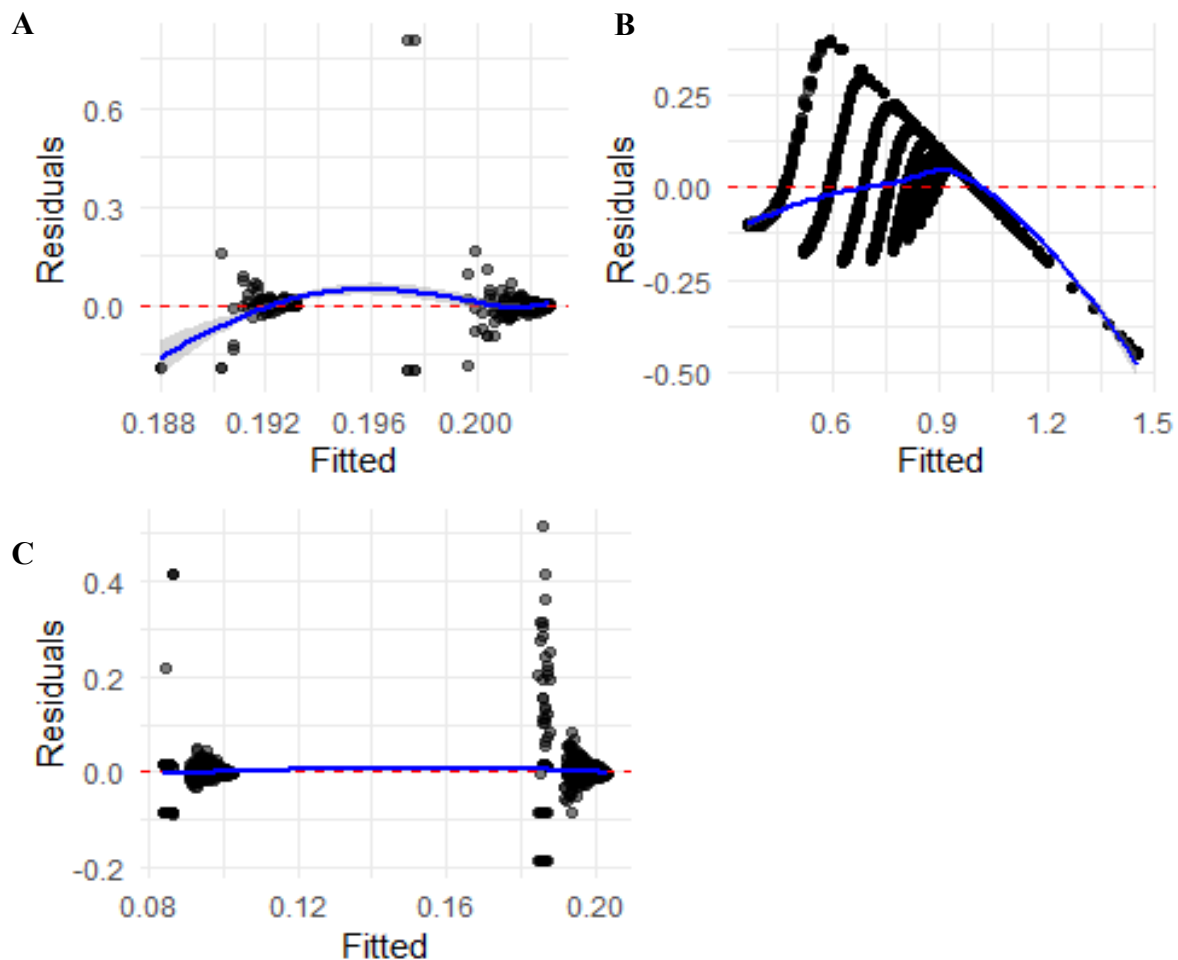
Multicollinearity

To assess whether there are correlations between individual predictors, the variance inflation factor (VIF) for each model is computed with the R package *car* (Fox & Weisberg, 2011). For the neutral change model, all variance inflation factors are equal to 1, which indicates that the independent variables are not correlated with each other. The same is true for both other models, the replicator selection model and the interactor selection model.

Bibliography

Atkinson, Q. D. (2011). Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science, 332*, 346–349.

Baxter, G. J., & Blythe, R. A. (2006). Utterance selection model of language change. *Physical Review E, 73*(4), 1-21.

Baxter, G. J., Blythe, R. A., Croft, W. B., & McKane, A. J. (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, *21*, 257 - 296.

Beckner, C., Blythe, R. A., Bybee, J., Christiansen, M. H., Croft, W. B., Ellis, N. C., Holland, J. H., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position Paper. *Language Learning, 59*, 1-26.

Bentley, R. A., Lipo, C. P., Herzog, H. A., & Hahn, M. W. (2007). Regular rates of popular culture change reflect random copying. *Evolution and Human Behaviour, 28*, 151-158.

Bentz, C. (2018). *Adaptive languages: An information-theoretic account of linguistic diversity.* De Gruyter Mouton.

Blythe, R. A. & McKane, J. (2007). Stochastic models of evolution in genetics, ecology, and linguistics. *Journal of Statistical Mechanics*.

