



**HAL**  
open science

## How Network Structure Shapes Languages: Disentangling the Factors Driving Variation in Communicative Agents

Mathilde Josserand, Marc Allasonnière-tang, François Pellegrino, Dan Dediu,  
Bart de Boer

► **To cite this version:**

Mathilde Josserand, Marc Allasonnière-tang, François Pellegrino, Dan Dediu, Bart de Boer. How Network Structure Shapes Languages: Disentangling the Factors Driving Variation in Communicative Agents. *Cognitive Science*, 2024, 48 (4), pp.e13439. 10.1111/cogs.13439 . hal-04547802

**HAL Id: hal-04547802**

**<https://hal.science/hal-04547802>**

Submitted on 16 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# How network structure shapes languages: disentangling the factors driving variation in communicative agents

Mathilde Josserand<sup>1\*</sup>, Marc Allasonnière-Tang<sup>1,2</sup>, François Pellegrino<sup>1</sup>, Dan Dediu<sup>3,4,5</sup>, & Bart de Boer<sup>6</sup>

<sup>1</sup> Laboratoire Dynamique du Langage, Université Lyon 2 - CNRS UMR 5596, Lyon, France

<sup>2</sup> Muséum National d'Histoire Naturelle, Paris, France

<sup>3</sup> Department of Catalan Philology and General Linguistics, University of Barcelona, Barcelona, Spain

<sup>4</sup> University of Barcelona Institute for Complex Systems (UBICS), Barcelona, Spain

<sup>5</sup> Catalan Institute for Research and Advanced Studies (ICREA), Barcelona, Spain

<sup>6</sup> Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium

\* [mathilde.josserand@univ-lyon2.fr](mailto:mathilde.josserand@univ-lyon2.fr)

**keywords:** language evolution, network structure, iterated learning, interindividual variation, Bayesian agents, communicative networks

## ABSTRACT

Languages show substantial variability between their speakers, but it is currently unclear how the structure of the communicative network contributes to the patterning of this variability. While previous studies have highlighted the role of network structure in language change, the specific aspects of network structure that shape language variability remain largely unknown. To address this gap, we developed a Bayesian agent-based model of language evolution, contrasting between two distinct scenarios: language change and language emergence. By isolating the relative effects of specific global network metrics across thousands of simulations, we show that global characteristics of network structure play a critical role in shaping inter-individual variation in language, while intra-individual variation is relatively unaffected. We effectively challenge the long-held belief that size and density are the main network structural factors influencing language variation, and show that path length and clustering coefficient are the main factors driving inter-individual variation. In particular, we show that variation is more likely to occur in populations where individuals are not well-connected to each other. Additionally, variation is more likely to emerge in populations that are structured in small communities. Our study provides potentially important insights into the theoretical mechanisms underlying language variation.

## 1. INTRODUCTION

Variation is ubiquitous not only in the natural world, but also in culture. There are intriguing patterns of variation ranging from within to between individuals (in most biological, psychological, cognitive and linguistic characteristics) and groups (most obvious in various aspects of culture). In the language domain, one can think about how two different persons would pronounce the phoneme /r/ in different ways, even if they speak the same language (Sankoff & Blondeau, 2007), or how grammatical constructions vary between English dialects (Trudgill & Chambers, 2017), or even about lexical variation (different use of 'firefly' and 'lightning bug' words in English, for example). Why this variation exists, why it is patterned the

This is the accepted version of the following article:

Josserand, M., Allasonnière-Tang, M., Pellegrino, F., Dediu, D. and de Boer, B. (2024), How Network Structure Shapes Languages: Disentangling the Factors Driving Variation in Communicative Agents. *Cognitive Science*, 48: e13439, which has been published in final form at <https://doi.org/10.1111/cogs.13439>

This article may be used for non-commercial purposes in accordance with the Wiley Self-Archiving Policy [<http://olabout.wiley.com/WileyCDA/Section/id-828039.html>].

44 way it is, and what processes govern its dynamics are extremely important and non-trivial  
45 scientific questions.

46 Language variation arises from divergences in the evolution of linguistic features over time  
47 among different populations. It is customary to distinguish the emergence of an innovation  
48 (the so-called “initiation”) from the dynamics of its diffusion in the speaker community  
49 (“actuation”) (Croft, 2008; Dediu & Moisik, 2019; Solé & Vives, 2012; Stevens & Harrington,  
50 2014; Yu, 2013). Initiation is usually seen as an individual-level process, while actuation is  
51 essentially a community-level process involving competing linguistic variants (Blythe & Croft,  
52 2012; Fagyal et al., 2010). These processes can take place in two main scenarios: when the  
53 language used by the population is being created *de novo* (language *emergence* scenario, as  
54 is arguably the case for emergent sign languages such as Al-Sayyid Bedouin Sign Language  
55 and Kaka Kolok; Zeshan & de Vos, 2012), and when a language is already in use and  
56 undergoing change (language *change* scenario, by far the most frequent and studies case).  
57 The selection of variants during the actuation phase can be influenced by many different  
58 factors, such as aspects of the environment (e.g., the effect of environment type on the  
59 phonemic inventories of languages; Everett et al., 2015; Maddieson & Coupé, 2015), cultural  
60 and historical factors (degree of contact with outsiders, or number of L2 speakers; Lupyan &  
61 Dale, 2010), or factors related to the structure of the network of linguistic interaction itself  
62 (Wray & Grace, 2007; Trudgill, 2011a). Indeed, interactions between individuals in human  
63 societies are embedded in social networks with specific characteristics (Beckner et al., 2009;  
64 Newman & Park, 2003), which would constrain the transmission of linguistic features  
65 (Chambers, 1995). While all human networks tend to share a few features, such as a small  
66 average path length, social network structures exhibit a great diversity across populations  
67 (Newman, 2003; Nichols, 1992). Our goal in this paper is to disentangle what characteristics  
68 in network structures are instrumental to understand how variation unfolds through evolution.

69 It is, however, challenging to carefully disentangle the relative effects of network structure from  
70 those of the other factors involved, prompting several researchers to introduce experimental,  
71 computer modeling and observational paradigms. The rich tradition of sociolinguistic studies  
72 uses fine-grained methods that provide insight into the role of network structure (among many  
73 other factors) on language change (Labov, 1972, 2001, 2010; Milroy & Milroy, 1985; Yu, 2013).  
74 Following a different approach, Raviv et al. (2019, 2020) tackle this question with an  
75 experimental approach involving the learning and transmission of an artificial miniature  
76 language within a carefully controlled network structure. Network structure was even put  
77 forward to have an influence on cultural evolution (Derex & Boyd, 2016; Migliano et al., 2020).  
78 These laboratory experiments emphasize the role of population size and network structure on  
79 language variability. However, despite their inherent advantages, laboratory experiments are  
80 strongly constrained in terms of population size, network structure and aspects of language  
81 considered, limiting the direct extrapolation of these results to understanding the dynamics of  
82 actual linguistic communities.

83  
84 Artificial multi-agent models do provide ways to address some of these issues, in particular  
85 concerning the number of linguistic agents, the structure of their communicative networks, and  
86 the usually long time scale necessary for innovations to spread in large populations. Some  
87 studies based on the use of such models highlight the role of population structure in shaping  
88 convergence (Lee et al., 2015), the formation of categories (Zubek et al., 2017; Gong et al.,  
89 2012), the diffusion of innovations (Ke et al., 2008), and linguistic structure (Reali et al., 2014;

90 Lou-Magnuson et al., 2018). Despite using different communication algorithms and  
91 addressing various aspects of language, these models share a common approach: they all  
92 compare different types of pre-existing network structures (such as fully-connected, star, or  
93 small-world networks) and then observe whether language evolves differently in these distinct  
94 network structures. For example, Lee et al. (2015) examined populations of learners in fully-  
95 connected, linear, Neumann lattice, and bridge network structures and showed that these  
96 distinct network typologies result in different patterns of grammar convergence. Departing from  
97 these pre-existing network structures, in the field of cultural evolution, Kobayashi et al. (2016)  
98 aimed to untangle two key metrics: population size and social connectedness. They propose  
99 that social connectedness might exert a stronger influence on the level of culture compared to  
100 population size. While most of these models address the dynamics of language change and  
101 do not specifically explore language variation, Mudd (2022) and Meir et al. (2012) also  
102 emphasize the role of social structure on lexical variation in emergent sign languages. Notably,  
103 Mudd (2022) developed computational models grounded in observations from the Kata Kolok  
104 community, focusing on various aspects of their social structure, such as connectivity and size.  
105 To explain why some sign languages witness more lexical variation than others, she proposed  
106 that population size is an important factor influencing lexical variability in these languages.  
107 However, while it appears that network structure influences language in both scenarios of  
108 emergence and change, it remains unclear whether the impact of network structure differs in  
109 these two contexts.

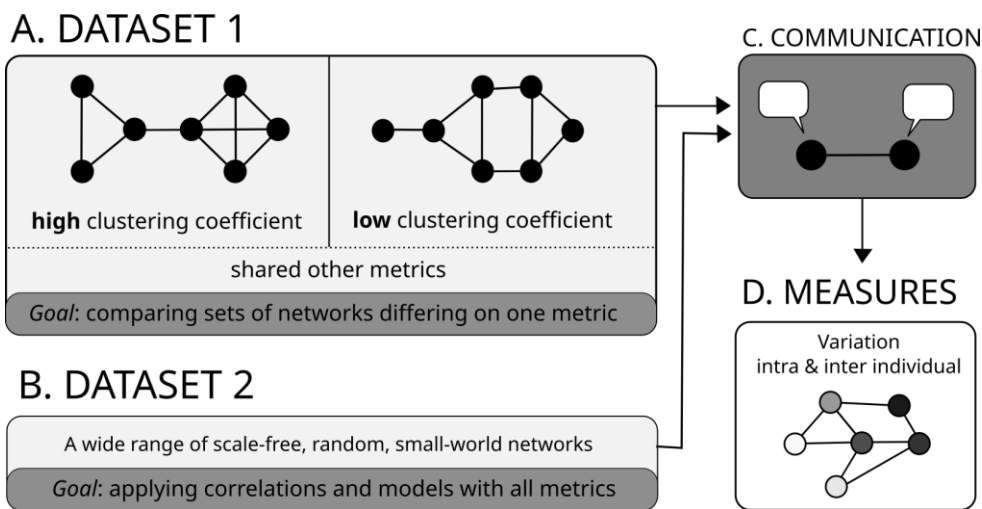
110  
111 Thus, a wide range of approaches strongly suggests that network structure likely influences  
112 the evolution of many linguistic features. But what exactly is meant by “network structure”?  
113 The networks used in such studies are either networks observed in the real world, or are  
114 automatically generated with specific properties (such as, for example, being fully connected,  
115 scale-free, or small-world). Any network, whether “natural” or “artificial”, can be described  
116 using a set of measures (called *metrics*) that summarize various aspects of the network,  
117 varying between the relatively global to the quite local (see Illustration 1). As Raviv et al. (2020)  
118 observed, it is very difficult to tease apart the contribution of each network characteristic in  
119 real-world observational networks. In other words, comparing real-world populations raises  
120 the question of understanding which network structural factors are involved in the actuation of  
121 linguistic innovations, and in which manner. In contrast, multi-agent models allow a much finer  
122 degree of control over these factors. As a consequence, most agent-based modeling studies  
123 to date have investigated the separate role of various characteristics of the communicative  
124 network. However, this approach has two main limitations: first, when focusing on specific  
125 factors in isolation, it is very difficult to compare their relative influence, and to understand the  
126 interactions between them. Secondly, working with automatically generated networks has not  
127 led to a fine-grained understanding of which network metrics behind these networks actually  
128 affect the dynamics of actuation. Indeed, there are intrinsic relationships between various  
129 metrics resulting from the type of network being generated. As an example, if communities  
130 structured as scale-free networks produce more stable variation than those structured as  
131 small-world networks, is this due to their lower centrality, lower clustering coefficient, or any  
132 other metric that intrinsically differs between these two types of networks?

