Doctoral Dissertation

博士論文

A mathematical model of cultural evolution on a population network and its application to the analysis of spatial distribution of dialects

(集団ネットワーク上の文化進化の数理モデルと、方言の空間的分布の解析への応用）

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2020

令和２年１２月　博士（理学）申請

Department of Biological Sciences, Graduate School of Science
The University of Tokyo

東京大学大学院理学系研究科

生物科学専攻

Takuya Takahashi

高橋　拓也

Contents

Abstract

In analogy with genetic traits, cultural traits diffuse from one population to another through interaction and learning process of individuals. To deduce how newly invented traits spread from their original populations and to analyze how observed geographic pattern of cultural traits has evolved, network is a convenient mathematical model. This thesis, composed of three substantial chapters enumerated Chapter 2, 3, and 4, aims to shed light to the cultural evolution among multiple populations with the aid of a network whose nodes represent populations and edges represent transmission of cultural traits. As an application of this model, I specifically treat the spatial pattern of dialects, or lexical variants.

In Chapter 2, I discuss a peculiar spatial distribution of dialects; some dialect words are shared among geographically distant groups of people without close interaction. Such a pattern may indicate the current or past presence of a cultural center exerting a strong influence on peripheries. For example, concentric distributions of dialect variants in Japan may be explicable by repeated inventions of new variants at Kyoto, the ancient capital, with subsequent outward diffusion. In Chapter 2, I develop a model of linguistic diffusion within a population network to quantify the distribution of variants created at the central population. Equilibrium distributions of word ages are obtained for idealized networks and for a realistic network of Japanese prefectures. My model successfully replicates the observed pattern, supporting the notion that a center-periphery social structure underlies the emergence of concentric patterns. Unlike what has previously been claimed, my model indicates that a novelty bias in linguistic transmission is not always necessary to account for the concentric pattern, whereas some bias in the direction of transmission between populations is needed to be consistent with the observed absence of old words near the central population. My analysis on the realistic network also suggests that the process of linguistic transmission is not much affected by between-prefecture differences in population size.

In Chapter 3, I further generalize the model used in Chapter 2 and consider the dynamics of cultural macroevolution, which concerns a long-term evolutionary process involving transmission of non-genetic or cultural traits between populations as well as birth and death of populations. To understand the spatial dynamics of macroevolution, I present a network model of cultural transmission in which any population may innovate a novel trait. Borrowing the method of ancestral backward process from population genetics, my model explores the genealogy of a cultural variant sampled in the present

generation. Mathematical analysis of the model enables us to predict the distribution of cultural age in each population of the network, investigate the frequencies of cultural variants originating from given populations, discuss the time it takes for a cultural variant to diffuse between a given pair of populations, and calculate the frequency of each variant in each population under a finite-variant assumption. I also perform numerical analysis on random networks of populations to investigate the effect of network topology and innovation rate on the age and origin of cultural variants in each population. The results suggest that a cultural variant tends to be maintained at a higher frequency if its original population is characterized by more innovations, more influence on other populations, and/or less influence from other populations. My finding also includes that a cultural variant invented in a given population tends to spread more rapidly if the population is more influential to and/or more influenced by other populations.

In Chapter 4, I investigate the spatial distribution of lexical variants of Japanese language based on both empirical data and the mathematical model presented in Chapter 3, focusing mainly on the relationship between linguistic distance (LD) and geographic distance (GD). For empirical data, I calculate the edit (Levenshtein) distance among lexical variants recorded at 2400 localities surveyed in the Linguistic Atlas of Japan (LAJ) project, showing that linguistic distance between localities is strongly correlated with geographic distance. Since a pair of words derived from the same source usually bears a marked similarity, simulating when and where lexical variants were invented is an essential way to examine the linguistic distance between localities. For this purpose, I develop a network of the surveyed localities of LAJ, based on their geographic positions and population sizes, and apply the model of Chapter 3 to quantify the origin and expected age of variants. I show that variants are likely to be originated from localities with large population sizes, whereas variants originating from an arbitrary population occupy every population with a relatively similar probability, unless cultural transmission occurs exclusively between proximate populations. In addition, simulating the linguistic distance for locality pairs on the network, I show that linguistic distance correlates strongly with the logarithm of geographic distance (Séguy’s curve) if variants transmit mostly between closely positioned localities. Conversely, if variants can also transmit between remote localities, linguistic distance correlates more strongly with geographic distance itself, resulting in a linear relationship between LD and GD. Moreover, my simulation reveals that accumulation of linguistic distance is not spatially isotropic when lexical variants can only transmit between proximate localities. Within the framework of approximate Bayesian computation with Markov Chain Monte Carlo (MCMC) sampling, we infer parameter values of our model based on empirical data in LAJ.

Chapter 1. General introduction

Humans are characterized by their wide variety of cultures [1,2]. Cultures or cultural traits are a broad concept defined as traits inherited from individual to individual via non-genetic transmission, such as archaeological traits, skills, information, and languages. Cultural traits are invented through individual learning and transmitted through social learning, and their dynamics have been studied as an analogy to evolution of genetic traits.

Cultural evolution is explained by two processes according to its scale: cultural microevolution and cultural macroevolution. While the former treats the frequency change of cultural traits within a single population, the latter considers the transmission of cultural traits among multiple populations along with divergence, convergence, birth, and death of populations [3]. As for cultural macroevolution, a growing body of literature examines the spatial dynamics of cultural transmission both empirically and theoretically. For example, phylogenic methods have been used to infer the historical relationship of cultural traits of multiple populations [4]. Transmission of hinoeuma superstition among Japanese prefectures was examined along with the fertility decline rate of each group [5]. Reaction-diffusion systems were applied to model the spatial range-expansion of modern humans coupled with diffusion of modern skills [6] and farming [7]. In addition, transmission of social information has been studied through a network model called voter model [8].

Here, in this thesis, I will explore the transmission of cultural traits among multiple populations, which plays a pivotal role in the dynamics of cultural macroevolution. For this purpose, we develop a network model whose nodes represent populations and whose edges represent the existence of inter-population transmission of cultural traits. Unlike voter model, which often treats the dynamics of extant traits and calculates the fixation time and probability of one trait, I will consider a case where novel cultural variants are continuously invented and old ones are replaced on the network. Emphasis will be placed on the age and origin of cultural variants, or in other words, I will establish a theory to infer when and where the traits were invented.

Beyond theoretical analysis of the mathematical model, I will apply the network model to the spatial distribution of dialects, which is a typical example of cultural traits in humans. Spatial patterns and diffusion of languages have been examined in linguistic geography, and classical research recorded the distribution of linguistic items on linguistic atlases [9, 10], based on interviews with local volunteers as informants. Starting from Seguy's [11] work which was based on the linguistic atlas of France [12], quantitative research into linguistic atlas correlated the linguistic distance with geographic distance, often finding a sublinear increase of linguistic distance as a function of geographic distance [11, 13]. In addition, spatial evolution of linguistic traits has been theoretically explored by means of mathematical models, such as voter model [14,15] and surface tension model [16].

The purpose of this thesis is two-fold; First, I will establish a mathematical model of cultural evolution which assumes the transmission rate and innovation rate of each population and infer the spatial and temporal dynamics of cultural variants. Second, having obtained the mathematical formulae to estimate age and origin of cultural variants, I will apply it to the diffusion of linguistic traits and discuss whether the previously observed phenomena in linguistic geography can be explained by our model.

This thesis consists of three main chapters. In Chapter 2, I will treat a phenomenon that some dialect words in Japanese mainland are arranged in a concentric ring-like pattern around Kyoto, where novel words occupy the center and ancient words occupy the periphery [17]. The theory, posited by Kunio Yanagita [18], is explained by consecutive innovation of new words in Kyoto and subsequent diffusion to its periphery. In Chapter 2, to provide a mathematical support for this theory, I will introduce a network model and explore how words created in a central population like Kyoto will diffuse into its periphery. In this chapter, I will introduce a notion of word age and examine its spatial pattern, based on both theoretical and realistic networks of populations. Using a variety of networks, I will analyze how network topology, such as direction of transmission and existence of barriers hindering transmission, exerts influence on the diffusion of dialect words.

In Chapter 3, I will further generalize the model used in Chapter 2, which is specialized for the case of center-periphery structure, and develop a network model in which any population has the chance to invent a new cultural variant. Considering the backward ancestral process and cultural genealogy [19], I will derive mathematical formulae which give the origins of variants in each population, as well as the expected time it takes for a cultural variant to transmit between two populations. Besides analytical approach, I will perform a numerical analysis using a random network to infer the effect of topological structure and innovation rate on the age and origin of cultural variants. In addition, I will establish a model of a finite number of variants and quantify the frequency, age, and origin of each variant for every node (population) of the network.

In Chapter 4, I will elucidate the spatial distribution of Japanese lexical variation relying on both empirical data and my network model. I will apply the mathematical formulae derived in Chapter 3 to the network whose nodes represent the survey locations of LAJ (Linguistic Atlas of Japan) project [9] and simulate the linguistic diffusion of Japanese language. Moreover, I will calculate empirical linguistic distance for every locality pair based on the lexical data archived in LAJDB (Linguistic Atlas of Japan Database) [20] and obtain the empirical pattern of Japanese lexical variants. I will thereby analyze the relationship between geographical and linguistic distances based on both empirical data and simulation. Finally, comparing the observed and simulated linguistic distances, I will infer parameter values of the model within the framework of approximate Bayesian computation (ABC).

Reference

[1] Boyd R, and Richerson PJ. 1985 Culture and the Evolutionary Process. University of Chicago Press, Chicago.

[2] Cavalli-Sforza LL, Feldman MW. 1981 Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton.

[3] Mesoudi A. 2011 Cultural Evolution: How Darwinian Theory Can Explain Human Culture & Synthesize the Social Sciences. University of Chicago Press, Chicago.

[4] Nunn CL, Mulder MB, Langley S. 2006 Comparative methods for studying cultural trait evolution: A simulation study. Cross-Cultural Research 40, 2, 177-209.

[5] Tamura K, Ihara Y. 2017 Quantifying cultural macro-evolution: a case study of the hinoeuma fertility drop. Evolution and Human Behavior 38, 117-124.

[6] Wakano JY, Gilpin W, Kadowaki S, Feldman MW, Aoki K. 2018 Ecological range-expansion scenarios for the replacement or assimilation of Neanderthals by modern humans. Theoretical Population Biology 119, 3-14

[7] Aoki K, Shida M. 1996 Travelling wave solutions for the spread of farmers into a region occupied by hunter-gatherers. Theoretical Population Biology 50, 1-17.

[8] Sood V, Antal T, Redner S. 2008 Voter models on heterogeneous networks. Physical Review E 77, 041121.

[9] National Institute for Japanese Language and Linguistics. 1966-1974 Linguistic Atlas of Japan 1-6.

[10] Orton H, Sanderson S, Widdowson J. 1998 The Linguistic Atlas of England.

[11] Séguy J. 1971 La relation entre la distance spatiale et la distance lexicale. Rev. Linguist. Romane 35, 335–357.

[12] Gilliéron J, Edmont E. 1903–1910. Atlas linguistique de la France.

[13] Nerbonne J. 2010 Measuring the diffusion of linguistic change. Phil. Trans. R. soc. B 365, 3821-3828.

[14] Hadzibeganovic T, Stauffer D, Schulze C. 2008 Boundary effects in a three-state modified voter model for languages. Physica A 387. 3242-3252.

[15] Fagyal Z, Swarup S, Escobar AM, Gasser L, Lakkaraju K. 2010 Centers and peripheries: Network roles in language change. Lingua 120. 2061-2079.

[16] Burridge J. 2018 Unifying models of dialect spread and extinction using surface tension dynamics. Royal Society Open Science 5, 171446.

[17] Matsumoto O. 1993 Zenkoku aho-baka bumpu-kou. Ohta Shuppan.

[18] Yanagita K. 1927 Kagyuko (1)-(4). Jinruigaku Zasshi 42, 125-135 (162-172, 223-233, 273-284).

[19] Aguilar E, Ghirlanda S. 2015 Modeling the genealogy of a cultural trait. Theoretical Population Biology 101, 1-8.

[20] Linguistic Atlas of Japan Database (LAJDB) https://lajdb.org/TOP.html

Chapter 2.

Quantifying the spatial pattern of dialect words spreading from a central population

1. Introduction

A dialect is a variant of a language that is spoken by a distinct group of people, where regional dialects may differ from each other in terms of phonology, lexicon, morphology and syntax [1-3]. On the level of linguistic variation within a language family, similarities between languages have been used to reconstruct the phylogenetic relationship among human populations [4-6] based on the premise that populations linguistically more similar to each other are likely to have diverged more recently from an ancestral population (i.e., cultural macroevolution; [7]). Similarly, on the level of dialect variation, much quantitative research has shown that the linguistic distance of each locality is to some extent explained by the geographic distance, although its correlation coefficient varies depending on how geographic distance is measured [8,9]. These studies give impression that the similarity of language reflects the phylogeny of human groups, but it is also often the case that the same dialect variant of a word is documented in phylogenetically distant local groups [10, 12], which is likely due to diffusion of words between groups.

On the basis of extensive documentation of Japanese dialects, Yanagita described peculiar geographic distributions of words within the country [10]. In particular, he pointed out that the same dialect variants of the word for snail (*kagyu*) were seen in both ends of the east-west stretch of the land, while they were absent in the middle. Similar patterns of dialect words were found in the nation-wide project of Linguistic Atlas of Japan (LAJ) [1], in which words meaning face epitomize this distribution (available at https://mmsrv.ninjal.ac.jp/laj_map/data/laj_map/LAJ_106.pdf). To account for these patterns, Yanagita posited that dialect forms in Japanese may exhibit a concentric distribution centred at Kyoto, the old capital in the middle Japan. According to this theory, new words were repeatedly invented in Kyoto and diffused gradually outward to periphery, leaving concentric traces. Underlying assumptions are that new words were preferentially adopted by people, perhaps owing to Kyoto's prestige as the capital, and that the diffusion was slow relative to creation of words, which is plausible given the absence of modern technologies such as television or the Internet.

Concentric distribution of word variants is not unique to Japanese language but occurs in other places where populations are socially or geographically structured into centres and peripheries (hereafter the centre-periphery structure). For example, research based on linguistic atlases of Breton and French languages [3, 11] has revealed that several word forms are distributed in a concentric pattern in Lower-Brittany, highlighting a profound impact of economically and culturally important towns on the spread of word variants [12]. Despite the ubiquity of similar concentric patterns, most previous studies have merely proposed verbal explanations of the phenomenon without presenting any quantitative analysis. In particular, there is a dearth of mathematical rationale to unveil the underlying factors of the concentric patterns in dialects.

This chapter aims to provide a quantitative support to the theory concerning the evolution of concentric distribution of dialects [10]. For this purpose, we will develop a mathematical model and examine whether and how observed and hypothesized concentric patterns are replicated from simplistic assumptions of the model.

For a mathematical treatment of geographical patterning of dialect variants in the presence of the centre-periphery structure, we need a model considering linguistic influences among multiple groups of people. One commonly used framework is the gravity model [13], in which the mutual influence of two centres (towns, cities, etc.) is assumed to be proportional to the product of their populations and inversely proportional to the squared distance between them. This model predicts that linguistic features first diffuse from city to city, skipping the rural area in between. Kretzschmar [14] used cellular automaton (CA) as a computational model to investigate temporal changes in linguistic features across areas. Fagyal et al. [15] conducted an agent-based simulation to investigate the language change in a heterogeneous social network, in which highly connected and isolated agents constitute a centre-periphery structure. Burridge [16,17] has recently developed spatially explicit models of linguistic change, borrowing methods from statistical physics. Incorporating demographic data, he demonstrated the spread of words from a city, or densely populated area, to the peripheries. These models provide explanations for interesting linguistic phenomena, including temporal dynamics of dialect boundaries; however, they are silent about the possibility of concentric dialect distribution. This is because these models are designed to deal with a fixed number of pre-existent dialect variants and thus do not allow for repeated inventions of new words in a central population as presupposed by Yanagita [10].

A theoretical study that is more relevant to the current context is by Lizana et al. [18], who focused on the proposed concentric distribution of swear words in Japanese dialects, such as *aho* and *baka*, meaning a stupid person [19]. They ran a computer simulation on a two-dimensional lattice that represents the real geography of Japanese Archipelago, assuming that new words are repeatedly invented in Kyoto and then transmitted to neighbouring regions. A critical assumption was that there is a novelty bias in the transmission of words, so that a newer variant will invade and replace an older variant occupying a lattice site, but not vice versa. The simulation successfully reproduced two empirical features of the swear-word distribution: (1) the same variants are found both to the east and west of Kyoto; and (2) the geographic band within which a variant is found is broader when it is further from Kyoto. The same research also reported that the absence of the novelty bias results in the disappearance of the concentric pattern.

While Lizana et al.'s work [18], which is mostly numerical, has demonstrated that a concentric distribution of words can indeed be formed under a set of reasonable assumptions, a fuller mathematical analysis would shed more light on the processes of linguistic diffusion underlying the observed patterns of linguistic variation. To achieve the latter, this paper develops a mathematical model, assuming a network of populations with a central population from which every word derives, as a simplest representation of the centre-periphery structure. Our model differs from the previous study in three ways. First, we deliberately omit the novelty bias in the transmission of words. Although Lizana et al. suggested that the appearance of concentric pattern was conditional on the presence of the preference for novel words, we show that this assumption is not always necessary for the formation of a concentric distribution. Second, while only one variant occupies each lattice site in Lizana et al.'s model, multiple variants can coexist in a single population in our model. The frequency of individuals having a given variant is represented by a real number ranging from zero to one in each population. This assumption seems more realistic because speakers in a single population may use different words, or multiple dialects may be seen in the same group of people. In fact, a questionnaire research has reported that some respondent answered multiple *aho-baka* expressions prevalent within the same area [19]. Finally, and as a corollary to the second point, we do not define the distance from the central population for each variant. This is because in our formulation each variant may be used in different frequencies in multiple populations, which is unlike Lizana et al.'s model. Instead, we track changes in the distribution of word ages in each population. As every word is consecutively invented in the central population, different word age corresponds to a different variant, so we can indirectly deduce the distribution of words by quantifying the spatial pattern of word ages.

In what follows, we will first develop general formulae to calculate the distribution, mean, and standard deviation of word age in each population within a network of populations, under the assumption that populations are large (Section 2). In Section 3, we apply them to simplistic, schematic networks in order to grasp the general characteristics of the word-age pattern. In particular, we will treat the following idealized networks:

(1) One-dimensional lattice with unidirectional diffusion
(2) One-dimensional lattice with bidirectional diffusion
(3) Two-dimensional lattice
(4) One-dimensional lattice with a barrier
(5) Two-dimensional lattice with a barrier

Section 4 examines the distribution on a more realistic network, based on the network of Japanese prefectures.

2. Theory

Description of the model

We consider transmission of a linguistic trait within and between $n+1$ populations, $P_0, P_1, \dots, P_n$, each of which consists of a sufficiently large number of individuals, where the assumption of large population sizes is for the sake of mathematical simplicity. Innovations of words occur only in population P_0, which we call the central population. In every timestep, one novel word is invented and immediately spreads within P_0. We treat a polychotomous linguistic trait, such as multiple words meaning the same object, or different pronunciations and intonations for the same word. Thus, an individual can have only one variant at given time. This is analogous to the one-locus model in population biology. Members of populations other than P_0, which we call peripheral populations, may obtain a variant by learning socially from an individual in the same or other populations. After social learning, all individuals' variants are updated simultaneously at the beginning of the next timestep. We define $t = 0$ as the time when the central population emerges, and transmission starts.

Every single linguistic variant in P_k $(0 \leq k \leq n)$ derives from P_0, given it was created when $t \geq 0$, so we can distinguish the variants by their ages. Let $f_k(\rho, t)$ denote the frequency of the variant aged ρ (≥ 0) in population P_k at timestep t, where the age of a variant is measured by the number of timesteps elapsed since the variant was created and does not indicate any concrete time unit such as year, decade or generation. We have

$$\sum_{\rho=0}^{\infty} f_k(\rho, t) = 1 \quad (0 \leq k \leq n). \quad (1)$$

Here, $0 \leq \rho \leq t$ corresponds to the variants which were invented in P_0, whereas $\rho > t$ represents the ones that had already existed when the central population emerged at $t = 0$.

As for the central population P_0, the frequencies of word ages are written as

$$f_0(\rho, t) = \begin{cases} 1 & (\text{if } \rho = 0) \\ 0 & (\text{if } \rho > 0) \end{cases}, \quad (2)$$

which means that all individuals in the central population always have the latest variant. In peripheral populations, each individual chooses a role model from whom to learn a linguistic variant. In the choice of role model, a learner first chooses a population to which a potential role model belongs, where the probability with which a learner in P_i chooses P_j is denoted by a_{ij} (≥ 0) $(1 \leq i \leq n, 0 \leq j \leq n)$, and then chooses a role model from all individuals in the chosen population with equal probability. Since the population is sufficiently large for stochastic effect to be negligible, we can deterministically obtain the following recursive formula as regard to frequencies in the peripheral populations:

$$f_k(\rho, t) = \sum_{j=0}^{n} a_{kj} f_j(\rho - 1, t - 1) \quad (1 \leq k \leq n). \quad (3)$$

Note that $f_k(\rho, t)$ and $f_j(\rho - 1, t - 1)$ represent the frequencies of the same variant in different populations at different timesteps. We will refer to a_{ij} as the transmission rate from P_j to P_i, which may depend on the geographical proximity, population sizes, or social prestige of the populations. In particular, a_{ii} represents the transmission rate within one population, indicating to what extent the word stays the same between timesteps. The transmission rate to P_0 is not defined because the central population does not learn from other populations by assumption. Note that the transmission rates are the same for all variants regardless of their ages ρ or frequencies. In other words, transmission of words is assumed to be unbiased, and novelty bias or frequency bias (e.g., conformity to the local majority) is absent in this model.

The transmission rates characterize the topological structure of the network. Our model considers arbitrary networks in which words created in P_0 can be reached to all $P_k (1 \leq k \leq n)$.

The definition of transmission rate gives

$$\sum_{j=0}^{n} a_{kj} = 1. \quad (4)$$

Distribution of word age

Defining $\boldsymbol{f}(\rho, t) = \left(f_1(\rho, t) \ldots f_n(\rho, t)\right)^T$ and $\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$, we have the distribution of age frequency in the n peripheral populations:

$$\boldsymbol{f}(\rho, t) = \begin{cases} \boldsymbol{0} & (\rho = 0) \\ \boldsymbol{A}^{\rho-1} \begin{pmatrix} a_{10} \\ \vdots \\ a_{n0} \end{pmatrix} & (1 \leq \rho \leq t) \\ \boldsymbol{A}^t \boldsymbol{f}(\rho - t, 0) & (t < \rho) \end{cases} \quad (5)$$

where $\boldsymbol{f}(\rho, 0)$ is the initial distribution of word ages, defined for any $\rho > 0$, in n peripheral populations. Let $\boldsymbol{r}(t) = \left(r_1(t) \cdots r_n(t)\right)^T$ represent the vector whose kth element corresponds to the mean word ages in P_k. We have

$$\boldsymbol{r}(t) = \sum_{\rho=0}^{\infty} \rho \boldsymbol{f}(\rho, t) = \boldsymbol{A}^t \left(\boldsymbol{r}(0) - (\boldsymbol{E} - \boldsymbol{A})^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \right) + (\boldsymbol{E} - \boldsymbol{A})^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \quad (6)$$

For the equilibrium state, we have

$$\boldsymbol{r}(\infty) = (\boldsymbol{E} - \boldsymbol{A})^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad (7)$$

where $\boldsymbol{E}$ represents *n*-dimensional identity matrix. For the derivation of (5), (6), and (7), see **Appendix A**.

A measure of linguistic diversity within population is the standard deviation (SD) of word age. Let $\sigma_k(t)$ denote the standard deviation of word age in population P_k at timestep t. We can also calculate the equilibrium standard deviation of word age within population, $\boldsymbol{\sigma}(\infty) = \left(\sigma_1(\infty) \cdots \sigma_n(\infty)\right)^T$ (see **Appendix A**). In addition, we introduce another diversity measure $H_k(t)$ as follows:

$$H_k(t) = 1 - \sum_{\rho=0}^{\infty} f_k(\rho, t)^2. \quad (8)$$

Here, $H_k(t)$ is the heterozygosity of the words in P_k, or the probability that two randomly sampled variants are not identical, which is analogous to the genetic heterozygosity at a single locus. Whereas $\sigma_k(t)$ is used to deduce how words are quantitatively diverse in a population, $H_k(t)$ only considers the identity of variants. In computing the infinite series in (8), we take summation over ρ from zero to a sufficiently large integer called the "cut-off value". We choose this value so that $f_k(\rho, t)$ is negligibly small for every ρ that is larger than the cut-off.

Analytically tractable cases

To provide a further mathematical analysis, we focus on the case when the transmission rate from one population to another is either a or 0; that is, $a_{ij} = a > 0$ (const.) for some combinations of transmitting and receiving populations and $a_{ij} = 0$ for others. Note that this is assumed throughout the rest of this paper unless otherwise stipulated. Suppose further that populations $P_0, P_1, \dots,$ and P_n are aligned in this order to form a one-dimensional chain, so that the central population is situated on an edge (Figure 1a). We consider the following two cases.

First, when transmission is unidirectional from P_j to $P_{j+1} (0 \le j \le n-1)$ so that words diffuse toward populations farther from the central population, the transmission matrix is given by

$$\boldsymbol{A} = \begin{pmatrix} 1-a & & & \\ a & 1-a & & \\ & \ddots & \ddots & \\ & & a & 1-a \end{pmatrix}, \quad (9)$$

where zero elements are omitted for the sake of notational simplicity. Based on the matrix, we obtain

$$r_k(\infty) = \frac{k}{a}, \quad (10a)$$

and

$$\sigma_k(\infty) = \sqrt{\frac{k}{a}\left(\frac{1}{a} - 1\right)}. \quad (10b)$$

Detailed derivation of (10a) and (10b) is given in **Appendix B**. These expressions show that words are on average older and more diverse in populations that are located further from the central population (Figure 1b, c).

Secondly, we consider the case of bidirectional diffusion. Words are transmitted bidirectionally between adjacent populations with the exception of P_0, to which transmission from other populations does not occur. Transmission matrix is written as

$$\boldsymbol{A} = \begin{pmatrix} 1-2a & a & & \\ a & \ddots & \ddots & \\ & \ddots & 1-2a & a \\ & & a & 1-a \end{pmatrix}. \quad (11)$$

The mean and standard deviation of word ages at equilibrium are calculated as follows:

$$r_k(\infty) = \frac{k}{2a}(2n - k + 1), \quad (12a)$$

and

$$\sigma_k(\infty) = \sqrt{\frac{k}{6a^2}(2n - k + 1)(2n^2 - 2nk + k^2 + 2n - k + 1) - \frac{k}{2a}(2n - k + 1)}. \quad (12b)$$

Again, see **Appendix B** for more detailed derivation. As in the unidirectional model, $r_k(\infty)$ and $\sigma_k(\infty)$ increase with k (Figure 1b). In addition, they also increase with n, which means that word age depends not only on the distance from the central population, but on the length of the population chain. In the bidirectional model, old variants can diffuse from remote populations to ones that are nearer to the central population, so it is natural that words become on average older when the chain of populations is longer.

Figure 1d depicts the equilibrium distribution of word age within the same population. In both unidirectional and bidirectional transmission, there is a peak of word age in every population. While old variants are extremely rare in the case of unidirectional transmission, they are maintained at a relatively high frequency with bidirectional transmission. This is because with bidirectional transmission old variants come in not only from more central, but also from more peripheral neighbours, and hence are maintained in peripheries for a long time. Also, heterozygosity of variants increases with k for both unidirectional and bidirectional transmission (Figure 1e) and is larger in bidirectional transmission. It is therefore suggested that the amount of polymorphism is larger under the condition of bidirectional transmission, which is because old words are maintained in the populations.

3. Numerical analysis on schematic networks

For less simplified cases, we can numerically obtain the mean and standard deviation of word age at equilibrium. In this section, we describe two such examples.

Two-dimensional diffusion

We now allow bidirectional diffusion for both horizontal and vertical directions in the $m \times l$ rectangle of populations. Unlike in the previous one-dimensional model, the central population is not necessarily situated at a corner or edge of the rectangle.

Figure 2b, c shows 3D plots of the mean, $r_k(\infty)$, and standard deviation, $v_k(\infty)$, of word age at equilibrium over the $m \times l$ rectangle of populations. As anticipated, the mean and standard deviation are smallest at the central population and increase with increasing distance from the centre in all cases. Beyond this overall similarity, however, the precise pattern of increase depends on the position of the central population and the shape of the rectangle. First, consider the case when $m = l$ holds, so that populations form a square, and P_0 is at the centre of the square, $((m+1)/2,\ (m+1)/2)$, assuming m and l as an odd number. In this case, the changes in the mean and standard deviation of word age are symmetric in all four directions (see the top row of Figure 2b, c). Second, in contrast, when P_0 is placed closer to one of the four sides of the square, the mean and standard deviation of word age increases less rapidly toward that side than toward the opposite side (see the middle row in Figure 2b, c), as a result of which the mean and standard deviation of word age exhibit asymmetric contour lines. Third, when either the horizontal or vertical side of the rectangle is longer than the other ($m \neq l$), the changes in the mean and standard deviation of word age are faster along the longer side than along the shorter side (see the bottom row in Figure 2b, c).

Effect of a natural or cultural barrier

We have so far assumed a constant rate of transmission among populations, but the degree of their interdependence is changeable depending on geographical factors. Here we consider the presence of a barrier which inhibits human interactions and linguistic transmission for some geographical or cultural reason (e.g., mountains, deserted area, culturally conservative population, prohibition of movement, etc.).

First, as the baseline model, we adopt the one-dimensional bidirectional diffusion of words as discussed earlier. Two consecutive populations P_h and P_{h+1} are separated by a barrier (e.g., river, mountain, etc.), and we denote the transmission rate between the two populations by b. Assuming $0 < b < a$, cultural transmission is weaker between these populations than in other pairs of neighbouring populations.

As suggested by Figure 3a, $r_k(\infty)$ becomes larger in the presence of a barrier in populations for which $h < k$ holds, which means that the mean word age at equilibrium becomes older in populations beyond the barrier (from the perspective of the central population). Interestingly, $r_k(\infty)$ is not affected in the near side of the barrier (i.e., $k \leq h$) (Figure 3a), even though variants diffuse in both directions and thus the barrier is expected to have an impact on all populations. As for the diversity estimators, the standard deviation of word age becomes larger in both sides of the barrier (Figure 3b). Conversely, heterozygosity $H_k(\infty)$ increases where $h < k$ and decreases where $k \leq h$ (Figure 3c).

The results are interpreted as follows. Since a barrier inhibits the transmission of novel variants created in P_0 to remote populations, it is straightforward that $r_k(\infty)$ becomes larger in populations farther than P_h. Where $k \leq h$, the matter is more complicated. In fact, the existence of barrier affects the word age of the near populations in two ways. On the one hand, a barrier makes words in remote populations even older, which results in the influx of old variants into the near populations. On the other hand, as transmission from remote populations is partially insulated, the near populations receive relatively larger number of old variants. It seems that these opposite effects are cancelled out, and the mean word age stays unchanged in $P_k (k \leq h)$. This interpretation is consistent with the finding that in the presence of barrier, extremely old variants are maintained at low frequencies in $P_k (k \leq h)$ (Figure 3c), so $\sigma_k(\infty)$ increases between the central population and the barrier (Figure 3b). However, as the number of new words increases significantly, the heterozygosity drops in the near populations (Figure 3d). In conclusion, the presence of a barrier exerts the opposite influences on the two diversity estimators in populations between the central population and the barrier.

Secondly, we consider the barrier based on the two-dimensional diffusion model.

One of the two-dimensionally arranged populations is an isolated barrier (P_h), the transmission to/from which occurs at the rate $b(< a)$. Figure 4 indicates that $r_k(\infty)$ becomes smaller between P_0 and P_h, and larger on the other side of the barrier. Unlike the one-dimensional model, P_h marks a peak of $r_k(\infty)$ and $\sigma_k(\infty)$ for small values of b. As populations are aligned in a two-dimensional shape, words can be transmitted via multiple pathways, and as a consequence diffusion can detour the barrier. For this reason, the existence of barrier has less impact on remote populations than in the one-dimensional case.

4. Numerical analysis on realistic networks

We have so far analysed populations arranged in a chain or lattice. In this section, we consider an extended model that incorporates a more realistic network of populations reflecting the geography and demography of Japanese prefectures to examine the case of concentric dialect distributions centred at Kyoto.

Adaptation of the model to the network of prefectures

To reflect the geographical features of Japanese Archipelago, we regard 46 Japanese prefectures except Okinawa as $P_0, \ldots, P_{45}$ of our model. We exclude Okinawa because this prefecture is geographically and was historically isolated from other parts of Japan. On the network of 46 prefectures, we regard Kyoto prefecture as the centre, P_0, from which every linguistic variant derives.

One typical method for modelling the linguistic diffusion on a network of cities is to utilize the *gravity model*, in which the extent of interaction between two cities is assumed to be proportional to the product of their population sizes and the inversed square of the distance in between [13]. However, since this assumption would always give $a_{ii} = 1$ in our model, we instead follow Burridge [17] to incorporate a modified gravity model, or the *interaction density*, φ_{ij}, which is defined as the time people in P_i spend interacting with speakers in P_j. Here, we adapt his equation (2.3) to our model:

$$\varphi_{ij} = \frac{\pi_i \pi_j}{1 + d_{ij}^2/\gamma^2}, \quad (13)$$

where d_{ij} is the distance between P_i and P_j, and π_i denotes the population size of P_i. Constant γ represents the half-decay distance, that is, the distance at which the interaction density is halved relative to that within the same node, where words tend to spread farther when γ is larger. As with Newton's law of gravity, (13) has a long algebraic tail. Geolinguistics has been adopting a variety of measures for geographical distance, such as Euclidean distance [9], great-circle distance (shortest distance on a sphere surface) [8], travel distance [8, 9], and railway distance [20]. Here, we use the great-circle distance between prefectural government offices (buildings), summarized in [21]. We use the population data of each prefecture surveyed in 2018 [22]. Although the population size was different during the time of dialect diffusion, as we shall discuss later (see (15)), our model depends on the ratio of population sizes, so the modern population size seems to be a good proxy assuming that all the populations grew at a uniform rate.

