Vol. 17 No. 1 (2026) 1–7
LETTERS ON EVOLUTIONARY BEHAVIORAL SCIENCE

Multiagent Bayesian Iterated Learning Model Revealing the Necessary Conditions for Peripheral Distribution of Dialect

Seo Hachimaru[*], **Motohide Seki**

Kyushu University, 4-9-1 Shiobaru, Minami Ward, Fukuoka 815-8540, Japan

[*]Author for correspondence (seohachimaru@gmail.com)

Languages evolve through diffusion processes similarly to biological evolution, forming linguistic clusters based on geographic proximity. A few mathematical modeling studies have tested the classical theory on the formation of a peripheral distribution of dialects, which originally assumed two conditions: (i) new words are innovated exclusively at a cultural center and (ii) these words spread outward due to the prestige of the cultural center. However, it is known that these special conditions are often not met. We examined whether and how the presence or absence of each condition influences the outcome using an extended Bayesian Iterated Learning Model. Our mathematical analyses revealed that peripheral distributions can emerge not only when both conditions are present but also when one of the two is absent. Furthermore, the satisfaction of one or both conditions in a population can be predicted by investigating the word age distribution there.

Keywords
language evolution, cultural transmission, Bayesian inference, agent-based modeling

Introduction

There is variation in languages, more specifically variation in linguistic forms such as morphemes, words, and grammars, at a wide range of levels (from global to local). Similarity between languages typically correlates with geographic closeness (De Gregorio et al., 2024). This would reflect the process by which a new variant generated in one subpopulation spreads between subpopulations, similar to the gene flow observed in biological evolution (Bromham, 2025). As a result, a linguistic cluster is formed by geographically close languages.

However, some linguistic distributions deviate from this general pattern of geographic clustering. An example of such a distribution pattern is the "peripheral distribution" proposed in the theory of Hōgen Shūkenron (theory of peripheral distribution of dialectal forms; Yanagita, 1930). The theory is originally based on two assumptions: (1) new words are created exclusively at a cultural center (hereafter referred to as Centralized Lexical Innovation; CLI), and (2) information on linguistic form flows unidirectionally from the center side to the peripheral side probably because people in surrounding areas perceive a "prestige of center" for a word introduced from the center side and accept it (Unidirectional Transmission; UT). These assumptions explain how new words spread outward from the center, forming a peripheral distribution of each word, with older ones further from the cultural center.

So far, several studies have quantitatively evaluated the validity of the verbal argument by Yanagita (1930). Lizana et al. (2011) developed a mathematical model incorporating both of the aforementioned key conditions to obtain the peripheral distribution in their model. Takahashi & Ihara (2020) found that the concentric distribution can be formed only with the condition of CLI.

However, language change does not necessarily conform to CLI, and there are examples of linguistic innovations originating in peripheral regions and subsequently diffusing back into cultural centers. For instance, the Japanese word *uzattai*, originally used only in the western suburbs of Tokyo, later diffused inward to central Tokyo and eventually nationwide, providing evidence of linguistic innovation flowing from peripheral regions into cultural centers (Inoue, 2010). Therefore, it is worth investigating whether peripheral distributions emerge even when the center does not play a special role in lexical innovation.

To address this gap, we developed an extended Iterated Learning Model that can independently test each of the two proposed conditions. By examining scenarios in which CLI and UT operate separately or together, we clarify the necessary conditions for peripheral distribution formation and provide insights into the fundamental mechanisms governing language change across geographic space.

Model

The present model is based on the previous, Bayesian Iterated Learning Model (ILM; Reali & Griffiths, 2010). The model considers a discrete-generation population, in which individuals (agents) have two life stages: learner and speaker stages. At the learner stage, they acquire linguistic knowledge by referring to data produced by individuals in the previous generation. At the speaker stage, they produce data, which can be observed by individuals in the next generation. Every individual born at time t serves as learner and speaker at time t and $t + 1$, respectively.