133 Addressing these questions makes it essential to disentangle and bring a theoretical basis to  
134 the question of what aspects of network structure do, in fact, shape language. This has  
135 theoretical implications for understanding how languages adapt to fit their social environments  
136 and how network structure shapes language. In turn, it will help understand language

137 differentiation and help explain the great diversity observable in the world’s languages. To  
 138 answer these questions, we introduce here a multi-agent model based on the one in  
 139 Josserand et al. (2021). The agents use a Bayesian model of language learning in artificially  
 140 generated static networks where we control various metrics, and, across thousands of runs,  
 141 we effectively isolated the relative influence of several global metrics (average shortest path  
 142 length, clustering coefficient, global assortativity, degree distribution, and size) and of local  
 143 characteristics of the agents (measures of centrality and clustering coefficients). See  
 144 Illustration 1 for an example. We found that these metrics have different contributions to the  
 145 evolution of language in this model: while intra-individual variation is not shaped by network  
 146 structure, inter-individual variation is strongly affected by the path length and the clustering  
 147 coefficient.

148 The paper is structured as follows: in the Data and Methods, we present our Bayesian agent-  
 149 based model, along with the networks used, the metrics under consideration, and the overall  
 150 procedure of this analysis. In the Results, we investigate whether some of these metrics affect  
 151 inter- and intra-individual variability in different types of networks. We close by discussing the  
 152 limitations and implications of our findings, and suggest several future directions of study.

153



154

155 *Illustration 1. A. Dataset 1 comprises sets of networks differing on only one metric (here, clustering*  
 156 *coefficient, but we applied it to every metric in Table 1). B. Dataset 2 comprises a wide range of different*  
 157 *networks: scale-free, small-world, random networks of different sizes and parameters. C. Throughout*  
 158 *our study, agents interact using Bayesian communication for an extended period of time. D. Finally, we*  
 159 *investigate the language variability within and between agents in the network. To do so, we compare*  
 160 *language variability in networks that differ based on global metrics, using dataset 1. We also explored*  
 161 *the networks in dataset 2 by computing the correlations between different metrics and measuring their*  
 162 *relative contributions to variability in models. More information about the differences between the two*  
 163 *datasets will be presented in Data and Methods section.*

## 164 2. DATA AND METHODS

165 Here we develop a Bayesian model of language evolution (Griffiths & Kalish, 2007) based on  
 166 published work (Josserand et al., 2021), where agents exchange linguistic messages as  
 167 constrained by a communicative network. Briefly, the network’s nodes represent the agents,

168 which all have an internal representation of language, here reduced to a single feature. The  
169 agents that are connected via undirected edges in the network are called “neighbors” and can  
170 “talk” to each other. Upon “hearing” their neighbors’ utterances, an agent updates its internal  
171 representation of language to take into account the newly “heard” data using Bayes’ rule. We  
172 study how the language changes after multiple rounds of such interactions, and, in particular,  
173 the manner in which different network structures affect these dynamics in two types of  
174 scenarios: language emergence (agents interact without any initial language exposure) and  
175 language change (agents interact with an initial language exposure).

## 176 1.1 The language

177 The language in our model is a multinomial feature that can take one of  $k$  exclusive possible  
178 values. This simple representation is arguably adequate to approximate numerous aspects of  
179 human language, such as lexical choices (using one of several words or expressions to  
180 describe a given concept), allophony (different actual sounds for the same phoneme as  
181 conditioned by the phonetic environment, socio-linguistic/dialectal choices, or in free  
182 variation), or morpho-syntactic structures (different ways of organizing the sentence). Here,  
183 we call an *utterance* a token drawn from this multinomial distribution. For example, if the  
184 feature refers to the possible allophones of “r” in English, the possible utterances include the  
185 alveolar trill [r], the approximant [ɹ], and the retroflex approximant [ɻ], among others. Abstractly,  
186 we can represent the feature through the vector of its possible utterances  $u = \{u_1, u_2, \dots, u_k\}$ .  
187 We fixed the number of utterances  $k$  to 10 here, but simulations ran with different values of  $k$   
188 led to similar conclusions.

189 While we also generalized our feature to a continuous one (which can represent, for example,  
190 very fine-grained measurements of vowel height or voice-onset-time (VoT); presented for the  
191 Dataset 1 in the **Supplementary Materials** in Part 3.2.2), we decided to focus here on the  
192 case of the discrete valued feature. Indeed, using a continuous representation instead of the  
193 discrete one does not alter our conclusions, but presenting both would lead to clutter and  
194 unnecessary complexity in the presentation of the results.

## 195 1.2 How do agents represent language?

196 Agents are Bayesian learners (Griffiths & Kalish, 2007). In this framework, there is a fixed  
197 “universe” of languages,  $U$ , and each possible language  $h$  from this universe ( $h \in U$ ) has a  
198 certain probability  $p$  of being “the real one” from the agent’s “subjective” point of view. Before  
199 being exposed to any utterance, each agent has an *a priori* “belief” concerning the probabilities  
200 of each possible language, denoted  $p(h)$  and known as the *a priori probability distribution* of  
201 the languages (or the *prior*). However, each time an agent “hears” an utterance, denoted as  
202  $d$ , it updates its subjective distribution of the probabilities of the languages in light of this data  
203 using Bayes’ theorem:  $p(h|d) = \frac{p(d|h)p(h)}{p(d)}$ . Here,  $p(d|h)$  is the likelihood of observing data  $d$   
204 given that the language is  $h$ , while  $p(h|d)$  is the updated belief, or *posterior*, i.e. the probability  
205 of the language being  $h$  after both the *prior* and the observed data  $d$  are considered. The  
206 denominator  $p(d)$  acts as a normalization factor and is not of immediate concern here. While  
207 this mechanism is extremely general, we apply it here in a very particular and relatively simple  
208 setting.

209 More precisely, an agent’s internal representation of language is given by a vector  $\mathbf{p} =$   
210  $(p_1, p_2, \dots, p_k)$ , where each component  $p_i$  represents the probability of the  $i^{\text{th}}$  utterance  $u_i$ ; as  
211 the utterances are mutually exclusive and exhaust the space of possible utterances for this  
212 features,  $\sum_{i=1}^k p_i = 1$ . An agent has an *a priori* distribution that represents its initial belief in  
213 the probability of each utterance, which leads naturally to a *multinomial distribution* that  
214 represents the likelihood. More precisely, given  $n$  such utterances “heard” independently by  
215 an agent, the random variable  $\mathbf{X} = (X_1, X_2, \dots, X_k)$  distributed as *Multinomial*( $n, \mathbf{p}$ ) gives the  
216 expected number of times each possible utterance  $u_i$  appears. Here, the multinomial  
217 distribution models how frequent any one of the  $k$  possible utterances should be among the  $n$   
218 “heard” independent ones, if their probabilities are given by the vector  $\mathbf{p}$ .

219 To compute the posterior distribution after having seen the data, it is possible to use the very  
220 computationally expensive Markov Chain Monte Carlo technique. However, in particular  
221 cases, one may choose specific conjugate priors, which drastically simplify the application of  
222 Bayes’ theorem. The conjugate prior of a multinomial distribution is given by the *Dirichlet*  
223 *distribution*, which is defined by a vector of  $k$  parameters  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k) \in R^k$ . These  
224 parameters  $\alpha_i$  are related to the probability  $p_i$  of each utterance  $u_i$ , so that the vector of  
225 probabilities  $\mathbf{p} = (p_1, p_2, \dots, p_k)$  is distributed as *Dir*( $\boldsymbol{\alpha}$ ). The prior and posterior distributions of  
226 probabilities are modeled by Dirichlet distributions and, thus, characterized by prior (denoted  
227 here as  $\boldsymbol{\alpha}$ ) and posterior (denoted here as  $\boldsymbol{\alpha}'$ ) Dirichlet parameters. Using Bayes’ theorem,  
228 one can update the prior distribution *Dir*( $\boldsymbol{\alpha}$ ) into the posterior distribution *Dir*( $\boldsymbol{\alpha}'$ ) after seeing a  
229 number  $n$  of independent utterances (the data,  $d$ ): it reduces to trivial arithmetic:  $\alpha'_i = \alpha_i + X_i$ ,  
230 where, as above,  $X_i$  gives the number of times utterance  $u_i$  was “heard” in the data  $d$ . Thus,  
231 this reduces to a simple model where the parameters of the Dirichlet distribution reflect the  
232 number of “heard” utterances and the prior belief. It is important to note that, as expected for  
233 Bayesian models, both the prior beliefs and the observed data matter. Seeing enough data  
234 should be able to overwhelm even very strong *a priori* beliefs and the order in which the  
235 utterances in the data arrive does not matter. This generalizes the prior model studied in  
236 (Josserand et al., 2021) by allowing more than two possible utterances (i.e., from a Bayesian  
237 model with a binary feature using Binomial/Beta conjugate distributions, to a multinomial  
238 feature using Multinomial/Dirichlet conjugate distribution).