In this framework, however, we simultaneously observe the effects of both the population sizes of the prefectures and the topological structure of the network. To discuss these two effects separately, we additionally examine a population-independent model, in which case the interaction density, given by

$$\varphi_{ij} = \frac{1}{1 + d_{ij}^2/\gamma^2}, \quad (14)$$

is uniquely dependent on the distances of prefectures, irrespective of their population sizes.

Since a_{ij} represents the probability that a person in P_i learns a word from a role model in P_j, it is natural that a_{ij} be given as the interaction density between P_i and P_j divided by the total amount of interaction (s)he experiences. Therefore, we have

$$a_{ij} = \frac{\varphi_{ij}}{\sum_{l=0}^{45} \varphi_{il}}. \quad (15)$$

In the case of population-dependent interaction given by (13), transmission rate a_{ij} is proportional to the population size of P_j (the transmitting prefecture), while it decreases as the population size of P_i (the receiving prefecture) increases. Based on the model, we calculate numerically the mean word age in each prefecture at equilibrium.

Word age at equilibrium

Using equation (13), we examine the case in which the interaction density and the transmission rate are proportional to the product of the population sizes of the prefectures. Figure 5 suggests that words become on average older with the distance from Kyoto, but prefectures to the west of Kyoto tend to contain newer words compared to the eastern prefectures located at the same distance from Kyoto. The distribution, therefore, is not symmetric, and words diffuse westward more rapidly than eastward. In this example, in which γ is set to 10 km, there are two separated regions in the east side having similar word ages of between 550 and 600, namely, around Tokyo and the northern Tohoku area. Qualitatively similar results were obtained for $\gamma = 20$ and 50 km (data not shown).

On the other hand, considering the population-independent transmission rate represented by (14), words become older almost symmetrically in both sides of Kyoto (Figure 6). We also find a decelerating rate of change in mean word age with distance from the centre, indicating that the same word occupies more extensive areas as it goes farther away from the centre. These two features are in concordance with the case of one-dimensional bidirectional diffusion (Figure 1b).

Comparison of figures 5 and 6 leads to the conclusion that the asymmetric distribution of mean word age in Figure 5 is attributable to the heterogeneity in population size. Since transmission rate a_{ij} is proportional to the population of P_j (i.e., the prefecture to which the role model belongs) in (13), the population-dependent model assumes that people in a highly populated community are likely to learn a word within their own community, delaying the entry of newer words, and as a consequence play a role as a conservative "barrier." Since the Tokyo area harbours an especially large population, the relatively slow diffusion of words into East Japan as predicted in Figure 5 may well be interpreted as resulting from hindered diffusion of novel variants from Kyoto into this region.

5. Discussion

To understand the emergence of geographic patterns in linguistic variants and the underlying process of diffusion in the presence of a centre-periphery social structure, we have developed a model of linguistic diffusion between populations distributed over space. Using the model, we have quantified the expected frequency distribution of variants, mean and standard deviation of word ages, and amount of linguistic variation in each population. Implications from our main analysis are summarized as follows. First, the mean word age of a given population is expected to increase with its distance from the central population. This indicates that the emergence of a concentric word distribution such as documented in Japan [10, 18] and France [3, 11] can at least partially be explicable by the presence of a centre-periphery structure. Secondly, difference in the mean word ages between two adjacent populations tends to be highest near the central population and decreases with the distance from it. This finding is in accord with the observed geographic distribution of swear words in Japanese [18], which further supports the hypothesis that the centre-periphery structure underlies the concentric distribution of word variants.

Thirdly, even in a population with a relatively low mean word age, old variants are expected to be maintained at a considerable frequency (Figure 1d right). While this result implies the strong persistence of old words, we have been unable to find empirical support for this prediction. The only exception to the second and third rules is found in one-dimensional unidirectional diffusion, which approximates the case when populations are hierarchically organized in a way that populations closer to the centre bear higher social status. In this case, word age increases lineally in proportion to the distance from the central population (Figure 1b) and old variants are almost eliminated at equilibrium (Figure 1d). Therefore, the fact that our third prediction does not receive empirical support may mean that the linguistic diffusion at the time when the concentric distribution was created was only partially bidirectional being biased in favour of the direction from the centre to the periphery.

Lizana et al.'s remark [18] on the distribution of swear words is two-fold: swear words are arranged in a concentric shape (i.e., concentric distribution of words), and the spatial interval between adjacent words increases with the distance from Kyoto (i.e., extended interval of words). In our analysis, the former is seen most clearly in two-dimensional diffusion with the central population situated at the centre of a lattice. The average word age increases with the distance from the central population, reflecting the fact that newly invented words are prevalent near Kyoto and older words are gradually pushed outward. Strictly speaking, however, the result contradicts the alleged observation in the way that our model produces a mixture of several variants in the periphery, instead of an array of gradually older variants which distinctively dominate each area. The latter feature of the observed distribution is manifested as the decreasing rate of difference in the mean word ages between neighbouring populations. Intuitively, our model predicts that moving outward from the central population, one will initially encounter a drastic linguistic change within a short distance, but the change will be decelerated as moving farther away from the centre. It should be noted, however, that since our model permits the coexistence of multiple variants in the same population, discrete "boundaries" or "intervals" of words cannot be defined. In this regard, therefore, it is difficult to compare Lizana et al.'s and our results quantitatively.

The outcome of our extended model that assumes population-dependent cultural diffusion is qualitatively different from the outcome of the main model in that the former predicts a geographically asymmetric word distribution. In particular, our analysis on a network of populations incorporating geographic and demographic characteristics of the present-day Japan does not replicate the observed pattern, where the same dialect variants are used in the east and west ends of the country [6, 15]. Since those models that successfully replicate the observed pattern (i.e., ours and Lizana et al.'s [18]) assume population-independent cultural diffusion, it is suggested that the historical word diffusion in Japan may have occurred in a population-independent manner.

We do not explicitly incorporate the novelty-biased transmission as assumed in Lizana et al.’s simulation work [18], because it violates the independence of transmission of each word and makes it impossible to track the dynamics with a simple matrix notation. Although analytical approach is impossible, we numerically calculate the word age while integrating novelty bias in **Appendix D**, which shows words become on average newer with novelty bias and that mean word age increases almost linearly with the distance from central population when the bias is strong. Thus, in terms of novelty bias, our result is not in line with Lizana et al.’s work, which claims concentric distribution is not reproduced without novelty bias and that the interval between two words become larger with the distance from Kyoto. Note that although Lizana et al. [18] claims the feature of extended interval of swear word variants in remote areas, we could not find other clear examples of this characteristics [1]. It may be that the increase in spatial interval between adjacent words is usually so subtle that can hardly be detected empirically.

Our prediction from the two-dimensional diffusion model that old words persist in populations around the centre is at odds with the empirical observation that dialect words documented in the east and west peripheries of Japan, which were supposedly created in the centre in the past, are no longer found in Kyoto [1]. The observation is more similar to the outcome of our one-dimensional unidirectional diffusion model. This may indicate that the linguistic diffusion from Kyoto was not strictly bidirectional as our bidirectional model assumes and was at least partially unidirectional. As mentioned above, people's preference for novelty or prestige may have an effect similar to unidirectional diffusion.

The analysis of rectangular-shaped population implies that words tend to be older in the longer side of the land. Since the Japanese Archipelago is long and narrow from southwest to northeast, the shape of word distribution may be elliptic rather than circular. From Kyoto, the distance to seashore is much shorter in the south and north than east and west directions. Thus, our model predicts that word variants may be older in the west and east of Kyoto and relatively new in northern and southern part. Testing this prediction would be interesting if relevant data are available.

Our two measures of linguistic diversity, σ_k and H_k, partially contradict with each other, in particular in the analysis of barrier. When transmission is insulated in one-dimensional bidirectional diffusion, the standard deviation of word age, σ_k increases in populations located between the central population and the barrier, while the heterozygosity, H_k, decreases in the same populations. Therefore, the effect of barrier on these populations seems equivocal (i.e., linguistic diversity is indicated to either increase or decrease, depending on how it is measured). Which of the two measures is more appropriate depends on the nature of the linguistic trait of interest. If the relevant trait is a quantitative trait that is subject to only gradual and unidirectional changes of the trait value (e.g., different accents of the same word), the standard deviation of age would represent the polymorphism, because the time of creation directly corresponds to the amount of difference. On the other hand, if the trait of interest is a qualitative trait subject to discrete changes (e.g., synonyms with different etymology), the time of creation does not provide information of the variants, and thus the heterozygosity is the better measure of the polymorphism.

We discuss possible applications of the present study. Although we have focused on the transmission of dialects as a test case, our model may be applicable to other socially transmitted behaviours or *culture* originating from a culturally influential population. The transmission of human cultures is extensively studied in the discipline of cultural evolution [7, 23, 24], and spatial patterns of cultural traits have been treated using phylogenetic approaches [25-30]. As well as the transmission of dialects, some population plays a greater role in the transmission of culture in general. For example, observed geographic patterns in the prevalence of the "hinoeuma" superstition within the Japanese Archipelago is better explained by considering the presence of a cultural centre, or a single prefecture of prominent cultural influence [31]. Another theoretical study investigated the spread of information in conjunction with the appearance of cultural centre [32]. Our model can be extended beyond the linguistic traits and can treat the spatial pattern of other culturally transmit traits which spread from a single population. For example, archaeological records suggest that stone weapons and burial goods were transmitted from the Eurasian Continent to Japanese mainland via Korean peninsula and Kita-Kyushu areas and eventually diffused to the eastern part of Japan [33]. In this case, these areas can be seen as the cultural centre, from which these archaeological traits derive. It is intriguing to investigate whether our model is consistent with empirical archaeological data.

We discuss limitations of our model and present suggestions for future work. Firstly, we have assumed for mathematical convenience that new words or dialect variants are invented exclusively in the central population and transmitted to other populations without any modification. While the reality is less simple than that, changes in the model outcome caused by relaxing the assumption would be rather predictable. For example, if the central population is not always filled with the latest variant, or if the periphery can also influence the centre, word age will presumably increase because relatively smaller number of novel variants will diffuse toward periphery.

Secondly, our assumption is that only the central population creates new word variants. This is one of the simplest representations of the centre-periphery structure that we consider as a common feature underlying the observed cases of concentric word distribution. While we show that a concentric distribution is indeed predicted under this assumption, whether this holds true when the peripheral populations sometimes invent or modify words is yet to be investigated. It would be more realistic to integrate multiple centres which create new words at different rates depending on their respective prestige and population size. Such an investigation, however, would require a completely new mathematical framework, which is able to keep track of multiple variants created in the same timestep in different populations, and thus is beyond the scope of the present study.

Thirdly, while our analysis on a realistic network reflecting the distance and population size of Japanese prefectures enables a close comparison of expected and observed word distributions, it is still difficult to perform any rigorous quantitative test. Such a test would require estimates of the ages of word forms, which are not available in any linguistic atlases. Nonetheless, our model proves to be useful in inferring the mode of cultural diffusion during the formation of a concentric distribution.

Finally, even though we have considered the word ages in each population, age does not necessarily correspond to the degree of qualitative difference between linguistic variants. To analyse the difference explicitly, we need to model how rapidly words change over generations in the central population. If the latest variant is almost the same as the previous one, spatial variation of mean word age will merely correspond to a slight difference of words among peripheral populations. This is particularly crucial in applying the model to different type of variants beyond lexicons, as different linguistic features are reported to evolve at various rates [34, 35]. To clarify this, future work could incorporate the linguistic features and mutation into our model, and a possible mathematical framework is the 0,1-vector model [36]. As the number of mutation events through the diffusion process can be considered proportional to the word age, our model may be extended to calculate the spatial pattern of the amount of accumulating mutation. In this way, it may be possible to quantify the distribution of linguistic features and calculate the similarity or difference of culture among populations.

References

[1] National Institute for Japanese Language and Linguistics. 1966-1974. Linguistic Atlas of Japan 1-6.

[2] Orton, H., Sanderson, S., and Widdowson, J. 1998. The Linguistic Atlas of England.

[3] Gilliéron, J., and Edmont, E. 1903-1910. Atlas linguistique de la France.

[4] Gray, R., D., Drummond, A., J., and Greenhill, S., J. 2009. Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. Science 323, 479.

[5] Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A., V., Drummond, A., J., Gray, R., D., Suchard, M., A., and Atkinson, Q., D. 2012. Mapping the Origins of the Indo-European Language Family. Science 337, 957-960.

[6] Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., and Pagel, M. 2015. Bantu expansion shows that habitat alters the route and pace of human dispersals. PNAS 112, 13296-13301.

[7] Mesoudi, A. 2011. Cultural Evolution: How Darwinian Theory Can Explain Human Culture & Synthesize the Social Sciences. University of Chicago Press, Chicago.

[8] Jeszenszky, P., Hikosaka, Y., Imamura, S., and Yano, K. 2019. Japanese Lexical Variation Explained by Spatial Contact Patterns. International Journal of Geo-Information 8, 400.

[9] Szmrecsanyi, B. 2012. Geography is overrated. In Dialectological and Folk Dialectological Concepts of Space. Eds: Hansen, S., Schwarz, C., Stoeckle, P., and Streck, T. 215-231.

[10] Yanagita, K. 1927. Kagyuko (1)-(4). Jinruigaku Zasshi 42, 125-135 (162-172, 223-233, 273-284).

[11] Le Roux, P., 1924-1953. Atlas Linguistique de la Basse-Bretagne.

[12] Blanchard, N. 2008. Dialectologie et standarisation linguistique – Centres et Marges économiques et culturels en Basse-Bretagne. Port Acadie, (13-14-15), 45-61.

[13] Trudgill, P. 1974. Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. Language in society 2, 215-246.

[14] Kretzschmar, W. A. Jr., Jusso, I., and Bailey, C. T. 2014. Computer simulation of dialect feature diffusion. Journal of Linguistic Geography 2, 41-57.

[15] Fagyal, Z., Swarup, S., Escobar, A. M, Gasser, L., and Lakkaraju, K. 2010. Centers and peripheries: Network roles in language change. Lingua 120, 2061-2079.

[16] Burridge, J. 2017. Spatial evolution of human dialects. Physical Review X 7, 031008.

[17] Burridge, J. 2018. Unifying models of dialect spread and extinction using surface tension dynamics. Royal Society Open Science 5, 171446.

[18] Lizana, L., Mitarai, N., and Kim, S. 2011. Modeling the spatial dynamics of culture spreading in the presence of cultural strongholds. Physical Review E 83, 066116.

[19] Matsumoto, O. 1993. Zenkoku aho-baka bumpu-kou. Ohta Shuppan.

[20] Inoue, F. 2009. Year of first attestation of standard Japanese forms and gravity centre

by railway distance. Dialectologia et Geolinguistica, 17(1), 118-123.

[21] Geospatial Information Authority of Japan. https://www.gsi.go.jp/KOKUJYOHO/kenchokan.html (Accessed on 20/11/2019)

[22] Statistics Bureau of Japan. 2018. https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200524&tstat=000000090001&cycle=7&year=20180&month=0&tclass1=000001011679&stat_infid=000031807142 (accessed on 4/3/2020)

[23] Cavalli-Sforza, L. L., and Feldman, M. W. 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton.

[24] Boyd, R., and Richerson, P. J. 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.

[25] Mace, R., Holden, C. J., and Shennan, S. 2005. The Evolution of Cultural Diversity: A Phylogenetic Approach. University College London Press.

[26] Nunn, C. L., Mulder, M. B., and Langley, S. 2006. Comparative methods for studying cultural trait evolution: A simulation study. Cross-Cultural Research 40, 2, 177-209.

[27] Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., and Mace, R. 2010. Rise and fall of political complexity in island South-East Asia and the Pacific. Nature 467, 801-804

[28] Towner, M. C., Grote, M. N., Venti, J., and Mulder, M. B. 2012. Cultural macroevolution on neighbor graphs. Vertical and horizontal transmissions among western north American Indian societies. Human Nature 23, 283-305.

[29] Brown, S., Savage, P. E., Ko, A. M. S., Stoneking, M., Ko, Y. C., Loo, J. H., and Trejaut, A., J. 2014. Correlations in the population structure of music, genes and language. Proceedings of the Royal Society B 281, 20132072.

[30] Savage, P. E. and Brown, S. 2014. Mapping music: Cluster analysis of song-type frequencies within and between cultures. Ethnomusicology 133-155.

[31] Tamura, K., and Ihara., Y. 2017. Quantifying cultural macro-evolution: a case study of hinoeuma fertility drop. Evolution and Human Behavior 38, 117-124.

[32] Dybiec, B., Mitarai, N., and Sneppen, K. 2012. Information spreading and development of cultural center. Physical Review E 85, 056116.

[33] Nakamura, D. 2011. The diversity of mortuary practice acceptance at the beginning of the Yayoi Period. In: Matsumoto, N., Bessho, H., Tomii, M. (eds) Coexistence and Cultural Transmission in East Asia. Left Coast Press, Walnut Creek CA, 223–256.

[34] Greenhill, S., J., Wu., C., H., Hua, X., Dunn, M., Levinson, S., C., and Gray, R., D. 2017. Evolutionary dynamics of language systems. PNAS E8822-E8829.

[35] Pagel, M., and Meade, A. 2017. The deep history of the number words. Phil. Trans.

R. Soc. B 373: 20160517.

[36] Fogarty, L., Wakano, J. Y., Feldman, M. W., and Aoki, K. 2017. The driving forces of cultural complexity. Neanderthals, modern humans, and question of population size. Human Nature 28, 39-52.

Appendix A: Distribution of cultural age

We elaborate a more detailed derivation of frequency distribution, mean value, and the standard deviation of word age.

From (2) and (3), when $\rho \geq 2$, we have

$$f_k(\rho, t) = \sum_{j=1}^{n} a_{kj} f_j(\rho - 1, t - 1) \quad (1 \leq k \leq n, \rho \geq 2). \quad \text{(A1)}$$

Defining $\boldsymbol{f}(\rho, t) = \left(f_1(\rho, t) \dots f_n(\rho, t)\right)^T$, $\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$, we have

$$\boldsymbol{f}(\rho, t) = \boldsymbol{A}\boldsymbol{f}(\rho - 1, t - 1) \quad (\rho \geq 2). \quad \text{(A2)}$$

By using (A2), we calculate the age frequency for peripheral populations. Since words with age zero cannot exist in any population other than the central population, we have

$$\boldsymbol{f}(0, t) = \boldsymbol{0}, \quad \text{(A3)}$$

where $\boldsymbol{0}$ is the *n*-dimensional zero vector. For variants which are created after the emergence of central population at $t = 0$, we have

$$\boldsymbol{f}(\rho, t) = \boldsymbol{A}^{\rho-1}\boldsymbol{f}(1, t - \rho + 1) = \boldsymbol{A}^{\rho-1}\begin{pmatrix} a_{10} \\ \vdots \\ a_{n0} \end{pmatrix} \quad (1 \leq \rho \leq t). \quad \text{(A4)}$$

As for the variants which date back before the emergence of central population, we have

$$\boldsymbol{f}(\rho, t) = \boldsymbol{A}^{t}\boldsymbol{f}(\rho - t, 0) \quad (\rho > t), \quad \text{(A5)}$$

where $\boldsymbol{f}(\rho, 0)$ represents the initial word-age distribution in the peripheral populations. Here, (A3), (A4), and (A5) represent the age distribution of words. Using these equations,

$$\boldsymbol{r}(t) = \sum_{\rho=0}^{\infty} \rho \boldsymbol{f}(\rho, t) = \sum_{\rho=1}^{t} \rho \boldsymbol{A}^{\rho-1} \begin{pmatrix} a_{10} \\ \vdots \\ a_{n0} \end{pmatrix} + \sum_{\rho=t+1}^{\infty} \rho \boldsymbol{A}^{t} \boldsymbol{f}(\rho - t, 0)$$

$$= \sum_{\rho=1}^{t} \rho \boldsymbol{A}^{\rho-1} \begin{pmatrix} a_{10} \\ \vdots \\ a_{n0} \end{pmatrix} + \boldsymbol{A}^{t} \boldsymbol{r}(0) + t \boldsymbol{A}^{t} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \quad \text{(A6)}$$

Defining the *n*-dimensional square matrix $\boldsymbol{S}$ by

$$\boldsymbol{S} = \sum_{\rho=1}^{t} \rho \boldsymbol{A}^{\rho-1}, \quad \text{(A7)}$$

we have

$$\boldsymbol{SA} = \sum_{\rho=1}^{t} \rho \boldsymbol{A}^{\rho} = \sum_{\rho=2}^{t+1} (\rho - 1) \boldsymbol{A}^{\rho-1}. \quad \text{(A8)}$$

Subtracting (A8) from (A7),

$$\boldsymbol{S}(\boldsymbol{E} - \boldsymbol{A}) = \boldsymbol{E} + \sum_{\rho=2}^{t} \boldsymbol{A}^{\rho-1} - t\boldsymbol{A}^{t} = \boldsymbol{E} + \boldsymbol{A}(\boldsymbol{E} - \boldsymbol{A}^{t-1})(\boldsymbol{E} - \boldsymbol{A})^{-1} - t\boldsymbol{A}^{t}$$

$$= (\boldsymbol{E} - \boldsymbol{A}^{t})(\boldsymbol{E} - \boldsymbol{A})^{-1} - t\boldsymbol{A}^{t}. \quad \text{(A9)}$$

Hence, we have

$$\boldsymbol{S} = \sum_{\rho=1}^{t} \rho \boldsymbol{A}^{\rho-1} = [(\boldsymbol{E} - \boldsymbol{A}^{t})(\boldsymbol{E} - \boldsymbol{A})^{-1} - t\boldsymbol{A}^{t}](\boldsymbol{E} - \boldsymbol{A})^{-1}. \quad \text{(A10)}$$

Considering (4),

$$(\boldsymbol{E}-\boldsymbol{A})\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}=\begin{pmatrix}1-(a_{11}+\cdots+a_{1n})\\ \vdots\\ 1-(a_{n1}+\cdots+a_{nn})\end{pmatrix}=\begin{pmatrix}a_{10}\\ \vdots\\ a_{n0}\end{pmatrix}, \quad \text{(A11)}$$

which gives

$$(\boldsymbol{E}-\boldsymbol{A})^{-1}\begin{pmatrix}a_{10}\\ \vdots\\ a_{n0}\end{pmatrix}=\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}. \quad \text{(A12)}$$

Substituting (A10) for (A6) and using (A12), the mean word age is given by

$$\boldsymbol{r}(t)=[(\boldsymbol{E}-\boldsymbol{A}^t)(\boldsymbol{E}-\boldsymbol{A})^{-1}-t\boldsymbol{A}^t]\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}+\boldsymbol{A}^t\boldsymbol{r}(0)+t\boldsymbol{A}^t\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$$

$$=\boldsymbol{A}^t\left[\boldsymbol{r}(0)-(\boldsymbol{E}-\boldsymbol{A})^{-1}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}\right]+(\boldsymbol{E}-\boldsymbol{A})^{-1}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}. \quad \text{(A13)}$$

Now we consider the equilibrium state. (A3), (A4), and (A5) give

$$\boldsymbol{f}(\rho,\infty)=\begin{cases}\boldsymbol{0} & \text{if } \rho=0\\ \boldsymbol{A}^{\rho-1}\begin{pmatrix}a_{10}\\ \vdots\\ a_{1n}\end{pmatrix} & \text{otherwise}\end{cases}, \quad \text{(A14)}$$

which in turn gives

$$\boldsymbol{r}(\infty)=\sum_{\rho=0}^{\infty}\rho\boldsymbol{f}(\rho,\infty)=(\boldsymbol{E}-\boldsymbol{A})^{-1}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}, \quad \text{(A15)}$$

$$\boldsymbol{q}(\infty)=\sum_{\rho=0}^{\infty}\rho^2\boldsymbol{f}(\rho,\infty)=(\boldsymbol{E}-\boldsymbol{A})^{-1}(\boldsymbol{E}-\boldsymbol{A})^{-1}(\boldsymbol{E}+\boldsymbol{A})\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}, \quad \text{(A16)}$$

where $\boldsymbol{q}(t)=\left(q_1(t)\cdots q_n(t)\right)^T$ denotes the second-order moment, from which the variance of word age, $\boldsymbol{v}(t) = \left(v_1(t) \cdots v_n(t)\right)^T$, is obtained. Standard deviation of word age $\boldsymbol{\sigma}(t) = \left(\sigma_1(t) \cdots \sigma_n(t)\right)^T$ is readily obtained by taking the square root of the variance. Note that every absolute value of eigenvalue of $\boldsymbol{A}$ is strictly smaller than 1 (see **Appendix C**), so that the infinite series in (A15) and (A16) converge.

Appendix B: Derivation of the formulae in one-dimensional diffusion

We derive the formulae to compute the mean and standard deviation of cultural age in one-dimensional unidirectional diffusion (i.e., (9) for unidirectional, (11) for bidirectional diffusion).

As for the unidirectional diffusion model, (9) gives

$$(\boldsymbol{E} - \boldsymbol{A})^{-1} = \frac{1}{a}\begin{pmatrix} 1 & & \\ \vdots & \ddots & \\ 1 & \dots & 1 \end{pmatrix}. \quad \text{(B1)}$$

Thus, (S15) and (S16) give

$$r_k(\infty) = \frac{k}{a}, \quad \text{(B2)}$$

$$q_k(\infty) = \frac{k^2}{a^2} + \frac{k}{a^2} - \frac{k}{a}. \quad \text{(B3)}$$

We can calculate the standard deviation by

$$\sigma_k(\infty) = \sqrt{q_k(\infty) - r_k(\infty)^2} = \sqrt{\frac{k}{a}\left(\frac{1}{a} - 1\right)}. \quad \text{(B4)}$$

In the case of bidirectional diffusion, (11) gives

$$(\boldsymbol{E}-\boldsymbol{A})^{-1}=\frac{1}{a}\begin{pmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & 2 & 2 & \cdots & 2\\ 1 & 2 & 3 & \cdots & 3\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & 2 & 3 & \cdots & n\end{pmatrix}. \quad \text{(B5)}$$

Thus, using (A15) and (A16), we have

$$r_k(\infty)=\frac{k}{2a}(2n-k+1), \quad \text{(B6)}$$

$$q_k(\infty)=\frac{k}{12a^2}\{(k^2-1)(k-4n-2)+4n(n+1)(2n+1)\}$$
$$-\frac{k}{2a}(2n-k+1). \quad \text{(B7)}$$

Hence, we have

$$\sigma_k(\infty)=\sqrt{q_k(\infty)-r_k(\infty)^2}$$
$$=\sqrt{\frac{k}{6a^2}(2n-k+1)(2n^2-2nk+k^2+2n-k+1)-\frac{k}{2a}(2n-k+1)}. \quad \text{(B8)}$$

Appendix C: On the eigenvalues of A

Here we prove the absolute value of every eigenvalue of transmission matrix $\boldsymbol{A}$ is strictly smaller than 1.

Let λ be any eigenvalue of $\boldsymbol{A}$, and an eigenvector corresponding to λ is denoted by $\boldsymbol{p}=(p_1 \ \cdots \ p_n)^T \neq \boldsymbol{0}$. We define p_k as the element of $\boldsymbol{p}$, such that $|p_k|=\max_j |p_j|$. As $\lambda\boldsymbol{p}=\boldsymbol{A}\boldsymbol{p}$, we have

$$\lambda p_k=\sum_{j=1}^{n} a_{kj}p_j. \quad \text{(C1)}$$

Therefore,

$$|\lambda||p_k| \leq \sum_{j=1}^{n} a_{kj}|p_j| \leq \sum_{j=1}^{n} a_{kj}|p_k| = (1 - a_{k0})|p_k| \leq |p_k|. \quad \text{(C2)}$$

As $|p_k| > 0$,we have $|\lambda| \leq 1$.

Now we will disprove the case $|\lambda| = 1$. Assuming $|\lambda| = 1$, (C2) gives

$$a_{k0} = 0, \text{and } a_{kj}\left(|p_k| - |p_j|\right) = 0 \text{ for any } j \ (1 \leq j \leq n). \quad \text{(C3)}$$

There exists an array of integers $C_0, \dots, C_l$, such that

$$C_0 = 0, C_l = k, \text{and } a_{C_iC_{i-1}} > 0. \quad \text{(C4)}$$

Substituting c_{l-1} into j in the second expression of (C3), we have $|p_{c_{l-1}}| = |p_{c_l}| = \max_j |p_j|$. This in turn gives $a_{c_{l-1}0} = 0$ and $|p_{c_{l-2}}| = \max_j |p_j|$ with the same discussion as (C2). Similarly, we can conclude

$$a_{c_l0} = a_{c_{l-1}0} = \cdots = a_{c_10} = 0 \quad \text{(C5)}$$

by mathematical induction. However, $a_{c_10} = 0$ contradicts the expressions (C4) for $i = 1$. Therefore, $|\lambda| < 1$.

Appendix D: Effect of novelty bias

To examine the effect of novelty bias on the distribution of word age, we extend the assumption concerning the learning process of individuals in peripheral populations. In choosing a role model, a learner in P_i $(1 \leq i \leq n)$ first chooses $P_j (0 \leq j \leq n)$ with probability a_{ij}, which stays the same as the main part of this chapter. After choosing a population to which their potential role model belongs, the learner acquires a word with a probability proportional to the word's attractiveness. Here, attractiveness of word aged ρ is given by

$$u(\rho) = \begin{cases} (1-s)^{\rho} & (\rho < \rho_{max}) \\ 0 & (\rho \geq \rho_{max}) \end{cases}, \quad \text{(D1)}$$

where $0 < s < 1$ signifies the strength of novelty bias and ρ_{max} denotes the cutoff value of word age assigned for the sake of numerical analysis. Note that word age ρ in (D1) is measured at the previous timestep. Given an individual at time t learns a word from P_j, the probability that (s)he acquires a word aged ρ is given by

$$\frac{u(\rho-1)f_j(\rho-1,t-1)}{\sum_{\rho'=1}^{\rho_{max}} u(\rho'-1)f_j(\rho'-1,t-1)}. \quad \text{(D2)}$$

It should be noted that one has to learn a word aged $\rho-1$ at the previous timestep, to acquire a word with age ρ. Thus, the recursive formulae as regard to $f_k(\rho,t)$ (cf., equation (3)) is given by

$$f_k(\rho,t)=\sum_{j=0}^{m} a_{kj}\frac{u(\rho-1)f_j(\rho-1,t-1)}{\sum_{\rho'=1}^{\rho_{max}} u(\rho'-1)f_j(\rho'-1,t-1)}. \quad \text{(D3)}$$

In this case, we cannot analytically obtain the equilibrium distribution of word age, so we numerically obtained the equilibrium distribution by consecutively computing (D3) for sufficiently long timesteps.

We use one-dimensional bidirectional diffusion (see equations (11) and (12)) as the baseline of the analysis and integrate novelty bias. Figure 7 depicts the case of $n=9, a=0.1, \rho_{max}=5000$ for four different values of s, strength of novelty bias. The figure shows that the frequency distribution of word age displays a steeper peak with stronger preference for novel words (see Figure 7a, b), and words become on average newer (Figure 7c). To test the linear relationship between computed mean word age $r_k(\infty)$ and distance from central population k, we calculate the correlation coefficient of $r_k(\infty)$ and k for the four values of s depicted in Figure 7c. The correlation coefficient is 0.975 for $s=0.0001$, 0.984 for $s=0.001$, 0.990 for $s=0.01$, and 0.993 for $s=0.1$. It is therefore suggested that mean word age increases more and more linearly with the distance from central population, as novelty bias becomes stronger.

Figure legends

Figure 1

(a) A schematic representation of cultural diffusion in a one-dimensional chain of populations $(n = 4)$. The arrows indicate the presence and direction of word transmission from one population to an adjacent population in the case of (above) unidirectional and (below) bidirectional transmission. Note that recursive arrows (loops) are omitted in this panel. (b) The mean word age at equilibrium in a chain of populations. The horizontal and vertical axes represent the distance from central population and the mean word age in each population, respectively. The blue and orange lines represent the unidirectional and bidirectional transmission models, respectively. Parameter values: $a = 0.1$, $n = 9$ (total number of populations: 10). (c) The standard deviation of word age at equilibrium in a chain of populations. The horizontal and vertical axes represent the distance from central population and the standard deviation of word age in each population, respectively, for (left) unidirectional and (right) bidirectional transmission. Parameter values: $a = 0.1$, $n = 9$ (total number of populations: 10). (d) Frequency distribution of word age ρ at equilibrium in each population. The blue, orange, and green lines represent the populations P_1, P_5, and P_9, respectively, for (left) unidirectional and (right) bidirectional transmission. Parameter values: $a = 0.1$, $n = 9$ (total number of populations: 10). (e) Heterozygosity of words in each population with unidirectional (left) and bidirectional transmission (right). Parameter values: $a = 0.1, n = 9$ (Total number of populations: 10), and cut-off value $\rho = 1000 \text{ and } 10000$ for unidirectional and bidirectional transmission, respectively.

Figure 2

(a) A schematic representation of the two-dimensional diffusion model. Bidirectional cultural diffusion occurs between horizontally or vertically neighbouring populations. In this example, population at (4, 3) plays the role as the central population (denoted by P_0). Parameter values: $m = 5$, $l = 4$. (b) Mean word age at equilibrium in the two-dimensional diffusion model. (Top) $m = l = 15$ and P_0 is located at (8, 8). (Middle) $m = l = 15$ and P_0 is assigned to a population not at the center of the square, specifically, (5, 8). (Bottom) $m = 15$ and $l = 5$ and P_0 is at (8, 3). (c) Standard deviation of word age at equilibrium in the two-dimensional diffusion model. The shape of the rectangle and the location of the central population are the same as in (b).