The learning process is formalized as Bayesian inference (Kirby, 2001; Kirby et al., 2007). Suppose that there are K variants in linguistic form (such as sounds, words, or grammatical constructions) $v_1, v_2, \dots, v_K$,

doi: 10.5178/lebs.2026.130
Received 28 April 2025.
Accepted 21 January 2026.

and consider two types of K-dimensional vectors: $\boldsymbol{x} = [x_1, x_2, \dots, x_K]$ and $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_K]$, where x_k and θ_k denote sampled frequency and the estimated probability, respectively, of a variant v_k. The sum of x_k, $\sum_{k=1}^{K} x_k$, is fixed to N. The parameter N represents total times of valid observation and thus would be correlated with size or activity level of the population. The inferred distribution of θ after receiving data $\boldsymbol{x}$, $p(\boldsymbol{\theta} \mid \boldsymbol{x})$ is calculated as

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\boldsymbol{x} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}, \tag{1}$$

where $p(\boldsymbol{\theta})$ is the prior distribution, indicating the innate biases common in the focal population. The likelihood $p(\boldsymbol{x} \mid \boldsymbol{\theta})$ is a multinomial distribution, which is the probability of observed data $\boldsymbol{x}$ from the parameter $\boldsymbol{\theta}$. We assume that the prior distribution is a symmetric K-dimensional Dirichlet distribution, which means that individuals have no preference for a specific form. The distribution is determined only by parameter α and K:

$$p(\boldsymbol{\theta}) \propto \prod_{k=1}^{K} \theta_k^{\frac{\alpha}{K}-1}. \tag{2}$$

The parameter α moderates the learner's preference for diversity: the larger α / K value represents a tendency to retain the greater number of variants. Under the above assumptions, probability that a variant v_k is produced is calculated as

$$\frac{x_k + \frac{\alpha}{K}}{N + \alpha}. \tag{3}$$

See Reali & Griffiths (2010) for derivation of equation (3). We focus on the situation that mutation always produces a variant that has never been observed, in which case every mutation can be called lexical innovation. The number of variants K is potentially infinite, which yields an infinite dimensional Dirichlet distribution for the prior distribution. The probability that an already existing variant v_k is reproduced and that of lexical innovation is given by

$$\frac{x_k}{N + \alpha}, \tag{4a}$$

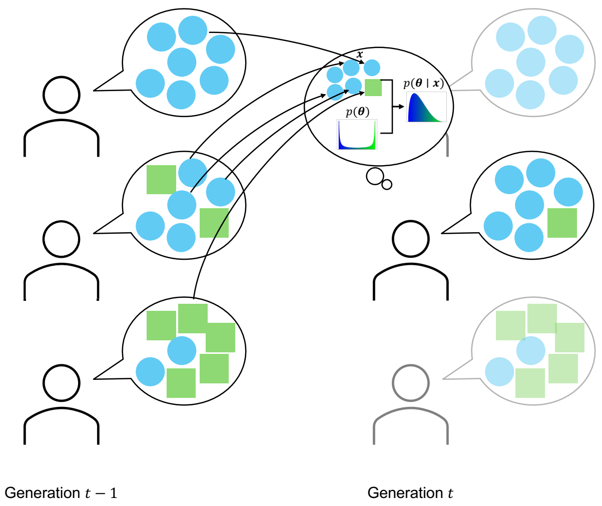
$$\frac{\alpha}{N + \alpha}, \tag{4b}$$

respectively. In this limit every mutation provokes lexical innovation, and the probability of mutation is correlated with the parameter α. Note that the Dirichlet prior does not allow the extreme case of $\alpha = 0$. However, equation (2) approaches the Haldane prior $\prod_{k=1}^{K} \theta_k^{-1}$ as $\alpha \to 0$. In this study, we define that the prior follows Dirichlet distribution when $\alpha > 0$ and Haldane distribution when $\alpha = 0$. In the case $\alpha = 0$, the posterior distribution $p(\boldsymbol{\theta} \mid \boldsymbol{x})$ follows $\prod_{k=1}^{K} \theta_k^{x_k - 1}$, and the production probability for variant v_k simplifies to x_k / N. In this case, the model is equivalent to the Wright-Fisher neutral model without mutation in genetics.