### 239 **1.3 The initial language exposure: emergence vs. change**

240 We considered two types of agent populations, differentiated by their “initial language  
241 exposure”, which represents the type of scenario (language emergence versus language  
242 change). In the populations in a language emergence scenario, the agents are born *without*  
243 initial language exposure. Such agents are naïve regarding the language, with all possible  
244 utterances having the same probability of occurrence. This is modeled by a Dirichlet  
245 distribution with all parameters  $\alpha_i$  being equal,  $\boldsymbol{\alpha} = (1, 1, \dots, 1)$ . It represents the situation  
246 where agents are born in a language-less community, but it may as well represent a  
247 community with a language where all possible utterances are equally frequent.

248 In contrast, in the language change scenario, agents are born in a population *with* initial  
249 language exposure, i.e. born in a community with a pre-existing language that favors a certain  
250 utterance at the expense of the others; say, utterance 4,  $\alpha_4$  is 6 times more frequent than the  
251 others (for various reasons, including past internally-motivated language change, language  
252 contact, or top-down language policies). This situation is represented by a Dirichlet distribution

253 with parameters  $\alpha = \{1, 1, 1, 6, 1, \dots, 1\}$ . Given how the parameters of the Dirichlet distribution  
 254 reflect the number of “heard” utterances, this can be seen as modeling the fact that agents are  
 255 expecting, before they engage in communication with other agents, that the utterance  $u_4$  is  
 256 much more frequent than all the others. A value of 6 is supposed to represent an initial  
 257 language with a very skewed frequency of the utterances, but not too strong (however, we  
 258 also investigated the effect of using 21 instead of 6 in the **Supplementary Materials**; see part  
 259 3.2.1.8).

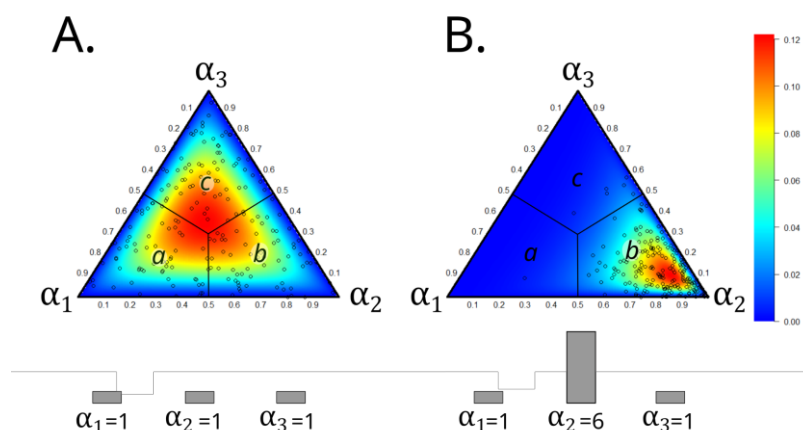
## 260 1.4 The communication process

261 As described above, the agents are “born” with their initial Dirichlet distribution, which  
 262 represents their bias concerning the language feature. The simulation happens in discrete  
 263 timesteps (*rounds*), and in each round, the agents “speak” (i.e., produce utterances) and  
 264 “hear” (i.e., receive) the utterances produced by their neighbors.

265 To produce utterances, an agent uses the following procedure: first, it normalises the  $k$   
 266 parameters of its Dirichlet distribution,  $\alpha$  to a probability vector  $\mathbf{m} = (m_1, m_2, \dots, m_k)$ , where  
 267  $m_i = \frac{\alpha_i}{\sum_j \alpha_j}$ . Then, it draws a single value  $1 \leq j \leq k$  proportional to its probability  $m_j$ , such that  
 268 each index  $j$  can be chosen depending on its probability relative to the other  $k-1$  indices in  $\mathbf{m}$ .  
 269 This is equivalent to the use of random multinomial generator,  $j \sim \text{Multinomial}(\mathbf{m})$ , as  
 270 implemented by, for example, R’s `rmultinom(n, size, prob)`, where `n=1`, `size=1` and  
 271 `prob=m`, or Stan’s `multinomial_rng(vector theta, int N)`, where `theta` is the  
 272 simplex vector `theta=m` and `N=1`. Finally, this value  $j$  represents the utterance that is actually  
 273 produced, namely  $u_j$  (see Illustration 2). So, we use what is known as a sampler (or ‘SAM’)  
 274 strategy (Griffiths & Kalish, 2007; Josserand et al., 2021) in which the agent picks any of the  
 275 possible  $k$  utterances proportional to its subjective probability (as given by the agent’s current  
 276 Dirichlet distribution).

277 Conversely, upon hearing the utterances produced by its neighbors, the agent updates its  
 278 current Dirichlet distribution as described above. This update consists of an addition, so that  
 279 the Dirichlet hyper-parameters become  $\alpha'_i = \alpha_i + X_i$ , where  $X = \{X_1, X_2, \dots, X_k\}$  represents  
 280 the number of times each utterance appears among those that the agent just heard.

281



282



283 *Illustration 2. Representation of the Dirichlet distribution for an agent without initial language exposure*  
 284 *(language emergence, panel A) and for an agent with an initial language exposure towards  $\alpha_2$*   
 285 *(language change, panel B). For visualization purposes, we plotted here  $k=3$  utterances instead of the*  
 286  *$k=10$  actually used in the study. The dots represent the random draws from the Dirichlet distribution: if*  
 287 *the random draw occurs in the **a** part, the agent will produce an utterance  $u_1$ ; if it occurs in the **b** part, it*  
 288 *will produce an utterance  $u_2$ , and it will produce an utterance  $u_3$  when drawn in the **c** region. In the*  
 289 *condition without initial language exposure, random draws happen equally probably for all three*  
 290 *utterances, but with initial language exposure, they occur mostly in the **b** region, resulting in many more*  
 291  *$u_2$  utterances. The bar plot below each triangle is an alternative way of representing the values of the*  
 292 *Dirichlet distribution. The color gradient reflects the density of random draws, with red denoting high*  
 293 *density and blue indicating low density.*

## 294 1.5 The networks: types and metrics

295 The agents form the nodes of a network with a static structure, fixed for the duration of a  
 296 simulation, while the edges between the nodes determine the agents that can talk to each  
 297 other. Given that we focus on the structure of these networks, our discussion in this section is  
 298 framed in terms of nodes and edges, instead of agents and their communicative exchanges.  
 299 Here, we use three classes of network topology: random, small-world, and scale-free  
 300 networks. The random networks are generated using Erdős & Rényi (1959)'s algorithm, where  
 301 each node has a fixed probability of connecting with other nodes in the network. We kept only  
 302 networks without any isolated components. The small-world networks are generated using the  
 303 Watts-Strogatz algorithm (Watts & Strogatz, 1998): this algorithm starts with a ring of nodes,  
 304 where each node is connected to a fixed number  $N$  of neighbors on either side, followed by a  
 305 rewiring with a fixed probability  $p$ . This process leads to the generation of networks with many  
 306 real-world properties (Kenett et al., 2018; Kitsak et al., 2010), such as the presence of short  
 307 average path lengths (see Milgram (1967)'s "Six degrees of separation"). The scale-free  
 308 networks are generated using the preferential attachment algorithm of Barabási et al. (2000),  
 309 and exhibit a power-law degree distribution: most nodes have a limited number of neighbors,  
 310 while few nodes have very many (Albert, 2005; Albert et al., 1999), resembling real offline and  
 311 online social networks. The algorithm gradually adds new nodes to a network such that the  
 312 probability  $p_i$  that the newly added node is connected to node  $i$  is  $p_i = \frac{d_i}{\sum_j d_j}$  where  $d_i$  is the  
 313 degree of node  $i$ , and the sum is over all pre-existing nodes  $j$ . We use these different classes  
 314 of network topology to investigate the effects of specific network metrics (see Table 1).  
 315

Name of the global network metric	Short name	Description	Mathematical description	Type of network
Average node degree	Degree	Average number of neighbors for all nodes	$D = \frac{\sum_{i=1}^n k_i}{n}$ where $k$ is the degree of a node $i$ and $n$ is the total number of nodes in the network	Constant for Barabasi-Albert scale-free networks with a fixed size
Network size	Size	Total number of nodes	$n$	All

Average shortest path length	Path length	On average, how close (defined by the number of intermediate nodes) are two random nodes	The mean of the shortest path length between all pairs of nodes: $l_G = \frac{\sum_{i,j} dist(v_i, v_j)}{n \cdot (n-1)}$ , where $dist$ is the shortest path length between two nodes	All, since we removed isolated components in random networks
Global assortativity coefficient	Assortativity	Preference for a network's nodes to attach to others that are similar (here, in terms of nodes's degree)	The mean of the local assortativity based on the node degree, where local assortativity is defined as $\rho = \frac{j(j+1)(\bar{m}_j - \mu_q)}{2M\sigma_q^2}$ , where $M$ is the number of edges in the network, $j$ is the excess degree of a particular node and $\bar{m}_j$ is the average excess degree of its neighbors. (The excess degree distribution $j$ is defined as $j(k) = \frac{(k+1)P(k+1)}{D}$ ). Thus, assortativity can vary between -1 (highly disassortative) to 1 (highly assortative).	All
Average clustering coefficient	Clustering	Measure of the degree to which nodes in a network tend to cluster together	The average local clustering coefficient of all nodes (as in Watts & Strogatz, 1998), where the local clustering is defined as: $C_i = \frac{\lambda_G(v)}{\tau_G(v)}$ , where $\lambda_G(v)$ is the number of triangles on $v \in V(G)$ , $V(G)$ being the set of all nodes in the undirected graph $G$ , and $\tau_G(v)$ is the number of triangles on $v \in G$ . The clustering coefficient varies between 0 (the friends of my friends are never connected) and 1 (the friends of my friends are all connected)	Constant in scale-free networks
Degree distribution: the shape of the power law distribution	Degree distribution	Shape of the degree distribution for power-law distributions: how strongly hubs gather all connections	A power-law distribution is defined as $P(k) \sim k^{-\alpha}$ where $\alpha$ is the exponent of the distribution. The larger the exponent, the rarer the large values are. The exponent is estimated by applying "log" to the degree distribution and looking at the slope of the obtained new graph. Exponents in scale-free networks obtained with Barabasi-Albert algorithm usually present some random fluctuations and vary between 2 and 3.	Only networks exhibiting a power-law distribution

316 In this study, we aim to isolate the effect of the different metrics in order to understand how  
317 they may influence language variability. Please note that we focus on these metrics because  
318 they capture essential properties of communicative networks that have been claimed, or may  
319 arguably contribute to language dynamics within networks. Moreover, some of these metrics  
320 are very hard to disentangle and, therefore, their effects might be confounded in most previous  
321 studies. To overcome those difficulties, we generated hundreds of thousands of networks, and  
322 then selected two sets of networks, each containing 100 networks: for a given metric, we  
323 selected the two sets of networks guaranteeing that the averages for the other metrics in the  
324 two sets were similar except for the metric under consideration (see Table 2). As significance  
325 can be easily attained when it is possible to generate a high number of networks, we ensured  
326 that the difference between the two sets was much larger (at least two orders of magnitude)  
327 for the metric considered than for the "equalized" metrics. We used scale-free networks to  
328 analyze the independent effects of the network size, the path length, the assortativity, and the  
329 degree distribution; small-world networks for the clustering coefficient; and random networks  
330 for the average degree and path length.