Figure 3

Bidirectional transmission in a chain of population with a barrier. The barrier divides populations P_5 and P_6 $(h = 5)$. Parameter values: $n = 9$, $a = 0.1$. (a) The mean word age at equilibrium when (blue) $b = 0.1$ (i.e., no barrier), (orange) $b = 0.05$, and (gray) $b = 0.01$. The three curves are overlapped when $k \leq 5$. (b) The standard deviation in word age at equilibrium when (blue) $b = 0.1$ (i.e., no barrier), (orange) $b = 0.05$, and (gray) $b = 0.01$. (c) The heterozygosity of variants in each population with cut-off value $\rho = 10000$ when (blue) $b = 0.1$ (i.e., no barrier) and (orange) $b = 0.01$. For the sake of presentation, the vertical axis shows $\log_{10}\left(1 - H_k(\infty)\right)$, which decreases as the heterozygosity increases. (d) Equilibrium frequency distribution of word ages in (left) P_5 and (right) P_6 when (blue) $b = 0.1$ (i.e., no barrier) and (orange) $b = 0.01$.

Figure 4

Effect of a barrier in the two-dimensional diffusion model. Populations form a 15×15 square, with the central population being at (8, 8) and a barrier at (11, 8). (a) The mean word age and (b) the standard deviation in word age. Parameter value: $b = 0.001$.

Figure 5

Population-dependent diffusion on a network of Japanese prefectures. (Above) Regions of Japan. Kyoto is coloured red, and Okinawa is not shown on this map. (Bottom-left) Rank order of the population size in each prefecture, the least and most populated prefectures being 1 and 46, respectively. (Bottom-right) Mean word age at equilibrium in each population on the basis of equation (13). Parameter value: $\gamma = 10$ km.

Figure 6

Population-independent diffusion on a network of Japanese prefectures, calculated by (14). (a) (Left) Equilibrium mean word age in each prefecture. (Right) Equilibrium mean word age as a function of distance from Kyoto. (b) (Left) Equilibrium standard deviation of word age in each prefecture. (Right) Equilibrium standard deviation of word age as a function of distance from Kyoto. Parameter value: $\gamma = 10$ km.

Figure 7

Effect of novelty bias on the word age in one-dimensional bidirectional diffusion with parameter values $n = 9, a = 0.1, and\ \rho_{max} = 5000$. (a) distribution of word age in $P_1, P_5, and\ P_9$ when $s = 0.1$. (b) distribution of word age in $P_1, P_5, and\ P_9$ when $s = 0.01$ (c) mean word age in all the peripheral populations for four values of s plotted against the distance from central population.

Figure 1

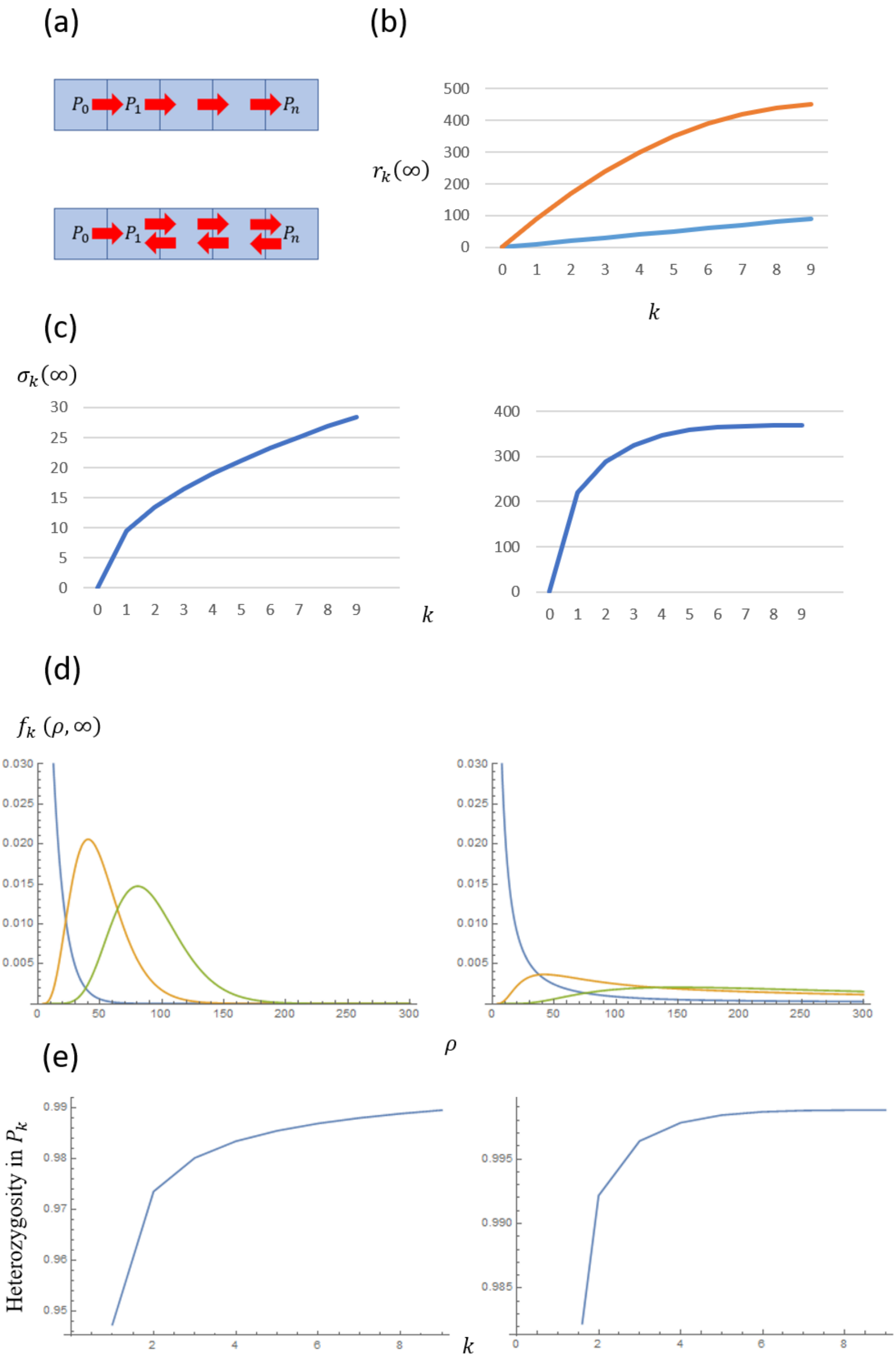
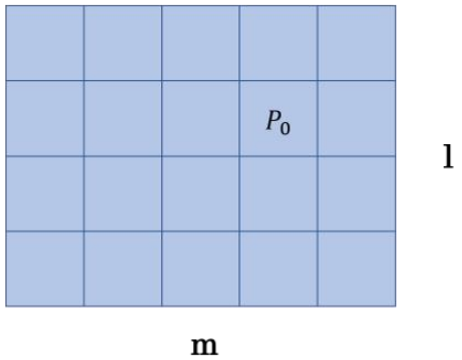
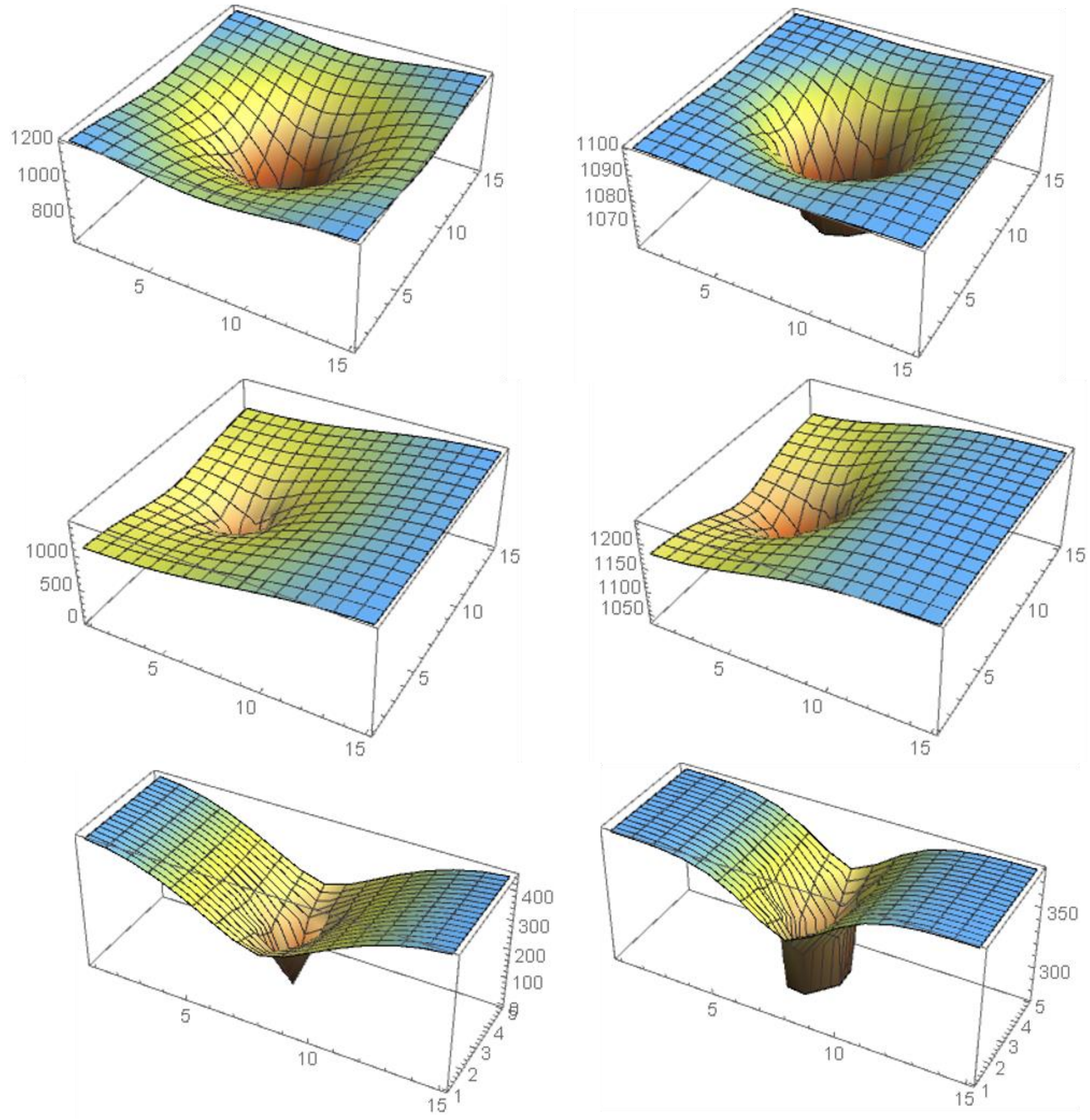

(a)
P_0
P_1
P_n
P_0
P_1
P_n
(b)
$r_k(\infty)$
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9
k
(c)
$\sigma_k(\infty)$
30
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9
k
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9
(d)
$f_k\ (\rho, \infty)$
0.030
0.025
0.020
0.015
0.010
0.005
0 50 100 150 200 250 300
0.030
0.025
0.020
0.015
0.010
0.005
0 50 100 150 200 250 300
ρ
(e)
Heterozygosity in P_k
0.99
0.98
0.97
0.96
0.95
2 4 6 8
k
0.995
0.990
0.985
0 2 4 6 8

Figure 2

(a)




(b) Mean Word Age

(c) SD of Word Age

Figure 3

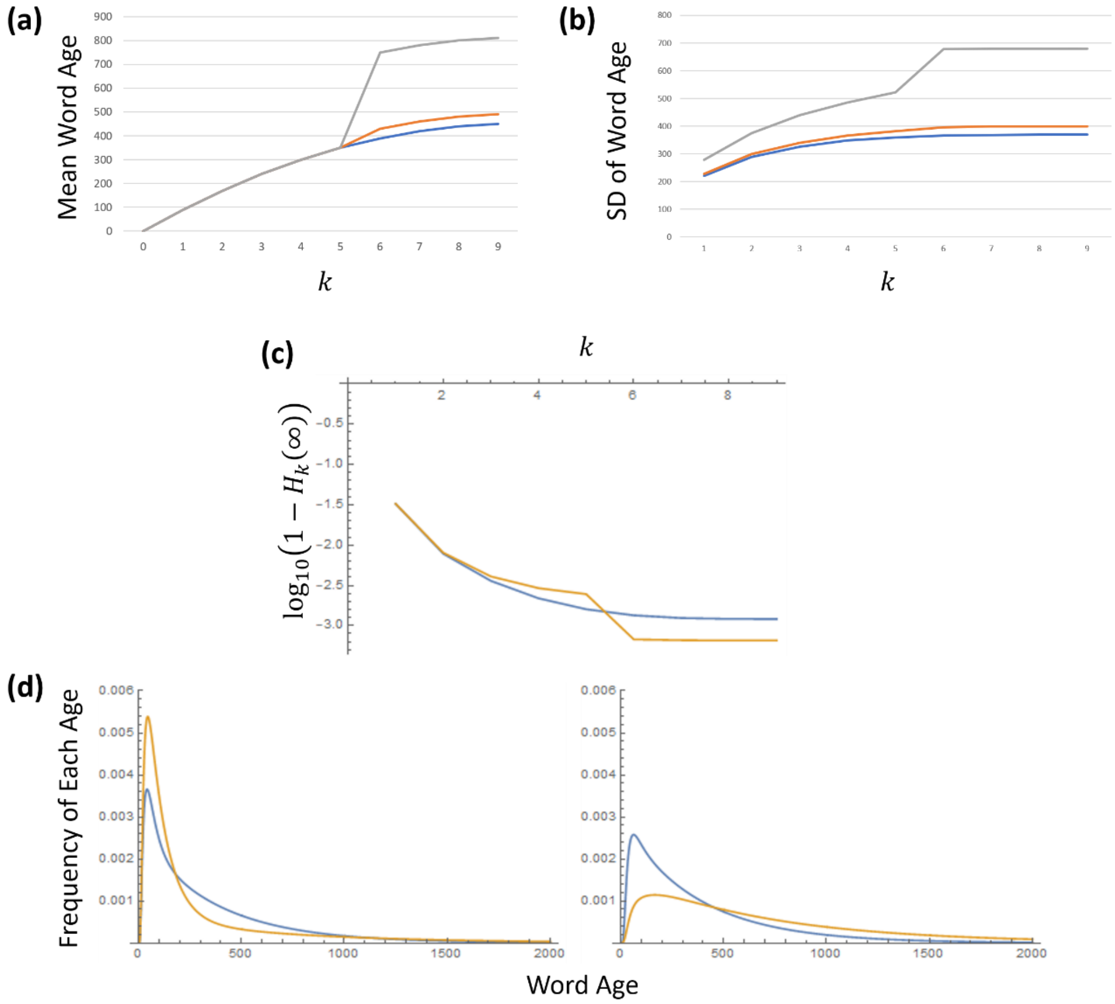
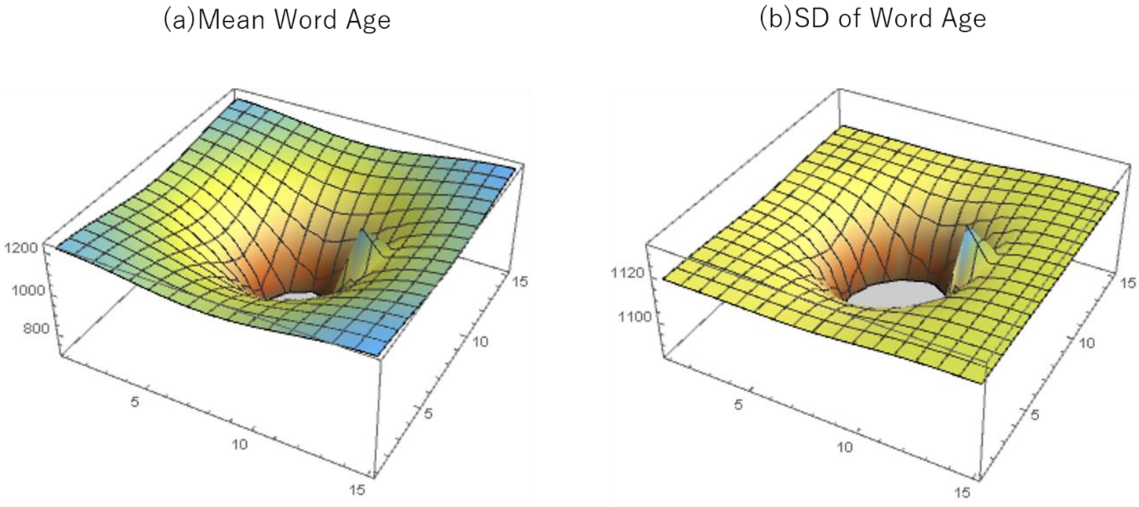

(a)
Mean Word Age
900
800
700
600
500
400
300
200
100
0
0
1
2
3
4
5
6
7
8
9
k
(b)
SD of Word Age
800
700
600
500
400
300
200
100
0
1
2
3
4
5
6
7
8
9
k
(c)
k
2
4
6
8
$\log_{10}(1 - H_k(\infty))$
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
(d)
Frequency of Each Age
0.006
0.005
0.004
0.003
0.002
0.001
0
500
1000
1500
2000
0.006
0.005
0.004
0.003
0.002
0.001
0
500
1000
1500
2000
Word Age

Figure 4



(a)Mean Word Age

(b)SD of Word Age

Figure 5

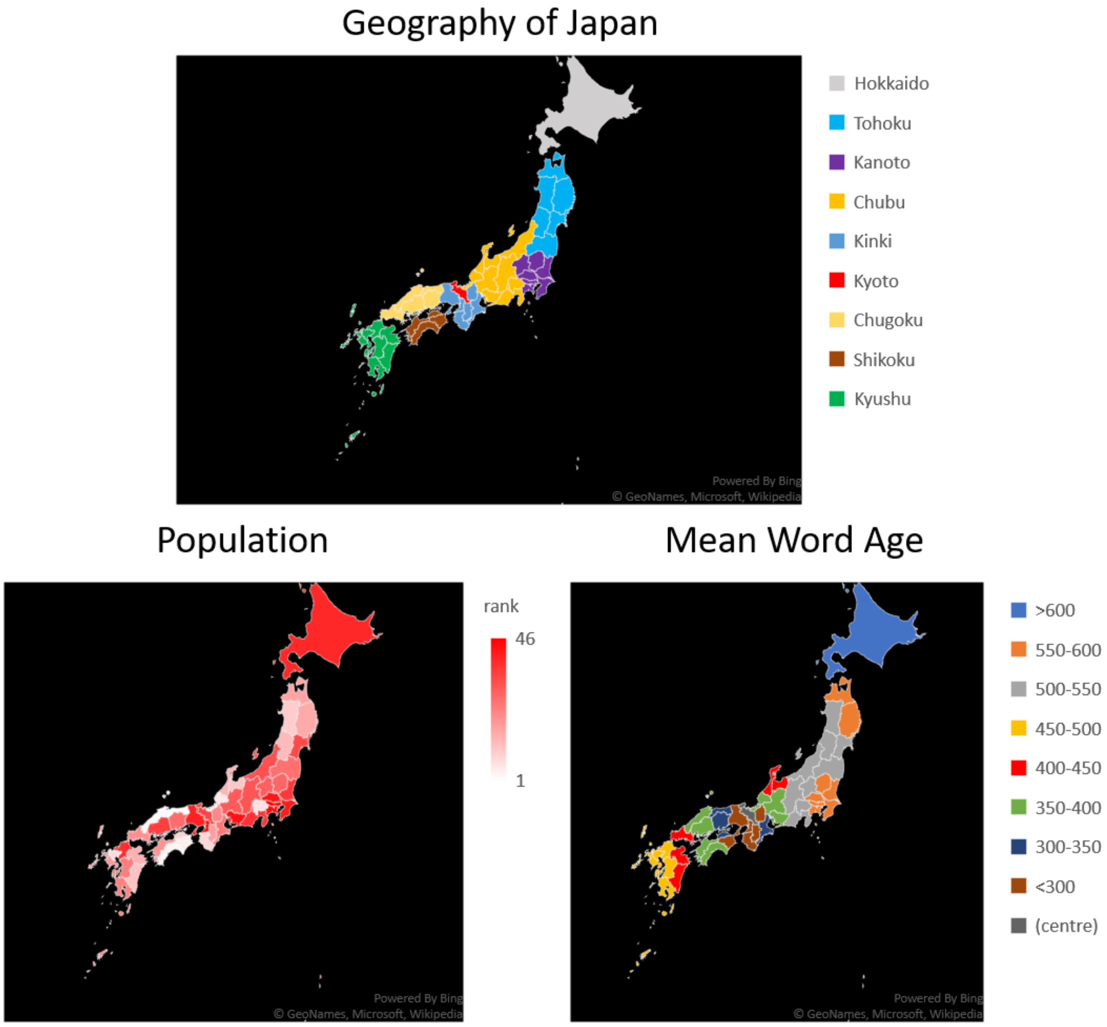

Geography of Japan
Hokkaido
Tohoku
Kanoto
Chubu
Kinki
Kyoto
Chugoku
Shikoku
Kyushu
Powered By Bing
© GeoNames, Microsoft, Wikipedia
Population
rank
46
1
Powered By Bing
© GeoNames, Microsoft, Wikipedia
Mean Word Age
>600
550-600
500-550
450-500
400-450
350-400
300-350
<300
(centre)
Powered By Bing
© GeoNames, Microsoft, Wikipedia

Figure 6

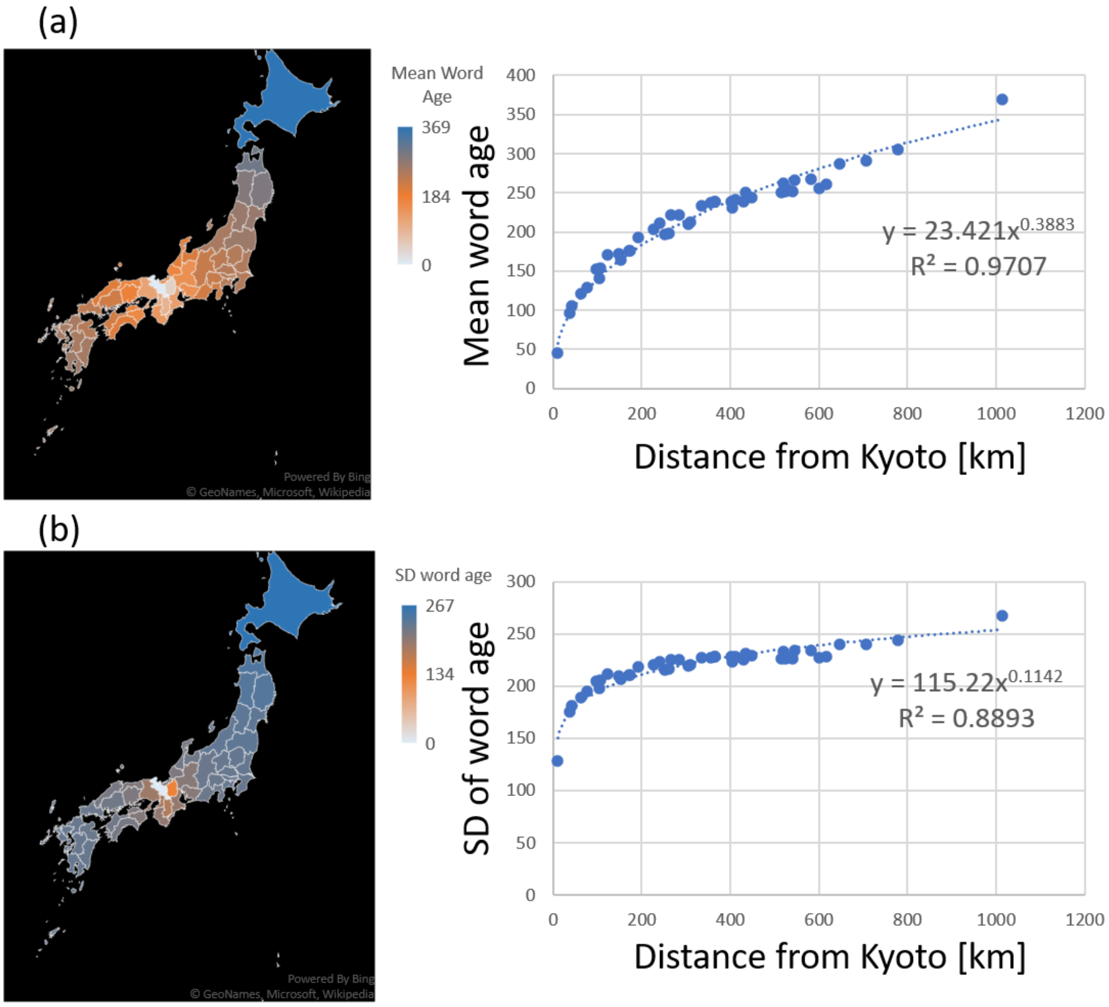
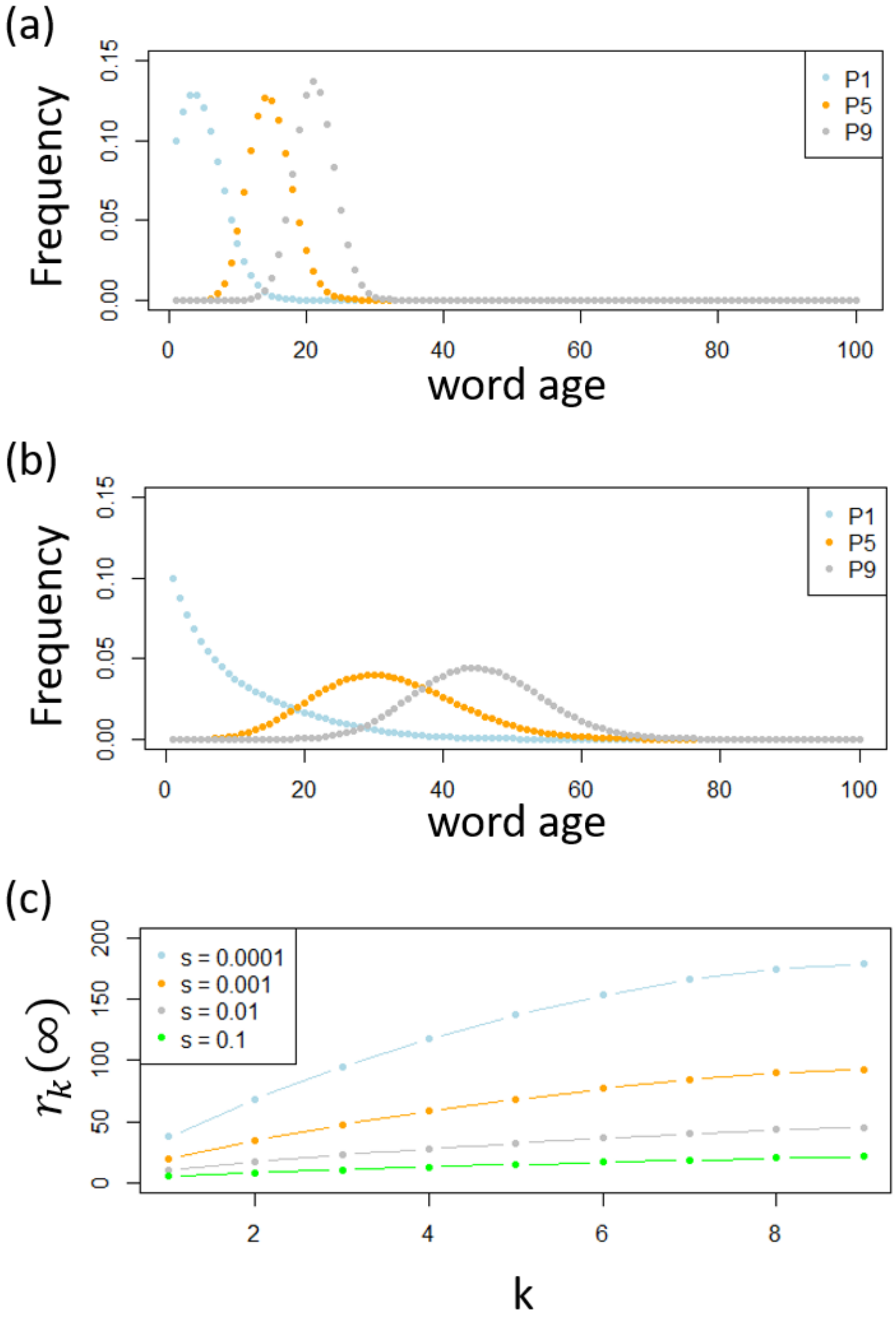

(a)
Mean Word Age
369
184
0
Powered By Bing
© GeoNames, Microsoft, Wikipedia
Mean word age
400
350
300
250
200
150
100
50
0
0
200
400
600
800
1000
1200
Distance from Kyoto [km]
y = 23.421x^0.3883
R² = 0.9707
(b)
SD word age
267
134
0
Powered By Bing
© GeoNames, Microsoft, Wikipedia
SD of word age
300
250
200
150
100
50
0
0
200
400
600
800
1000
1200
Distance from Kyoto [km]
y = 115.22x^0.1142
R² = 0.8893

Figure 7


(a)
Frequency
0.00
0.05
0.10
0.15
0
20
40
60
80
100
word age
P1
P5
P9
(b)
Frequency
0.00
0.05
0.10
0.15
0
20
40
60
80
100
word age
P1
P5
P9
(c)
$r_k(\infty)$
0
50
100
150
200
2
4
6
8
k
s = 0.0001
s = 0.001
s = 0.01
s = 0.1

Chapter 3.

Application of a Markovian ancestral model to the temporal and spatial dynamics of cultural evolution on a population network.

1. Introduction

Cultural evolution, as an analogy to biological evolution, is a process in which compositions in populations of non-genetic traits or "cultures" are modified over time by such factors as innovation and differential transmission between individuals [1,2]. Two different levels of the process can be distinguished: cultural microevolution and cultural macroevolution [3]. While the former concept covers the dynamics of cultural traits within a single population, the latter deals with the transmission of cultural traits among multiple populations along with the divergence, convergence, birth, and death of populations.

Both empirical and theoretical studies on cultural macroevolution frequently place emphasis on the spatial pattern of cultural traits, and a variety of techniques has been conceived for analysis. For example, cultural macroevolution has been studied with phylogenetic techniques [4-9]. In a different vein, spatially explicit diffusion-reaction models were developed to capture the spread of farming [10] and skill-dependent demic expansion [11]. These models use partial differential equations, assuming that human populations are arranged on a one-dimensional number line, to explore the spatial and temporal dynamics of cultural traits. However, real human populations are not arranged in a homogeneous space. Furthermore, certain pairs of human populations may be socially more connected to each other than other pairs, so that they would be considered close to each other in a conceptual space relevant to cultural transmission. To treat the spatial dynamics of cultural trait, therefore, it is more appropriate to use a network model which represents the positions and degree of interdependence of human populations.

As for research into the network models, a huge body of literature in physics has considered the *voter model*, in which individuals represented by vertices of a graph receive social information from their neighbors [12]. In conjunction with its reversed process [13], voter models have been examined to deduce the effect of network topology on the spatial evolution and fixation of cultural traits. As an application, the model has been adopted to investigate the evolution and transmission of languages on a network [14,15].

In chapter 2, we conceived a spatially explicit model of linguistic diffusion using a network of multiple populations, showing that a simple model of cultural transmission can reproduce the observed concentric pattern of Japanese dialects [16]. The model considers transmission of words within and between populations under the assumption that there exists a “central population,” a highly prestigious group of people which exerts a huge cultural impact on others. To gain analytical tractability, the model in chapter 2 makes two assumptions at the cost of generality. First, the model assumes that only one population (central population) on the network can create new cultural variants, which limits the application of the model to the case when all cultural variants are originated from the same place. Second, the model enables the calculation of the average word age (i.e., time elapsed since the invention) in each population of the network but precludes the analysis on the relative prevalence of words with different origins (i.e., places of the invention) in each population. However, to fully understand the spatial distribution of cultural traits beyond linguistic variants, more comprehensive analysis, treating when and where the present culture was made, is required. In the present research, we further extend the network model in the previous chapter to obtain more general formulae of cultural age and origin.

To establish a generalized model, we incorporate an ancestral process, which has been extensively explored in theoretical studies of population genetics for half a century [17, 18]. Classical ancestral process in population genetics starts with sampling multiple genes on a single locus from the current population and traces back the genealogy of the sampled genes in the past to obtain inference about the topological structure and branch length of the genealogical tree. Widely known as coalescent theory, this class of model has been exploited to estimate the amount of genetic polymorphism and to statistically detect natural selection on a given DNA sequence [19]. More recent studies introduced the framework of ancestral process to treat the dynamics of cultural transmission. First, Aguilar and Ghirlanda [20] simulated the genealogy of a cultural trait, introducing the concept of time to the most recent unique ancestor (MRUA), which makes a stark contrast with the idea of the most recent common ancestor (MRCA) in population genetics. In addition, Kobayashi et al. [21] developed a finite-population model of cultural genealogy to obtain the age-frequency spectrum within the samples. Their research also proved the duality of the ancestral backward process and the time-forward process, which observes the change of the frequency of a trait generation by generation.

Ancestral process of either genetic or cultural trait has been represented by a Markov chain model, where the state variable is the number of ancestors in the genealogy starting from multiple samples. However, in this study, we will introduce a novel mathematical framework, which is markedly different from the previous models in the following points. First, as we are interested in the spatial dynamics of cultural macroevolution, we will consider a network of multiple interdependent populations, instead of a single population. Secondly, ancestral models in previous research have usually assumed a finite population size, whereas we will assume an infinite number of individuals, neglecting any stochastic effect or drift. Thirdly, and most significantly, the ancestral process in our model starts from a single sample. The process, therefore, experiences no coalescence or branching events, and always finishes with an invention event of the sampled culture. Also, the state variable of the Markovian process is the population to which the ancestor belongs in each past generation, instead of the number of ancestors the genealogical tree contains.

By assigning the transmission rates between every pair of two populations and the social learning rate in each population, our model will calculate three values about cultural diffusion on a network. First, we calculate the *mean cultural age* in each node (i.e., population). Secondly, we consider the *origin* of cultural variants; focusing on one population X, we calculate the proportion of cultural variants in X that were created in each population of the network. Finally, we calculate the *conditional mean cultural age*. Focusing again on two arbitrary populations X and Y, we calculate the mean age of cultural variants in X, given that the variant was invented in Y. Therefore, the conditional mean cultural age is a proxy of the time it takes for a variant to be transmitted from Y to X. These quantities are fundamental in cultural diffusion on a network, as they provide inference about both temporal and spatial pattern of transmission. In addition, we will introduce an additional model with finite cultural variants to examine the frequency, age, and origin of specific cultural variants.