We further extended the model to describe dynamics in a structured population. In the present model, an individual belongs to one of M. subpopulations throughout his/her lifetime. A learner probabilistically chooses a subpopulation from which he/she samples a variant. We assume that the size of data pool produced by the speakers in each subpopulation is sufficiently large, allowing to regard the data received by each learner to be independently produced. The probability of receiving an already produced variant v_k is given by

Figure 1

Schematic of the Iterated Learning Model (ILM)



Note. In each generation, a learner observes linguistic data from a speaker of the previous generation. The learner updates its internal hypothesis based on this data via Bayesian inference. It then becomes the speaker for the subsequent generation, producing data from its updated hypothesis.

$$\sum_{j=1}^{M} W_{ij} \frac{x_k^{(j)}(t-1)}{N + \alpha^{(j)}}, \tag{5}$$

where $x_k^{(j)}(t-1)$ is the number of times that v_k is produced by speakers born at time $t - 1$ in subpopulation j, $\alpha^{(j)}$, is the value of α (innovation rate) shared by individuals in subpopulation j, and W_{ij} is the probability that a learner in subpopulation i adopts a variant produced by a speaker in subpopulation j. the probability of receiving a novel variant is given by

$$\sum_{j=1}^{M} W_{ij} \frac{\alpha^{(j)}}{N + \alpha^{(j)}}. \tag{6}$$

To examine the effects of specific conditions on the formation of peripheral distribution, we considered four scenarios, each of which was with or without the following two conditions:

Centralized Lexical Innovation (CLI)

New words are innovated exclusively at the cultural center.

Unidirectional Transmission (UT)

Learners can accept linguistic information provided by an outgroup speaker only when the speaker comes from the subpopulation on the center side for some reason(s) (e.g., individuals in surrounding areas perceive a "prestige of center" for a word introduced from the center side), resulting in unidirectional information flow from center to periphery.

We analyzed a structured population consisting of M subpopulations (subpopulation 1, 2, … , M, where we assign an odd number to M) that are arranged on a one-dimensional lattice. The following numerical examples were obtained with seven subpopulations (i.e., $M = 7$). One subpopulation, c, is defined to serve as the cultural center. Here we assumed that it is the same as the geometric center (i.e., $c = (M + 1)/2 = 4$). Inter-subpopulation interaction is restricted to occur between immediate neighbors. In a population without CLI, $\alpha^{(i)} = \alpha$ for all i is assumed. In a population with CLI, $\alpha^{(c)} = \alpha$ and $\alpha^{(i)} = 0$ for $i \neq c$ are assumed. In a population without UT, the interaction is symmetric: learners in subpopulation i can receive data from speakers in both of the neighboring subpopulations $i - 1$ and $i + 1$, with an adjustment at the boundaries. The interaction weight matrix is defined as follows:

$$W_{ij} = \begin{cases} 1 - \frac{m}{2}, & \text{if } i = j = 1 \text{ or } i = j = M \\ 1 - m, & \text{if } i = j \text{ and } 1 < i < M \\ \frac{m}{2}, & \text{if } |i - j| = 1 \\ 0, & \text{otherwise} \end{cases}, \tag{7}$$

where m represents relative magnitude of interaction ($0 \leq m \leq 1$; here we restrict the range to $0 < m \leq 1/2$ so that a learner is most affected by speakers in the same subpopulation). With UT, there is asymmetry in the interaction. The weight matrix in this case is defined as:

$$W_{ij} = \begin{cases} 1, & \text{if } i = j \text{ and } i = c \\ 1 - m, & \text{if } i = j \text{ and } i \neq c \\ m, & \text{if } j \leq c \text{ and } i = j - 1 \\ m, & \text{if } j \geq c \text{ and } i = j + 1 \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

with which learners do not receive data from the peripheral-side subpopulation. Note that, in this case, learners in subpopulation c receive data exclusively from speakers in the same subpopulation (c) regardless of the value of m.

A peripheral distribution is said to be realized when at least one non-center subpopulation i has a non-center partner j that exists beyond the center c and is linguistically closer to i than c is. To measure linguistic dissimilarity between subpopulations, we apply Nei's standard genetic distance (Nei, 1972). In this study, we refer to this metric as the linguistic distance. The Nei's distance is defined as:

$$D_{ij} = -\ln \frac{J_{ij}}{\sqrt{J_{ii} \cdot J_{jj}}}, \tag{9}$$

where J_{ij} is the probability that two randomly chosen data elements—one from subpopulation i and the other from subpopulation j—are copies of the same variant:

$$J_{ij} = \mathrm{E}\left[\sum_{k=1}^{K} \frac{x_k^{(i)}}{N} \frac{x_k^{(j)}}{N}\right]. \tag{10}$$