Metric value <sup>1</sup>	Difference between the two sets varying in the network metric (in <b>bold</b> ) <sup>2</sup>					
	Size	Path length <sup>3</sup>	Degree distribution	Assortativity	Clustering	Node degree
<i>Size</i>	( <b>100</b> / 50) <b>50</b>	(50 / 50) 0	(50 / 50) 0	(50 / 50) 0	(150 - 150) 0	(50 - 50) 0
<i>Path length</i>	(4.49 / 4.50) 0.01	( <b>4.99</b> / <b>3.37</b> ) <b>1.62</b>	(4.09 / 4.10) 0.01	(4.02 / 4.01) 0.01	(1.68 - 1.68) 0	(1.78 / 1.79) 0.01
<i>Degree distribution</i>	(2.85 / 2.85) 0.00	(2.79 / 2.80) 0.01	( <b>3.04</b> / <b>2.55</b> ) <b>0.49</b>	(2.75 / 2.75) 0	--	--
<i>Assortativity</i>	(-0.29 / -0.30) 0.01	(-0.31 / 0.31) 0	(-0.35 / -0.35) 0.01	( <b>-0.16</b> / <b>-0.48</b> ) <b>0.32</b>	(-0.01 / -0.01) 0	(0.02 / 0.01) 0.01
<i>Clustering</i>	(0 / 0) 0	(0 / 0) 0	(0 / 0) 0	(0 / 0) 0	( <b>0.58</b> - <b>0.32</b> ) <b>0.26</b>	(0.25 / 0.25) 0.01
<i>Degree</i>	(1.98 / 1.96) 0.02	(1.96 / 1.96) 0	(1.96 / 1.96) 0	(1.96 / 1.96) 0	(48 / 48) 0	(12.69 / 11.98) <b>0.71</b>

332 These metrics reflect the *global* network structure, but there is *local* heterogeneity between  
333 nodes, which is captured by node-level metrics. Using the same model, we investigated the  
334 influence of *local* network heterogeneity on language change (Fagyal et al., 2010).  
335 Specifically, we extended the results from our previous paper (Josserand et al., 2021). In this  
336 earlier work, we found that language adapts more to a group of biased agents—namely, those  
337 born with a preference for a specific utterance—when these biased agents have higher  
338 centrality. While in the prior study, we defined centrality as betweenness centrality, we  
339 complemented these results here by exploring whether various other node-level metrics (local  
340 clustering coefficient, node degree, eigenvector centrality, and closeness centrality) might  
341 exert different influences. These findings, available only in the **Supplementary Materials** (see  
342 part 3.4), not only expand upon but also validate the results presented by Josserand et al.  
343 (2021).

## 344 Measurements

345 Previous studies (see Figure 13 in Josserand et al., 2021) have shown that in a type of  
346 Bayesian model of language change similar to the ones implemented here but where the  
347 linguistic feature was binary, language stabilized before 1000 rounds in all network types (see  
348 our **Supplementary Materials** part 3.1 for visualization). Here, we introduced a conservative

<sup>1</sup> Each row shows the absolute difference between the two sets in the appropriate metric given by the names. The parenthesis indicates the mean value for each set of networks (high / low).

<sup>2</sup> Each column refers to the two sets of networks varying only in the shown metric.

<sup>3</sup> For example, the second column shows the differences between the two sets of networks differing by design in path length: these two sets contain 100 networks each, have the same size, clustering coefficient, average degree, and almost the same degree distribution and assortativity, but (as intended) markedly different path length. These numbers should be interpreted with caution, as the strength of a difference of “1” or “0.01” may vary depending on the metric being analyzed.

349 margin and analyzed the inter- and intra-individual variability after 3000 rounds (when all  
350 agents had produced 3000 utterances) in order to make sure that language would have  
351 stabilized.

352

353 First, we define *inter-individual variability* by how different the agents are regarding their  
354 internal language representation. For example, if agents tend to use different utterances  
355 (some preferentially  $u_4$  while some others preferentially  $u_3$ ), the inter-individual variability will  
356 be high. We estimate inter-individual variation using the Kullback-Leibler divergence, which is  
357 a measure of the statistical distance between two probability distributions (as coded in the  
358 *entropy* package in R; Hausser & Strimmer, 2021). To do so, we computed for all possible  
359 pairs of agents the pairwise Kullback-Leibler divergence between their Dirichlet distributions,  
360 and then averaged the resulting pairwise Kullback-Leibler divergences to obtain an average  
361 measure of the inter-individual variation in the network.

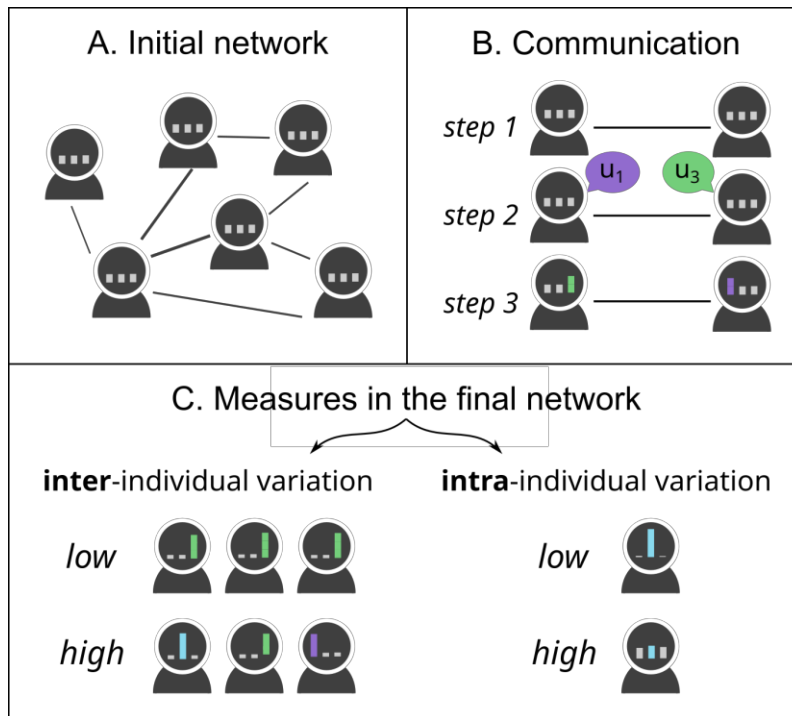
362

363 Secondly, *intra-individual variability* captures how consistent an agent is in its own language  
364 productions. For example, if an agent uses equally likely two utterances ( $u_2$  and  $u_8$  half of the  
365 time each, for example), its intra-individual variability will be high. On the contrary, an agent  
366 that always uses the same utterance has a low intra-individual variation. We estimate intra-  
367 individual variation as the entropy of an agent's Dirichlet distribution (as coded in the *entropy*  
368 package, using Jeffreys prior<sup>4</sup>; Hausser & Strimmer, 2021), which is bounded between 0 and  
369  $\log_2(k)$ , where here  $k=10$ . The *average* of this intra-individual variation across all agents  
370 captures their propensity to consistently use the same utterance (a high average indicates that  
371 agents switch between different utterances, while a low average suggests that agents tend to  
372 use the same utterance consistently), while its *standard deviation* across all agents captures  
373 to what degree there are differences between the agents regarding their intra-individual  
374 variation (a high standard deviation suggests that while some agents always use the same  
375 utterance, others show a high variability, and a low standard deviation suggests that all agents  
376 tend to be equally consistent or not in their language production). See Illustration 3 for a  
377 visualization of the entire process.

378

---

<sup>4</sup> This ensures that the resulting entropy value is robust and not overly influenced by the specific parameterization of the probability distribution. We also selected other priors, but conclusions were the same.



379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390

*Illustration 3.* A schematic representation of the multi-agent model. (A) Agents are “born” embedded in a network with a fixed structure and with given characteristics. Each agent has an internal language representation given by the parameters of the Dirichlet distribution (see gray bar plot in Illustration 2). For visualization purposes, we decided to represent  $k=3$  utterances here instead of the  $k=10$  utterances used in the paper. (B) **Step 1.** Each agent randomly extracts a number from its Dirichlet distribution: here, the probability of picking any of the three utterances is the same. **Step 2.** Each agent produces the chosen utterance (here,  $u_1$  for the agent on the left and  $u_3$  for the one on the right). **Step 3.** Each agent hears the utterance(s) of its neighbor(s) and updates its own Dirichlet distribution accordingly. (C) After each individual talks 3000 times, we measure the inter- and intra-individual variation among the agents in the final network.