The rest of this chapter is structured as follows. In "Model" section, we first describe the infinite-variant model emphasizing the ancestral process of one sampled cultural variant and derive formulae to calculate the mean cultural age, the frequencies of cultural variants with different origins, and the conditional mean cultural age. Based on the general formulae, the effect of network topology and innovation rate on these quantities are numerically examined in "Numerical analysis on a random graph" section, and further mathematical analysis is carried out for special cases in "Analytically tractable cases" section. In "Inferring network structure" section, we discuss ways to detect the presence of inter-population transmission and deduce the transmission rate (i.e., network topology). Finally, in “Finite-variant model” section, we extend the model to treat a finite number of distinct cultural variants, deriving the frequency, age, and origin of each variant type.

2. Model

2-1 Cultural transmission on a population network

Consider a network of n culturally interdependent populations $P_1, \dots, P_n$, each of which is composed of an infinite number of individuals. Every individual in the populations bears one cultural variant. At the beginning of a generation, every newborn individual acquires a cultural variant either socially or individually (Figure 1). A newborn in P_k $(1 \leq k \leq n)$ invents a completely new cultural variant (i.e. individual learning) with probability b_k; otherwise, the newborn engages in social learning. When learning socially, learners simply copy one cultural parent chosen from the immediately preceding generation; specifically, they choose a population, with a probability proportional to the transmission rate as described below, from which an individual is drawn at random to be the cultural parent. Note that the cultural transmission in our model is unbiased, meaning that a learner adopts a randomly selected individual's cultural variant without exerting any preference that depends on, for example, the prevalence of variants or prestigiousness of individuals. The probability that a learner in P_i learns from an individual in P_j is denoted by a_{ij}, which represents the transmission rate from P_j to P_i. We have

$$b_k + \sum_{j=1}^{n} a_{kj} = 1, \qquad (1)$$

for every k. We only consider the equilibrium state, where cultural variants have been transmitted for a sufficiently long time. Symbols used in this chapter are summarized in Table 1.

2-2 Cultural genealogy of a single sample

We take a single sample of cultural variant from one of the present populations in the network and track back its genealogy (backward ancestral process). As cultural transmission always occurs from one individual to another, the cultural genealogy contains one ancestor at every generation. The genealogy disappears when it eventually reaches the creator of the sampled cultural variant (Fig. 2).

The backward process constitutes a Markov chain whose state variable is the identity of the population to which the cultural ancestor at a given generation belongs. Let S_k $(1 \leq k \leq n)$ denote the state where the ancestor is in population P_k, and S_0 be the state where no ancestor exists. The backward process is Markovian with total number of $n+1$ states, $\{S_0, \dots, S_n\}$ being the state space. The probability transition matrix $\boldsymbol{\Pi}$ is given by,

$$\boldsymbol{\Pi} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ b_1 & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ b_n & a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \boldsymbol{b} & \boldsymbol{A} \end{pmatrix}, \qquad (2)$$

where

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \qquad (3)$$

is the matrix whose elements are the transmission rates among populations, $\boldsymbol{b} = (b_1 \dots b_n)^T$, and $\mathbf{0}$ is the n-dimensional row vector whose elements are all zero. Note that S_0 is the unique absorbing state of the process, and the backward process inevitably ends up with the invention event. Here, we assume that for every i there exists at least one array $u_1, \dots, u_L$, such that $a_{iu_1} a_{u_1 u_2} \cdots a_{u_{L-1} u_L} b_{u_L} > 0$ holds, so that the backward process has at least one pathway to the absorbing state. Figure 2 illustrates the ancestral process starting from a sample taken from P_1. In this case, the genealogy ends in P_2 three generations ago, meaning that the sampled cultural variant was invented in P_2 three generations ago.

2-3 Calculation of cultural age

We consider the distribution of cultural age in each population. Let $f_k(\rho)$ denote the sum of the frequencies of cultural variants with age ρ in P_k, which is equivalent to the probability with which a variant randomly sampled from P_k has experienced the invention event ρ generations ago. Therefore, the distribution of cultural age is equivalent to the probability distribution of absorption time of the backward process. Apparently, we have

$$\sum_{\rho=0}^{\infty} f_k(\rho) = 1, \quad (4)$$

for every k. Let $\boldsymbol{e_k}$ denote the n-dimensional row unit vector whose k-th element is one and the others are zero. In the backward process starting from one sample in P_k, the probability that its ancestor ρ generations ago is in P_j is given by the j-th element of $\boldsymbol{e_k A^{\rho}}$. Hence, we have

$$f_k(\rho) = \boldsymbol{e_k A^{\rho} b}. \qquad (5)$$

Denoting the vector $\left(f_1(\rho), \cdots, f_n(\rho)\right)^T$ by $\boldsymbol{f}(\rho)$, we have

$$\boldsymbol{f}(\rho) = \boldsymbol{A}^{\rho}\boldsymbol{b}. \qquad (6)$$

Equation (6) is the most general formula which gives the distribution of cultural age given the network structure (transmission rates) and invention rates of the populations. The average cultural ages in every population $\boldsymbol{r} = (r_1, \cdots, r_n)^T$ is given by

$$\boldsymbol{r} = \sum_{\rho=0}^{\infty} \rho\, \boldsymbol{f}(\rho) = \sum_{\rho=0}^{\infty} \rho\, \boldsymbol{A}^{\rho}\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{E}-\boldsymbol{A})^{-1}(\boldsymbol{E}-\boldsymbol{A})^{-1}\boldsymbol{b} = (\boldsymbol{E}-\boldsymbol{A})^{-1}\begin{pmatrix} 1-b_1 \\ \vdots \\ 1-b_n \end{pmatrix}. \ (7)$$

Similarly, the second order moment of cultural age distribution in every population $\boldsymbol{q} = (q_1, \cdots, q_n)^T$ is given by

$$\boldsymbol{q} = \sum_{\rho=0}^{\infty} \rho^2\, \boldsymbol{f}(\rho) = \sum_{\rho=0}^{\infty} \rho^2\, \boldsymbol{A}^{\rho}\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{E}-\boldsymbol{A})^{-\mathbf{1}}(\boldsymbol{E}-\boldsymbol{A})^{-\mathbf{1}}(\boldsymbol{E}-\boldsymbol{A})^{-\mathbf{1}}(\boldsymbol{E}+\boldsymbol{A})\boldsymbol{b}$$

$$= (\boldsymbol{E}-\boldsymbol{A})^{-\mathbf{1}}(\boldsymbol{E}-\boldsymbol{A})^{-\mathbf{1}}(\boldsymbol{E}+\boldsymbol{A})\begin{pmatrix} 1-b_1 \\ \vdots \\ 1-b_n \end{pmatrix}. \ (8)$$

Variance of cultural age in every population $\boldsymbol{v} = (v_1, \cdots, v_n)^T$ can be calculated by

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} - \begin{pmatrix} (r_1)^2 \\ \vdots \\ (r_n)^2 \end{pmatrix}. \quad (9)$$

Note that for derivation of (7) and (8), we have used the formula

$$(\boldsymbol{E} - \boldsymbol{A})^{-1}\boldsymbol{b} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad (10)$$

which is shown by

$$(\boldsymbol{E} - \boldsymbol{A})\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 - \sum_{j=1}^{n} a_{1j} \\ \vdots \\ 1 - \sum_{j=1}^{n} a_{nj} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \boldsymbol{b}. \quad (11)$$

Here, we consider a specific case, where every population has the same individual learning rate (i.e., $b_k = b$ for every n). In this case, frequency of cultural age follows the geometric distribution:

$$f_k(\rho) = (1 - b)^{\rho} b. \quad (12)$$

This is proved as follows. As $b_k = b$ for every k, summation of the n elements in any row of $\boldsymbol{A}$ is $1 - b$ (see (1)). Therefore, the n elements in any row of $\boldsymbol{A}^{\rho}$ sum up to $(1 - b)^{\rho}$. As $\boldsymbol{b} = (b, \cdots, b)^T$ clearly holds, equation (6) readily gives (12). This result gives an intriguing feature that cultural age distribution only depends on the individual learning rate, irrespective of the topological structure of the network or the intensity of intra- or inter-population transmission. Equation (12) gives

$$r_k = \frac{1 - b}{b}. \quad (13)$$

2-4 Origin of cultural variants

Let us call the population where a cultural variant was originally invented the origin of the cultural variant. The goal of this subsection is to apportion the frequencies of cultural variants currently existing in each population according to their origins. Let x_{ij} be the sum of the frequencies of cultural variants in P_i whose origin is P_j. Clearly, we have

$$\sum_{j=1}^{n} x_{ij} = 1, \quad (14)$$

for every i. Recalling that the ancestor ρ generations ago of one sample taken from P_k is in P_j with a probability equal to the j-th element of $\boldsymbol{e_k A^\rho}$, we have

$$(x_{k1}, \cdots, x_{kn}) = \sum_{\rho=0}^{\infty} \boldsymbol{e_k A^\rho} diag(b_1, \dots, b_n) = \boldsymbol{e_k}(\boldsymbol{E} - \boldsymbol{A})^{-\mathbf{1}}\boldsymbol{B} \quad (15)$$

for every k, where $\boldsymbol{B} = diag(b_1, \dots, b_n)$ is the diagonal matrix with invention rates of all the populations. Hence,

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nn} \end{pmatrix} = (\boldsymbol{E} - \boldsymbol{A})^{-1}\boldsymbol{B}. \quad (16)$$

Equation (16) indicates that we can calculate the spectrum of cultural origins for every single population.

If cultural transmission is symmetric (i.e., $a_{ij} = a_{ji}$ for every i and j), (16) indicates $x_{ij} = x_{ji}$ because both matrices $(\boldsymbol{E} - \boldsymbol{A})^{-\mathbf{1}}$ and $\boldsymbol{B}$ are symmetric. In this case, if one arbitrary population contains cultural variants originated from another arbitrary population at a certain frequency, the latter contains the same frequency of cultural variants invented in the former.

2-5 Origin-dependence of age (Calculation of "conditional mean cultural age")

We have so far calculated the distribution of cultural ages and origins separately. In this section, we will further subdivide the age frequency spectrum $\boldsymbol{f}(\rho)$ (see (6)) according to the origins. First, let $f_{ij}(\rho)$ be the frequency of cultural variants in P_i which were invented ρ generations ago in P_j. Therefore, $f_{ij}(\rho)$ is the frequency spectrum of variants, currently present in P_i, subdivided according to when and where it was created. We have

$$\sum_{\rho=0}^{\infty}\sum_{j=1}^{n} f_{ij}(\rho) = 1, \quad (17)$$

$$\sum_{j=1}^{n} f_{ij}(\rho) = f_i(\rho), \quad (18)$$

$$\sum_{\rho=0}^{\infty} f_{ij}(\rho) = x_{ij}. \quad (19)$$

As $\boldsymbol{e_k A}^{\rho}$ specifies the probability that the ancestor ρ generations ago of a cultural sample taken by P_k belongs to a given population, we have

$$(f_{k1}(\rho) \quad \cdots \quad f_{kn}(\rho)) = \boldsymbol{e_k A}^{\rho}\boldsymbol{B}, \quad (20)$$

which readily gives

$$\boldsymbol{F}(\rho) = \begin{pmatrix} f_{11}(\rho) & \cdots & f_{1n}(\rho) \\ \vdots & \ddots & \vdots \\ f_{n1}(\rho) & \cdots & f_{nn}(\rho) \end{pmatrix} = \boldsymbol{A}^{\rho}\boldsymbol{B}. \quad (21)$$

We can recover the frequency of ages and origins: taking the summation of $f_{ij}(\rho)$ in (21) with regard to j and ρ, we obtain (6) and (16), respectively.

Now, we consider the age frequency of cultural variants which were invented in P_j and which are currently in P_i. Considering the probabilistic description of $f_{ij}(\rho)$ and x_{ij}, we have

$$\begin{aligned} &P\big(age\ \rho | invented\ in\ P_j\ \cap\ sampled\ in\ P_i\big) \\ &\qquad = \frac{P\big(age\ \rho \cap invented\ in\ P_j\ |\ sampled\ in\ P_i\big)}{P\big(invented\ in\ P_j | sampled\ in\ P_i\big)} = \frac{f_{ij}(\rho)}{x_{ij}}. \quad (22) \end{aligned}$$

Let r_{ij} denote the *conditional mean cultural age*, or the average age of cultural variants in P_i, given they were made in P_j. Equation (22) gives

$$r_{ij} = \sum_{\rho=0}^{\infty} \frac{\rho f_{ij}(\rho)}{x_{ij}} = \frac{r'_{ij}}{x_{ij}}, \quad (23)$$

where

$$r'_{ij} = \sum_{\rho=0}^{\infty} \rho f_{ij}(\rho). \quad (24)$$

Using (21), r'_{ij} is calculated by

$$\boldsymbol{R}' = \begin{pmatrix} r'_{11} & \cdots & r'_{1n} \\ \vdots & \ddots & \vdots \\ r'_{n1} & \cdots & r'_{nn} \end{pmatrix} = \sum_{\rho=0}^{\infty} \rho \boldsymbol{F}(\rho) = \sum_{\rho=0}^{\infty} \rho \boldsymbol{A}^{\rho} \boldsymbol{B} = (\boldsymbol{E}-\boldsymbol{A})^{-1}(\boldsymbol{E}-\boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{B}. \quad (25)$$

By substituting (16) and (25) for (23), we obtain the conditional mean cultural age. This can be used to approximate the expected time it takes for a variant to transmit from P_j to P_i, although r_{ij} is different from the expected time because a cultural variant currently present in P_i may have stayed there for a considerably long time or may have already visited P_i multiple times. We use r_{ij} as a proxy to relative transmission time between two populations. Similar to the case of x_{ij}, we have $r_{ij} = r_{ji}$ if $a_{ij} = a_{ji}$ holds for every i and j.

3. Numerical analysis on a random graph

We have so far obtained the formulae to find, for any given population, the mean cultural age, frequencies of cultural variants with different origins, and conditional mean cultural age for each original population. To examine how the topological structure of the network and innovation rates of populations affect these values, we perform a numerical analysis on a random graph.

3-1 Generating a random directed graph

We generate a random directed graph of populations with the following algorithm. First, we have n isolated populations $P_1, \cdots, P_n$. Second, for every ordered pair of different populations (P_i, P_j) $(i \neq j)$, we assign a directed edge starting at P_i and ending at P_j with probability p. Third, we assign $a_{ij} = a = const. > 0$ $(i \neq j)$ if there is an edge directing from P_i to P_j, and $a_{ij} = 0$ otherwise. Fourth, we assign innovation rate b_k to every population. Fifth, we assign the value of transmission rates within the same population (a_{ii}) following equation (1).

In this way, we obtain a random network of n populations whose transmission rates from one population to another is a with probability p and 0 otherwise. Note that we do not assign self-loops, even though a_{ii} always takes a positive value. We provide an example of random network in Figure 3. The *in-degree* of a population P_i or $indeg(P_i)$ is the number of edges directed to P_i, representing the number of populations which learns from P_i. Conversely, the *out-degree* of a population P_i or $outdeg(P_i)$ is the number of edges originating from it, or the number of populations from which cultural variants transmit to P_i. In other words, the out-degree and in-degree of P_i measure how many other populations culturally influence and are influenced by P_i, respectively. Note that the edges of the graph are directed in the same direction as the movement of the sampled cultural variant through the backward ancestral process. Instead of directly assigning n and p in generating a random network, we specify n and $M = (n-1)p$, the latter of which equals to both the expected in-degree and out-degree per vertex, representing the density of the network. The analysis was performed using the R package *igraph*.

3-2 Effect of network topology on the age and origin of cultural variants

We generate a random graph using parameter values $a = 0.01$ and $b_k = b = 0.01$ for every k. We consider networks that are either small ($n = 200$) or large ($n = 1000$) and either sparse ($M = 5$) or dense ($M = 50$). Since innovation rate is constant in every population, (13) gives $r_k = (1-b)/b = 99$ for all k. Using equations (16), (23), and (25), we calculate x_{ij} and r_{ij}, and examine their correlation with topological variables of the network.

We first consider x_{ii} and r_{ii}, the variables assigned to each population. Using the dataset of all the populations (200 or 1000) of the network, we calculate the correlation coefficients between those variables (x_{ii} or r_{ii}) and the topological variables (in-degree or out-degree of P_i) (Table 2a). The result shows that both x_{ii} and r_{ii} are negatively correlated with the out-degree of P_i, although the latter does not hold with the parameter set $n = 200,\ M = 50$. This result is quite straightforward because higher out-degree of population P_i by assumption implies (see (1)) lower transmission rate within P_i, which results in lower frequency and younger age of the variants created by P_i, reducing $x_{ii}$and

r_{ii}. On the other hand, x_{ii} is found to have only a weak correlation, if any, with the in-degree of P_i. In addition, r_{ii} has a strong positive correlation with the in-degree of P_i in the case of dense network $(M = 50)$.

We also examine the variables assigned to each pair of populations, x_{ij} and r_{ij}. Although we have $n(n-1)$ ordered pairs of populations in the network, not every population is reachable from every other population through the ancestral process (i.e., the graph is not strongly connected). As r_{ij} is defined only if cultural variants made by P_i can reach P_j, we only treat the ordered pairs (P_i, P_j) which satisfy this condition. Table 2b shows the correlation coefficient between the topological variables of the ordered pairs (P_i, P_j) with both x_{ij} and r_{ij}.

First, x_{ij} is negatively correlated with the distance d_{ij} (length of the shortest path from P_i to P_j) and out-degree of P_j. As for the distance d_{ij}, it is quite intuitive that the frequency in P_i of cultural variants made by P_j decreases with the distance of the two populations. Interestingly, the correlation is stronger with $log(d_{ij})$ than with the distance itself with the exception of the case $n = 200, M = 50$, suggesting that the frequency of cultural variants decreases in an exponential manner with the distance from the population that created the variants. The fact that x_{ij} is negatively correlated with the out-degree of P_j is explained by the negative correlation between x_{jj} and $outdeg(P_j)$ (see Table 2a). More specifically, it is intuitive that the frequency in any population of cultural variants originating from P_j decreases if P_j contains only a low frequency of variants invented by the population itself. Note that the negative correlation between x_{ij} and $outdeg(P_j)$ becomes stronger when the network is dense $(M = 50)$. Conversely, Table 2b indicates that x_{ij} is positively correlated with the in-degree of P_j, which is as expected because the cultural variants deriving from P_j will plausibly increase in P_i if more populations learn from P_j. Again, this positive correlation becomes stronger in a dense network $(M = 50)$.

Second, r_{ij} is positively correlated with the distance d_{ij}, indicating that the cultural variants that are currently in P_i and was invented by P_j become on average newer when the two populations are located closer. This is totally straightforward because shorter distance means fewer times of inter-population transmission, which is rare particularly when a is low. On the other hand, r_{ij} is negatively correlated with the out-degree of P_i especially in the case of $M = 5$. The variable is also negatively correlated with both in-degree and out-degree of P_j, although the correlation coefficient does not take large values in any parameter set. These results are interpreted in a quite similar way to the case of x_{ij}, which we have just mentioned: if more populations learn from P_j, cultural variants invented in the population will more rapidly reach P_i. Also, cultural variants created by P_j tend to become on average newer in P_j when the population receives cultural transmission from other populations (see Table 2a), so it is natural that r_{ij} have a negative correlation with the out-degree of P_j.

3-3 Effect of innovation rate on the age and origin of cultural variants

We generate a random graph using the parameters $n = 200$ or 1000, $M = 5$ or 50, and $a = 0.01$. This time, we assign the innovation rates b_k following the uniform distribution with its minimum and maximum values being 0 and 0.02, respectively. The values are assigned independently to each population, so the innovation rates are not equal. Again, we calculate the values r_k, x_{ij}, and r_{ij} and explore the correlation between these values and innovation rate.

Table 3a summarizes the correlation coefficients of r_i, x_{ii}, and r_{ii} with the innovation rate of P_i. It is shown that both mean cultural ages r_i and r_{ii} are negatively correlated with b_i, although the correlation is extremely weak in the latter case, confirming that culture becomes on average newer when individuals innovate more frequently. On the other hand, x_{ii} is positively correlated with b_i, which is also in concordance with an intuitive explanation that the frequency of cultural variants created in the same population increases when the population invents more variants. This correlation is larger when the network is dense $(M = 50)$.

Focusing on two populations P_i and P_j, we discuss how the innovation rates of these two populations exert influence on the values x_{ij} and r_{ij}. In calculating the correlation coefficients, we only treat the ordered pairs $\left(P_i, P_j\right)$, such that cultural variants invented in P_j can reach P_i. Table 3b shows that x_{ij} is positively correlated with b_j, which is in particular strong in the case of dense network $(M = 50)$.

4. Analytically tractable cases

In Model section, we derived general formulae to calculate the cultural age, origin, and origin-dependent age in the population network. Although these formulae give numerical solutions for arbitrary networks, further mathematical analysis is only possible for a more restricted class of networks. Here, we present such special cases.

4-1 Recursive formulae for $r_k, x_{ij}, and\ r_{ij}$

Before moving on to special cases, we present the recursive formulae of r_k, x_{ij}, and r'_{ij} with regard to k and i as follows:

$$r_k = \sum_{j=1}^{n} a_{kj}(r_j + 1), \quad (26)$$

$$x_{ij} = \sum_{l=1}^{n} a_{il}x_{lj} + \delta_{ij}b_i, \quad (27)$$

$$r'_{ij} = \sum_{l=1}^{n} a_{il}(r'_{lj} + x_{lj}). \quad (28)$$

Note that $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ otherwise. While the proof of expression (27) is easily done by considering the relationship between the frequencies in the previous generation and those in the current generation, the other expressions (26) and (28) are relatively difficult to prove, and their derivation is provided in **Appendix A**. General formulae (7), (16) and (25) can be alternatively obtained by (26), (27) and (28), respectively (**Appendix B**), which may facilitate the understanding of mathematical property of our model.

4-2 One-dimensional unidirectional transmission in the presence of a central population. We consider the case where populations $P_1, \dots, P_n$ are linearly arranged in this order. We assume that P_1 is a "central population", which is a highly innovative population, exerting a huge cultural influence on others. More specifically, we assume $b_1 = 1$ so that P_1 is always filled with the cultural variants which are created in the current generation. For other populations, which we call "peripheral populations", we assign $b_k = b\ (\text{const.})$ for $k \neq 1$. Transmission rates $a_{ij}\ (i \geq 2)$ is given by $1 - a - b$ and a when $i = j$ and $i = j + 1$, respectively, and zero otherwise, which means that cultural transmission occurs only from a certain population to the neighboring population that is further from the central population. This is an extension of the unidirectional diffusion which we modeled in Chapter 2.

Under this specific condition, we derive the mean cultural age in each population. From (26), we have the following system of recursive formulae:

$$\begin{cases} r_1 = 0 \\ r_k = (1 - a - b)(r_k + 1) + a(r_{k-1} + 1) \quad (k \geq 2) \end{cases}, \quad (29)$$

which is reduced to

$$r_k = \begin{cases} \dfrac{1}{a}(k-1) & (b=0) \\ \dfrac{1-b}{b}\left\{1-\left(\dfrac{a}{a+b}\right)^{k-1}\right\} & (b>0) \end{cases}. \quad (30)$$

Equation (30) suggests that the distribution of mean cultural ages changes substantially depending on whether peripheral populations create new cultural variants or not. When P_1 is the only population that innovates new variants, cultural age on average increases in proportion to the distance from the central population, which is obviously the same result as equation (10a) of Chapter 2. On the other hand, when peripheral populations also innovate, the mean cultural age becomes larger with the distance from the central population and converges to $(1-b)/b$.

As for the cultural origins, we concentrate our analysis on the frequency of cultural variants that derive from the central population, which is represented by x_{i1}. Equation (27) gives

$$\begin{cases} x_{11} = 1 \\ x_{i1} = (1-a-b)x_{i1} + ax_{i-1,1} \end{cases}, \quad (31)$$

which is reduced to

$$x_{i1} = \left(\frac{a}{a+b}\right)^{i-1}. \quad (32)$$

Equation (32) shows that cultural variants deriving from P_1 decreases in frequency exponentially with the distance between the current and original populations. The frequency increases with transmission rate (a) and decreases with the innovation rate in the peripheral populations (b). When peripheral populations do not innovate $(b=0)$, all the cultural variants derive from the central population.

Finally, we consider the mean age of cultural variants that derive from the central population, r_{i1}, which works as a proxy to the time it takes for a variant to transmit from P_1 to P_j. Equation (28) gives

$$\begin{cases} r'_{11} = 0 \\ r'_{i1} = (1-a-b)(r'_{i1} + x_{i1}) + a(r'_{i-1,1} + x_{i-1,1}) \end{cases}. \quad (33)$$

Using (32), (33) is reduced to

$$r'_{i1} = \frac{1}{a+b}\left(\frac{a}{a+b}\right)^{i-1}(i-1). \quad (34)$$

Equations (32) and (34) give

$$r_{i1} = \frac{r'_{i1}}{x_{i1}} = \frac{1}{a+b}(i-1). \quad (35)$$

Equation (35) shows an interesting characteristic of cultural transmission: variants invented in the central population become on average older in proportion to the distance between the current and central populations. Moving away from P_1 by one population, cultural variants created in P_1 become older on average by $1/(a+b)$. Intuitively speaking, cultural variants spreading from a strong center diffuse at this constant velocity even under the influence of innovation in peripheral populations.

4-3 Transmission on a complete graph

In this subsection, we consider the case when cultural transmission occurs equally between any pair of two populations, meaning that $a_{ij} = a$ (including the case of $i = j$) and $b_i = b = 1 - an$, where $0 < a < 1/n$. In this case, social learners in a given population learn a variant from any population with the same probability. The transmission matrix is written as

$$\boldsymbol{A} = \begin{pmatrix} a & \cdots & a \\ \vdots & \ddots & \vdots \\ a & \cdots & a \end{pmatrix}. \quad (36)$$

We also have

$$\boldsymbol{B} = \begin{pmatrix} 1-an & & \\ & \ddots & \\ & & 1-an \end{pmatrix}. \quad (37)$$

Note that zero elements are omitted for the sake of notational simplicity. Now we consider the origins of cultural variants using (16). Letting α_{ij} be the element of $(\boldsymbol{E}-\boldsymbol{A})$ at *i*-th row and *j*-th column, we have

$$\alpha_{ij}=\begin{cases}-a & (if \quad i\neq j)\\ 1-a & (if \quad i=j)\end{cases}. \quad (38)$$

Letting β_{ij} be the element of $(\boldsymbol{E}-\boldsymbol{A})^{-1}$ at *i*-th row and *j*-th column yields

$$\beta_{ij}=\begin{cases}\dfrac{a}{1-an} & (if \quad i\neq j)\\ \dfrac{1-(n-1)a}{1-an} & (if \quad i=j)\end{cases}. \quad (39)$$

Equation (39) is proved in **Appendix C**. Using (16) and (37), we have

$$x_{ij}=\begin{cases}a & (if \quad i\neq j)\\ 1-(n-1)a=a+b & (if \quad i=j)\end{cases}. \quad (40)$$

Note that x_{ij} is always larger when $i=j$ than when $i\neq j$, showing that each population contains a larger number of cultural variants created by itself, in comparison with the variants deriving from any other population. This difference is obviously attributable to the effect of individual learning. We also have

$$x_{ij}\to\frac{1}{n} \quad (\text{as } b\to 0). \quad (41)$$

Now we explore r_{ij}. Letting γ_{ij} denote the element at *i*-th row and *j*-th column of $(\boldsymbol{E}-\boldsymbol{A})^{-1}(\boldsymbol{E}-\boldsymbol{A})^{-1}$, we have

$$r'_{ij}=a(1-an)\sum_{l=1}^{n}\gamma_{il}. \quad (42)$$

Using (39),

$$\sum_{l=1}^{n} \beta_{il} = \frac{1}{1 - an} \qquad (43)$$

is a constant value regardless of i, so we have

$$\sum_{l=1}^{n} \gamma_{il} = \left(\sum_{l=1}^{n} \beta_{il}\right)^2 = \frac{1}{(1 - an)^2}. \qquad (44)$$

Equation (42) gives

$$r'_{ij} = \frac{a}{1 - an} = \frac{a}{b}. \qquad (45)$$

Therefore,

$$r_{ij} = \frac{r'_{ij}}{x_{ij}} = \begin{cases} \dfrac{1}{b} & (i \neq j) \\ \dfrac{a}{b(a+b)} & (i = j) \end{cases}. \qquad (46)$$

Equation (46) indicates that cultural variants made in the same population is on average newer than the ones that were created in a different population and then transmitted to the current population.

The most remarkable characteristic of the complete graph is its high symmetricity: the mean cultural age is uniform among all populations, and the frequencies and the conditional mean ages of cultural variants other than those currently belong to the original population are constant irrespective of the identities of the current and original populations.

5. Inferring network structure

In Model section, we obtained the frequencies of cultural variants with different origins in each population of a given network (see equation (16)). Conversely, we introduce a method to calculate the transmission rates among populations for given frequencies of variants with different origins. In other words, we infer, from a dataset of the frequencies of originating populations, the underlying topological structure of the population network.

Starting from the equation (16), we have

$$\boldsymbol{X}^{-1} = \boldsymbol{B}^{-1}(\boldsymbol{E} - \boldsymbol{A}) = \begin{pmatrix} 1/b_1 & & \\ & \ddots & \\ & & 1/b_n \end{pmatrix} \left(\boldsymbol{E} - \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \right). \quad (47)$$

Let y_{ij} denote the element at i-th row and j-th column of $\boldsymbol{X}^{-1}$. Comparing each element in both hands of (47), we have

$$y_{ij} = \frac{\delta_{ij} - a_{ij}}{b_i}, \quad (48)$$

which is readily reduced to

$$a_{ij} = \delta_{ij} - b_i y_{ij}. \quad (49)$$

Note that $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ otherwise. Here, as x_{ij} is given for every i and j, y_{ij} is numerically calculated. Hence, if we know the rate of individual learning in P_i (i.e., b_i), we can calculate the rates of cultural transmission to P_i. Furthermore, even when the rates of individual learning are unavailable, we can calculate the relative rate of cultural transmission as follows:

$$\frac{a_{ij}}{a_{ik}} = \frac{y_{ij}}{y_{ik}} \ (i \neq j, k). \quad (50)$$

The left-hand side of (50) represents the influence that P_i receives from P_j relative to that from P_k.

6. Finite-variant model

We have so far considered the total frequency of cultural variants having the same age or origin, not the frequency of each variant. While this does not hamper analysis of cultural traits whose variants can be identified from the time and place of invention, the distribution of variants is not well represented if the same variant can convergently evolve in multiple populations. As a solution to this problem, we employ a model which allows for creation of different variants within a generation and recurrent creation of a variant in different generations to calculate the frequency, age, and origin of each cultural variant.

6-1 Model description

Again, we consider the network of n populations $P_1, \dots, P_n$. In this model, there exist m types of cultural variant $C_1, \dots, C_m$, and each individual bears only one of them. Here, an individual in P_k acquires the variant C_l by individual learning with probability c_{kl}. This parameter, representing the innovation rate of the specific variant in each population, satisfies

$$\sum_{l=1}^{m} c_{kl} = b_k \quad (51)$$

and

$$\sum_{j=1}^{n} a_{ij} + \sum_{l=1}^{m} c_{il} = 1 \quad (52)$$

for every i. Note that transmission rate a_{ij} is assumed to be common to all variants and that a variant does not mutate into another during the process of cultural transmission.

As in the infinite-variant model, we can consider a Markovian ancestral process, starting from one sampled variant in any population, where the ancestral tree contains only one ancestor in each generation and ends with an invention event.