Note that the sampling is with replacement when $i = j$. To obtain the steady-state values of J_{ij} we derive the set of difference equations describing their temporal dynamics. For each pair of subpopulations $i \neq j$, we have:

$$J_{ij}(t+1) = \sum_k \sum_l W_{ik} W_{jl} (1 - \mu_k)(1 - \mu_l) J_{kl}(t), \tag{11}$$

where $\mu_k = \alpha^{(k)}/(N + \alpha^{(k)})$ is the probability that a novel variant is produced in subpopulation k. The term inside the summation describes the probability that (i) parents of the first and second samples are elements of subpopulation k and l, respectively, at time t ($W_{ik} W_{jl}$), (ii) those parents are copies of an identical variant ($J_{kl}(t)$, and (iii) no innovation has occurred in the course of learning ($(1 - \mu_k)(1 - \mu_l)$). For $i = j$, the equation includes an additional term accounting for self-similarity:

$$J_{ii}(t+1) = \frac{1}{N} + \left(1 - \frac{1}{N}\right) \sum_k \sum_l W_{ik} W_{il} (1 - \mu_k)(1 - \mu_l) J_{kl}(t). \tag{12}$$

The first term of the right-hand side in equation (12) represents the probability that the same element is sampled twice. When different elements are sampled, the variant type of each element is determined in the same manner as equation (11). By solving this system of equations algebraically, we obtain the steady-state values of J_{ij}, which can then be used to compute linguistic distances between subpopulations.

We additionally employed the Bray–Curtis dissimilarity (Bray & Curtis, 1957) to verify robustness and investigate the frequency distribution of concept-wise distance. It is defined as:

$$d_{\mathrm{BC}}(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \frac{\sum_{k=1}^{K} \left| x_k^{(i)} - x_k^{(j)} \right|}{2N}. \tag{13}$$

All numerical simulations and mathematical analyses were implemented using Python 3.11.6. The core libraries utilized for computation included NumPy (version 2.3.4) and SciPy (version 1.16.3).

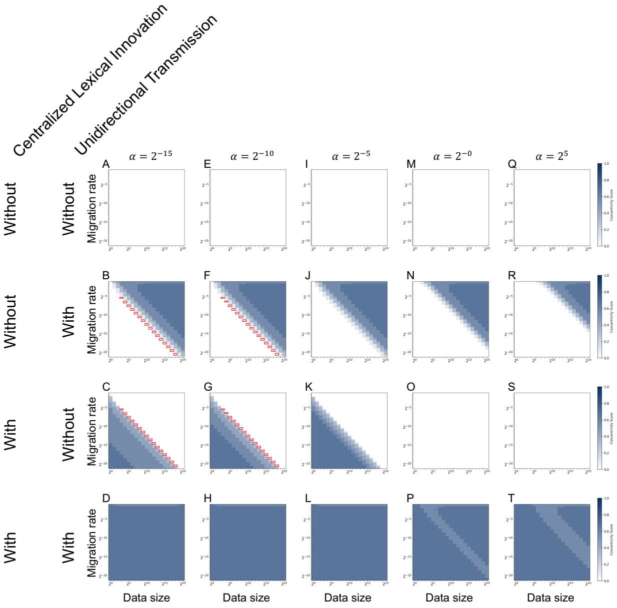
Results

To systematically examine the conditions under which peripheral distribution can be formed, we analyzed the parameter space of the innovation rate α, migration rate of linguistic form m, and data size N (Figure 2). With neither Centralized Lexical Innovation (CLI) nor Unidirectional Transmission (UT), no parameter sets yield peripheral distribution (Figure 2A–E). When both are present, peripheral distribution emerges across all explored parameters (Figure 2P–T). For the single-condition cases (Figure 2F–O), when $\alpha \ll 1$ and $N \gg 1$, the boundaries depend solely on mN (appearing as lines with slope –1 in the log-log plot), with the two conditions exhibiting opposite dependencies: with CLI and without UT, peripheral distribution is observed when mN is below a threshold (Figure 2K–O). With UT and without CLI, peripheral distribution is found in the parameter region in which mN exceeds a threshold (Figure 2F–J). For both of the single-condition scenarios, smaller α values expand the parameter regions where peripheral distribution emerges. The robustness of the above findings was confirmed by Monte-Carlo simulations using Bray–Curtis dissimilarity for between-subpopulation similarity (Supplementary Figure S1).