391

## 1.6 The network generation procedure

392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405

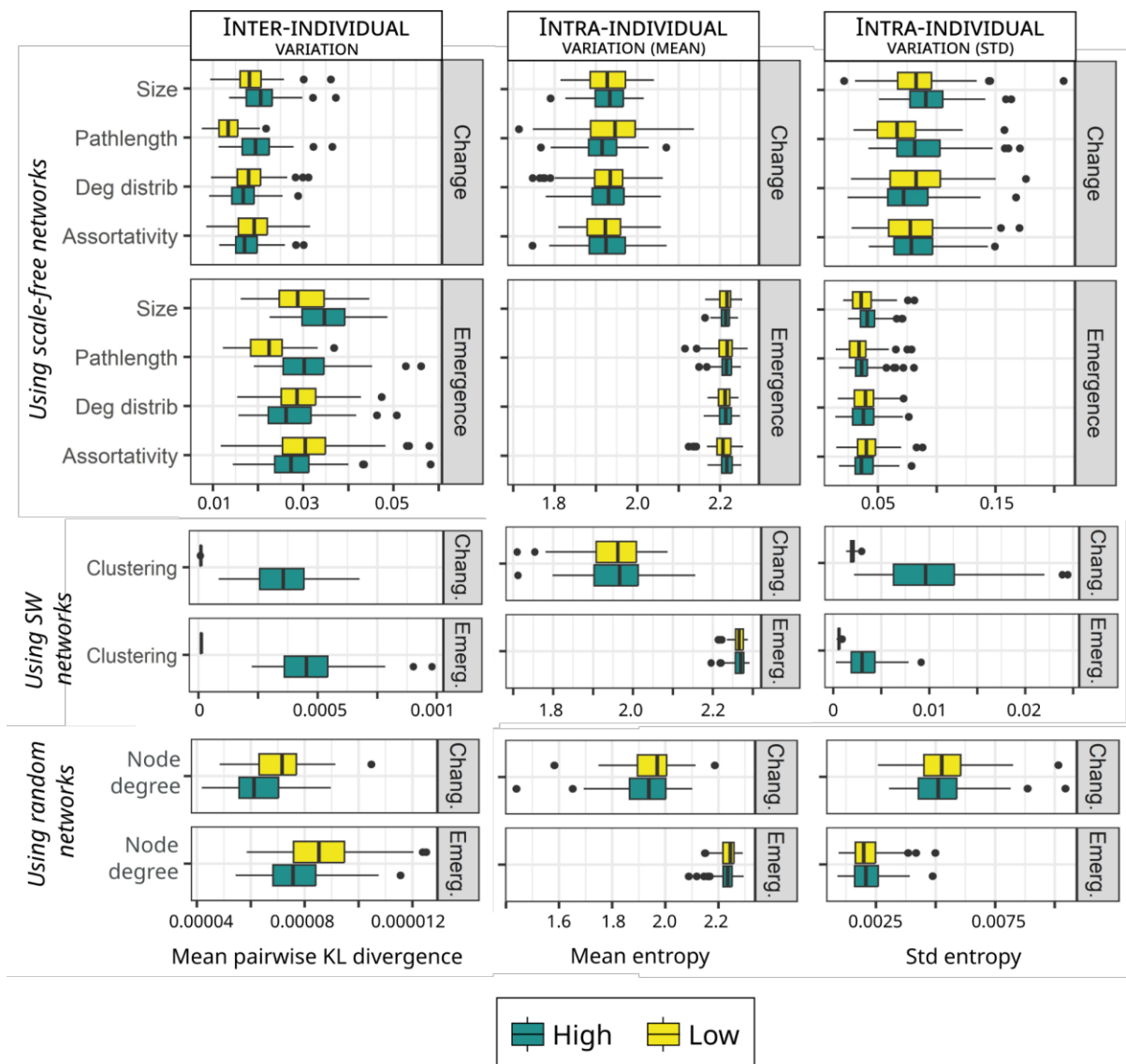
We conducted simulations using Netlogo (version 6.2.2; Wilensky, 1999) and generated two datasets. In the first (which we will call *dataset 1*), we systematically compared two sets of networks differing in exactly one metric, as described above. To achieve this, we first generated 100 networks for each combination of parameters which resulted in 80,000 scale-free networks (with varying sizes), 240,000 random networks (with varying connection probability to control for density and pathlength), and 600 small-world networks (with high and low rewiring probability, to examine the influence of clustering). Detailed information about the specific parameters is available in the **Supplementary Materials** (see Part 2.2.1). Following the generation of these networks, and using graphic visualization, we manually selected sets of networks that differed in only one metric. Subsequently, to compare the amount of variation in the two sets of networks and assess whether the difference was meaningful, we used Cohen’s  $d$ , a statistical metric that measures the difference between two groups. This metric is computed by dividing the difference between the means of the two groups by the pooled standard deviation of the groups.

406 With dataset 2, we ran a series of simulations with random, scale-free, and small-world  
407 networks using a wide range of parameters. For each network type, we generated networks  
408 with sizes of 50, 150, and 300. In the case of random networks, we explored connection  
409 probabilities ranging from 0.05 to 0.95 in increments of 0.05. For small-world networks, we  
410 created networks with varying average numbers of neighbors (4, 8, 16, 48) and different  
411 rewiring probabilities (0.1, 0.3, 0.5, 0.7, 0.9). Initially, we examined how the metrics co-varied  
412 across a broad range of values using Spearman's and Pearson's correlations. Then, we  
413 modeled how these metrics influenced the final variability in the communicative network. We  
414 z-scored all the metrics and we predicted inter-individual, mean intra-individual and standard-  
415 deviation of intra-individual variation using all our metrics as predictors. To do so, we used  
416 three machine learning techniques. First, we measured the predictor importance with *random*  
417 *forests* (using package `randomForest`; Liaw & Wiener, 2002) using an accuracy-based index  
418 and a gini-based Index. Second, we used *conditional forests* (using `cforest` function in  
419 `partykit` package; Hothorn & Zeileis, 2015) to measure the unconditional predictor  
420 importance index. Last, we used *Support Vector Machines* (SVM; `fit` function in `rminer`  
421 package; Cortez, 2010) for sensitivity analysis. Additionally, for all of them, we measured the  
422 goodness of fit using  $R^2$  and RMSE. To prevent overfitting, we trained the model on 80% of  
423 the data (stratified by size) and tested the model on the 20% remaining data using 50  
424 replications. For easier plotting, we z-scored and then aggregated the results for importance  
425 by method (accuracy-based, gini-based, unconditional, and sensitivity), by network type, and  
426 by predictor. The values for each method and each predictor, as well as additional details on  
427 the exact parameters used in the models, are available in the **Supplementary Materials** (see  
428 Part 3.3.4).

## 429 3. RESULTS

### 430 3.1 Dataset 1

431 To produce sets of networks that vary only in one metric, we needed to use multiple types of  
432 networks. We used scale-free networks to investigate the influence of assortativity, degree  
433 distribution, path length, and size. Small-world networks were used to examine the impact of  
434 the clustering coefficient, while random networks were used to investigate the effect of the  
435 average node degree. The inter- and intra-individual variation (mean and standard-deviation)  
436 values are presented in Figure 1, while Figure 2 indicates the effect sizes and significance for  
437 each set.  
438  
439



440  
 441 *Figure 1. Measure of inter-individual variation (left column), mean of intra-individual variation (column*  
 442 *in the middle), and standard deviation of intra-individual variation (right column). Each row shows the*  
 443 *results for one metric: for example, the first row shows the results for 100 networks that have a low size*  
 444 *on average (in blue), and 100 networks with a high size (in yellow) (see Table 2), in networks where*  
 445 *there is an initial language exposure (language change, upper panel) and networks without initial*  
 446 *language exposure (language emergence, lower panel). Depending on the network type used to study*  
 447 *the metric of interest, the x axis may vary: the x axis of inter-individual variation shows the mean pairwise*  
 448 *Kullback-Leibler divergence, while the two other columns show respectively the mean and the standard*  
 449 *deviation of the entropy for all agents, computed using the final Dirichlet distribution for each agent. The*  
 450 *quantification of the effect size between the two sets is indicated in Figure 2.*  
 451

452 **Clustering coefficient.** Both networks in language emergence (Cohen's  $d = 4.2$ ) and  
 453 language change (Cohen's  $d = 4.2$ ) scenarios exhibit a strong influence of clustering  
 454 coefficient on inter-individual variability, as shown in Figures 1 and 2. Networks without a  
 455 clique structure tend to display lower levels of inter-individual variation. Similar to all other  
 456 network structure metrics, the clustering coefficient has no effect on the mean intra-individual  
 457 variation. However, higher clustering coefficients lead to more differences among agents in  
 458 terms of standard deviation of intra-individual variation (with some agents consistently using  
 459 the same utterance and others switching between several different utterances), which is

460 observed in networks with language emergence or change scenarios (Cohen's  $d = 2.0$  and  
461 Cohen's  $d = 2.2$ , respectively).

462

463 **Path length.** The path length strongly affects inter-individual variability, in both emergence  
464 (Cohen's  $d = 1.5$ ) and change (Cohen's  $d = 1.5$ ) scenarios. Thus, networks where agents are  
465 easily reachable by others are likely to show low levels of inter-individual variation. Path length  
466 does not affect the mean amount of intra-individual variation, but it does slightly positively  
467 affect the standard deviation of intra-individual variation in networks in language change  
468 scenario only (Cohen's  $d = 0.7$ ).

469

470 **Size.** The network size also influences language variability, but its effect quickly disappears  
471 with higher initial language exposure. Big networks show higher inter-individual variation in  
472 networks with language emergence (Cohen's  $d = 0.9$ ) and language change scenarios  
473 (Cohen's  $d = 0.5$ ). However, an additional analysis suggests that the stronger the initial  
474 language exposure, the less the network size will play a role in shaping inter-individual  
475 variation. More specifically, Cohen's  $d$  drops to 0.3 when the initial language exposure is  
476 strong (namely, if the initial language is skewed with 21 utterances instead of 6; see  
477 **Supplementary Materials** part 3.2.1.8). Thus, the size of the population may only affect inter-  
478 individual variation in emergent languages. As for intra-individual variation, size does not  
479 appear to have any impact on the amount and standard-deviation of intra-individual variation  
480 observed.

481

482 **Average node degree.** The average node degree has a discernible but moderate impact on  
483 the degree of language variation among individuals, compared to other metrics. In networks  
484 with language emergence (Cohen's  $d = 0.7$ ) and language change (Cohen's  $d = 0.7$ )  
485 scenarios, networks with a low average node degree typically exhibit high inter-individual  
486 variation. However, the average node degree does not affect any type of intra-individual  
487 variation.

488

489 **Assortativity and degree distribution.** Finally, both assortativity and degree distribution  
490 have no observable effect on inter-individual nor intra-individual variation. Networks where  
491 individuals are preferentially connected to individuals who have a similar degree, and networks  
492 with very skewed degree distributions, do not show specific patterns of variation.