6-2 Frequency, age, and origin of each variant

First, we calculate the frequency of each variant in each population. Let $\varphi_k(l,\rho)$ denote the frequency of C_l with age ρ in P_k, which is equivalent to the probability that the variant sampled in P_k is C_l and is created ρ generations ago, or formally,

$$\varphi_k(l,\rho) = P(invented\ as\ C_l \cap age\ \rho | sampled\ in\ P_k). \quad (53)$$

The probability is calculated by two steps. First, the probability that the ancestor ρ generations ago of the sampled variant belongs to P_j is given by the *j*-th element of $\boldsymbol{e_k A^{\rho}}$. Second, given the ancestor is in P_j, the probability that the ancestral variant was invented as type C_l in this generation is given by c_{jl}. Since this holds for all j, we have

$$\varphi_k(l,\rho) = \boldsymbol{e_k A^\rho c_l}, \quad (54)$$

where $\boldsymbol{c_l} = (c_{1l} \cdots c_{nl})^T$ represents the invention rate of the variant C_l in each population. In matrix form, (54) is written as

$$\begin{pmatrix} \varphi_1(1,\rho) & \cdots & \varphi_1(m,\rho) \\ \vdots & & \vdots \\ \varphi_n(1,\rho) & \cdots & \varphi_n(m,\rho) \end{pmatrix} = \boldsymbol{A}^\rho \boldsymbol{C}, \quad (55)$$

where

$$\boldsymbol{C} = \begin{pmatrix} c_{11} & \cdots & c_{1m} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nm} \end{pmatrix} \quad (56)$$

is a matrix each of whose element represents the invention rate of a given cultural variant in a given population. Therefore, the frequency of C_l in population P_k, denoted by $\varphi_k(l)$, is given by

$$\begin{pmatrix} \varphi_1(1) & \cdots & \varphi_1(m) \\ \vdots & & \vdots \\ \varphi_n(1) & \cdots & \varphi_n(m) \end{pmatrix} = \sum_{\rho=0}^{\infty} \begin{pmatrix} \varphi_1(1,\rho) & \cdots & \varphi_1(m,\rho) \\ \vdots & & \vdots \\ \varphi_n(1,\rho) & \cdots & \varphi_n(m,\rho) \end{pmatrix} = (\boldsymbol{E} - \boldsymbol{A})^{-1} \boldsymbol{C}. \quad (57)$$

In addition to the frequency of each variant, equations (55) and (57) give the age distribution of each variant and population. Among the variants C_l that is present in P_k, the proportion of variants aged ρ is calculated by $\varphi_k(l,\rho)/\varphi_k(l)$. Thus, mean age of C_l in population P_k is given by

$$r_k(l) = \sum_{\rho=0}^{\infty} \rho \frac{\varphi_k(l,\rho)}{\varphi_k(l)} = \frac{1}{\varphi_k(l)} \sum_{\rho=0}^{\infty} \rho \varphi_k(l,\rho). \quad (58)$$

Using (55), the summation in the third side of equation (58) is calculated by

$$\sum_{\rho=0}^{\infty}\rho\begin{pmatrix}\varphi_1(1,\rho) & \cdots & \varphi_1(m,\rho)\\ \vdots & & \vdots\\ \varphi_n(1,\rho) & \cdots & \varphi_n(m,\rho)\end{pmatrix}=\sum_{\rho=0}^{\infty}\rho\boldsymbol{A}^{\rho}\boldsymbol{C}=(\boldsymbol{E}-\boldsymbol{A})^{-1}(\boldsymbol{E}-\boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{C}\quad(59)$$

Now we consider the origin of each variant. Letting

$$\varphi_{ij}(l)=P\big(invented\ in\ P_j\cap invented\ as\ C_l|sampled\ in\ P_i\big)\quad(60)$$

be the proportion of variants derived from P_j among the variants C_l in P_i, backward ancestral process gives

$$\big(\varphi_{i1}(l),\cdots,\varphi_{in}(l)\big)=\sum_{\rho=0}^{\infty}\boldsymbol{e_i}\boldsymbol{A}^{\rho}diag(c_{1l},\dots,c_{nl})=\boldsymbol{e_i}(\boldsymbol{E}-\boldsymbol{A})^{-\mathbf{1}}diag(c_{1l},\dots,c_{nl}).\,(61)$$

Hence, we have

$$\begin{pmatrix}\varphi_{11}(l) & \cdots & \varphi_{1n}(l)\\ \vdots & \ddots & \vdots\\ \varphi_{n1}(l) & \cdots & \varphi_{nn}(l)\end{pmatrix}=(\boldsymbol{E}-\boldsymbol{A})^{-\mathbf{1}}diag(c_{1l},\dots,c_{nl}).\quad(62)$$

Note that this equation is quite similar to equation (16), which is indeed recouped by taking the summation of (60) as regard to l. Let $x_{ij}(l)$ denote the proportion of the variants which were created in P_j among the variant type C_l in P_i. We have

$$x_{ij}(l)=P\big(invented\ in\ P_j|sampled\ in\ P_i\cap invented\ as\ C_l\big)=\frac{\varphi_{ij}(l)}{\varphi_i(l)},\quad(63)$$

which can be calculated by (55) and (62).

7. Discussion

In this chapter, assuming a network of multiple populations, we have established a mathematical model of cultural evolution and derived the formulae to deduce the age and origin of cultural variants in each population. Using these formulae, we have numerically investigated the effect of the network topology and innovation rates on the spatial distribution of cultural variants with different ages. We have also obtained simple mathematical expressions describing spatial distribution of cultural variants for special network topologies. Conversely, the model can also estimate the transmission rates, given the proportion of cultural variants derived from each node of the network. In addition, the analysis of finite-variant model enabled us to know the distribution and frequency of respective distinct variant type, as well as mean age and expected origins for each variant.

Implications from the numerical analysis on a random network can be summarized as follows. Consider cultural variants deriving from a single population P_j. First, the frequency of cultural variants invented by P_j tends to be high in P_i if (1) two populations are closely located, (2) P_j influences many populations, (3) P_j is influenced by few populations, and/or (4) P_j invents new variants frequently. Second, cultural variants that are in P_i and were invented in P_j tend to be on average older if (1) two populations are located far away, and/or (2) P_i is influenced by few populations. Although the mean age is also negatively correlated with the number of populations which influence and are influenced by P_j, the correlation stays quite weak. We also found that some of the effects of both network topology and innovation rate on the origin and mean age of cultural variants depend upon the network density.

As for the mathematical property of the model, we analyze a backward ancestral process and trace when and where a variant was invented. Our ancestral model is starkly different from that of previous studies, either the classic coalescent models of population biology [17,18] or recent studies of cultural genealogy [20,21]. While the models of previous research take a sample of multiple gene or trait copies and track their genealogical tree, the ancestral process in our model starts with a single sample, and therefore the backward process experiences no coalescence and is destined to finish with the invention by the creator. Unlike Kobayashi et al. [21] model of cultural genealogy, individuals in our model learn from one role model in the immediate generation and cannot receive influence from more than one individual. Therefore, the number of ancestors of the sample does not increase as we trace back the genealogy to the past.

Our model has many mathematical properties in common with the voter model. Usually, the voter model regards each vertex as an individual, possessing one piece of social information or trait (i.e., idea, opinion, language), who accepts information from one of its neighbors (*one-vertex-one-variant model*). On the other hand, our model regards the vertices as populations composed of numerous individuals who may potentially have different cultural variants (*one-vertex-many-variants model*). Although these two classes of models differ in terms of the number of variants occupying each vertex, our model can easily be adapted into the *one-vertex-one-variant* version. In this case, each population has only one variant and transmission rates a_{ij} and b_i

respectively represent the probability that the population P_i learns the variant of P_j and invents a new variant. Quantities we calculated such as r_k, x_{ij}, r_{ij}, etc. are now interpreted in a probabilistic way (see Table 4).

We discuss the application of our model to the spatial evolution of real cultural traits. As we examined in chapter 2, one of the most promising fields of application is linguistics and dialectology, where recent theoretical research has been introducing network models of word diffusion [15]. As our model includes parameters of transmission and innovation rates in each population, we can assign these parameters from geographic and demographic dataset of real populations. Assuming a formula which gives transmission rates and innovation rates as a function of population sizes, geographical proximity, and the social status of the populations, we obtain the network topology and subsequently distribution of cultural ages and origins expected by this model. Comparison between empirical and expected distribution of cultural variants may enable us to know what factors affect the transmission and innovation rates.

It is also interesting to compare our analysis with phylogenic approach, which quantifies the cultural distance of each pair of the populations. It is intriguing to examine the correlation between phylogenic distance between two populations with the values x_{ij} and r_{ij} obtained from our model. Phylogenic approach gives the taxonomical proximity among cultural variants in several populations and we can deduce the expected origin and time duration of the diffusion. Comparing the result of our model with that obtained through phylogenic methods, we would further deepen the understanding of how cultural traits are transmitted among populations.

Finally, we discuss the limitation of the model and suggestion for future research. First of all, our model does not include the learning bias [2], whereby learners acquire cultural variants selectively according to such factors as innate preference, frequency, novelty, or the social status of the role model. Previous theoretical studies proved that learning bias during the cultural transmission exerts substantial influence on the evolution of cultural traits [22,23]. Under the framework of the current study, it is impossible to introduce bias in learning process, as it would violate the independence of each cultural variant during the ancestral process, and the model would no longer be considered as a Markovian process. This class of research would require a different model, which probably compromises the simple matrix notation and the analytical tractability to a certain extent.

Even though we have explored a finite-variant model, in which the frequency of each variant is explicitly calculated, the model lacks the mutual mutation among the variants. In our model, once cultural variants are invented, they will never change into other variants through the repeated series of social learning over generations. This drawback would probably be overcome by integrating a *mutation-rate matrix* among the variants, each element of which represents the mutation rate through social learning from one variant to another. In this way, the finite-variant model would be applied to the diffusion of cumulative cultural traits which have been alleged to characterize the cultures of modern humans. Among the m variants $C_1 \dots C_m$, we would assume that C_1 is the most primitive and C_m is the most sophisticated variant (i.e., skills), and mutation among these variants would represent the improvement or degradation of the cumulative culture. Again, one possible problem of this research is that we cannot examine the effect of learning bias toward high-level culture or skills, which are assumed in much theoretical research into cumulative culture [24-26]. Also, our model assumes that each cultural variant is selectively neutral, indicating that the variants do not affect the survivorship or reproductive success of their bearers, which might be unrealistic in the context of cumulative skills of modern humans.

References

[1] Cavalli-Sforza, L. L., and Feldman, M. W. 1981. Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton.

[2] Boyd, R., and Richerson, P. J. 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.

[3] Mesoudi, A. 2011. Cultural Evolution: How Darwinian Theory Can Explain Human Culture & Synthesize the Social Sciences. University of Chicago Press, Chicago.

[4] Mace, R., Holden, C. J., and Shennan, S. 2005. The Evolution of Cultural Diversity: A Phylogenetic Approach. University College London Press.

[5] Nunn, C. L., Mulder, M. B., and Langley, S. 2006. Comparative methods for studying cultural trait evolution: A simulation study. Cross-Cultural Research 40, 2, 177-209.

[6] Currie, T. E., Greenhill, S. J., Gray, R. D., Hasegawa, T., and Mace, R. 2010. Rise and fall of political complexity in island South-East Asia and the Pacific. Nature 467, 801-804.

[7] Towner, M. C., Grote, M. N., Venti, J., and Mulder, M. B. 2012. Cultural macroevolution on neighbor graphs. Vertical and horizontal transmissions among western north American Indian societies. Human Nature 23, 283-305.

[8] Brown, S., Savage, P. E., Ko, A. M. S., Stoneking, M., Ko, Y. C., Loo, J. H., and Trejaut, A., J. 2014. Correlations in the population structure of music, genes and language. Proceedings of the Royal Society B 281, 20132072.

[9] Savage, P. E. and Brown, S. 2014. Mapping music: Cluster analysis of song-type

frequencies within and between cultures. Ethnomusicology 133-155.

[10] Aoki, K., and Shida, M. 1996. Travelling wave solutions for the spread of farmers into a region occupied by hunter-gatherers. Theoretical Population Biology 50, 1-17.

[11] Wakano, J. Y., Gilpin, W., Kadowaki., S., Feldman, M. W., and Aoki, K. 2018. Ecological range-expansion scenarios for the replacement or assimilation of Neanderthals by modern humans. Theoretical Population Biology 119, 3-14.

[12] Sood, V., Antal, T., and Redner, S. 2008. Voter models on heterogeneous networks. Physical Review E 77, 041121.

[13] Castellano, C. 2005. Effect of network topology on the ordering dynamics of voter models. AIP Conference Proceedings 779. 114.

[14] Hadzibeganovic, T., Stauffer, D., Schulze, C. 2008. Boundary effects in a three-state modified voter model for languages. Physica A 387. 3242-3252.

[15] Fagyal, Z., Swarup, S., Escobar, A., M., Gasser, L., Lakkaraju, K. 2010. Centers and peripheries: Network roles in language change. Lingua 120. 2061-2079.

[16] Yanagita, K. 1927. Kagyuko (1)-(4). Jinruigaku Zasshi 42, 125-135 (162-172, 223-233, 273-284).

[17] Kingman, J. F. C. 1982. The coalescent. Stochastic Processes and their Applications 13, 3, 235-248.

[18] Tajima, F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics 105, 437-460.

[19] Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 3, 585-595.

[20] Aguilar, E., and Ghirlanda, S., 2015. Modeling the genealogy of a cultural trait. Theoretical Population Biology 101, 1-8.

[21] Kobayashi, Y., Wakano, J. Y., and Ohtsuki, H. 2018. Genealogies and ages of cultural traits: An application of the theory of duality to the research on cultural evolution. Theoretical Population Biology 123, 18-27.

[22] Aoki, K., Lehmann, L., Feldman, M., W. 2011. Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission. Theoretical Population Biology 79, 192-202.

[23] Fogarty, L., Wakano, J., Y., Feldman, M., W., and Aoki, K. 2017. The driving forces of cultural complexity. Neanderthals, modern humans, and question of population size. Human Nature 28, 39-52.

[24] Henrich, J. 2004. Demography and cultural evolution: how adaptive cultural processes can produce maladaptive losses – the Tasmanian case. Am. Antiquity 69. 197-214.

[25] Kobayashi, Y. and Aoki, K. 2012. Innovativeness, population size, and cumulative cultural evolution. Theoretical Population Biology 82, 38-47.
[26] Ohtsuki, H., Wakano, J., W., and Kobayashi, Y. 2016. Inclusive fitness analysis of cumulative cultural evolution in an island-structured population. Theoretical Population Biology 115, 13-23.

Appendix A: Derivation of the recursive formulae

Here we prove the expressions (26) and (28), which are the recursive formulae for r_k and r'_{ij}.

$$r_k = \sum_{\rho=0}^{\infty} \rho f_k(\rho) = \sum_{\rho=1}^{\infty} \rho \sum_{j=1}^{n} a_{kj} f_j(\rho-1) = \sum_{j=1}^{n} a_{kj} \sum_{\rho=1}^{\infty} \rho f_j(\rho-1)$$

$$= \sum_{j=1}^{n} a_{kj} \left(\sum_{\rho=1}^{\infty} (\rho-1) f_j(\rho-1) + \sum_{\rho=1}^{\infty} f_j(\rho-1) \right)$$

$$= \sum_{j=1}^{n} a_{kj} \left(r_j + 1 \right). \quad (A1)$$

$$r'_{ij} = \sum_{\rho=0}^{\infty} \rho f_{ij}(\rho) = \sum_{\rho=1}^{\infty} \rho \sum_{l=1}^{n} a_{il} f_{lj}(\rho-1) = \sum_{l=1}^{n} a_{il} \sum_{\rho=1}^{\infty} \rho f_{lj}(\rho-1)$$

$$= \sum_{l=1}^{n} a_{il} \left(\sum_{\rho=1}^{\infty} (\rho-1) f_{lj}(\rho-1) + \sum_{\rho=1}^{\infty} f_{lj}(\rho-1) \right)$$

$$= \sum_{l=1}^{n} a_{il} \left(r'_{lj} + x_{lj} \right). \quad (A2)$$

Appendix B: Another derivation of $\boldsymbol{r}$, $\boldsymbol{X}$, and $\boldsymbol{R}'$

In the main part of this chapter, we have calculated $\boldsymbol{r}$, $\boldsymbol{X}$, and $\boldsymbol{R}'$ by using the infinite series of matrices (see expressions (7), (16), and (25)). However, we can derive the same expressions using the recursive formulae (26), (27) and (28), which have been proven in **Appendix A**. Here, we present the different way to derive (7), (16), and (25), which may help better understanding of the mathematical property of the model.

(26) gives

$$r_k = \sum_{j=1}^{n} a_{kj} r_j + \sum_{j=1}^{n} a_{kj} = \sum_{j=1}^{n} a_{kj} r_j + 1 - b_k, \quad (B1)$$

which readily gives

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{r} + \begin{pmatrix} 1 - b_1 \\ \vdots \\ 1 - b_n \end{pmatrix}. \quad (B2)$$

Therefore,

$$\boldsymbol{r} = (\boldsymbol{E} - \boldsymbol{A})^{-1} \begin{pmatrix} 1 - b_1 \\ \vdots \\ 1 - b_n \end{pmatrix}. \quad (B3)$$

As for the origin of cultural variants, (27) gives

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X} - \boldsymbol{B}. \quad (B4)$$

Therefore, we have

$$\boldsymbol{X} = (\boldsymbol{E} - \boldsymbol{A})^{-1}\boldsymbol{B}. \quad (B5)$$

Similarly, (28) gives

$$\boldsymbol{R}' = \boldsymbol{A}\boldsymbol{R}' + \boldsymbol{A}\boldsymbol{X}. \quad (B6)$$

Combining (B5) and (B6), we have

$$\boldsymbol{R}' = (\boldsymbol{E} - \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{X} = (\boldsymbol{E} - \boldsymbol{A})^{-1}(\boldsymbol{E} - \boldsymbol{A})^{-1}\boldsymbol{A}\boldsymbol{B}. \quad (B7)$$

Appendix C: Proof of equation (39)

We provide a proof for the equation (39), showing the product of two matrices with elements (38) and (39) is equivalent to the identity matrix. When $i \neq j$, (38) and (39) give

$$\sum_{l=1}^{n} \alpha_{il}\beta_{lj} = (n-2)(-a)\frac{a}{1-an} + (1-a)\frac{a}{1-an} + (-a)\frac{1-(n-1)a}{1-an} = 0. \quad (C1)$$

When $i = j$, we have

$$\sum_{l=1}^{n} \alpha_{il}\beta_{lj} = (n-1)(-a)\frac{a}{1-an} + (1-a)\frac{1-(n-1)a}{1-an} = 1. \quad (C2)$$

Table 1. Symbols used in this article.

Symbols	Meanings
n	Number of populations (i.e., size of network).
P_k	*k*-th population.
a_{ij}	Probability that an individual in P_i learns a cultural variant from an individual in P_j in each generation.
b_k	Probability that an individual in P_k invents a new cultural variant in each generation.
$\boldsymbol{A}$	Square matrix of order n whose *i*-th row and *j*-th column corresponds to a_{ij}.
$\boldsymbol{b}$	*n*-dimensional vector whose *k*-th element is given by b_k.
$\boldsymbol{B}$	Diagonal matrix of order n whose *k*-th diagonal element is b_k.
$\boldsymbol{E}$	Identity matrix of order n.
$\boldsymbol{e}_k$	*k*-dimensional row unit vector, whose *k*-th element is 1 and the other elements are 0.
$f_k(\rho)$	Frequency of cultural variants with age ρ in P_k.
r_k	Mean age of cultural variants in P_k.
$\boldsymbol{r}$	*n*-dimensional vector whose *k*-th element corresponds to r_k.
q_k	Second-order moment of the cultural age in P_k.
v_k	Variance of the cultural age in P_k.
x_{ij}	Proportion of cultural variants invented in P_j to the entire variants present in P_i.
$\boldsymbol{X}$	Square matrix of order n whose *i*-th row and *j*-th column corresponds to x_{ij}.
$f_{ij}(\rho)$	Proportion of cultural variants invented ρ generations ago in P_j to the entire variants present in P_i.
$\boldsymbol{F}(\rho)$	Square matrix of order n whose *i*-th row and *j*-th column corresponds to a_{ij}.
r_{ij}	Mean age of cultural variants in P_i which are invented in P_j.
m	Number of cultural variants (finite-variant model).
C_l	*l*-th cultural variant (finite-variant model).
c_{kl}	Probability that an individual in P_k invents C_l.
$\boldsymbol{C}$	Matrix with *n* rows and *m* columns, with the element in the *k*-th row and *l*-th column being c_{kl}.

$\varphi_k(l,\rho)$	Proportion of C_l invented ρ generations ago to the entire variants present in P_k.
$\varphi_k(l)$	Frequency of C_l in P_k.
$r_k(l)$	Mean age of C_l in P_k.
$\varphi_{ij}(l)$	Proportion of C_l invented in P_j to the entire variants present in P_i.
$x_{ij}(l)$	Proportion of C_l invented in P_j to all the C_l variants in P_i.

Table 2. Correlation coefficient between topological variables and origin and conditional mean cultural age.

(a)

		$indeg(P_i)$	$outdeg(P_i)$
$n = 200$	x_{ii}	0.060	-0.770
$M = 5$	r_{ii}	0.112	-0.771
$n = 200$	x_{ii}	0.059	-0.961
$M = 50$	r_{ii}	0.981	0.030
$n = 1000$	x_{ii}	0.011	-0.847
$M = 5$	r_{ii}	0.027	-0.847
$n = 1000$	x_{ii}	0.010	-0.979
$M = 50$	r_{ii}	0.898	-0.359

(b)

		d_{ij}	$log(d_{ij})$	$indeg(P_i)$	$outdeg(P_i)$	$indeg(P_j)$	$outdeg(P_j)$
$n = 200$	x_{ij}	-0.540	-0.646	-0.002	0.048	0.230	-0.271
$M = 5$	r_{ij}	0.878	0.882	0.018	-0.320	-0.134	-0.201
$n = 200$	x_{ij}	-0.242	-0.242	-0.001	0.018	0.651	-0.607
$M = 50$	r_{ij}	0.973	0.973	-0.010	-0.082	-0.016	-0.077
$n = 1000$	x_{ij}	-0.404	-0.541	0.000	0.026	0.124	-0.118
$M = 5$	r_{ij}	0.890	0.887	0.013	-0.342	-0.130	-0.191
$n = 1000$	x_{ij}	-0.347	-0.399	0.000	0.013	0.618	-0.614
$M = 50$	r_{ij}	0.720	0.860	-0.002	-0.051	-0.011	-0.042

Table 3. Correlation coefficient between innovation rates and origin and conditional mean cultural age.

(a)

		r_i	x_{ii}	r_{ii}
$n = 200, M = 5$	b_i	-0.160	0.482	-0.119
$n = 200, M = 50$	b_i	-0.988	0.966	-0.058
$n = 1000, M = 5$	b_i	-0.705	0.484	-0.088
$n = 1000, M = 50$	b_i	-0.981	0.958	-0.001

(b)

		b_i	b_j
$n = 200$	x_{ij}	-0.034	0.278
$M = 5$	r_{ij}	-0.003	-0.117
$n = 200$	x_{ij}	-0.015	0.934
$M = 50$	r_{ij}	0.003	0.004
$n = 1000$	x_{ij}	-0.020	0.142
$M = 5$	r_{ij}	-0.011	-0.028
$n = 1000$	x_{ij}	-0.010	0.924
$M = 50$	r_{ij}	-0.002	-0.003

Table 4. Symbols used in the one-vertex-one-variant model.

Symbols	Meanings
a_{ij}	Probability that P_i adopts the cultural variant occupying P_j in each generation.
b_k	Probability that a new cultural variant is invented and occupies P_k in each generation.
$f_k(\rho)$	Probability that a cultural variant occupying P_k was invented ρ generations ago.
r_k	Expected age of cultural variants occupying P_k.
x_{ij}	Probability that the cultural variant occupying P_i was invented in P_j.
$f_{ij}(\rho)$	Probability that the cultural variant occupying P_i was invented in P_j, ρ generations ago.
r_{ij}	Expected age of the cultural variant occupying P_i given that the variant was invented in P_j.
c_{kl}	Probability that P_k invents C_l. (finite-variant model)

$\varphi_k(l,\rho)$	Probability that the cultural variant occupying P_k is C_l and that was invented ρ generations ago.
$\varphi_k(l)$	Probability that C_l occupies P_k.
$r_k(l)$	Expected age of the cultural variant occupying P_k given that the variant is C_l.
$\varphi_{ij}(l)$	Probability that the cultural variant occupying P_i is C_l and that was invented in P_j.
$x_{ij}(l)$	Probability that the cultural variant occupying P_i was invented in P_j, given that the variant is C_l.

Figure legend.

Figure 1

Example of cultural transmission among populations. (blue ellipses) populations. (circles) cultural variants. Different colors correspond to different variants. (arrows) cultural transmission. This example illustrates a simple case of $n = 2$ through three consecutive generations with parameters $a_{11} = a_{22} = 0.6, a_{12} = a_{21} = 0.2, b_1 = b_2 = 0.2$. Blue, red, yellow, and green circles indicate new cultural variants invented after generation t. Although the figure displays only five variants in each population, our model assumes infinite number of variants.

Figure 2

Example of backward ancestral process. (blue ellipses) populations. (red circles) sampled variant and its ancestors. (green arrows) cultural transmission. This example illustrates the case of $n = 3$.

Figure 3

Example of a random network with $n = 5$. (blue circles) populations. (red arrows) cultural transmission. Each population is labeled with two integers within the round brackets, the first and second integers of which represent the in-degree and out-degree, respectively. Note that the direction of an arrow corresponds to the direction in the ancestral process. For example, the arrow directed from P_2 to P_3 signifies that P_2 learns from P_3, that P_3 influences P_2, or that a cultural variant in P_2 may have been in P_3 in the previous generation.

Figure 1

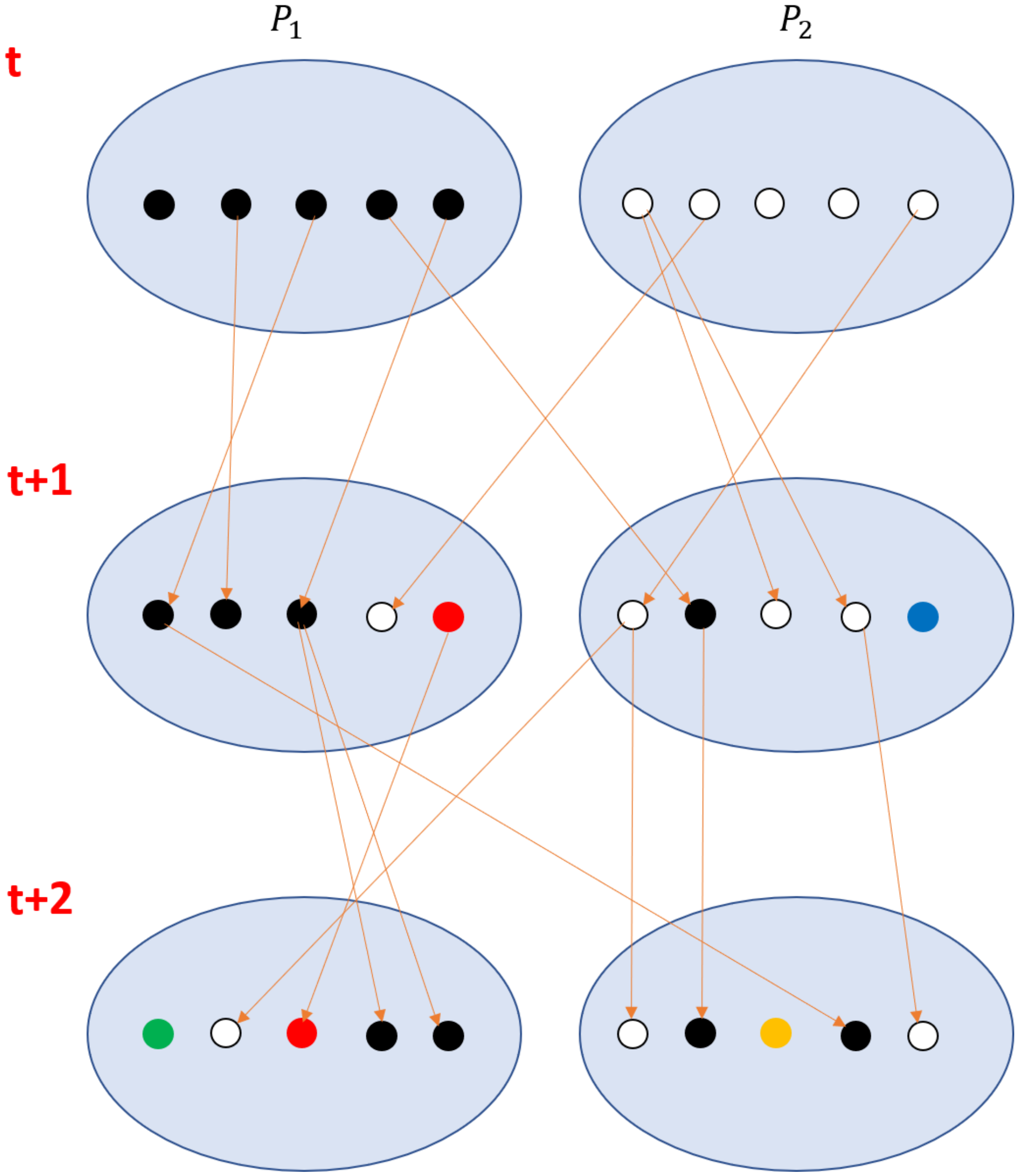


Figure 2

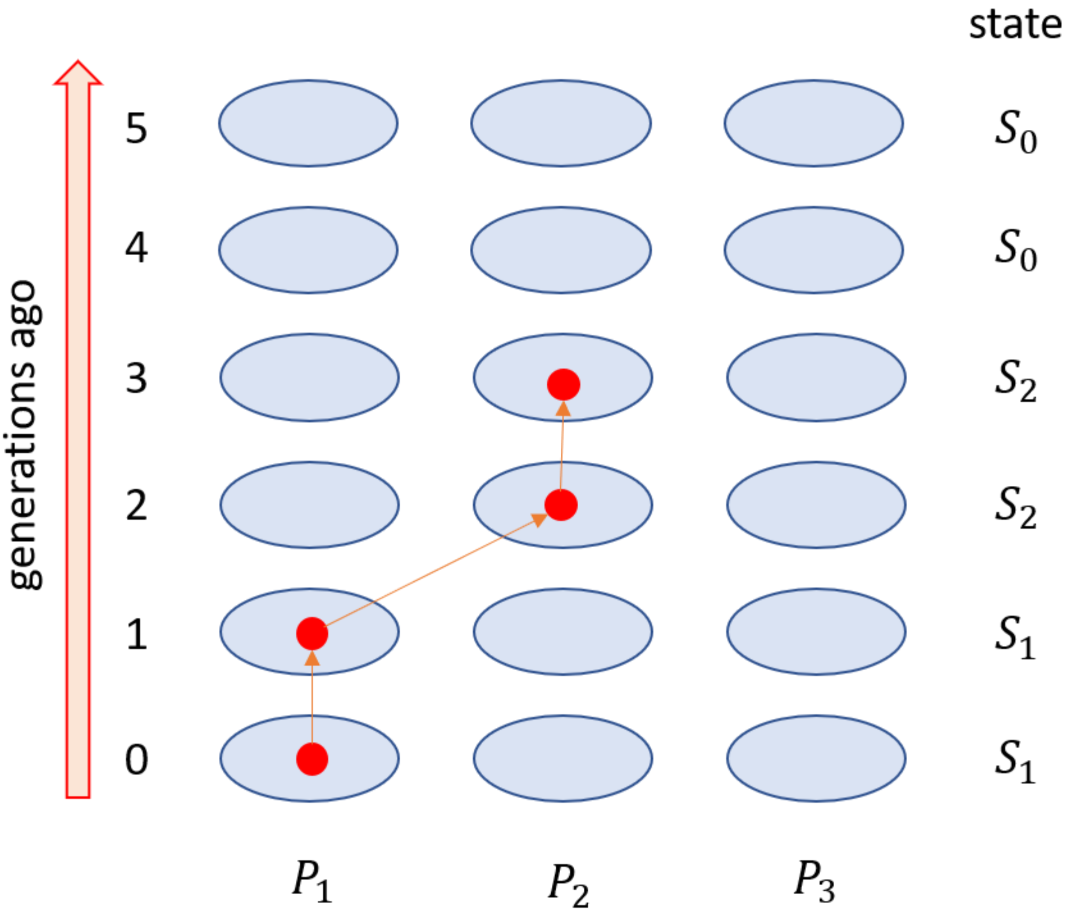
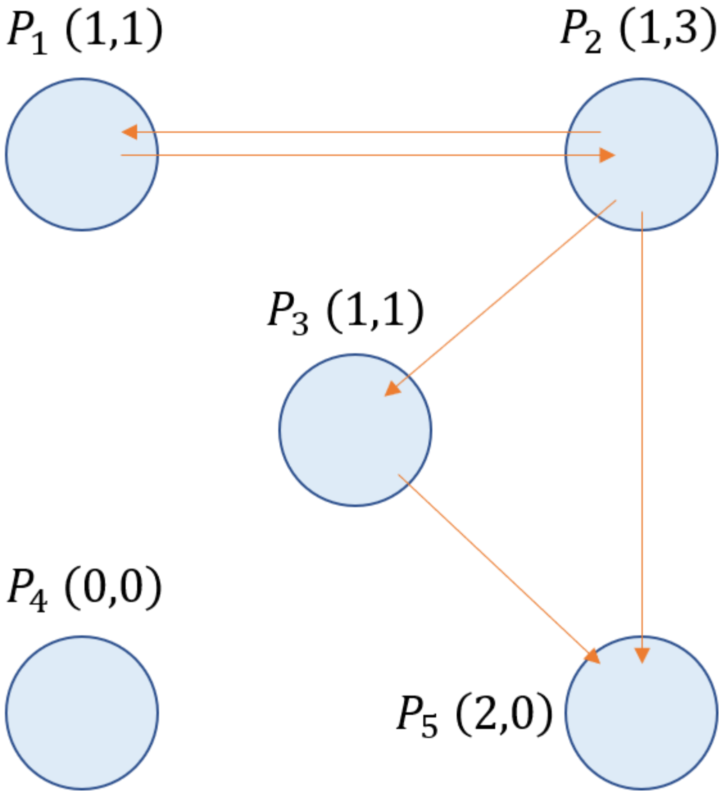


Figure 3



Chapter 4.

Analyzing the spatial evolution of Japanese lexical variation through network model

1. Introduction

We observe an extremely wide diversity in the languages all over the world. In addition to the variation on the scale of languages, each language contains dialect words which differ from one another in terms of phonology, syntax, morphology, and lexicon. A large body of classical linguistic literature records the geographic distribution pattern of linguistic variation on linguistic atlases based on nation-wide surveys, displaying the variants of linguistic features on a map [1-3].

Two processes underlie the spatial evolution of linguistic traits: migration and spatial interaction. As for the former, when a large population diverges and its subgroup inhabits in a new area, the newly arisen population accumulates the mutation of linguistic features and its language becomes more and more distant from the original population. Indeed, linguistic distance between varieties of Ryukyuan language was shown to correlate with time since divergence of populations [4]. As for the spatial interaction, on the other hand, individuals adjust their language use to adapt their speech patterns to their speech partners in their respective speech community [5]. Since human interaction occurs most frequently among individuals with geographical proximity, linguistic traits spread gradually from one place to its neighbors, resulting in a spatially continuous spectrum of language distribution. Trudgill's *gravity model* [6], which is analogous to Newton's law of gravity, quantifies linguistic interaction between localities, assuming that the influence of one city on another is given by the product of the population sizes of two cities divided by the squared distance between them. The model predicts a characteristic phenomenon called *urban jumping*, in which a newly invented linguistic variant first spreads from a large city to another, reaching the smaller towns and villages with a certain time lag, even if they are geographically more proximate to the place of invention (see also [7]).