Further numerical examination revealed that different CLI/UT conditions form different linguistic distance patterns (Figure 3). Although peripheral distribution can be formed with either one of CLI and UT, the underlying mechanisms generate qualitatively different distance patterns (compare Figure 3C and 3F). In contrast, when both conditions are present simultaneously, the resulting patterns show similar characteristics regardless of which mechanism dominates (Figure 3D and 3H). With neither condition, linguistic distance correlates with geographic distance as was well investigated by Nei (1972), and no peripheral distribution was observed (Figure 3A and 3E). The effects of CLI and UT are not simply additive

(Figure 3D and 3H). Detailed distance patterns from other subpopulations are provided in Supplementary Figure S2.

Figure 2
Parameter space for peripheral distribution emergence



Note. Rows represent the four conditions: (A, E, I, M, Q) neither CLI nor UT; (B, F, J, N, R) with UT; (C, G, K, O, S) with CLI; and (D, H, L, P, T) with both. Columns correspond to the parameter α. Horizontal and vertical axes show data size N and migration rate m, respectively (log scale). Colored regions indicate the degree of peripheral distribution (defined as the proportion of subpopulation pairs showing peripheral pattern: pairs where a non-central subpopulation b finds a subpopulation r on the opposite side of the center linguistically closer than the center itself). Red borders in (B, F, C, G) indicate parameter regions in which peripheral distribution is formed both with Centralized Lexical Innovation (CLI) without Unidirectional Transmission (UT) and with UT without CLI. When neither condition is present, no parameter combination yields peripheral distribution; when both conditions are present, peripheral distribution emerges across all explored parameters. For the single-condition cases, when $\alpha << 1$ and $N >> 1$, the boundaries depend solely on mN (appearing as lines with slope −1 in the log-log plot), with the two conditions exhibiting opposite dependencies: CLI produces peripheral distribution when mN is below a threshold, while UT produces it when mN exceeds a threshold. For both single-condition scenarios, smaller α values expand the parameter regions where peripheral distribution emerges. Parameter ranges: $\log_2 \alpha \in [-15, 5]$ with step size 5, $\log_2 m \in [-21, -1]$ with step size 0.25, $\log_2 N \in [0, 20]$ with step size 1, $M = 7$.

To clarify the distinct mechanisms by which CLI and UT generate peripheral distribution, we obtained frequency distribution of concept-wise Bray–Curtis dissimilarities (Supplementary Text). Here shown are the frequency distributions for two representative pairs: one between subpopulation 1 and the cultural center 4, and the other between subpopulation 1 and subpopulation 5 (Figure 4). With neither condition, both pairs spent most of the time at the maximum possible Bray–Curtis dissimilarity 1, indicating divergence (see Supplementary Figure S3 for joint distributions).

CLI led both pairs to have concepts with Bray–Curtis dissimilarity of 0, which is more prominent for the 1–5 pair. With CLI, a variant that has once fixed in the center and spread to peripheral regions is more likely to be kept in peripheral subpopulations than the center because the center generates innovations to replace the existing one, causing distant subpopulations to share identical variants for extended periods. In contrast, UT resulted in pairs having fewer concepts with the maximum dissimilarity of 1, particularly for the 1–5 pair. UT enhances transmission efficiency, reducing chance of complete divergence by enabling peripheral subpopulations to consistently share variants diffused from the center.

When both conditions were present, their effects combined: the 1–4 pair has more concepts with dissimilarity 0, while the 1–5 pair has more concepts near the maximum dissimilarity compared to the 1–4 pair. CLI and UT thus create peripheral distribution through distinct pathways, fixation and efficient diffusion, respectively, yet both ultimately amplify the cultural center's influence on peripheral linguistic composition.

The tendency for peripheral areas to retain older words was found to be driven solely by CLI and not by UT alone (Figure 5). However, the distribution with CLI only and that with both conditions are different; UT increases the age of words linearly with distance from the center.

The expected value of age can be obtained analytically. Let $a_i(t)$ be the expected age of the variants possessed by subpopulation i at time t. We have

$$a_i(t+1) = \sum_j W_{ij} \left(0 \cdot \mu_j + \left(a_j(t) + 1\right)\left(1 - \mu_j\right)\right). \tag{14}$$

Let us introduce the vectors $\boldsymbol{a} = (a_1, a_2, \dots, a_M)^{\mathrm{T}}$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_M)^{\mathrm{T}}$ and denote by $\mathbf{1}$ the M-dimensional vector with all entries equal to one. Define the diagonal matrix as

$$\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu}) = \begin{pmatrix} 1-\mu_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1-\mu_M \end{pmatrix}. \tag{15}$$