Blythe, R. A. (2011). Neutral evolution: A null model for language dynamics. *Advances in Complex Systems*.

Blythe, R. A., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, *88*, 269 - 304.

Bromham, L., Hua, X., Fitzpatrick, T. G., & Greenhill, S. J. (2015). Rate of language evolution is affected by population size. *Proceedings of the National Academy of Sciences, 112*, 2097 - 2102.

Bybee, J. (2001). Phonology and language use. *Cambridge Studies in Linguistics, 94*.

Cohen, J. E. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Lawrence Erlbaum Associates.

Croft, W. B. (2000). *Explaining language change: An evolutionary approach*. Pearson Education Limited

Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. Harper & Row.

Donohue, M., & Nichols, J. (2011). Does phoneme inventory size correlate with population size? *Linguistic Typology*, *15*, 161-170.

Fenk-Oczlon, G., & Pilz, J. (2021). Linguistic complexity: Relationships between phoneme inventory size, syllable complexity, word and clause length, and population size. *Frontiers in Communication*, *6*, 1-7.

Foramitti, J. (2021). AgentPy: A package for agent-based modeling in Python. *Journal of Open Source Softw*are, *6*(62).

Fox, J., & Weisberg, S. (2019): *An R companion to applied regression*.

Frank, S., & Smith, K. (2020). Natural population growth can cause language simplification. *Proceedings of Evolang*.

Gong, T., Tamariz, M., & Jäger, G. (2012). Individual and social effects on linguistic diffusion. *The Evolution of Language*, 142-149.

Greenhill, S. J., Hua, X., Welsh, C. F., Schneemann, H., & Bromham, L. (2018). Population size and the rate of language evolution: A test across Indo-European, Austronesian, and Bantu languages. *Frontiers in Psychology, 9*.

Hagberg, A. A., Schult, D. A., Swart, P., & Hagberg, J. (2008). *Exploring network structure, dynamics, and function using NetworkX*.

Hay, J. L., & Bauer, L. (2007). Phoneme inventory size and population size. *Language, 83*, 388-400.

Hull, D. L. (1988). *Science as a process: An evolutionary account of the social and conceptual development of science.* The University of Chicago Press.

Katz, E., & Lazarsfeld, P. (1955). *Personal influence*. Free Press.

Kauhanen, H. (2016). Neutral change. *Journal of Linguistics*, *53*, 327-358.

Ke, J., Gong, T., & Wang, W. (2008). Language change and social networks. *Communications in Computational Physics*, *3*, 935-949.

Labov, W. (2001). *Principles of linguistic change, Volume 2: Social factors*. Wiley-Blackwell.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE, 5*(1).

Kazil, J. L., Masad, D., & Crooks, A. T. (2020). Utilizing Python for Agent-Based Modeling: The Mesa Framework. In T. Robert, B. Halil, D. Christopher, H. Ayaz & H. Muhammad (Eds.), *Social, cultural, and behavioral modeling* (pp. 308-317). Springer International Publishing.

Moran, P. A. (1963). *The statistical processes of evolutionary theory*. Clarendon Press.

Moran, S., McCloy, D. R., & Wright, R. A. (2012). Revisiting population size vs. phoneme inventory size. *Language, 88*, 877-893.

Nettle, D. (1999a). Is the rate of linguistic change constant. *Lingua*, *108*, 119-136.

Nettle, D. (1999b). Using Social Impact Theory to simulate language change. *Lingua, 108*, 95-117.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review, 97*, 362-376.

Selya, A. S., Rose, J. S., Dierker, L., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Frontiers in Psychology*, *3*.

Spike, M. (2018). Language complexity as an interaction between social structure, innovation, and simplicity. *The Evolution of Language. Proceedings of the 12th International Conference on the Evolution of Language (Evolang12)*, 489-491.

Trudgill, P. W. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press.

Van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual.* Centrum voor Wiskunde en Informatica Amsterdam.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*, 440-442.

Weisberg, S., & Fox, J. (2010). *An R companion to applied regression*. SAGE Publications.

Wichmann, S., Stauffer, D., Schulze, C., & Holman, E.W. (2007). Do language change rates depend on population size? https://arxiv.org/abs/0706.1842

Wichmann, S., & Holman, E.W. (2009). Population Size and Rates of Language Change. *Human Biology: The Official Publication of the American Association of Anthropological Genetics.*

Wichmann, S., Rama, T., & Holman, E.W. (2011). Phonological diversity, word length, and population sizes across languages: The ASJP evidence. *Linguistic Typology*, 15, 177-197.

R Core Team (2022). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing.

Raviv, L., Meyer, A. S., & Lev-Ari, S. (2020). The role of social network structure in the emergence of linguistic structure. *Cognitive science, 44*.

Reali, F., Chater, N., & Christiansen, M. H. (2014). The paradox of linguistic complexity and community size. In E. A. Cartmill, S. Roberts, H. Lyn & H. Cornish (Eds.), *The evolution of language: Proceedings of the 10th International Conference* (pp. 270-277). World Scientific.

Rogers, E.M. (1995). *Diffusion of innovations*. 4th edn. Free Press.