493

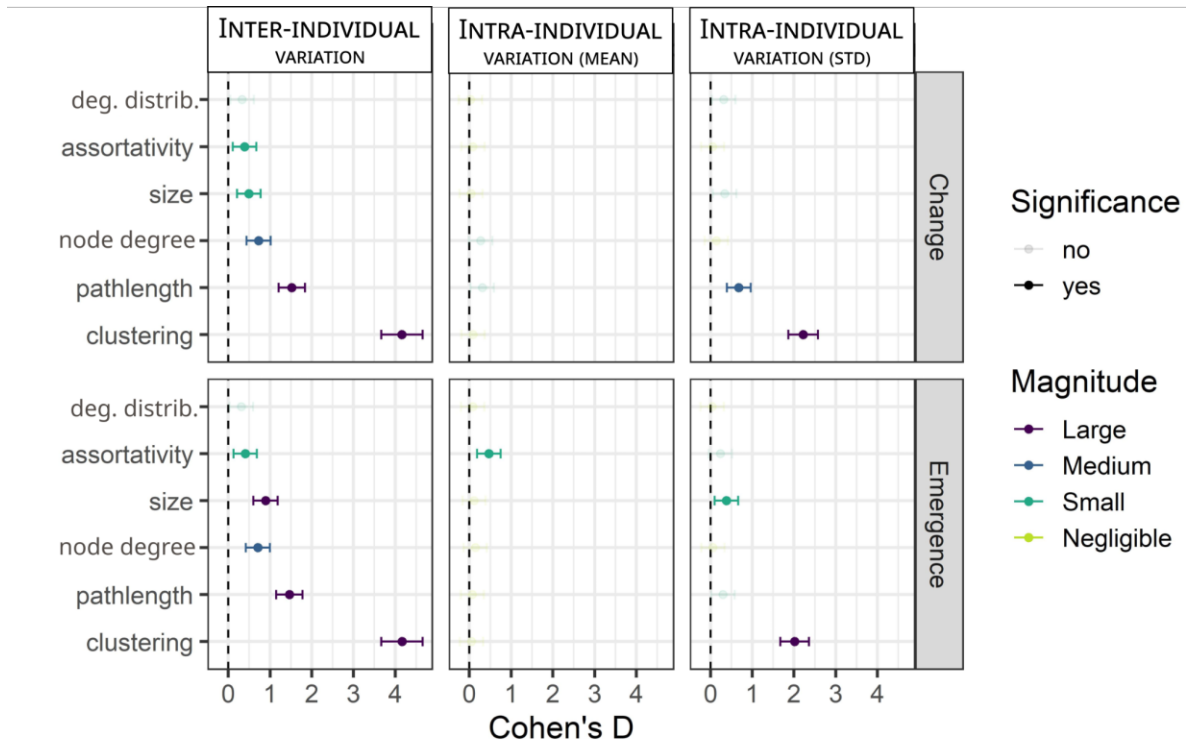
494 **Initial language exposure.** The initial language exposure, which reflects whether agents are  
495 born in language emergence or language change scenario, affects both inter- and intra-  
496 individual variation. Networks in language emergence scenarios exhibit greater inter- and  
497 mean intra- individual variation: as expected, agents communicating in a language change  
498 scenario exhibit more consistency (lower entropy on average) than agents involved in a  
499 language emergence scenario. Notably, the amount of intra-individual variability is solely  
500 affected by the initial language exposure. However, the standard-deviation of intra-individual  
501 variation is lower in language emergence scenario.

502

503 **Summary.** In summary, our findings indicate that clustering coefficient and path length are  
504 the main factors impacting both inter-individual variation and the standard deviation of intra-  
505 individual variation. Size and density moderately affect inter-individual variation, specifically in  
506 language emergence scenarios. Notably, the mean intra-individual variation remains



507 unaffected by any of our network structure measures. The initial language exposure, reflecting  
 508 language change or emergence scenarios, also strongly influences all types of variation.  
 509

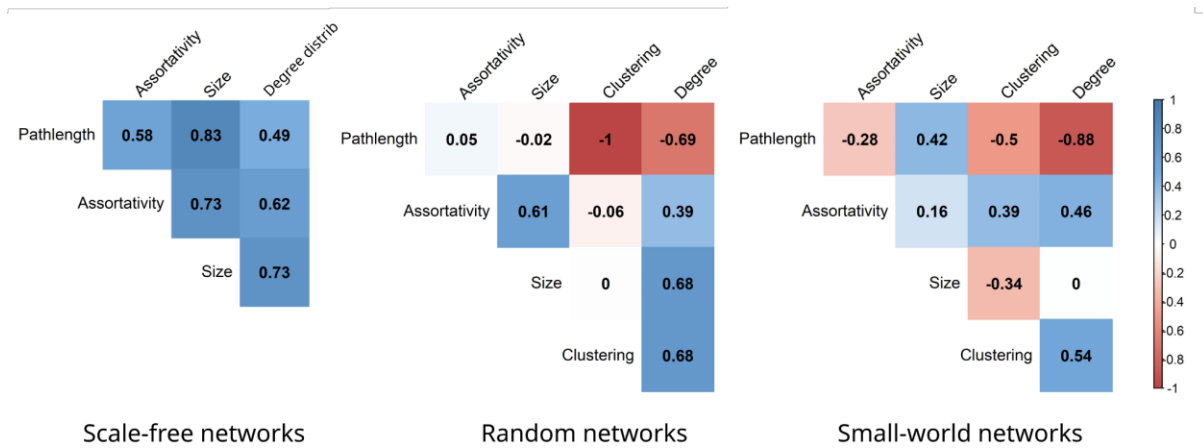


510  
 511 *Figure 2. Measure of inter-individual variation (left column), mean intra-individual variation (middle*  
 512 *column), and standard deviation of intra-individual variation (right column) in networks with initial*  
 513 *language exposure (language change; upper panel) and without initial language exposure (language*  
 514 *emergence; lower panel). The Cohen's d index for each predictor indicates the effect size of the*  
 515 *difference in variation between the two sets (high and low) for each predictor along with its confidence*  
 516 *interval. Gray indicates a non-significant metric, while significant metrics are in shades of color.*

### 517 3.2 Dataset 2

518  
 519 Here, we also use simulations obtained using automatically generated random, small-world,  
 520 and scale-free networks (all generated networks were kept without any specific selection  
 521 criteria<sup>5</sup>). We analyzed the Spearman's correlations between the metrics (see Figure 3). The  
 522 clustering coefficient, the node degree, and the path length are highly correlated with each  
 523 other. Thus, the more neighbors a given agent has on average, the more likely the friends of  
 524 this agent will be connected to each other, and the lower the average path length will be in the  
 525 network. Network size also correlate to the other metrics in scale-free networks: bigger scale-  
 526 free networks have higher path length and higher assortativity. These results reflect the  
 527 necessity of the sampling procedure used above to study the effect of the different metrics  
 528 separately.

<sup>5</sup> We only excluded random networks that possessed isolated component, in order to be able to compute the path length.



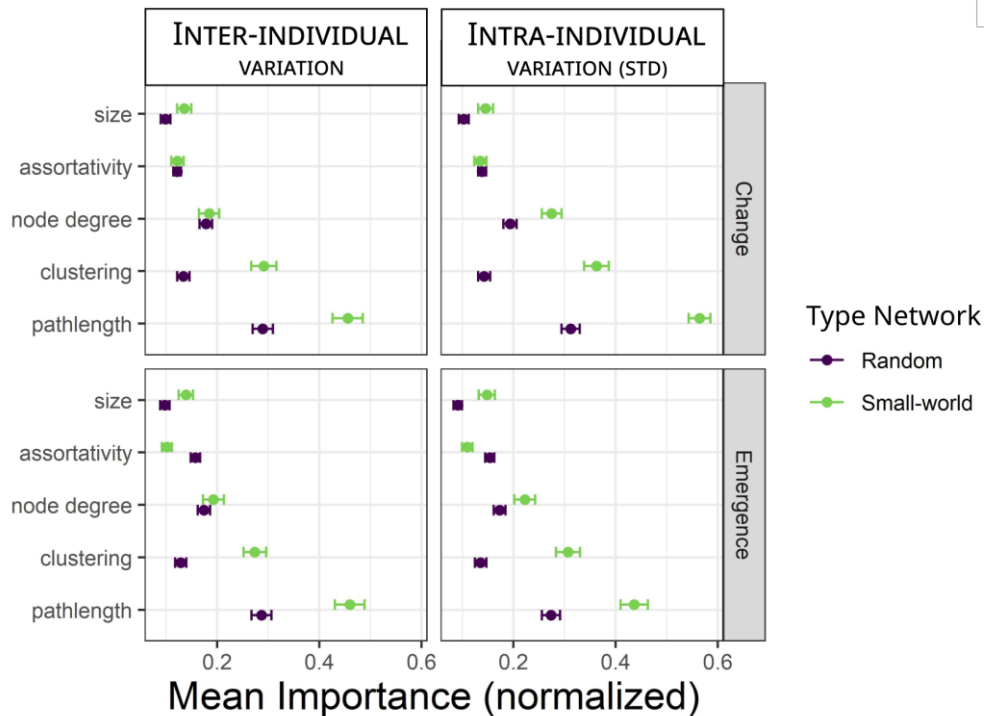
529 *Figure 3. Correlation matrix with the different metrics in scale-free (left), small-world (middle), and*  
 530 *random (right) networks.*  
 531  
 532

533 Using this same set of simulations, we applied several machine learning techniques to predict  
 534 variability in different network types using our metrics as predictors (see Figure 4) in both  
 535 language emergence and language change scenarios. We observed that the  $R^2$  for the mean  
 536 intra-individual variation was notably low, with an average  $R^2$  of 35% for language emergence  
 537 and of 2% in language change scenarios (these means were computed for all methods and  
 538 all types of networks; please note that there are high differences between the  $R^2$  obtained in  
 539 the different methods). Furthermore, our predictors in models using scale-free networks poorly  
 540 predicted any types of variations, with  $R^2$  around 37% for inter-individual variation and 4% for  
 541 the standard deviation of intra-individual variation. However, in small-world and random  
 542 networks, and in both language scenarios, our predictors effectively predicted inter-individual  
 543 variation and the standard deviation of intra-individual variation, with  $R^2$  values for these  
 544 models around 92% for inter-individual variation and 76% for standard-deviation of intra-  
 545 individual variation. Thus, we focus our analysis on these two types of networks and types of  
 546 variation. The exact  $R^2$  and RMSE for each method, network type and type of variation are  
 547 available in the **Supplementary Materials** (see Part 3.3.4).  
 548