On the spatial pattern of linguistic variation, much research performed more quantitative analysis on empirical data and found a positive correlation between geographic distance (GD) and linguistic distance (LD), demonstrating that linguistic traits between two localities become more and more different with the geographic distance in-between. For example, the dissimilarity between Japanese lexical variants was shown to correlate with geographical distance [4, 8], especially when the latter was calculated by travel time between localities instead of naïve linear distance [8]. In addition, based on phonological variation in vowels among English dialects, [9] performed a cluster analysis and showed geographically proximate locations tend to belong to the same cluster. Other research discovered that linguistic variation increases with geographical distance, consistently in six different countries and areas (reviewed in [10]). Starting from well-known *Séguy's curve* [11], most research above reported sublinear growth of linguistic distance as a function of geographic distance [8, 10]. However, variation of syntactic and morphological features of British English was reported to be only slightly explained by the geographical distance [12]. Nevertheless, it is safe to say that the correlation of geographical and lexical distances is a ubiquitous phenomenon observed in a number of languages and countries.

Although the empirical studies mentioned above have discovered the relationship between geography and language in a statistical and quantitative way, they do not quantitatively explain the temporal evolution of language distribution. Since corpora on which research in linguistic geography relies, such as linguistic atlases, are merely a snapshot taken at one moment and do not explain how the spatial pattern has evolved. Considering the tremendous costs that nation-wide linguistic research incurs, it is almost impossible to record the linguistic variants extensively in both temporal and spatial manner.

To investigate the spatial and temporal dynamics of linguistic variation, mathematical models have often been used. For example, Burridge [7, 13] established models of language diffusion based on the geographic contact pattern between populations, showing the temporal evolution of isoglosses on a real map. The models are a breakthrough in the sense that they mathematically demonstrate the *urban jumping* theory assuming the influence between localities as a function of population sizes and geographic distance. From a different perspective, transmission of linguistic traits is often simulated on a network of individuals [14, 15] or populations (Chapter 2), and its relationship with the topological variables is numerically investigated.

One crucial limitation of mathematical models is that they are not always tested by means of empirical data, which makes it difficult to establish the correspondence with reality. Although mathematical models may give insight into the spatial and temporal dynamics of language variation, it often remains to be theoretical. To fully understand the dynamics of dialect distribution, therefore, it is necessary to combine an empirical study and mathematical model, compensating for the limitations of each of them.

In this chapter, we will study the evolution of the geographic distribution of

Japanese lexical variation based on both empirical data and a mathematical model and discuss how the observed spatial pattern is explained by the model. In particular, the main purpose of this paper is to capture the underlying factors which form séguy's curve [11], that is, the sublinear growth of linguistic distance with geographic distance [10]. For empirical data, we utilize Linguistic Atlas of Japan (LAJ) [1] and its electronic version called Linguistic Atlas of Japan Database (LAJDB) [16], from which Jeszenvsky et al. [8] correlated linguistic distance and geographic distance calculated by multiple metrics. To take historical relationship of linguistic variants into consideration, they calculated linguistic distance based on the variant categories which class similar lexical variants together for every survey item. Their method enables us to reflect historical relation of lexical variants but may fail to eliminate subjective bias in judging the distance among them. Indeed, the category does not necessarily correspond to the cognate groups of lexicons, and the way to group lexicons is inconsistent among questions [1]. In our research, we will calculate the linguistic distance among localities by edit (Levenshtein) distance [17], which has been extensively used in calculating linguistic distance in dialectology [4].

After discussing the relationship between geographic and linguistic distances obtained from empirical data, we will perform a numerical analysis and computer simulation of language diffusion based on a learning model on the network of the 2400 survey localities in LAJ project. Lexical variants originating from a common ancestral word tend to have a smaller Levenshtein distance than pairs of non-cognate words, so analyzing the origin of lexical variants is an effective method which sheds light to the relationship between geographic and linguistic distances. For this purpose, we will apply the network model of Chapter 3, which estimates origin and expected age of cultural traits. Application of this model to the network of LAJ localities helps us to infer when and where lexical variants were invented as well as the velocity at which a lexical variant diffuses. In this model, which treats the dynamics of cultural traits [18] within the framework of backward ancestral process [19], each node (locality) of the network may either invent a new lexical variant or learn an extant variant from any locality.

In addition to discussing variants' origin and speed of diffusion, we will simulate linguistic distance on the network of the survey localities of LAJ project. To represent the dialect form at each locality, we will assign one string to each node of the network and simulate the spatial and temporal evolution of the strings over generations. We will discuss again the relationship between linguistic distance (LD) and geographic distance (GD) among localities and infer what parameter influences the spatial distribution of linguistic traits.

The structure of this chapter is as follows. We will first analyze the empirical lexical data on the basis of LAJDB in “Analysis of empirical lexical data” section, calculating the linguistic distance among localities. In “Mathematical model and simulation” section, we will apply the model of Chapter 3 and perform numerical analysis and simulation to quantify the origin and origin-dependent age of lexical variants, as well as linguistic distance of every locality. Finally, in “Estimation of parameter values” section, we will perform approximate Bayesian computation (ABC) to infer parameter values of the mathematical model based on the empirical data.

2. Analysis of empirical lexical data

2-1 Material

We analyze the lexical data recorded in Linguistic Atlas of Japan (LAJ) [1]. LAJ, a linguistic atlas published in 6 volumes during the period of 1966 to 1974, is an extensive project of language investigation in which 285 linguistic items were surveyed at 2400 localities throughout the Japanese Archipelago, including Okinawa and small islands (see Figure 1). Most of the surveyed items are lexical variation including nouns, verbs, and adjectives, such as name of animals, plants, and body parts, which are frequently used in daily life. The survey was conducted in the form of interview, and one respondent was interviewed at each of the 2400 localities. Respondents were mostly elderly men (2392 males, 8 females).

In the current study, we analyze the lexical data of 103 LAJ-items available on Linguistic Atlas of Japan Database (LAJDB) [16], an electronic database which archives the lexical data of LAJ.

2-2 Calculation of linguistic distances among localities

Based on the recorded lexical variants, we calculate the linguistic distances among 2400 localities. Jeszenvsky et al.’s [8] study, which was also based on LAJDB, calculated the distance between variants, relying on the category of variants displayed on LAJ maps [1] to consider the historical relationships of variants. For example, words meaning *face* are classified into three categories *kao*, *cura*, and *omote*, each of which includes several variants (see https://mmsrv.ninjal.ac.jp/laj_map/data/laj_map/LAJ_106.pdf). Jeszenvsky et al. [8] assigned a linguistic distance of 1 to variant pairs which belong to different categories, while linguistic distance between different variants belonging to the same category was considered 0.2. However, the category does not necessarily correspond to the cognate groups of lexicons, and the way of grouping seems inconsistent among items. To eliminate subjective bias in calculating the distance between variants, we instead use normalized Levenshtein distance [17], which is often employed in studies of dialectology [4]. Normalized Levenshtein distance, defined for any pair of two strings, is the minimum number of one-letter substitutions, deletions, and additions which are necessary to change one string into the other, divided by the length of the longer string.

Letting $L_1 \dots L_{2400}$ denote the 2400 surveyed localities of LAJ, linguistic distance between two localities L_i and L_j are the mean value of normalized Levenshtein distance calculated for the 103 items. We will thereby obtain a 2400-dimensional distance matrix.

2-3 Clustering analysis

Having obtained the 2400×2400 linguistic distance matrix, we performed multidimensional scaling (MDS) and plotted the 2400 localities on a two-dimensional plain (Figure 2a), which enables us to visualize the linguistic similarity among localities. The Japanese Archipelago is conventionally divided into nine regions. The figure shows that dialects within the same region are clustered together, and that positions indicated by the MDS are largely in concordance with the real geography of Japan (Figure 2b), suggesting that geographically distant localities are linguistically distant as well.

An outstanding exception to this finding is Hokkaido, which is geographically positioned next to the Tohoku region but linguistically closer to the Chubu and Kanto regions according to the MDS analysis. The result is quite straightforward because a mass immigration occurred from Japanese mainland to Hokkaido near the end of 19th century. Indeed, principal component analysis based on genetic polymorphism of Japanese populations has revealed that Hokkaido is genetically close to prefectures in the Kanto region [20].

2-3 Relationship between linguistic and geographic distance

The histogram of linguistic distance between every locality pair is shown in Figure 3a, with mean linguistic distance being $0.442 \pm 0.106\ (SD)$. The histogram shows the highest peak in the range of 0.40-0.45 and smaller peak in 0.70-0.75, whereas locality pairs with linguistic distance less than 0.2 are quite rare. In Figure 3b, the linguistic distances (LD) between all the pairs of localities are plotted against geographical distances (GD), or more precisely, great circle distances, the shortest distance on the sphere surface of the Earth. As Figure 3b shows, linguistic distance increases with the geographic distance, and their correlation coefficients, shown in Table 1, indicates a strong correlation between geographic and linguistic distances. Figure 3b suggests a sublinear growth of linguistic distance as a function of geographic distance, which is in line with previous studies that drew a logarithmic curve to regress the relation between geographic and linguistic distance [8,10]. However, the results indicate that correlation coefficient stays almost the same whether we logarithmically transform geographic distance or not ($r = 0.689$ with GD, and $r = 0.678$ with log(GD)), which contradicts [8]. To visualize the linguistic distance from a fixed locality which we call *reference site*, we select the most populated locality in Tokyo and Kyoto prefectures and plot the linguistic and geographic distances from these two localities (Figure 4). The figure shows that growth of linguistic distance from Tokyo is visibly asymmetric depending on whether going eastward (plotted blue) or westward (plotted red) from the reference site (Figure 4a). In the eastward direction, linguistic distance from Tokyo is maximal around 600 km of geographic distance (Tohoku region), but localities to the west of Tokyo shows a lower value of linguistic distance at 600 km.

Considering each region separately, correlation between geographic and linguistic distances is strong in Kyushu and Okinawa and weak in Hokkaido and Shikoku areas (Table 1). As for Kyushu, Okinawa, and Hokkaido, this feature is quite in line with Jeszenvsky et al.'s [8] work, which is also based on LAJDB. Modest correlation in Hokkaido is plausibly explained by recent migration [21]; lexical features vary due to the origin of immigrants' ancestors, which decreases the variance of linguistic distance explained by geographic distance. In contrast, weak correlation in Shikoku entirely contradicts Jeszenvsky et al.'s [8] result, in which Shikoku scored the largest correlation coefficient between linguistic distance and great circle distance ($r = 0.7391$ between LD and GD, and $r = 0.7824$ between LD and log(GD)).Since both studies are founded on LAJDB, this discrepancy is due to either the size of dataset (this study: 103 items, [8]: 37 items) or measure of linguistic distance (this study: normalized Levenshtein distance, [8]: variant category in LAJ map). As for the former, we calculated normalized Levenshtein distance from the same set of 37 LAJ-items analyzed in Jeszenvsky et al.'s [8] work but still observed a low correlation with geographic distance in the Shikoku area ($r = 0.326$ with GD and $r = 0.412$ with log(GD)). Thus, the difference of two studies is attributable to the different measures of linguistic distance, even though it is unknown why such a stark discrepancy was observed exclusively in the Shikoku region.

3. Mathematical model and simulation

In the previous section, we analyzed the empirical data of Japanese lexical variation, placing emphasis on the relationship between geographic and linguistic distances. In this section, on the other hand, we will introduce a mathematical model and simulate how lexical variants diffuse among localities. Since a pair of cognate variants originating from a common ancestor bears a higher similarity than a pair of variants invented independently, it is of great importance to quantify where and when lexical variants were invented in the network. To explore factors underlying the formation of Séguy's curve, we will first consider the origin and diffusion time of lexical variants, and subsequently simulate the diffusion of lexical variants represented by strings.

3-1 Description of network model

We apply the network model of Chapter 3 to the network of 2400 LAJ localities denoted $L_1, \dots L_{2400}$, each of which is occupied by one lexical variant. Lexical variants represent different word forms referring to the same object or meaning, such as different words for a single surveyed item of the LAJ project. This is analogous to alleles on a single locus in population genetics. While the model of Chapter 3 originally assumed an infinite number of cultural variants occupying each vertex of the network, the model can also be interpreted as a one-vertex-one-variant model (see Discussion of Chapter 3). In each generation of the model, every locality updates its lexical variant either by learning an extant variant from one locality, including itself, or by innovating a novel variant. The probability with which L_i learns a variant from L_j is denoted by a_{ij}, which we call *transmission rate* from L_j to L_i, and the probability that L_i invents a new lexical variant, which we refer to as *innovation rate*, is denoted by b_i. By definition, we have $a_{i1} + \cdots a_{in} + b_i = 1$ for every *i*.

We assume the transmission rates a_{ij}, which specifies the topological structure of the network, depend on geographical distance and population size. Namely, we follow Burridge's [13] theory of *interaction density*, which assumes a Gaussian function of geographic distance weighted by population size; an individual dwelling at L_i interacts with individuals at L_j with frequency given by

$$\varphi_{ij} = \frac{P_j}{2\pi\sigma^2} exp\left(-\frac{d_{ij}^2}{2\sigma^2}\right), \quad (1)$$

where P_j denotes the population size of L_j, and d_{ij} denotes the geographic distance between the localities. Parameter σ, the interaction range, is associated with the dependence of interaction density on geographic distance; when σ is small, individuals tend to interact only with people living nearby while higher value of σ permits a long-distanced interaction. As it is plausible that speakers are linguistically influenced by localities with frequent interaction, transmission rate from L_j to L_i is proportional to the ratio of φ_{ij} to the whole interaction that L_i experiences. Formally, we have

$$a_{ij} = \frac{\varphi_{ij}}{\sum_{l=1}^{2400} \varphi_{il}}(1 - b_i), \qquad (2)$$

from which we obtain a 2400×2400 transmission matrix $\boldsymbol{A} = (a_{ij})$, consisting of the transmission rates between every locality pair.

To estimate relative population sizes of LAJ localities, we use the mesh data of a demographic survey carried out in 1995 (see https://www.e-stat.go.jp/), which is the earliest data set of this kind. Although the population sizes in 1995 were certainly different from those at the time of word diffusion, what is relevant to (1) and (2) is relative, rather than absolute, population size. We assume that modern population size provides a sufficiently good approximation, which is true if the population at each locality has grown at a similar rate. Mesh data record the number of residents living in every $1km \times 1km$ square of the land, and we regard P_i as the mean value of population sizes of $25\ (= 5 \times 5)$ squares centered at L_i (Figure 5).

We assume innovation rate $b_i = b$ is constant for every locality, and to make the generation in our model comparable with realistic generation as the length of human lifecycle, we use the per-year innovation rate of lexicons from previous studies. Based on a phylogenic method, Pagel and Meade [22] estimated the replacement rate of fundamental vocabulary on Swadesh list [23] was on average 0.00020 per year in the Indo-European family, which is converted to 0.005 per generation, provided one generation is approximately 25 years. Although replacement rate varies from word to word [22] depending on the frequency of word use [24], we basically use $b_i = b = 0.005$ unless otherwise stipulated to consider the average scenario of language change.

3-2 Origin and age of lexical variants

The dynamics will reach an equilibrium state. Letting x_{ij} denote the equilibrium probability that the lexical variant used in L_i was invented in L_j, Chapter 3 showed

$$\boldsymbol{X} = (x_{ij}) = (\boldsymbol{E} - \boldsymbol{A})^{-1} diag(b_1 \ldots b_{2400}), \quad (3)$$

where $\boldsymbol{E}$ represents 2400-dimensional identity matrix. Following (3), we calculate the value x_{ij} for every ordered pair of i and j, and the probability that the variant originates from the most populated locality of Tokyo and Kyoto prefectures is plotted against geographic distance in Figure 6. L_j being fixed to either one of the two reference sites, the figure shows that x_{ij} drops with geographic distance but stays relatively constant when σ is large and distanced transmission is possible. Intriguingly, when we fix L_j to a locality in Tokyo, x_{ij} drops more rapidly in the west than in the east (Figure 6a, b), which is most prominently observed with small values of σ. Conversely, when we fix L_j to a locality in Kyoto, x_{ij} drops more rapidly in the east than in the west (Figure 6e, f). Hence, diffusion of lexical variants is not isotropic, and localities in Kyoto and Tokyo are not frequently occupied by a lexical variant originating from each other. This is interpreted as follows; since Tokyo and Kyoto-Osaka areas are densely populated (Figure 5), nearby localities often learn lexical variants from these localities. Thus, if a lexical variant invented in Tokyo diffuses westward, they are likely to be overridden by a variant originating from Kyoto and Osaka areas, and the opposite is also true if a lexical variant invented in Kyoto diffuses eastward. On the other hand, when σ is large, the figure suggests x_{ij} is largely uniform in Japanese mainland but still sharply decreases in Amami and Okinawa islands (plotted black), indicating that any locality on the mainland contains variants invented in Tokyo and Kyoto with a similar probability but lexical variants in Amami and Okinawa islands are less likely to derive from Tokyo and Kyoto.

We visualize the effect of population size on the value of x_{ij} in Figure 7. In addition to the locality with largest population size of Tokyo and Kyoto (Figure 7a, c), we choose less populated localities (Figure 7b, d) as reference sites. For Tokyo, we select a locality with population size 250.4 situated at 56.4 km away from the most populated locality (Figure 7b), and for Kyoto, we select a locality with population size 198.8 situated 18.4 km away from the most populated locality (Figure 7d). Figure7 clearly shows that variants are likely to derive from densely populated localities in comparison with localities with a smaller population size.

We correlate x_{ij} with the distance between two localities and their population sizes (Table 2). The result shows that x_{ij} tends to be large when L_i and L_j are located close to each other, and that x_{ij} positively correlates with the population size of L_j. In other words, a lexical variant is likely to derive from a locality which is located nearby or/and which is of a large population size. In contrast, we find almost no correlation between x_{ij} and population size of L_i ($|r| < 10^{-6}$ for every parameter set). Table 2 also shows that the logarithm of x_{ij} is negatively correlated with geographic distance indicating that the frequency of variants invented at a fixed locality decreases exponentially with geographic distance. Strength of correlation depends on the parameter values, and correlation between $\log(x_{ij})$ and geographic distance tends to be stronger with larger value of b, meaning that origin of linguistic variants is heavily influenced by geographic distance when localities invent new variants frequently. In contrast, correlation between x_{ij} and population size of L_j is strong if σ is large, or in other words, if variants can transmit between distant localities.

The variable s_j calculated by

$$s_j = \sum_{i=1}^{2400} x_{ij} \qquad (4)$$

indicates the expected number of localities whose lexical variant originates from L_j. We plot the value of s_j against population size P_j (Figure 8). As the figure depicts, s_j and P_j have a strong linear relationship $(r = 0.956 \text{ when } \sigma = 150, b = 0.005)$.

Besides origin of linguistic variants, we discuss the time it takes for a variant to transmit between two localities. Let r_{ij} represent the expected age (elapsed time since invention event) of the variant occupying L_i, given it was invented in L_j. At the equilibrium state, Chapter 3 showed

$$\boldsymbol{R} = (r_{ij}) = (\boldsymbol{E} - \boldsymbol{A})^{-1}(\boldsymbol{E} - \boldsymbol{A})^{-1}\boldsymbol{A}diag(b_1 \dots b_{2400}) \oslash \boldsymbol{X}, \qquad (5)$$

where $\oslash$ notates Hadamard (element-wise) division of matrices. Hence, (3) and (5) enable us to compute the value of r_{ij} for every ordered pair of i and j. This value is an estimator of diffusion time from L_j to L_i, but strictly speaking, the value represents the expected age of the variant in L_i given that it was invented in L_j, so it may deviate from the diffusion time because the variants may have stayed in L_i for many generations. Figure 9 and 10 shows the expected age of lexical variant given that the variant was invented at a fixed reference site. Expected age of variants originating at a fixed locality increases gradually toward its periphery and shows a concentric pattern (Figure 10). However, as with the value of x_{ij}, growth of r_{ij} is asymmetric especially when σ is small; variants invented in Tokyo diffuse more rapidly eastward than westward (Figure 9a), and those invented in Kyoto diffuse rapidly to the west (Figure 9e). Unlike x_{ij}, expected age of variants hardly depends on the population size of the reference site.

Again, we correlate the expected age r_{ij} with variables of each locality (Table 3a) and found an extremely strong correlation between r_{ij} and geographic distance. In contrast, population sizes of L_i (locality where the variant currently exists) and L_j (locality where the variant was invented) are almost irrelevant to the expected age. We perform a simple linear regression assuming the model:

$$r_{ij} = \beta_0 + \beta_1 d_{ij}, \quad (6)$$

where coefficients β_0 *and* β_1 are estimated in Table 3b. Based on the estimated coefficient, the value $1/\beta_1$, distance with which expected age increases by one generation, is shown to be larger if innovation rate is smaller or if variants can transmit between distant localities. The value is used as an estimator of the distance which lexical variants advance in one generation. Thus, variants advance rapidly when transmission occurs between remote localities and when new variants are invented frequently.

3-3 Simulation of linguistic distance

Besides origin and expected ages of variants, we will simulate the linguistic distance for each locality pair. In this simulation, a linguistic variant is represented by a string of Latin alphabets, which is invented and transmitted as described in subsection 3-1. In the initial generation, each locality is occupied by an independently generated random string, whose length follows the uniform integer distribution of 1-12. When an invention occurs in a locality, the old string is replaced by another random string. If a locality learns an extant variant from any locality, the string experiences a one-letter substitution, deletion, or addition with probability μ for each (i.e., total rate of mutation: 3μ). However, deletions and additions which push the string length out of the range of 1 to 12 do not occur.

We simulate the diffusion of variants and record normalized Levenshtein distance of strings for every pair of localities 100 times at the interval of $5/b$ generations to let the system approach equilibrium. This simulation is practically equivalent to computing Levenshtein distance of words for 100 different meanings, such as 100 surveyed items of LAJ, assuming that they all diffuse independently. Simulated linguistic distance, defined to be the mean value of normalized Levenshtein distances calculated 100 times, is then plotted on a map from a reference site situated in Tokyo (Figure 11). The figure shows that linguistic distance increases concentrically from the reference site, especially when lexical variants cannot transmit between remote localities (i.e., small value of σ). However, the simulation does not replicate the empirically observed fact that Hokkaido is more linguistically similar to Tokyo than the Tohoku region (see Figure2).

Figure 12 depicts a scatter plot of linguistic distance (LD) for every locality pair against geographic distance (GD). When σ is small, linguistic distance increases with geographic distance in a decelerating manner (Figure 12a), which is observed in the analysis of LAJ data (see Figure 3b). On other hand, when σ is large, LD stays largely constant independent of GD, and LD for majority of locality pairs are within the range of 0.6 to 0.8 (Figure 12b). Linguistic distance from the most populated locality in Tokyo and Kyoto is plotted in Figure 13, with three values of σ. When $\sigma = 50km$, accumulation of linguistic distance is not isotopic in eastward and westward directions; the locality in Tokyo tends to be linguistically closer to eastern localities than western localities, and the locality in Kyoto tends to be linguistically closer to western localities (i.e., Chugoku, Shikoku, and Kyushu regions) than eastern ones. This finding is in line with the distribution of x_{ij} in Figure 6. Variants invented in Tokyo is not as likely to occupy western localities as eastern localities (Figure 6a), so it is suggested that variants at western localities are not likely to be cognate with the variant of Tokyo and thereby score a large value of Levenshtein distance.

Unsurprisingly, linguistic distance (LD) is positively correlated with geographic distance (GD) in every parameter set, but the correlation coefficient varies depending on the value of parameters (Figure 14). Namely, assuming $\mu = 10^{-4}$, correlation coefficient is maximum at $\sigma = 50$. It is also suggested by the result that LD correlates with log(GD) more strongly than with GD when σ is small, and vice versa when σ is large.

4. Estimation of the parameter values of the model

We have so far discussed the spatial distribution of linguistic variants empirically and simulated the temporal and spatial evolution of variants using a network model. To integrate empirical and simulated research, we will estimate the parameter values of our mathematical model (see subsection 3-3), based on the empirical linguistic distance matrix obtained in section 2.

4-1 ABC-rejection algorithm with MCMC sampling

We estimate the parameter values of the model under the framework of approximate Bayesian computation (ABC) with Markov Chain Monte Carlo (MCMC) sampling [25]. Our model has three parameter values (i.e., $\sigma, \mu,$ and b), but we use a new parameter $m = 3\mu$, total rate of mutation, instead of μ itself. We use a uniform distribution of the range 10 to 500 (km) as the prior distribution of σ and exponential distribution of the range 10^{-5} to 0.1 as the prior distributions of b and m. Based on the prior distributions and MCMC sampling, we sequentially generate parameter sets and conduct simulation.

For each parameter set, we simulate the transmission of lexical variants (strings) as described in 3-1, but to reduce the time complexity, simulation is conducted via ancestral process, whereby we trace the ancestors of 2400 variants in the present. For each parameter set, we calculate Levenshtein distances among variants 100 times and assign linguistic distance of locality pairs as the mean value of 100 simulations, so the simulation is equivalent to investigating the lexical variation with regard to 100 meanings. As it is costly to compute linguistic distance for all $2400 \times 2399/2 \approx 2.88 \times 10^6$ locality pairs, we randomly sample 10^5 locality pairs and compute their linguistic distances. Let $\widehat{D_i}$ denote the simulated linguistic distance of i-th sampled locality pair and D_i be its observed linguistic distance calculated from the empirical data of Linguistic Atlas of Japan Database (LAJDB) in section 2. Here, we calculate the mean square error (MSE):

$$MSE = \frac{1}{10^5}\sum_{i=1}^{10^5}\left(D_i - \widehat{D_i}\right)^2.$$

Accepting parameter sets which satisfy $MSE < 0.01$, we estimate the posterior distribution of parameter values.

4-2 Result

We performed MCMC sampling using five independent chains, each of which sequentially samples 5000 parameter sets. Hence, linguistic distance was simulated based on 25000 parameter sets, and 16156 parameter sets were accepted according to the error level of $MSE < 0.01$. Within the accepted parameter sets, σ, interaction range, shows little correlation with innovation and mutation rates ($r = 0.00045$ between σ and b, $r = -0.11$ between σ and m). Thus, we independently drew the posterior distribution of σ in Figure 15a, which displays two peaks around 70 km and 140 km, respectively. The distribution shows a median value of 116 km and 95% confidence interval was 59 to 178 km.

On the other hand, innovation and mutation rates which satisfied $MSE < 0.01$ were closely related, and Figure 15b represents the scatter plot of b and m which were accepted by ABC-rejection algorithm. The figure shows that the lower innovation rate is, the higher mutation rate must be in order that the simulated linguistic distance is consistent with empirical data.

5. Discussion

In this chapter, we explored the relationship between geographic and linguistic distances by means of both empirical and theoretical methods. First, analysis of lexical data in LAJDB led to conclusion that linguistic distance was strongly correlated with geographic distance, and geographically close localities formed clusters through multidimensional scaling based on empirical linguistic distance (Figure 2). On the other hand, we applied the mathematical model which we developed in Chapter 3 to a network of the LAJ localities and numerically obtained the probability of lexical variant deriving from each population and origin-dependent age of lexical variants at each locality. Subsequently, we simulated the diffusion of lexical variants (strings) and analyzed the linguistic distance between every locality pair and correlated it with geographic distance (GD). Finally, we estimated the parameter values of the model by ABC (approximate Bayesian computation) algorithm.

As for the empirical study of LAJDB, large part of our results is in line with the previous research [8]. Novelty that the current study offers is rather the metrics of linguistic distance (i.e., normalized Levenshtein distance) which is different from Jeszenszky et al.'s work [8] based on the variant categories of LAJ map. One discrepancy between our and Jeszenszky et al.'s [8] research is the correlation between LD and GD in Shikoku region; our research indicated only a modest correlation (Table 1) whereas [8] reported the highest correlation coefficient in this region. Also, our study into LAJ data did not show a higher correlation of LD with logarithm of GD compared to that with GD itself (Table.1), which was reported in multiple studies [8, 10]. Nevertheless, sublinear increase of LD as a function of GD was observed (Figure 3b), which is reminiscent of Séguy's curve [11].

Concerning the numerical analysis of origins and expected age of variants, our model suggested that expected age of variants increases concentrically from the place of invention. It was suggested that expected age increases quite linearly and the speed at which variants advance was estimated through regression (Table 3b). This result is inconsistent with Chapter 2 and another previous study which simulated the dynamics of Japanese dialects spreading from Kyoto [26], because these models predicted that expected age of words increases rapidly near the origin but more slowly after the words have diffused far away. The difference is probably due to the center-and-periphery structure which the Chapter 2 and the previous study [26] assumed: in their study, new words arise only at one location of network (Chapter 2) or 2D-lattice [26] and diffuse without interference of words invented at different places. In contrast, although localities are weighted by their population size, they all have the same probability to invent new words in our model. It may be suggested that the diffusion speed may be uniform when all the populations create new words with the same probability, whereas the diffusion may become slow in periphery when the network has a center-and-periphery structure.

Simulating the transmission of strings have revealed what factors exert influence on the correlation between GD and LD; the correlation is large when the interaction range is moderate. This is interpreted as follows; when localities only interact with the nearby localities (small value of σ), each locality tends to contain variants originating from a nearby locality, so LD steeply skyrockets even for small GD and stays constant for sufficiently large GD. Conversely, when localities interact with remote localities independently of GD (large value of σ), they often accept variants originating from remote populations and LD does not depend much on GD. In either case, variance of GD does not explain much of the variance of LD, decreasing the value of correlation coefficient. The result also shows that LD correlates more strongly with logarithm of GD than GD itself when σ is small, suggesting the appearance of Séguy's curve [11] is conditional on variants transmitting only between proximate localities.

Regarding the emergence of Séguy's curve, we posit that the sublinear growth of linguistic distance is associated with the exponential decay of the value x_{ij} from a fixed origin or reference cite. When σ is small, $log(x_{ij})$ shows a strong negative correlation with geographic distance (see Table 2), and the value of x_{ij} drops sharply near the reference site but stays relatively constant in the periphery (Figure 6a, e). Since variants which derive from a common ancestral word usually have a small Levenshtein distance, exponential decrease of x_{ij} may in turn result in the sublinear increase of linguistic distance. However, note that x_{ij} denotes the probability that variant at L_i originates from L_j and is different from the probability with which two localities have variants deriving from a common ancestral variant created by a single invention event.

We found that spatial patterns of x_{ij}, r_{ij}, and simulated linguistic distance from a reference site are not symmetric to every direction when interaction range is small (Figure 6, 9, 13). As for the origin, lexical variants invented in Tokyo and Kyoto are likely to spread eastward and westward, respectively (Figure 6a, e), approaching both edges of the archipelago. In addition, simulated linguistic distance from Tokyo and Kyoto grows more rapidly in the west and east, respectively (Figure 13a, e). However, we cannot observe this feature in the scatter plot of linguistic and geographic distances calculated from the empirical data (Figure 4), and small values of $\sigma(< 50\,km)$ were mostly rejected by ABC-MCMC method (Figure 15). Although it is intriguing that small interaction range breaks the isotropy of linguistic distance, this finding remains theoretical and cannot be observed in the empirical data of Japanese lexical variation.

We discuss limitations of our research. First, we used Levenshtein distance as a metric to measure the distance among lexical variants, because unlike Jeszenszky et al.'s [8] linguistic distance based on variant category, Levenshtein distance can eliminate subjective bias, and because it can be computed in our simulation. Nevertheless, it is still questionable if this criterion is appropriate since some research shows that Levenshtein distance does not necessarily assess historical or phylogenic relationship of languages [27].

As for limitations of mathematical model, a network of 2400 LAJ localities might not represent the real geography of Japan because there are many more populations which may potentially invent or transmit lexical variants. Ideally, we should conduct numerical analysis and simulation on a network of lattice sites which divide Japanese Island into fine meshes and subsequently extract data of lattice sites corresponding to the LAJ localities (see [28] for this class of research). The problem is that calculation of (3) and (5) takes $O(n^3)$ time where n denotes the number of vertices in the network, so increasing the number of nodes will make numerical analysis infeasible. Although the network of LAJ localities may overleap the effect of populations which were not surveyed by the project, it does not seriously affect the result of our model because survey localities in LAJ project were selected at somewhat regular interval, if not perfectly equally spaced. Another limitation of our model is that lexical variants are assumed to diffuse only by interaction and learning of speakers, and the effect of migration is completely omitted. Indeed, the model fails to replicate the observed spatial pattern of variants in Hokkaido (Figure 2a), which was plausibly formed by the recent migration event [21]. In addition, formulae (3) and (5) assume the equilibrium state, by which lexical variants have transmitted for a sufficiently long time. As for simulation, we calculated Levenshtein distance several times at a regular interval, so that the system approaches the equilibrium state. However, it is still questionable whether the dynamics of empirical linguistic variants is at the equilibrium state, and a mathematical model which predicts the dynamics of a non-equilibrium state may better fit the empirical data.

We discuss suggestions for future research. Since Chapter 3 proposes mathematical formulae about origin and expected age for arbitrary network, numerical analysis and simulation of language diffusion in this chapter can be applied to any linguistic atlases which record lexical variants at multiple places. Japan is characterized as an island country and interaction between other nations has been relatively rare. It seems interesting to conduct this research using a linguistic atlas of Indo-European languages [2,3] to elucidate the effect of interaction between populations belonging to different language areas. For mathematical model, a model unifying invention, transmission, and migration is a possible extension of our study. In this way, we may understand the effect of migration on the spatial pattern of linguistic traits and infer the parameter values concerning these effects in combination with empirical data.

Reference

[1] National Institute for Japanese Language and Linguistics. 1966-1974 Linguistic Atlas of Japan 1-6.

[2] Orton H, Sanderson S, Widdowson J. 1998 The Linguistic Atlas of England.

[3] Gilliéron J, Edmont E. 1903–1910. Atlas linguistique de la France.

[4] Huisman JLA, Majid A, van Hout R 2019 The geographical configuration of a language area influences linguistic diversity. PLoS ONE 14(6): e0217363.

[5] Bloomfield L. 1933 Language. Holt, Rinehart & Winston: New York, NY, USA.

[6] Trudgill P. 1974 Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. Lang. Soc. 2, 215–246.

[7] Burridge J. 2018 Unifying models of dialect spread and extinction using surface tension dynamics. R. Soc. open sci. 5: 171446.

[8] Jeszenszky P, Hikosaka Y, Imamura S, Yano K. 2019 Japanese lexical variation explained by spatial contact patterns. ISPRS Int. J. Geo-Inf. 8, 400.