The stationary state is obtained by solving

$$\begin{aligned} &\boldsymbol{a} = W \,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})(\boldsymbol{a} + \mathbf{1}) \\ \leftrightarrow\ &\left(I - W \,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})\right)\boldsymbol{a} = W \,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})\mathbf{1}. \end{aligned} \tag{16}$$

where I is the identity matrix. If the matrix $(I - W\,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu}))$ is invertible, we can obtain the solution for $\boldsymbol{a}$ as

$$\boldsymbol{a} = \left(I - W \,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})\right)^{-1} W \,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})\mathbf{1}. \tag{17}$$

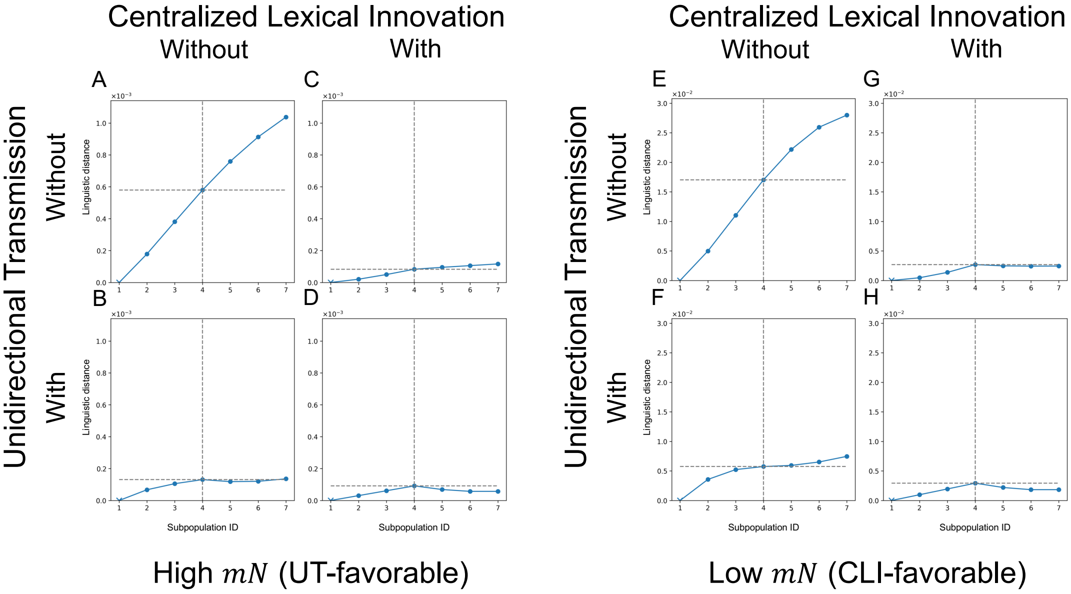
Without CLI, μ_i is uniform across all subpopulations (i.e., $\mu_i = \mu$ for all i). In the steady state, the expected age is likewise uniform across all subpopulations. Denoting this common value by a, the steady-state condition yields $a = (a + 1)(1 - \mu)$, and thus

$$a = \frac{1-\mu}{\mu}. \tag{18}$$

This result shows that in the steady state, the expected age of variants is inversely proportional to the mutation rate μ.

Figure 3

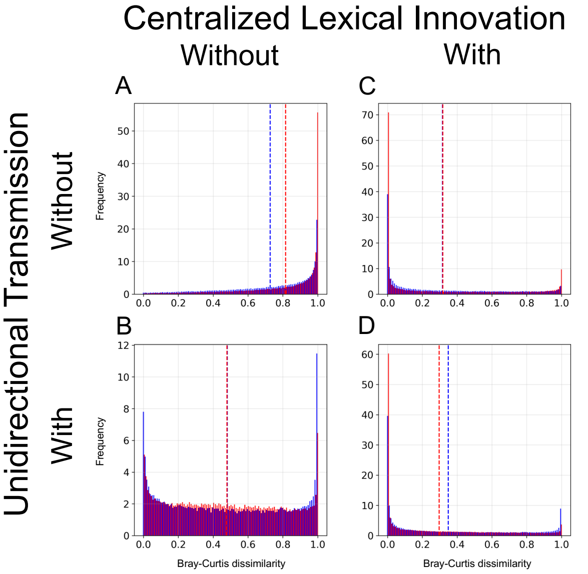
Linguistic distance patterns under different parameter conditions