549 **Inter-individual variation.** Encouragingly, these results concur with the previous findings  
 550 comparing sets of networks. The models consistently indicate that path length and clustering  
 551 coefficient are the most reliable predictors for inter-individual variation in both language  
 552 change and emergence scenarios. Notably, in this case, path length emerges as a stronger  
 553 predictor than the clustering coefficient. Interestingly, the models reveal a substantial impact  
 554 of average degree on inter-individual variation, particularly in random networks. The effect  
 555 sizes of the other metrics are very small, corroborating the idea that network size and  
 556 assortativity are not important for understanding inter-individual variability. There are no big  
 557 differences between the language change and language emergence scenarios regarding the  
 558 importance of predictors in both random and small-world networks.  
 559

560 **Mean intra-individual variation.** Similar to our previous observations, strong differences in  
 561 the mean intra-individual variation were identified between language change and language  
 562 emergence scenarios. However, our metrics do not contribute significantly to predicting this  
 563 type of variation, as evidenced by the relatively low  $R^2$ . This aligns with the findings of Dataset  
 564 1.  
 565

566 **Standard-deviation of intra-individual variation.** Concerning the standard deviation of intra-  
 567 individual variation, the importance of the predictors varies based on the type of network and  
 568 the initial language exposure. In both random and small-world networks, path length emerges  
 569 as the most crucial predictor. Notably, in small-world networks, it exerts a stronger influence  
 570 in language change scenarios compared to language emergence scenarios. In small-world  
 571 networks, the clustering coefficient takes the position of the second most important predictor,  
 572 followed by node degree. Within random networks, node degree holds the second-highest  
 573 importance as a predictor, closely followed by clustering coefficient and assortativity. In both  
 574 cases, assortativity and size exhibit the least importance.



575  
 576  
 577 *Figure 3. Mean importance for inter-individual variation (left column) and standard deviation of intra-*  
 578 *individual variation (right column) in language change scenario (upper panel) and language emergence*  
 579 *scenario (lower panel). Results for the mean intra-individual variation and for scale-free networks are*  
 580 *not presented, as the predictors exhibited poor prediction performance for all types of variation (low  $R^2$ ).*  
 581 *This measure of importance was computed by averaging the following z-scored predictor importance*  
 582 *measures: accuracy-based and Gini-based (extracted from random forests models), unconditional*  
 583 *(extracted from conditional forests), and sensitivity-based (extracted from support vector machines).*  
 584 *The dots represent the mean of these four measures, and the line represents the standard error*  
 585 *between these four measures. Results for random networks are shown in purple, while results for small-*  
 586 *world networks are displayed in green.*  
 587

#### 588 4. DISCUSSION AND CONCLUSIONS

589  
 590 Understanding the patterns of variability within a language provides important insights into the  
 591 great diversity observable in the world's languages. While many studies highlight the role of  
 592 network structure on language change, no differentiation was made between various network  
 593 characteristics in real world-populations or agent-based models, making it impossible to

594 disentangle their relative contributions. In this study, we applied several methods using a  
595 Bayesian agent-based model of language evolution to understand how specific characteristics  
596 of a network affect language in a language emergence and change scenarios.

597

598 We found that network structure affects inter-individual language variation in both language  
599 emergence and change scenarios. Although most earlier studies focused on grammatical  
600 complexity and other aspects of language, rather than linguistic variation per se, previous  
601 claims supported the idea that the effect of network structure on language was largely  
602 attributed to differences in the *size* and *density* of social networks (Dale & Lupyan, 2012;  
603 Trudgill, 2011b; Wray & Grace, 2007; Raviv et al., 2019). However, our results suggest that  
604 when considering language change, the size and density of a social network do not appear to  
605 be the most important factors. Instead, parameters such as path length and the clustering  
606 coefficient seem to be more influential in determining language variability. The confusion  
607 between size and density on the one hand, and clustering and path length on the other hand,  
608 can be easily understood given that these factors are highly correlated: bigger and well-  
609 connected populations tend to have shorter path lengths and higher clustering coefficients.  
610 Hampering the transmission of information and promoting the existence of distinct  
611 communities within a population will favor the emergence of variation in a language. Also, a  
612 population growing in size will not necessarily witness more variation, as long as the  
613 transmission of information is effective; this is, in particular, the case in populations with hubs,  
614 which facilitate the circulation of information. However, interestingly, in the first dataset and in  
615 the language emergence scenario only, it appears that the size of a network has an effect:  
616 larger populations tend to exhibit more variation. This emphasizes the intrinsic difference  
617 between the two scenarios and underscores the need to consider this factor when studying  
618 the evolution of language. Quite consistently, network density has a moderate effect in most  
619 analyses, but its exact role remains to be determined.

620

621 Additionally, the type of scenario (language change versus emergence) also plays an  
622 important role in predicting the amount of inter-individual variation present in a population.  
623 Thus, individuals born in a society where there is already a pre-existing language (language  
624 change scenario) would be more likely to use homogenous forms of language. In turn, this  
625 would explain why emerging sign languages in small remote villages (such as Al Sayyid  
626 Bedouin Sign Language sign language in Israel; Jaraisy & Stamp, 2022) usually exhibit more  
627 variation than official deaf community sign languages taught at schools, even though these  
628 official languages are characterized by sparser and larger populations of signers (Meir et al.,  
629 2012; Mudd et al., 2020). We suggest that the pressure for conformity (and more specifically,  
630 for using the dominant variant in the population) is the main factor underlying the differences  
631 between an established language (language *change* scenario) and an emerging language  
632 (language *emergence* scenario). Importantly, the strong influence of the type of scenario  
633 suggests that non-structural factors do play a very important role in shaping language  
634 variability, and that network structure cannot alone account for all the diversity present in real-  
635 world networks. For example, other extrinsic factors such as the role of a shared context (Mudd  
636 et al., 2022) and the openness to strangers (Dale & Lupyan, 2012) were not taken into account  
637 here, but they might have additional contributions to the effect of network structure.

638

639 On the contrary, intra-individual variability is not shaped by the network structure, but only by  
640 the type of scenario of the society in which an individual is born. An individual born in a society  
641 with low pressure to use a specific form of language (language *emergence* scenario) will use

642 different types of utterances to refer to the same concept. These findings relate to the study  
643 of very recent emergent sign languages, such as Kata Kolok, where high intra-individual  
644 variation was observed (Lutzenberger et al., 2021). However, differences between individuals  
645 regarding their level of intra-individual variation are affected by the network structure. Here  
646 too, a high path length and a high clustering coefficient increase the probability that the  
647 population possesses both individuals very confident in their language beliefs and individuals  
648 with low confidence in which language form to use. We suggest that both high path length and  
649 high clustering coefficients are related to the presence of structural communities (cliques),  
650 which then lead to linguistic communities. Individuals bridging these communities are then  
651 more likely to possess high intra-individual variation, changing their language according to  
652 whom they are speaking to, while individuals inside communities use the language spoken by  
653 their group unconditionally.

654  
655 Together, these results suggest that differences at the global level of network structure,  
656 especially for path length and clustering coefficient, influence language inter-individual  
657 variability. They also highlight the weak role of network structure on intra-individual variation.  
658 However, one potential limitation of our study is that the degree of contrast between the sets  
659 of networks generated based on a single metric may not have been strong enough to capture  
660 the full extent of the metric's influence. For instance, the magnitude of the difference between  
661 the high and low node degree networks may have been insufficient to fully explain the impact  
662 of node degree on language variability. Also, our "flat" Bayesian model contrasts with the  
663 human capacity to differentiate and track various language varieties depending on the  
664 interlocutors and contexts. This highlights the potential for elaborating more complex  
665 simulations (such as hierarchical Bayesian models) in future research, and meanwhile, our  
666 conclusions should be extrapolated with caution to real-world language networks. Additionally,  
667 the model only accounted for static networks, whereas real-world networks are dynamic and  
668 evolve through interactions. While interactions are the driving force behind network structure,  
669 further research could explore the mechanisms of preferential attachment to gain a deeper  
670 understanding of their impact on language variability.

671

## 672 5. DATA AVAILABILITY STATEMENT / SUPP MATERIALS

673

674 The datasets and the analysis presented in this study can be found in the online repositories  
675 [mathjoss/NetworkStructure\\_ABM](https://mathjoss.org/NetworkStructure_ABM).

## 676 6. AUTHOR CONTRIBUTIONS

677

678 BdB, DD, MA-T, and MJ designed the research. MJ performed the research, wrote and ran  
679 the simulations. MJ wrote the first draft with substantial contributions from DD and FP. All  
680 authors (BdB, DD, MA-T, MJ, and FP) read, contributed to, and approved the manuscript.

681

682

683

684 7. REFERENCES

- 685 Albert, R. (2005). Scale-free networks in cell biology. *Journal of Cell Science*, 118(21), 4947.  
686 <https://doi.org/10.1242/jcs.02714>
- 687 Albert, R., Jeong, H., & Barabási, A.-L. (1999). Diameter of the World-Wide Web. *Nature*, 401(6749),  
688 130–131. <https://doi.org/10.1038/43601>
- 689 Barabási, A.-L., Albert, R., & Jeong, H. (2000). Scale-free characteristics of random networks : The  
690 topology of the world-wide web. *Physica A: Statistical Mechanics and Its Applications*, 281(1), 1.  
691 [https://doi.org/10.1016/S0378-4371\(00\)00018-2](https://doi.org/10.1016/S0378-4371(00)00018-2)
- 692 Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Ke, J., Christiansen, M. H., Larsen-  
693 Freeman, D., Croft, W., & Schoenemann, T. (2009). Language is a complex adaptive system :  
694 Position paper. *Language Learning*, 59(Suppl 1), 1–26. <https://doi.org/10.1111/j.1467->  
695 9922.2009.00533.x
- 696 Blythe, R. A., & Croft, W. (2012). S-curves and the mechanisms of propagation in language change.  
697 *Language*, 88(2), 2. <https://doi.org/10.1353/lan.2012.0027>
- 698 Chambers, J. K. (1995). *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*.  
699 Blackwell Publishers.
- 700 Cortez, P. (2010). Data Mining with Neural Networks and Support Vector Machines using the R/rminer  
701 Tool. in *Advances in Data Mining – Applications and Theoretical Aspects, 10th Industrial Conference*  
702 *on Data Mining* (ed. Perner, P.) 572–583 (LNAI 6171, Springer, 2010).
- 703 Croft, W. (2008). Evolutionary Linguistics. *Annual Review of Anthropology*, 37, 219-234.
- 704 Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity : The linguistic  
705 niche hypothesis. *Advances in Complex Systems*, 15(03n04), 03n04.  
706 <https://doi.org/10.1142/S0219525911500172>
- 707 Dediu, D., & Moisik, S. R. (2019). Pushes and pulls from below : Anatomical variation, articulation and  
708 sound change. *Glossa: A Journal of General Linguistics*, 4(1), 1. <https://doi.org/10.5334/gjgl.646>
- 709 Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups.  
710 *Proceedings of the National Academy of Sciences*, 113(11), 11.  
711 <https://doi.org/10.1073/pnas.1518798113>
- 712 Erdős, P., & Rényi, A. (1959). On Random Graphs I. *Publicationes Mathematicae (Debrecen)*, 6,  
713 290-297.
- 714 Everett, C., Blasi, D. E., & Roberts, S. G. (2015). Climate, vocal folds, and tonal languages :  
715 Connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences*,  
716 201417413. <https://doi.org/10.1073/pnas.1417413112>
- 717 Fagyal, Z., Swarup, S., Escobar, A. M., Gasser, L., & Lakkaraju, K. (2010). Centers and peripheries :  
718 Network roles in language change. *Lingua*, 120(8), 2061-2079.  
719 <https://doi.org/10.1016/j.lingua.2010.02.001>