[9] Robert G, Shackleton Jr. 2007 Phonetic Variation in the Traditional English Dialects A Computational Analysis. Journal of English Linguistics. 35, 30-102.

[10] Nerbonne J. 2010 Measuring the diffusion of linguistic change. Phil. Trans. R. soc. B 365, 3821-3828.

[11] Séguy J. 1971 La relation entre la distance spatiale et la distance lexicale. Rev. Linguist. Romane 35, 335–357.

[12] Szmrecsanyi B. 2012 Geography is Overrated. In Dialectological and Folk Dialectological Concepts of Space—Current Methods and Perspectives in Sociolinguistic Research on Dialect Change; Hansen S, Schwarz C, Stoeckle P, Streck T, Eds. 215–231.

[13] Burridge J. 2017 Spatial Evolution of Human Dialects. Physical Review X 7, 031008.

[14] Hadzibeganovic T, Stauffer D, Schulze C. 2008 Boundary effects in a three-state modified voter model for languages. Physica A 387. 3242-3252.

[15] Fagyal Z, Swarup S, Escobar AM, Gasser L, Lakkaraju K. 2010 Centers and peripheries: Network roles in language change. Lingua 120. 2061-2079.

[16] Linguistic Atlas of Japan Database (LAJDB) https://lajdb.org/TOP.html

[17] Levenshtein VI. 1966 Binary Codes Capable of Correcting Deletions, Insertions and Reversals. Soviet Physics Doklady 10, 8, 707-709.

[18] Boyd R, Richerson P J, 1985 Culture and the Evolutionary Process. University of Chicago Press, Chicago.

[19] Aguilar E, Ghirlanda S, 2015 Modeling the genealogy of a cultural trait. Theoretical Population Biology 101, 1-8.

[20] Watanabe Y, Isshiki M, Ohashi J. 2020 Prefecture-level population structure of the Japanese based on SNP genotypes of 11,069 individuals. J Hum Genet.

[21] Hirai S. 2002 The productive process of emigrants to Hokkaido and overseas and the formation of the emigrating regions in the modern era of Japan. (Japanese title: *Kindai nihon ni okeru imin no soushutsu katei to tashutsu chiiki no keisei – Hokkaido imin to kaigai imin no hikaku kara.*) Historical Geography 44, 1, 20-36.

[22] Pagel M, Meade A. 2017 The deep history of the number words. Phil. Trans. R. Soc. B 373: 20160517.

[23] Swadesh M. 1952 Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. Proc. Am. Philos. Soc. 96, 452 – 463.

[24] Pagel M, Atkinson QD, Meade A. 2007 Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. Nature 449, 717– 720.

[25] Marjoram P, Molitor J, Plagnol V, Tavare S. 2003 Markov chain Monte Carlo without likelihoods. PNAS 100 (26) 15324-15328.

[26] Lizana L, Mitarai N, Sneppen K, Nakanishi H. 2011 Modeling the spatial dynamics of culture spreading in the presence of cultural strongholds. Physical Review E 83, 6, 066116.

[27] Greenhill SJ, 2011 Levenshtein distances fail to identify language relationships accurately. Computational Linguistics. 689-698.

[28] Burridge J, Blaxter T, 2020 Using spatial patterns of English folk speech to infer the universality class of linguistic copying. Physical Review Research 2, 043053.

Table legends

Table 1

Correlation coefficient (r) of linguistic distance (LD) with geographic distance (GD) and its logarithm (log(GD)) calculated separately in each region of Japan.

Table 2

Correlation coefficient (r) of x_{ij} and its logarithm with geographic distance (GD) and population sizes.

Table 3

(a) Correlation coefficient (r) of expected age r_{ij} with geographic distance (GD) and population sizes of two populations. (b) Values of intercept and coefficient of model (6) estimated by linear regression. We performed correlation and regression analysis based on all the ordered pairs of i and j which satisfy $x_{ij} > 0$.

Table 1

Region	Number of localities	r between GD and LD	r between log(GD) and LD
All the localities	2400	0.689	0.678
Hokkaido	83	0.348	0.350
Tohoku	431	0.623	0.687
Kanto	270	0.645	0.598
Chubu	475	0.499	0.565
Kinki	274	0.487	0.494
Chugoku	270	0.518	0.588
Shikoku	156	0.334	0.426
Kyushu	381	0.831	0.797
Okinawa	60	0.820	0.875

Table 2

Parameter values		Correlation Coefficient		
$\sigma\ (km)$	b	$log(x_{ij})$ and GD	x_{ij} and P_i	x_{ij} and P_j
50	0.005	-0.659	6.03e-8	0.584
100	0.005	-0.470	7.20e-8	0.630
150	0.005	-0.331	9.24e-8	0.747
200	0.005	-0.273	1.18e-7	0.918
150	0.05	-0.548	6.97e-8	0.512
150	0.01	-0.379	8.59e-8	0.686
150	0.001	-0.267	1.10e-7	0.897

Table 3

(a)

Parameter values		Correlation Coefficient		
$\sigma\ (km)$	b	r_{ij} and GD	r_{ij} and P_i	r_{ij} and P_j
50	0.005	0.938	-0.001	-0.001

100	0.005	0.815	0.014	0.014
150	0.005	0.734	0.013	0.013
200	0.005	0.656	0.022	0.022
150	0.05	0.935	-0.004	-0.004
150	0.01	0.796	0.013	0.013
150	0.001	0.616	0.010	0.010

(b)

Parameter values		Coefficients			
$\sigma\ (km)$	b	(intercept)	GD	R^2	$1/\beta_1$ (km)
50	0.005	92.1	3.01e-1	0.880	3.32
100	0.005	139.9	1.22e-1	0.664	8.22
150	0.005	164.7	6.95e-2	0.545	14.4
200	0.005	182.7	3.06e-2	0.430	32.7
150	0.05	11.9	2.06e-2	0.874	48.5
150	0.01	75.7	4.86e-2	0.634	20.6
150	0.001	934.9	1.28e-1	0.379	8.80

Figure legends.

Figure 1
2400 localities surveyed in the LAJ project. Each red point corresponds to one locality. Map data were extracted from the GADM database (www.gadm.org), version 2.8, November 2015 (same for Figures 2b, 5, 7, 10, and 11).

Figure 2
(a)Result of MDS based on the linguistic distance matrix. Each plot is colored according to the regions of Japan. (b)2400 localities colored according to the regions. (blue: Hokkaido, black: Tohoku, green: Kanto, yellow: Chubu, light blue: Kinki, purple: Chugoku, orange: Shikoku, pink: Kyushu, and red: Hokkaido.)

Figure 3
(a)Histogram of linguistic distance calculated for every pair of two localities. (b)Scatter diagram of geographic distance (horizontal axis) and linguistic distance (vertical axis), where the density of grayscale represents the number of locality pairs appearing in each rectangle.

Figure 4
Scatter diagram of geographic distance (GD) and linguistic distances from a fixed locality called *reference site*. (a)Reference site is the most populated locality in Tokyo. (b)Reference site is the most populated locality in Kyoto. (Blue) Localities situated to the east of reference site. (Red) Localities situated to the west of reference site. (Black) Localities in Amami and Okinawa. (Yellow) Localities in Hokkaido.

Figure 5
Population sizes (number of people within 1km square) of 2400 LAJ localities displayed by $log_{10}(population\ size)$.

Figure 6
Scatter plot of geographic distance and value of x_{ij}, where L_j is fixed to one reference site for each panel. (a)-(d)Reference site is the most populated locality in Tokyo. (e)-(h)Reference site is the most populated locality in Kyoto. (a)(e)$\sigma = 50km$. (b)(f)$\sigma = 100km$. (c)(g)$\sigma = 150km$. (d)(h)$\sigma = 200km$. (Blue) Localities situated to the east of reference site. (Red) Localities situated to the west of reference site. (Black) Localities in

Amami and Okinawa.

Figure 7

Probability that each locality has a variant originating from a fixed reference site, or formally, value of x_{ij} is plotted on the map where L_i corresponds to each point of the map and L_j is fixed in each map. Black cross pointed by a red triangle indicates the reference site L_j. (a) reference cite is in Tokyo with population size of 18649 per $1km^2$. (b) reference cite is in Tokyo with population size of 50. (c) reference cite in Kyoto with population size of 10608. (d) reference cite in Kyoto with population size of 40. Parameter values: $\sigma = 150km, b = 0.005$.

Figure 8

(Horizontal axis) Population size of L_j (number of people in one square kilometer). (Vertical axis) Expected number of localities which have a lexical variant originating from L_j (s_j). (Red line) Result of linear regression. $y = -0.173 + 0.00160x$ $(R^2 = 0.913)$.

Figure 9

Scatter plot of geographic distance and value of r_{ij}, where L_j is fixed to one reference site for each panel. (a)-(d)Reference site is the most populated locality in Tokyo. (e)-(h)Reference site is the most populated locality in Kyoto. (a)(e)$\sigma = 50km$. (b)(f)$\sigma = 100km$. (c)(g)$\sigma = 150km$. (d)(h)$\sigma = 200km$. (Blue) Localities situated to the east of reference site. (Red) Localities situated to the west of reference site. (Black) Localities in Amami and Okinawa.

Figure 10

Expected age of the lexical variant at each locality given it originates from a fixed reference site, or formally, value of r_{ij} is plotted on the map where L_i corresponds to each point of the map and L_j is fixed in each map. (a) reference cite is in Tokyo with population size of 18649 per $1km^2$. (b) reference cite in Kyoto with population size of 10608. Parameter values: $\sigma = 150km, b = 0.005$.

Figure 11

Simulated linguistic distance from the most populated locality in Tokyo (red plot). (a) $\sigma = 50km$ (b)$\sigma = 150km$. Other parameters: $b = 0.005,\ \mu = 0.0001$.

Figure 12

Diagram of linguistic distance (LD) against geographic distance (GD) for every pair of two localities. Grayscale represents the number of locality pairs in each rectangle. (a) $\sigma = 50km$ (b)$\sigma = 150km$. Other parameters: $b = 0.005,\ \mu = 0.0001$.

Figure 13

Scatter plot of geographic distance and simulated linguistic distance, where we display both distances from a fixed to one reference site for each panel. (a)-(d)Reference site is the most populated locality in Tokyo. (e)-(h)Reference site is the most populated locality in Kyoto. (a)(e)$\sigma = 50km.$ (b)(f)$\sigma = 100km.$ (c)(g)$\sigma = 150km.$ (d)(h)$\sigma = 200km.$ (Blue) Localities situated to the east of reference site. (Red) Localities situated to the west of reference site. (Black) Localities in Amami and Okinawa. Other parameters: $b = 0.005,\ \mu = 0.0001.$

Figure 14

Correlation coefficient between geographic distance and linguistic distance. Horizontal axes represent the value of σ. (blue) correlation coefficient between geographic distance (GD) and linguistic distance (LD). (orange) correlation coefficient between logarithm of geographic distance (log(GD)) and linguistic distance (LD). Parameter values: $b = 0.005, \mu = 0.0001.$

Figure 15

(a)Posterior distribution of σ inferred from approximate Bayesian computation. (b)Scatter plot of parameter sets b and m, which satisfied $MSE < 0.01$.

Figure 1

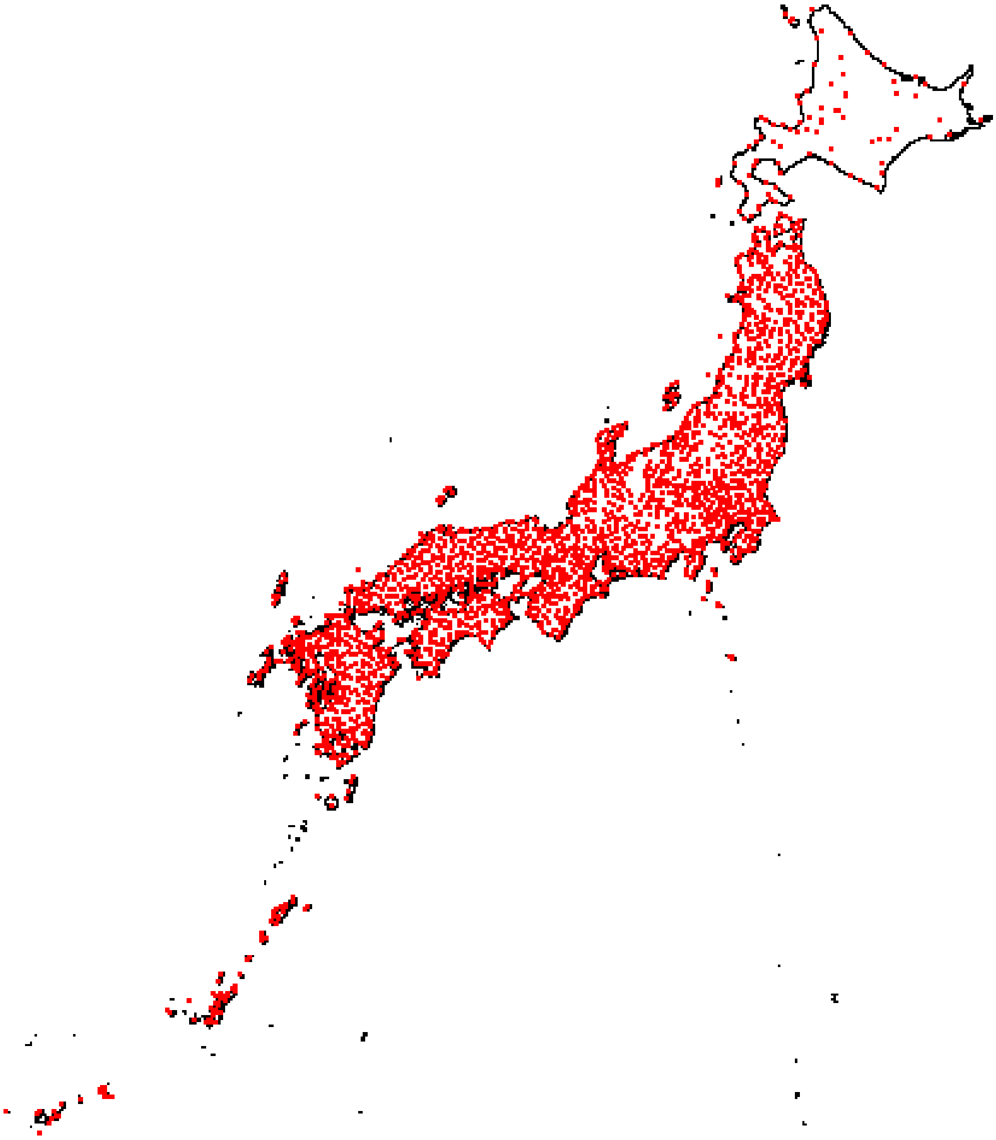


Figure 2

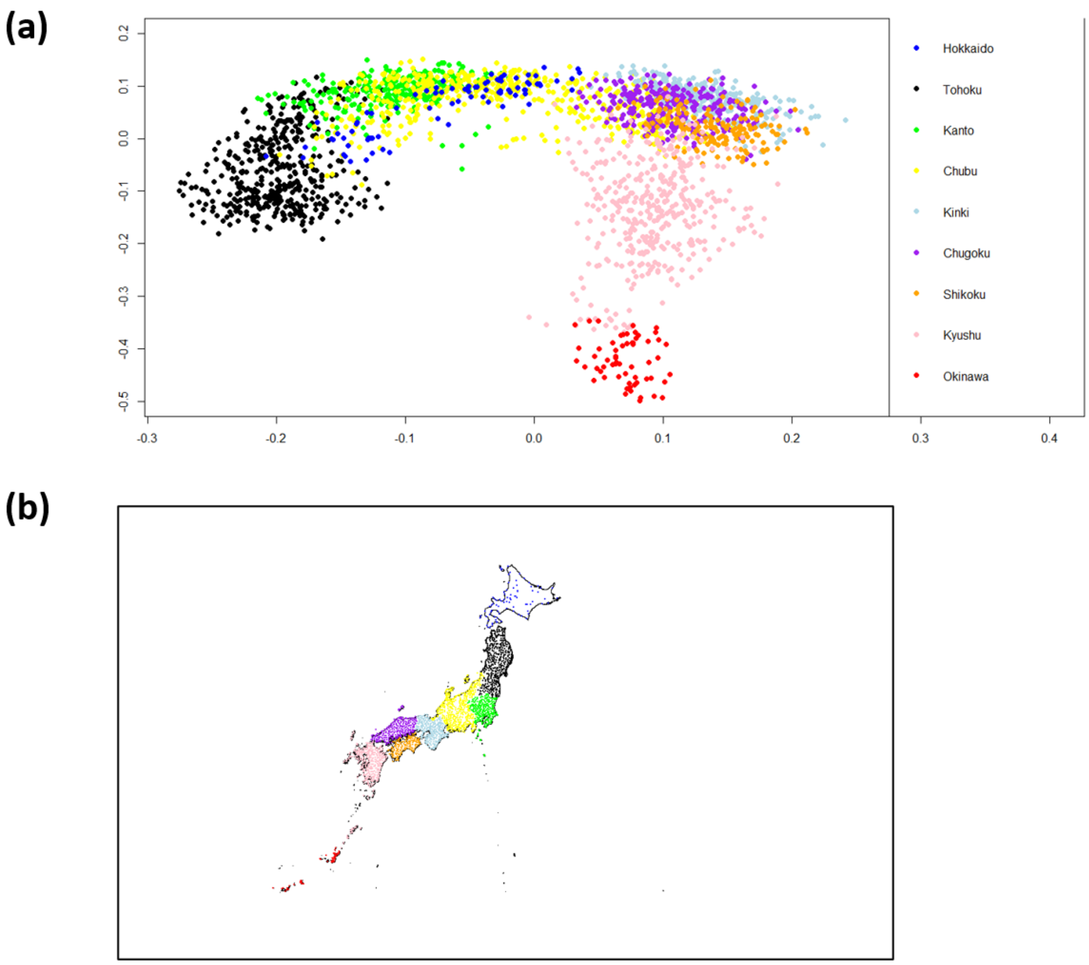

(a)
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
Hokkaido
Tohoku
Kanto
Chubu
Kinki
Chugoku
Shikoku
Kyushu
Okinawa
(b)

Figure 3

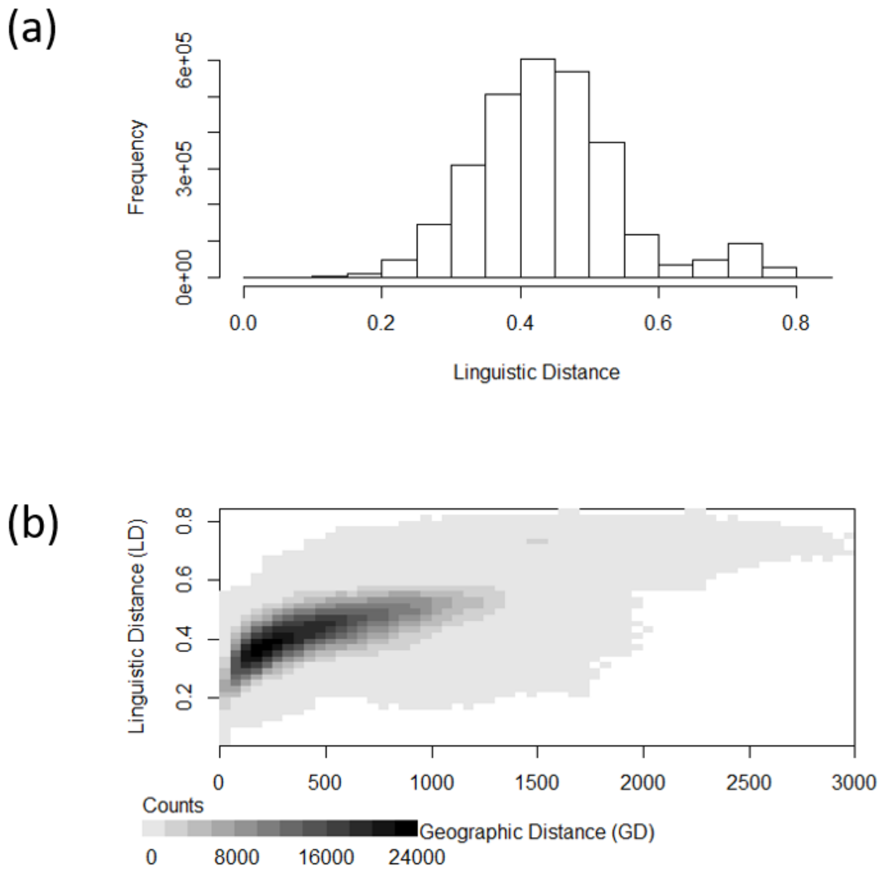
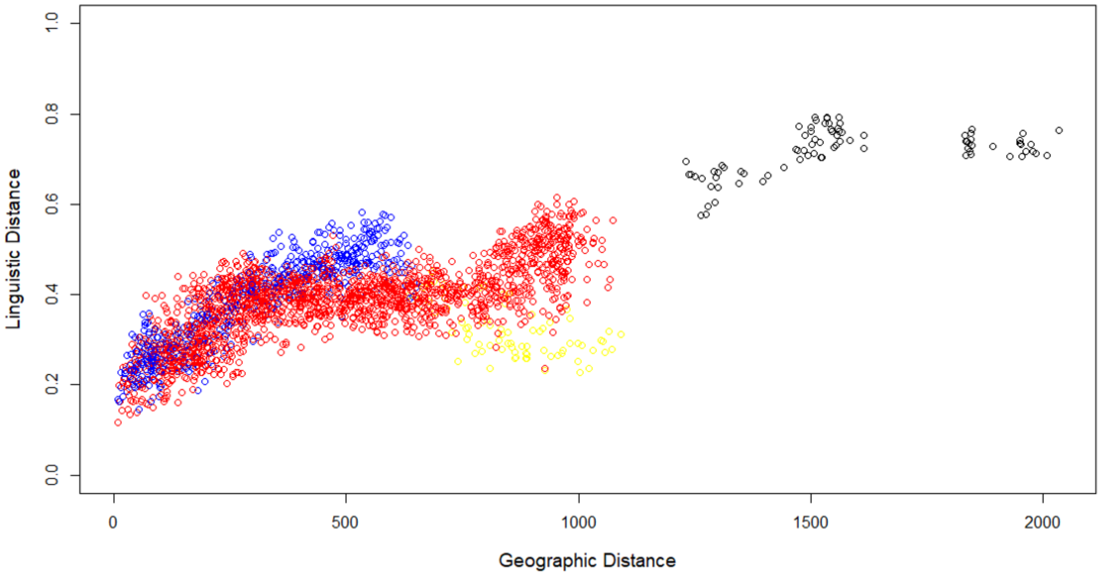

(a)
Frequency
0e+00
3e+05
6e+05
0.0
0.2
0.4
0.6
0.8
Linguistic Distance
(b)
Linguistic Distance (LD)
0.2
0.4
0.6
0.8
0
500
1000
1500
2000
2500
3000
Counts
Geographic Distance (GD)
0
8000
16000
24000

Figure 4

(a) Linguistic distance from Tokyo



(b) Linguistic distance from Kyoto

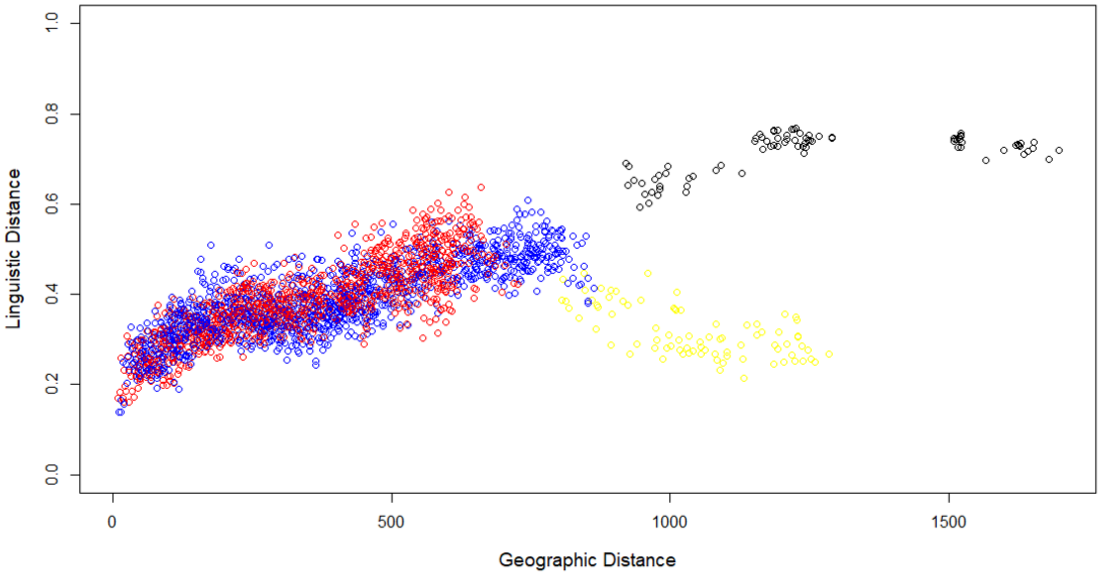


Figure 5

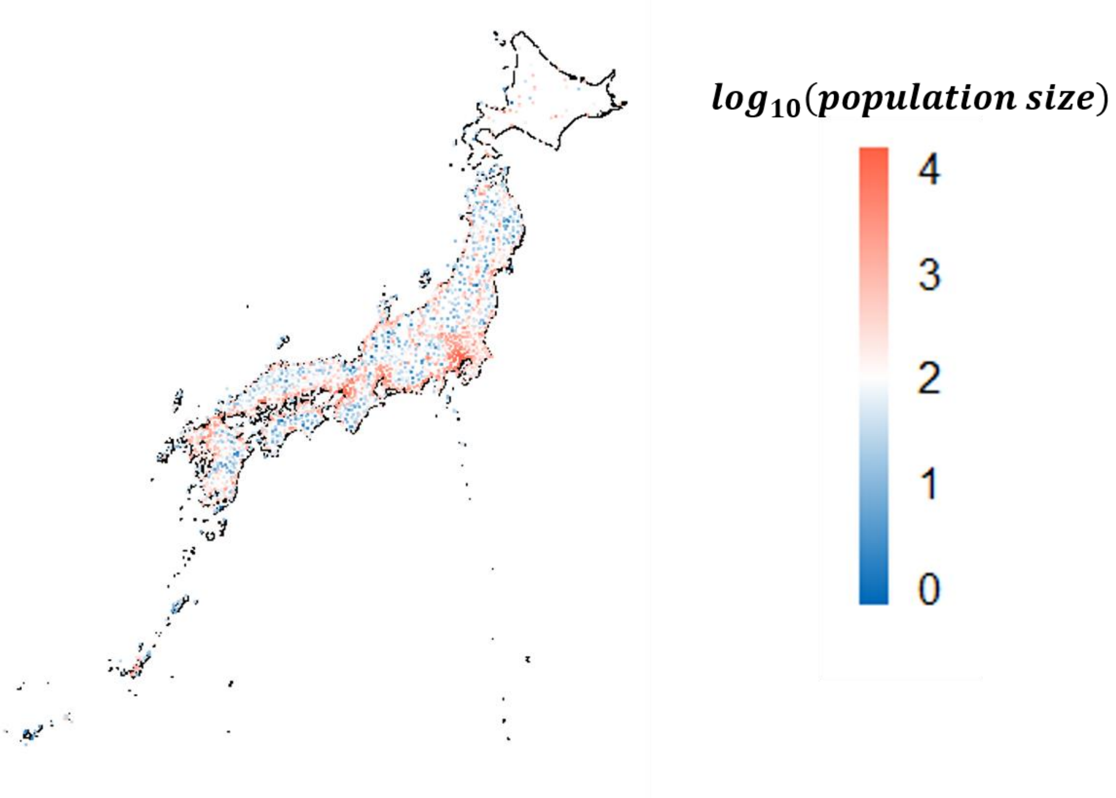


Figure 6

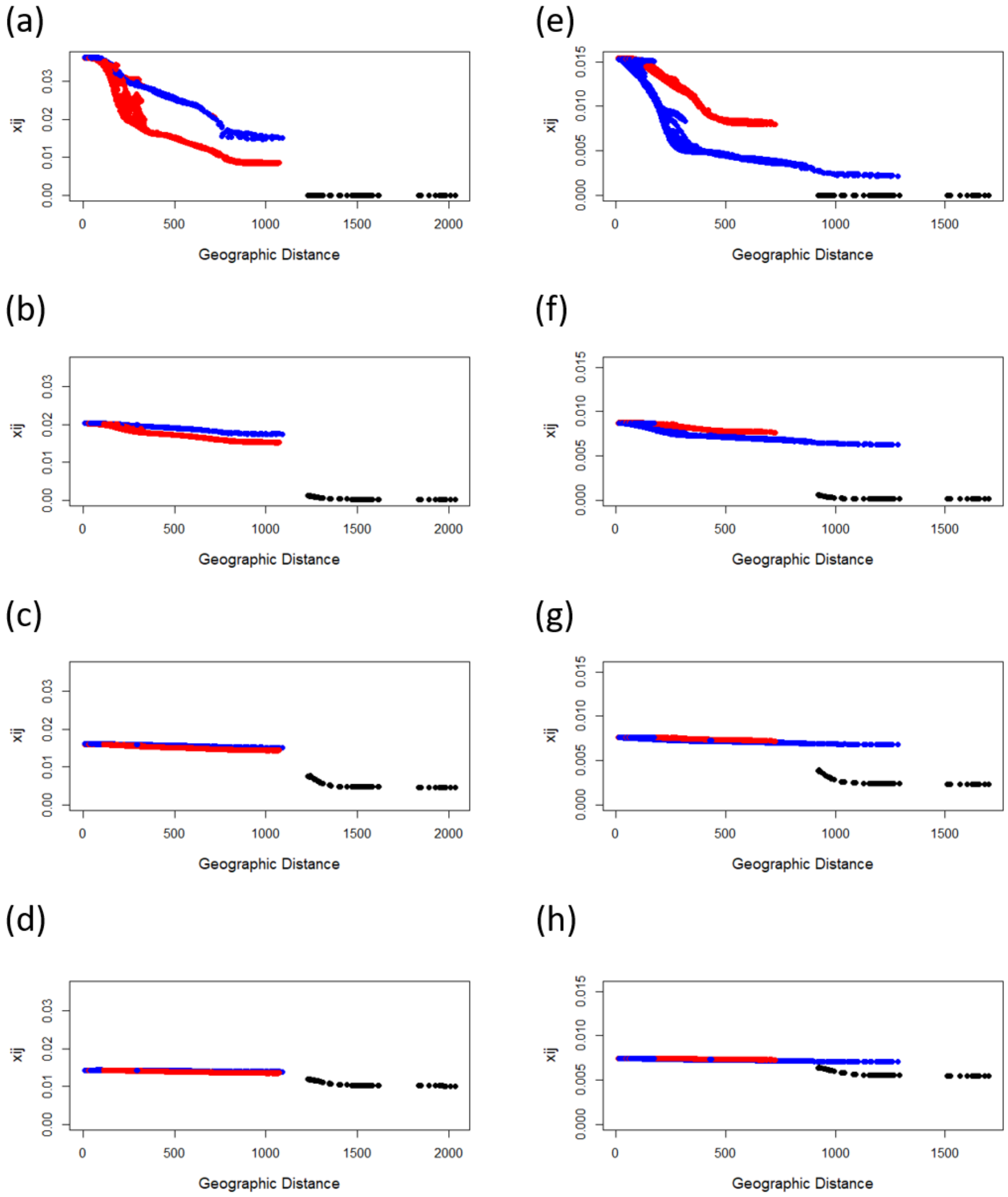

(a)
xij
0.00
0.01
0.02
0.03
0
500
1000
1500
2000
Geographic Distance
(b)
xij
0.00
0.01
0.02
0.03
0
500
1000
1500
2000
Geographic Distance
(c)
xij
0.00
0.01
0.02
0.03
0
500
1000
1500
2000
Geographic Distance
(d)
xij
0.00
0.01
0.02
0.03
0
500
1000
1500
2000
Geographic Distance
(e)
xij
0.000
0.005
0.010
0.015
0
500
1000
1500
Geographic Distance
(f)
xij
0.000
0.005
0.010
0.015
0
500
1000
1500
Geographic Distance
(g)
xij
0.000
0.005
0.010
0.015
0
500
1000
1500
Geographic Distance
(h)
xij
0.000
0.005
0.010
0.015
0
500
1000
1500
Geographic Distance