Note. (A–D) Linguistic distance from a subpopulation at one end of the lattice (subpopulation 1) to all other subpopulations under parameter conditions where UT alone produces peripheral distribution ($m = 2^{-10}$, $N = 2^{10}$, $\alpha = 2^{-15}$, $M = 7$). (E–H) The same analysis under parameter conditions where CLI alone produces peripheral distribution ($m = 2^{-10}$, $N = 2^{5}$, $\alpha = 2^{-15}$, $M = 7$). The x-axis represents the subpopulation index, and the y-axis represents the linguistic distance. The vertical dashed line indicates the position of the central subpopulation, while the horizontal dashed line shows the linguistic distance between subpopulation 1 and the central subpopulation. Peripheral distribution is indicated when subpopulations beyond the cultural center fall below the horizontal dashed line, meaning they are linguistically closer to subpopulation 1 than the center despite being geographically farther. These panels demonstrate that peripheral distribution can emerge through either UT or CLI mechanisms independently, and that the two mechanisms generate qualitatively different distance patterns.

Figure 4

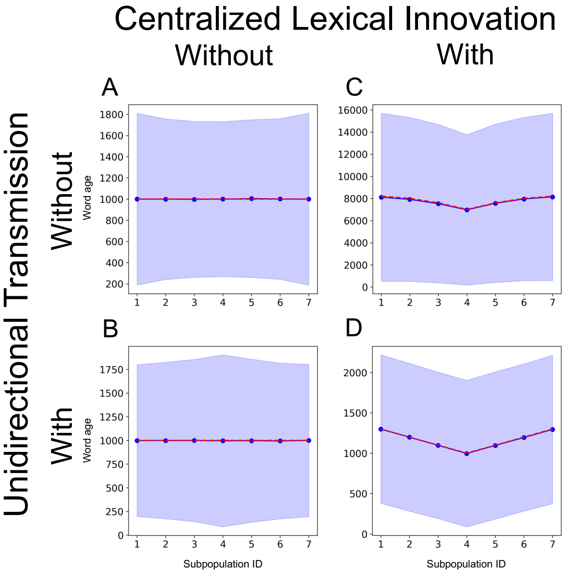
Histograms of concept-wise Bray–Curtis dissimilarity for (A) the population with neither Centralized Lexical Innovation (CLI) nor Unidirectional Transmission (UT), (B) that with UT/without CLI, (C) with CLI/without UT, and (D) with both CLI and UT



Note. For each condition, histograms show dissimilarity for two subpopulation pairs: between a peripheral subpopulation 1 and the cultural center 4 (blue), and between subpopulation 1 and a peripheral subpopulation on the opposite side 5 (red). The plots illustrate the frequency distribution of the dissimilarity for each pair, revealing how often specific dissimilarities occurred under the four experimental conditions. They highlight the distinct impacts of CLI and UT on the linguistic relationship between the center and the other subpopulations. Simulation parameters: $M = 7$, $m = 0.01$, $N = 100$, $\alpha = 0.1$.

Figure 5

Expected age of variants (i.e., the number of generations since the variant first emerged) across subpopulations for (A) the population with neither Centralized Lexical Innovation (CLI) nor Unidirectional Transmission (UT), (B) that with UT/without CLI, (C) with CLI/without UT, and (D) with both CLI and UT



Note. The blue solid line represents the time-averaged simulation results (with shaded regions indicating the standard error), while the red dashed line indicates the analytical solution. Simulation parameters: $M = 7$, $m = 0.01$, $N = 100$, $\alpha = 0.1$.

Discussion

This study investigated the dynamics of language change using an extended Iterated Learning Model that incorporates population structure. Specifically, we examined the conditions under which the theory of peripheral distribution of dialectal forms holds by simulating the diffusion of linguistic features from a cultural center to its periphery. Unlike previous mathematical models (Lizana et al., 2011; Takahashi & Ihara, 2020), which have inevitably assumed Centralized Lexical Innovation (CLI), the present model can examine cases without CLI. Our mathematical and numerical analyses revealed that peripheral distribution can be formed through multiple pathways—either through CLI alone, Unidirectional Transmission (UT) alone, or both conditions combined.

Our analyses show that a peripheral distribution emerges not only when CLI and UT are both present, but also when either condition is active alone (Figures 2 and 3). When both conditions are applied simultaneously, the effects combine to produce the most pronounced peripheral distribution patterns. One of our conclusions that peripheral distribution can be achieved with CLI and without UT aligns with Takahashi & Ihara (2020), despite the two studies utilized different measures for (dis) similarity between subpopulations. A novel finding that the peripheral distribution can be formed with UT and without CLI was also independent from distance measures, indicating robustness of this thesis.