- 720 Gong, T., Baronchelli, A., Puglisi, A., & Loreto, V. (2012). Exploring the roles of complex networks in  
721 linguistic categorization. *Artificial Life*, 18(1), 107-121. [https://doi.org/10.1162/artl\\_a\\_00051](https://doi.org/10.1162/artl_a_00051)
- 722 Griffiths, T. L., & Kalish, M. L. (2007). Language Evolution by Iterated Learning With Bayesian Agents.  
723 *Cognitive Science*, 31(3), 441-480. <https://doi.org/10.1080/15326900701326576>
- 724 Hausser, J., & Strimmer, K. (2021). *entropy: Estimation of Entropy, Mutual Information and Related*  
725 *Quantities*. <https://CRAN.R-project.org/package=entropy>
- 726 Hothorn T, & Zeileis A (2015). "partykit: A Modular Toolkit for Recursive Partytioning in R." *Journal of*  
727 *Machine Learning Research*, 16, 3905-3909. <https://jmlr.org/papers/v16/hothorn15a.html>.
- 728 Jaraisy, M., & Stamp, R. (2022). The Vulnerability of Emerging Sign Languages : (E)merging Sign  
729 Languages? *Languages*, 7(1), 49. <https://doi.org/10.3390/languages7010049>
- 730 Josserand, M., Allasonnière-Tang, M., Pellegrino, F., & Dediu, D. (2021). Interindividual Variation  
731 Refuses to Go Away : A Bayesian Computer Model of Language Change in Communicative  
732 Networks. *Frontiers in Psychology*, 12, 626118. <https://doi.org/10.3389/fpsyg.2021.626118>
- 733 Ke, J.-Y., Gong, T., & Wang, W. (2008). Language change in social networks. *Communications in*  
734 *Computational Physics (Cicp)*, v.3, 935-949 (2008), 3.
- 735 Kenett, Y. N., Levy, O., Kenett, D. Y., Stanley, H. E., Faust, M., & Havlin, S. (2018). Flexibility of  
736 thought in high creative individuals represented by percolation analysis. *Proceedings of the National*  
737 *Academy of Sciences*, 115(5), 5. <https://doi.org/10.1073/pnas.1717362115>
- 738 Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010).  
739 Identification of influential spreaders in complex networks. *Nature Physics*, 6(11), 11.  
740 <https://doi.org/10.1038/nphys1746>
- 741 Kobayashi, Y., Ohtsuki, H., & Wakano, J. Y. (2016). Population size vs. Social connectedness—A  
742 gene-culture coevolutionary approach to cumulative cultural evolution. *Theoretical Population Biology*,  
743 111, 87-95. <https://doi.org/10.1016/j.tpb.2016.07.001>
- 744 Labov, W. (1972). *Sociolinguic Patterns* (Blackwell).
- 745 Labov, W. (2001). *Principles of linguistic change Vol. 2: Social Factors*. Blackwell.
- 746 Labov, W. (2010). *Principles of Linguistic Change : Cognitive and Cultural Factors*. Wiley-Blackwell.  
747 <https://doi.org/10.1002/9781444327496>
- 748 Lee, Y., Collier, T., Stabler, E., & Taylor, C. (2005). The role of population structure in language  
749 evolution.
- 750 Liaw A, & Wiener M (2002). "Classification and Regression by randomForest." *R News*, 2(3), 18-22.  
751 <https://CRAN.R-project.org/doc/Rnews/>.
- 752 Lou-Magnuson, M., & Onnis, L. (2018). Social Network Limits Language Complexity. *Cognitive*  
753 *Science*, 42(8), 2790-2817. <https://doi.org/10.1111/cogs.12683>

- 754 Lupyan, G., & Dale, R. (2010). Language Structure Is Partly Determined by Social Structure. *PLoS*  
755 *ONE*, 5(1), 1. <https://doi.org/10.1371/journal.pone.0008559>
- 756 Lutzenberger, H., de Vos, C., Crasborn, O., & Fikkert, P. (2021). Formal variation in the Kata Kolok  
757 lexicon. *Glossa: a journal of general linguistics*, 6(1). <https://doi.org/10.16995/glossa.5880>
- 758 Maddieson, I., & Coupé, C. (2015). Human spoken language diversity and the acoustic adaptation  
759 hypothesis. *The Journal of the Acoustical Society of America*, 138(3), 1838–1838.  
760 <https://doi.org/10.1121/1.4933848>
- 761 Meir, I., Israel, A., Sandler, W., Padden, C. A., & Aronoff, M. (2012). The influence of community on  
762 language structure : Evidence from two young sign languages. *Linguistic Variation*, 12(2), 247-291.  
763 <https://doi.org/10.1075/lv.12.2.04mei>
- 764 Migliano, A. B., Battiston, F., Viguier, S., Page, A. E., Dyle, M., Schlaepfer, R., Smith, D., Astete, L.,  
765 Ngales, M., Gomez-Gardenes, J., Latora, V., & Vinicius, L. (2020). Hunter-gatherer multilevel sociality  
766 accelerates cumulative cultural evolution. *Science Advances*, 6(9), eaax5913.  
767 <https://doi.org/10.1126/sciadv.aax5913>
- 768 Milgram, S. (1967). The small-world problem. *Psychology Today*, 1(1), 1.
- 769 Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of*  
770 *Linguistics*, 21(2), 2. <https://doi.org/10.1017/S0022226700010306>
- 771 Mudd, K. (2022). *How social structure affects the persistence and features of shared sign languages*.  
772 Artificial Intelligence Lab, VUB University.
- 773 Mudd, K., de Vos, C., & de Boer, B. (2020). The effect of cultural transmission on shared sign  
774 language persistence. *Palgrave Communications*, 6(1), 1. <https://doi.org/10.1057/s41599-020-0479-3>
- 775 Mudd, K., de Vos, C., & de Boer, B. (2022). Shared Context Facilitates Lexical Variation in Sign  
776 Language Emergence. *Languages*, 7(1), 1. <https://doi.org/10.3390/languages7010031>
- 777 Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2),  
778 167-256. <https://doi.org/10.1137/S003614450342480>
- 779 Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks.  
780 *Physical Review E*, 68(3), 036122. <https://doi.org/10.1103/PhysRevE.68.036122>
- 781 Nichols, J. (1992). *Linguistic Diversity in Space and Time* (The University of Chicago Press).
- 782 Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages.  
783 *Proceedings of the Royal Society B: Biological Sciences*, 286(1907), 20191262.  
784 <https://doi.org/10.1098/rspb.2019.1262>
- 785 Raviv, L., Meyer, A., & Lev-Ari, S. (2020). The Role of Social Network Structure in the Emergence of  
786 Linguistic Structure. *Cognitive Science*, 44(8), e12876. <https://doi.org/10.1111/cogs.12876>
- 787 Realí, F., Chater, N., & Christiansen, M. H. (2014). The paradox of linguistic complexity and  
788 community size. In *The Evolution of Language* (p. 270-277). WORLD SCIENTIFIC.  
789 [https://doi.org/10.1142/9789814603638\\_0035](https://doi.org/10.1142/9789814603638_0035)



- 790 Sankoff, G., & Blondeau, H. (2007). Language Change across the Lifespan : /R/ in Montreal French.  
791 *Language*, 83(3), 560–588.
- 792 Solé, M.-J., & Vives, D. R. i. (2012). *The Initiation of Sound Change : Perception, Production, and*  
793 *Social Factors*. John Benjamins Publishing.
- 794 Stevens, M., & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens*,  
795 1(1), 1. <https://doi.org/10.3989/loquens.2014.003>
- 796 Trudgill, P. (2011a). *Sociolinguistic Typology : Social Determinants of Linguistic Complexity*. OUP  
797 Oxford.
- 798 Trudgill, P. (2011b). Social structure and phoneme inventories. *Linguistic Typology*, 15(2), 2.  
799 <https://doi.org/10.1515/lity.2011.010>
- 800 Trudgill, P., & Chambers, J. K. (2017). *Dialects of English : Studies in Grammatical Variation*.  
801 Routledge.
- 802 Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*,  
803 393(6684), 440–442. <https://doi.org/10.1038/30918>
- 804 Wilensky, U. (1999). NetLogo. <http://ccl.northwestern.edu/netlogo/>
- 805 Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers : Evolutionary corollaries  
806 of socio-cultural influences on linguistic form. *Lingua*, 117(3), 3.  
807 <https://doi.org/10.1016/j.lingua.2005.05.005>
- 808 Yu, A. C. L. (2013). *Origins of Sound Change : Approaches to Phonologization*. Oxford University  
809 Press.
- 810 Zeshan, U., & de Vos, C. (2012). *Sign Languages in Village Communities : Anthropological and*  
811 *Linguistic Insights*. Walter de Gruyter.
- 812 Zubek, J., Denkiewicz, M., Barański, J., Wróblewski, P., Rączaszek-Leonardi, J., & Plewczynski, D.  
813 (2017). Social adaptation in multi-agent model of linguistic categorization is affected by network  
814 information flow. *PloS One*, 12(8), e0182490. <https://doi.org/10.1371/journal.pone.0182490>