Figure 7

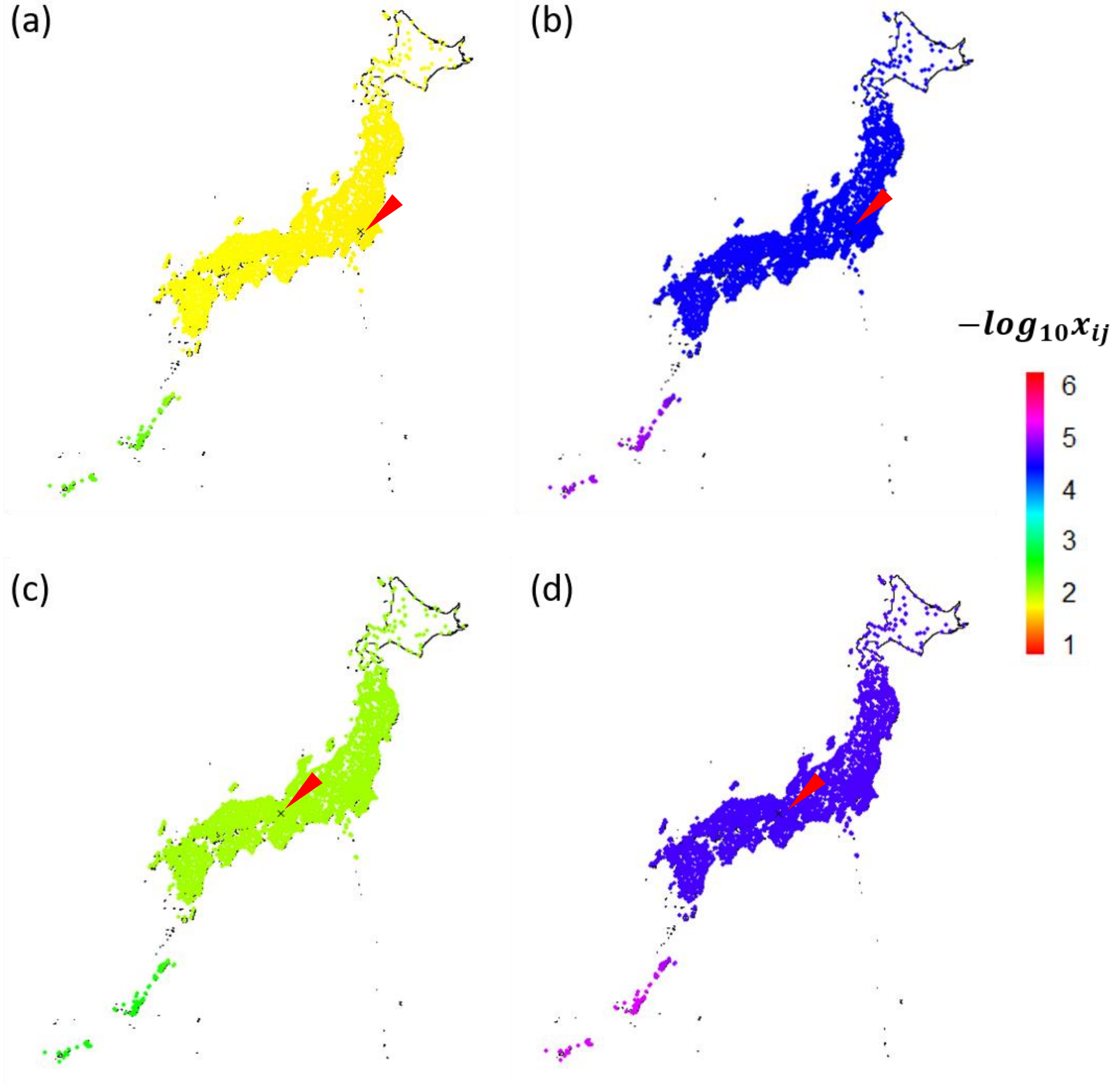


Figure 8

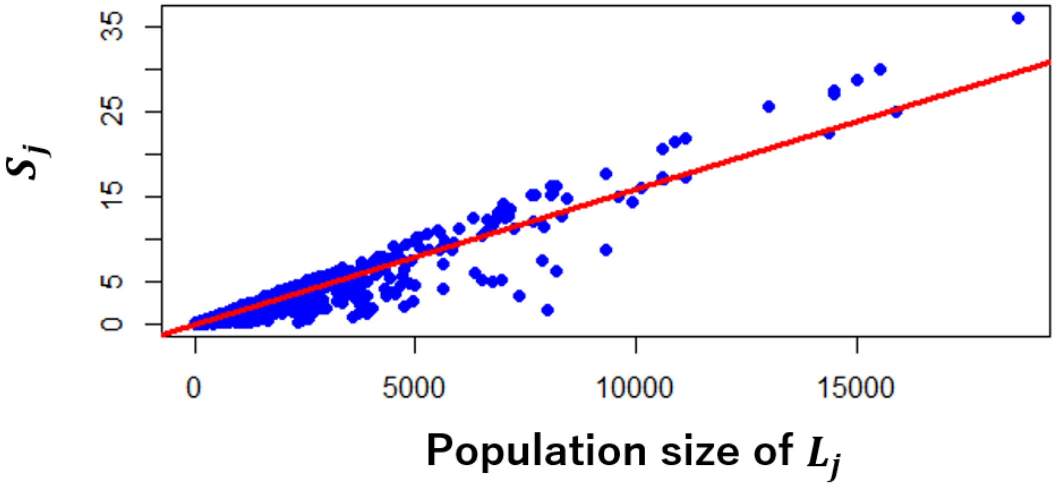


Figure 9

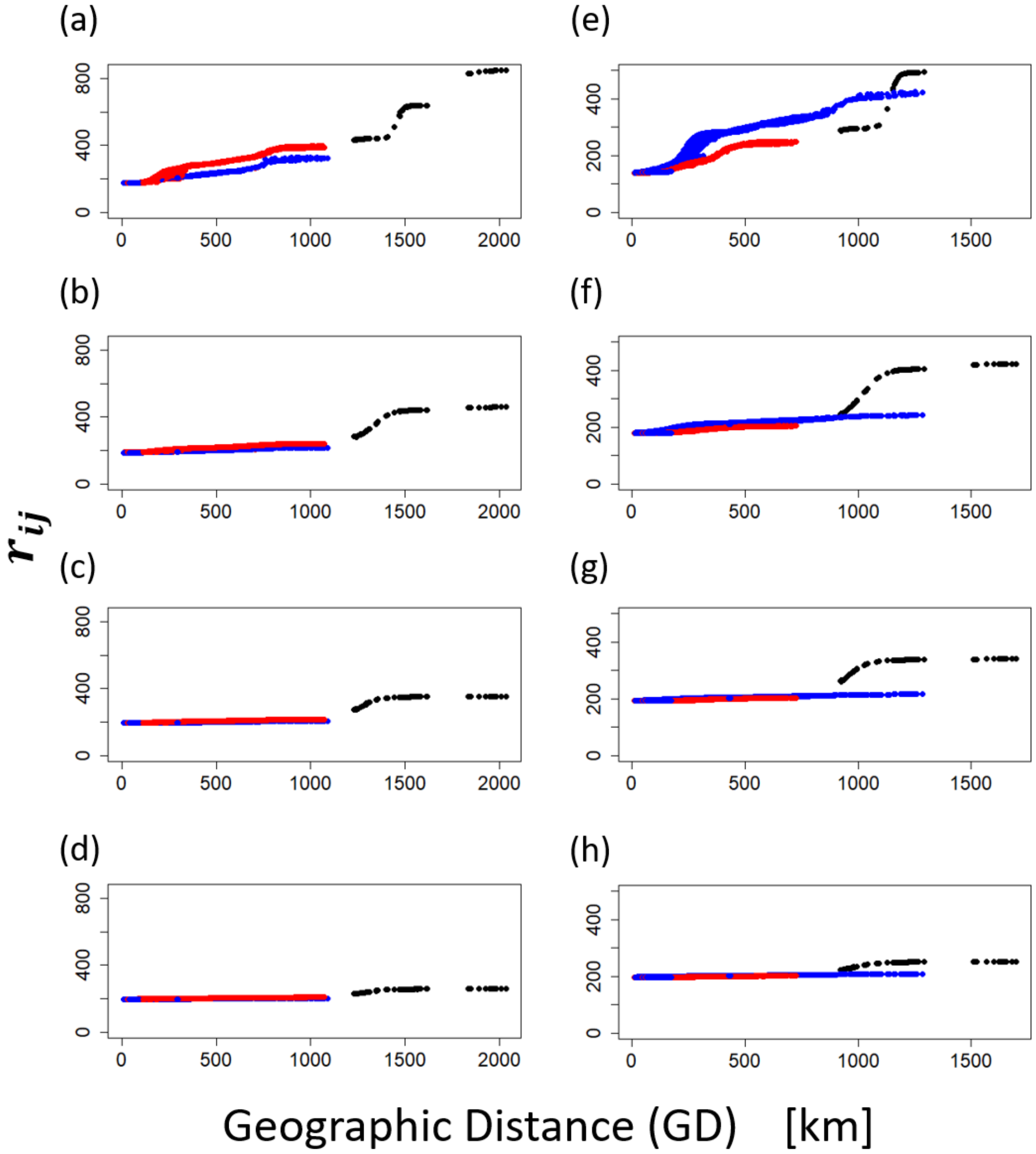

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
r_{ij}
0
400
800
200
500
1000
1500
2000
Geographic Distance (GD) [km]

Figure 10

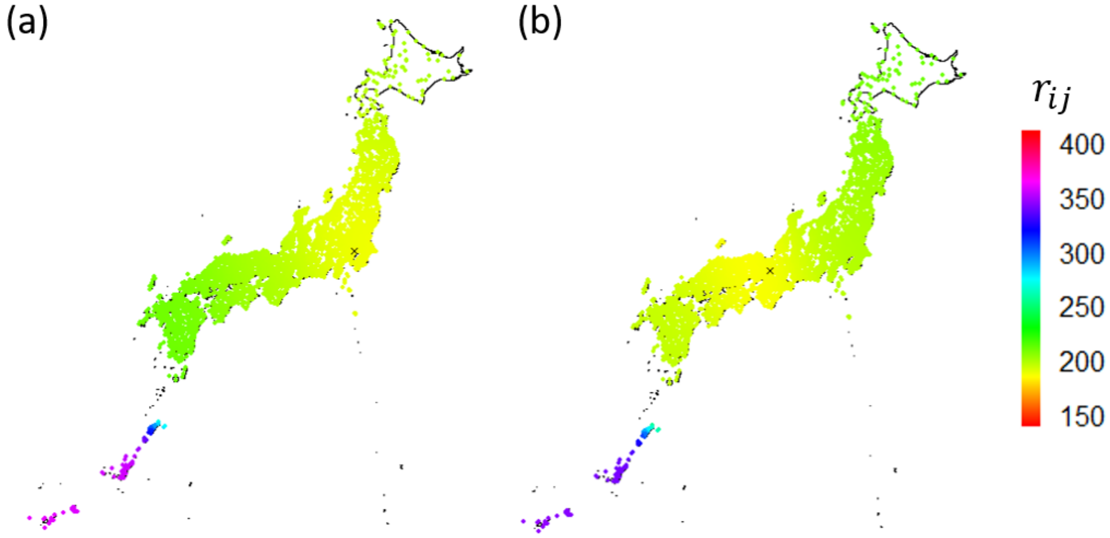


Figure 11

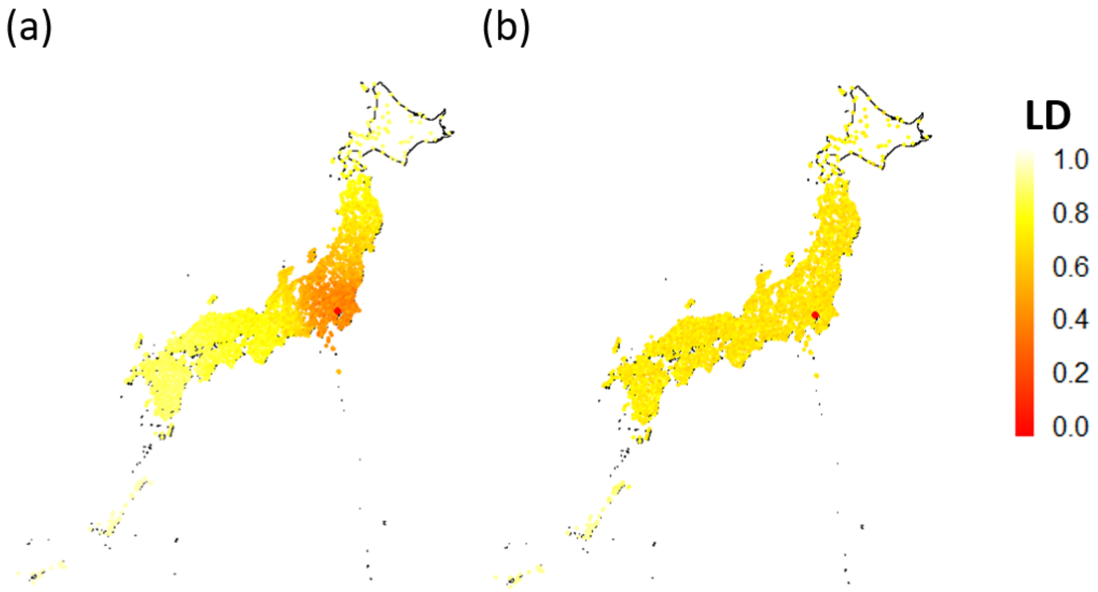


Figure 12

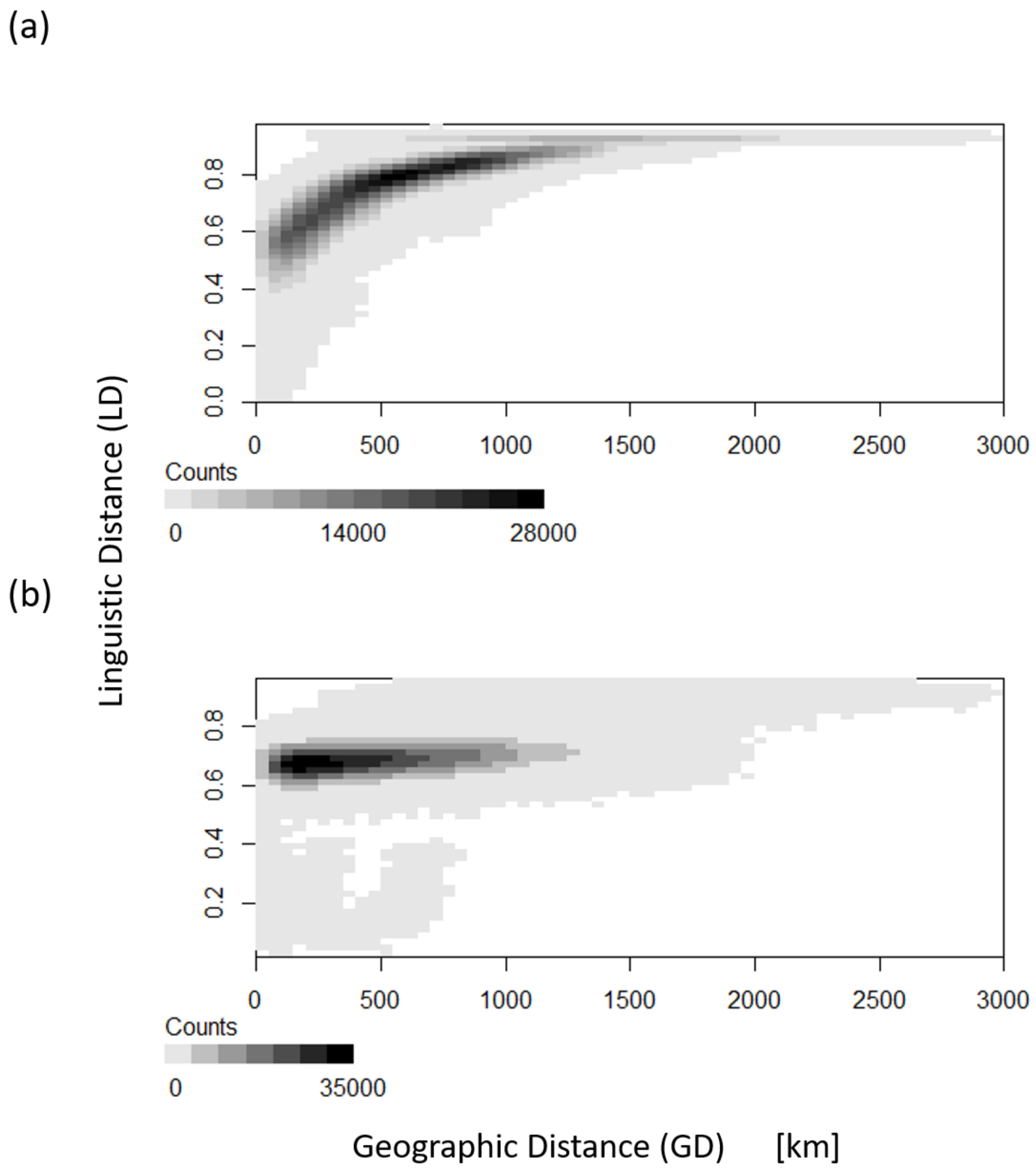

(a)
0.0
0.2
0.4
0.6
0.8
0
500
1000
1500
2000
2500
3000
Counts
0
14000
28000
(b)
0.2
0.4
0.6
0.8
0
500
1000
1500
2000
2500
3000
Counts
0
35000
Linguistic Distance (LD)
Geographic Distance (GD) [km]

Figure 13

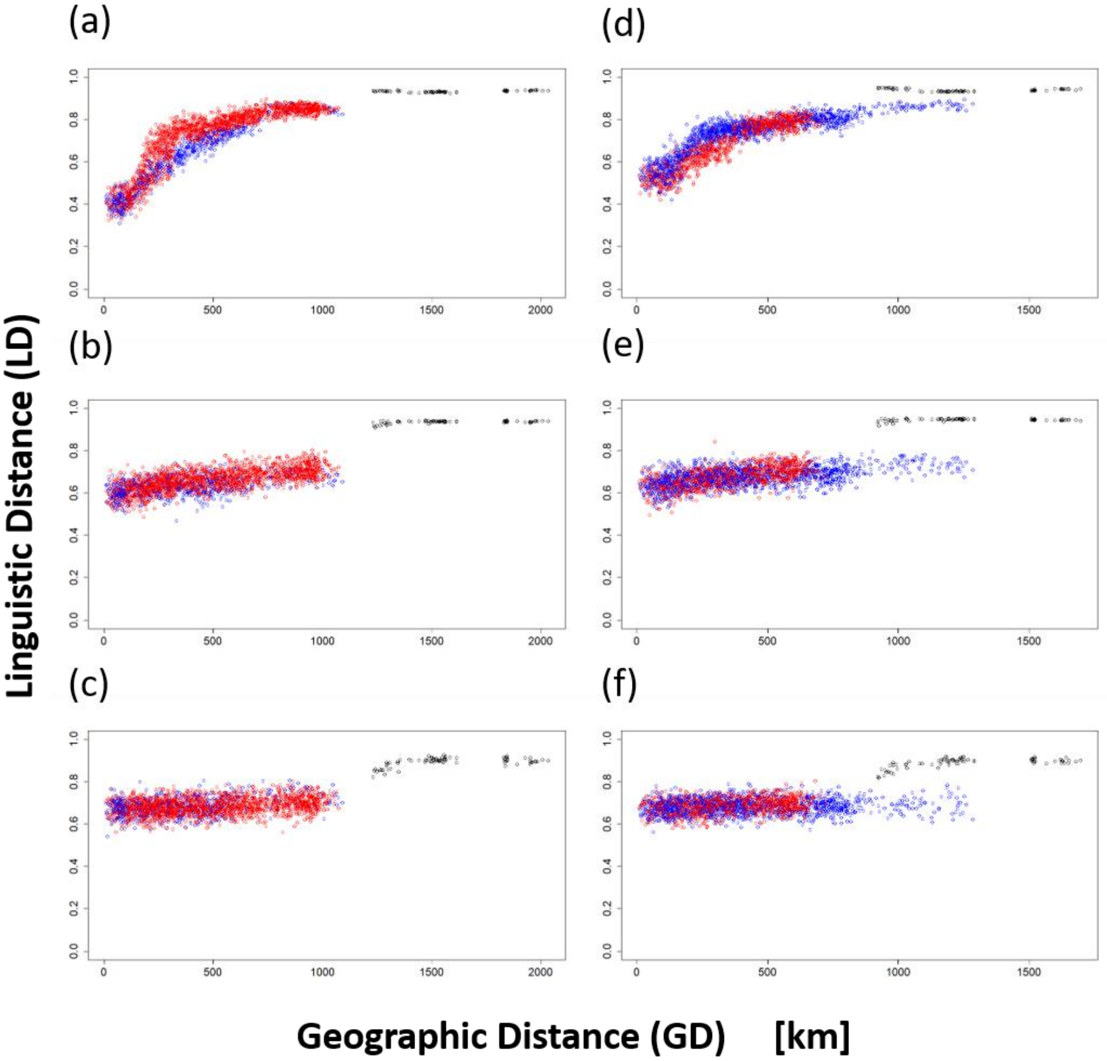

(a)
(d)
(b)
(e)
(c)
(f)
Linguistic Distance (LD)
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
Geographic Distance (GD) [km]

Figure 14

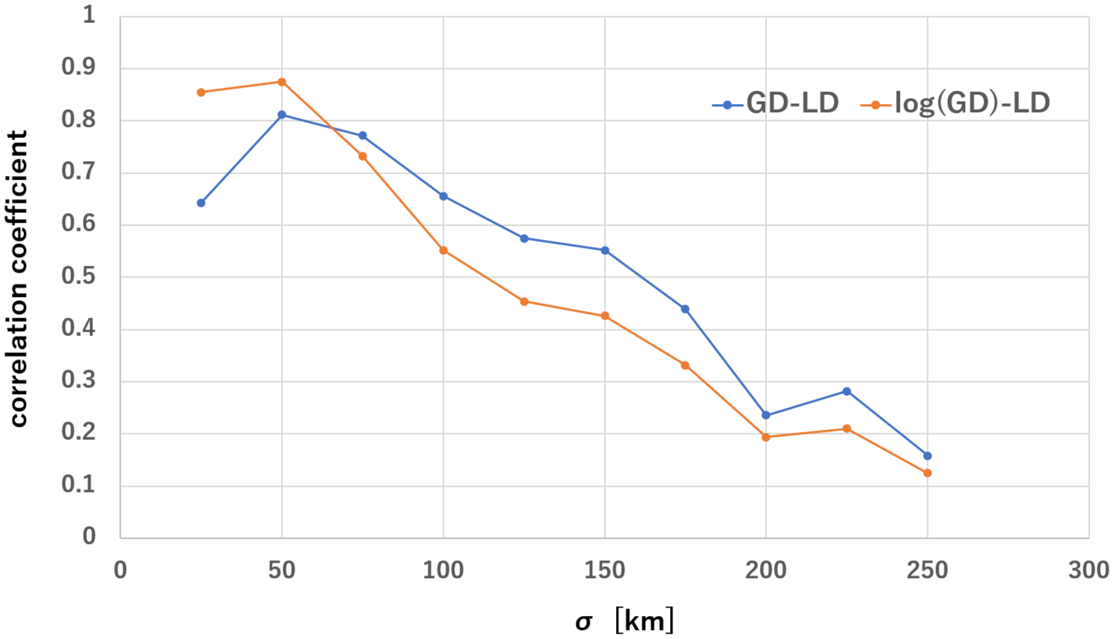
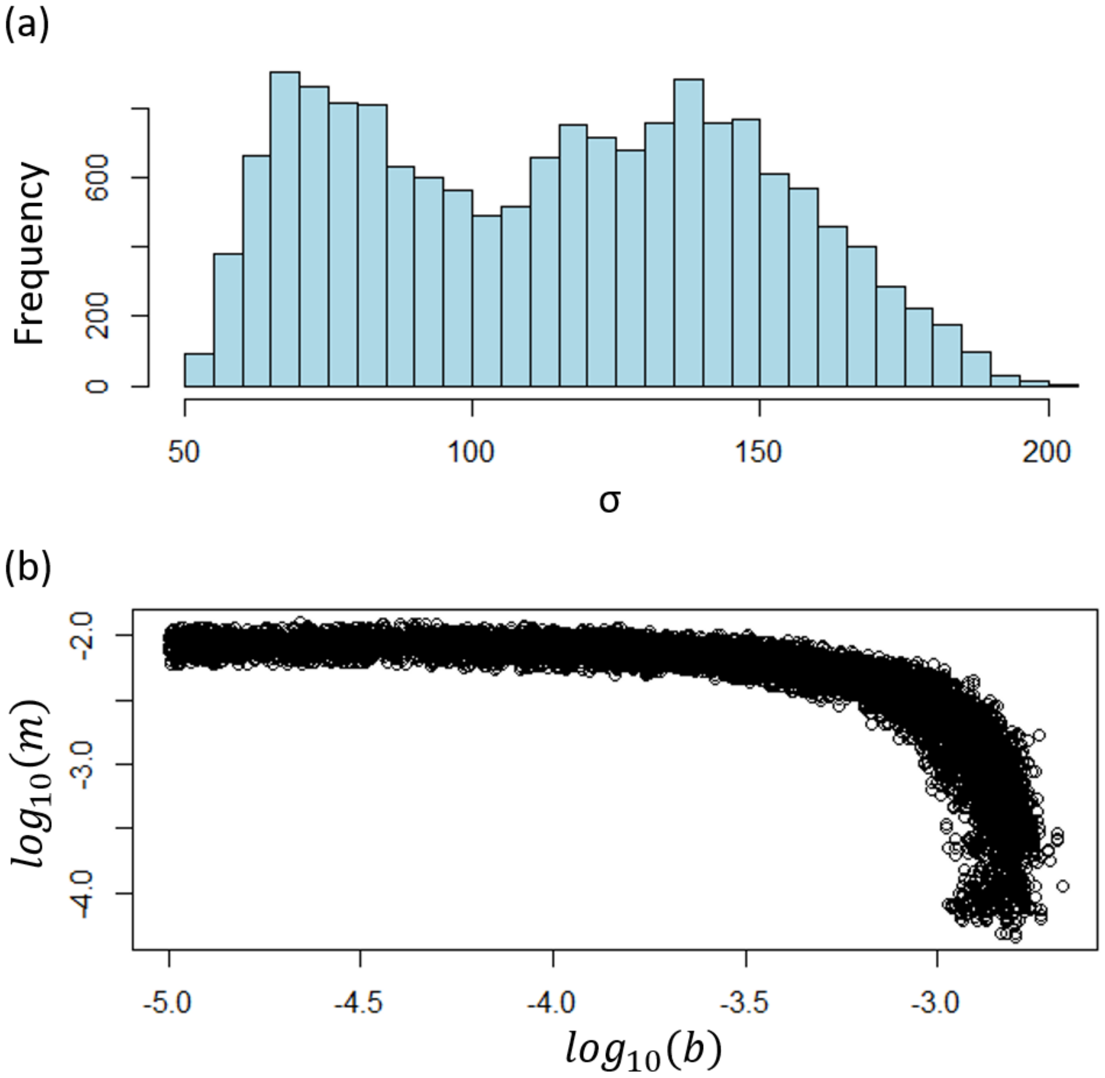

correlation coefficient
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250
300
σ [km]
GD-LD
log(GD)-LD

Figure 15


(a)
Frequency
0
200
600
50
100
150
200
σ
(b)
$log_{10}(m)$
-2.0
-3.0
-4.0
-5.0
-4.5
-4.0
-3.5
-3.0
$log_{10}(b)$

Chapter 5. General discussion

Three main chapters of the thesis consistently treated the spatial dynamics of cultural evolution between populations and diffusion of linguistic traits with the aid of network model. In chapter 2, I presented the minimum model, which assumed an exclusive innovation of dialect words by a central population, to provide mathematical support for the alleged concentric distribution of Japanese dialects [1,2]. Subsequently, to gain the general model of cultural evolution between populations, Chapter 3 extended the minimal model and introduced the notion of origin and average diffusion time (conditional mean age), incorporating the Markovian ancestral process of cultural genealogy [3]. I also performed a numerical analysis on a random network, to explore the effect of network topology and innovation rates on these variables. In Chapter 4, the focus was again placed on Japanese dialects, and the model in Chapter 3 was applied to the network of 2400 survey localities of Linguistic Atlas of Japan (LAJ) [4] to examine the relationship between geographic distance and linguistic distance (i.e., Séguy's curve [5]). Notably, unlike the two preceding chapters, I incorporated the empirical data of LAJDB [6] and compared between empirical and theoretical distribution of Japanese lexical variation.

Throughout the three chapters, I have analyzed the effect of central population (Chapter 2), natural or cultural barrier (Chapter 2), in-degree and out-degree of each node (Chapter 3), innovation rate (Chapter 3), population size (Chapter 2, 4), and geographic distance (Chapter 4) on the spatial distribution of cultural or lexical variants. The generalized model in Chapter 3 can be applied to any network which may have a complex topological structure, which enabled me to obtain insight into a wide variety of network structures. Although a huge body of literature in complex system has elucidated the diffusion of social traits on a network of individuals (i.e., voter model [7]), the novelty of my research consists in the notion of cultural age, origin, and diffusion time, as well as the mathematical formulae given in a simple matrix notation. Unlike previous research into voter model and its application to linguistics [8,9], which often treated a finite number of extant traits and calculated the fixation probability and time, my network model includes innovation (individual learning), and the number of variants is theoretically infinite.

As for application to dialect distribution, I introduced the notion of word age and replicated the observed concentric distribution of Japanese dialects. This method is different from Lizana et al's [10] work, which examined the concentric patten by directly measuring the distance from Kyoto. In addition, I found that transmission limited to short distance is a key factor which produces a sublinear relationship of linguistic and geographic distances. Since cognate words originating from the same ancestor are more similar than non-cognate pair of variants, my model, which estimates when and where variants were created, was helpful in deducing the accumulation of linguistic distance over space. My finding also includes the fact that diffusion of lexical variants is not isotropic, and the growth of linguistic distance may differ according to direction. Although network models have often been applied to linguistics [8,9], there was to my knowledge no previous study which examined the age and origin of lexical variants.

I discuss possible extensions and suggestions for future studies. Although this thesis sought the application field in linguistics and dialectology, the model I conceived in Chapter 3 is applicable to any trait which diffuses from population to population via social learning. Since cultural trait is a general concept [11,12], ranging from ancient potteries and stone tools to modern technology and industries, possible application field is quite abundant. For example, previous research into cultural evolution treated the spatial dynamics or distribution of music [13], superstition [14], and farming [15]. Besides mathematical modelling, incorporating empirical data is also possible given a dataset of multiple populations, such as linguistic atlas for dialectology, is available for the cultural trait of interest.

I consider limitations of this model. First, the model does not contain learning bias [11], which has been treated by much theoretical research of cultural evolution [16,17] and previous mathematical models of linguistic geography [10,18]. Second, stochastic noise is absent in this model. Models in Chapters 2 and 3 assumed infinite number of individuals and therefore infinite number of variants in each vertex of the network, so the model was completely deterministic. On the other hand, in Chapter 4, we adapted the generalized model of Chapter 3 into the *one-vertex-one-variant* model, so origin, diffusion time, and linguistic distance between localities were interpreted in a probabilistic way. Both of these assumptions seem extreme, and it is intriguing to develop a network model, whose vertices contain multiple but finite number of cultural variants and consider the effect of stochastic noise on frequency, age, and origin of the variants. Third, my model assumes variants diffuse only by social learning, and the effect of migration is completely neglected. Since cultural macroevolution is driven not only by social learning between populations but also by demic diffusion, divergence, and death of populations [19], a more synthetic model which unifies the effect of social learning and immigration may better represent the dynamics of cultural evolution.

Reference

[1] Matsumoto O. 1993 Zenkoku aho-baka bumpu-kou. Ohta Shuppan.

[2] Yanagita K. 1927 Kagyuko (1)-(4). Jinruigaku Zasshi 42, 125-135 (162-172, 223-233, 273-284).

[3] Aguilar E, Ghirlanda S. 2015 Modeling the genealogy of a cultural trait. Theoretical Population Biology 101, 1-8.

[4] National Institute for Japanese Language and Linguistics. 1966-1974 Linguistic Atlas of Japan 1-6.

[5] Séguy J. 1971 La relation entre la distance spatiale et la distance lexicale. Rev. Linguist. Romane 35, 335–357.

[6] Linguistic Atlas of Japan Database (LAJDB) https://lajdb.org/TOP.html

[7] Sood V, Antal T, Redner S. 2008 Voter models on heterogeneous networks. Physical Review E 77, 041121.

[8] Hadzibeganovic T, Stauffer D, Schulze C. 2008 Boundary effects in a three-state modified voter model for languages. Physica A 387. 3242-3252.

[9] Fagyal Z, Swarup S, Escobar AM, Gasser L, Lakkaraju K. 2010 Centers and peripheries: Network roles in language change. Lingua 120. 2061-2079.

[10] Lizana L, Mitarai N, Kim S. 2011 Modeling the spatial dynamics of culture spreading in the presence of cultural strongholds. Physical Review E 83, 066116.

[11] Boyd R, and Richerson PJ. 1985 Culture and the Evolutionary Process. University of Chicago Press, Chicago.

[12] Cavalli-Sforza LL, Feldman MW. 1981 Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton.

[13] Savage PE, Brown S. 2014 Mapping music: Cluster analysis of song-type frequencies within and between cultures. Ethnomusicology 133-155.

[14] Tamura K, Ihara Y. 2017 Quantifying cultural macro-evolution: a case study of the hinoeuma fertility drop. Evolution and Human Behavior 38, 117-124.

[15] Aoki K, Shida M. 1996 Travelling wave solutions for the spread of farmers into a region occupied by hunter-gatherers. Theoretical Population Biology 50, 1-17.

[16] Aoki K, Lehmann L, Feldman MW. 2011 Rates of cultural change and patterns of cultural accumulation in stochastic models of social transmission. Theoretical Population Biology 79, 192-202.

[17] Fogarty L, Wakano JY, Feldman MW, Aoki K. 2017 The driving forces of cultural complexity. Neanderthals, modern humans, and question of population size. Human Nature 28, 39-52.

[18] Burridge J, Blaxter T, 2020 Using spatial patterns of English folk speech to infer the universality class of linguistic copying. Physical Review Research 2, 043053.
[19] Mesoudi A. 2011 Cultural Evolution: How Darwinian Theory Can Explain Human Culture & Synthesize the Social Sciences. University of Chicago Press, Chicago.

Acknowledgements

I would like to express my sincerest gratitude to my supervisor Dr. Yasuo Ihara, who gave me tremendous support for this research and the opportunity to conduct research in his laboratory. I would also like to express my gratitude to every member of Evolutionary Anthropology Laboratory.

I am grateful to Dr. Kohei Tamura for sharing his expertise in cultural evolution and for teaching me the basic methodology of geographic information system (GIS) mapping. I thank Dr. Takenori Takada for teaching me mathematical properties of Markovian process. I thank Dr. Ayaka Onohara for her comments about the empirical research of Japanese lexical variation in Chapter 4 and for sharing her knowledge about linguistic geography. I am grateful to my former supervisor Dr. Fumio Tajima for teaching me population genetics and coalescent theory, which unexpectedly motivated me to apply the ancestral process on the network model in Chapter 3. I thank Drs. Thomas Currie and Alex Mesoudi for their valuable comments on this research. I also thank Dr. Marco Campenni for teaching me the basics of R language. I thank Ms. Michiko Ishi, who supported my research through her daily paperwork.

This research was supported by JSPS, Japan KAKENHI grant numbers JP17H06381 and JP20J13493.