The opposite dependencies on mN between the two single-condition scenarios reflect the fact that the two conditions help to form peripheral distribution through different mechanisms. With CLI, new variants originate exclusively at the center. When mN is small, these variants spread slowly and often fail to reach the periphery, becoming fixed locally near the center while older variants persist at the periphery. In contrast, with UT, variants can originate anywhere but those from the center spread preferentially. Larger mN values reduce drift, enabling peripheral regions to more faithfully reproduce the center's linguistic state with a time lag, thus creating peripheral distribution. These mechanisms also generate distinct patterns in linguistic distance distributions: CLI increases the frequency of zero-distance states between opposite peripheral regions as they share identical archaic variants, while UT reduces the frequency of maximum-distance states by maintaining continuous transmission from the center.

While both CLI and UT can generate peripheral distributions in the absence of the other condition, the present study also provided methods using empirical data to identify which of the two candidate conditions actually formed an observed peripheral distribution. Specifically, different conditions yield (1) different mean word age distributions across subpopulations and (2) different frequency distributions of concept-wise dissimilarity between a specific pair of subpopulations. When innovation is centralized, peripheral regions accumulate older variants, whereas uniform innovation produces uniform mean age distributions; UT modulates this by creating more linear age gradients. As for the frequency distributions of concept-wise dissimilarity, CLI characteristically increases the zero-distance frequency between opposite peripheries (i.e., the frequency of concepts each of which is called by the same word in the two subpopulations), while UT reduces the maximum-distance frequency (i.e., the frequency of concepts each of which is called differently in the two subpopulations). Joint examination of these complementary patterns may therefore allow researchers to determine whether a given peripheral distribution has arisen from CLI, UT, or both.

The assumption that linguistic innovation always occurs at a cultural center and diffuses outward is echoed in the Principle of Adjacent Distribution (Shibata, 1969), which posits that geographically adjacent communities using distinct words likely innovated them in geographic order. While this principle has traditionally been linked to peripheral distribution, our results demonstrate that peripheral distribution can be formed without CLI, enabling more nuanced interpretation of dialect data than previous approaches that implicitly assumed CLI.

Further studies using the present or an updated model are needed. Our model examined only unidirectional flow and symmetric bidirectional flow of linguistic data and simplified the distinction between cultural centers and surrounding regions. For example, cases with more than one center have not been examined. Language change can be influenced by additional factors such as heterogeneous social networks, socioeconomic status, group ideologies, and interactions with other linguistic communities (Hock, 1991). Future research should incorporate these factors and validate model outcomes with historical linguistic data. Additionally, the model's capability to represent language change in arbitrary network structures presents promising avenues for exploring language evolution within complex social relationships, such as those in online communities.

Acknowledgments

We would like to thank Kenichi Sawai for his helpful discussions and support throughout this research.

Author contribution

Seo Hachimaru: Conceptualization, Software, Formal analysis, Visualization, Writing - Original Draft. **Motohide Seki**: Writing - Review & Editing, Supervision.

Data accessibility & program code

https://github.com/seo-80/multi-agents-ilm

Supplementary material

Electronic supplementary materials are available online.

References

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. https://doi.org/10.2307/1942268

Bromham, L. (2025). Population size and language change: An evolutionary perspective. *Annual Review* *of Linguistics*, *11*, 183–208. https://doi.org/10.1146/annurev-linguistics-031422-123959

De Gregorio, J., Toral, R., & Sánchez, D. (2024). Exploring language relations through syntactic distances and geographic proximity. *EPJ Data Science*, *13*(1) , Article 61. https://doi.org/10.1140/epjds/s13688-024-00498-7

Hock, H. (1991). *Principles of historical linguistics* (2nd ed.). Mouton de Gruyter.

Inoue, F. (2010). Real and apparent time clues to the speed of dialect diffusion. *Dialectologia: Revista Electrònica*, *5*, 45–64. https://raco.cat/index.php/Dialectologia/article/view/198836

Kirby, S. (2001). Spontaneous evolution of linguistic structure - An iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110. https://doi.org/10.1109/4235.918430

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(12), 5241–5245. https://doi.org/10.1073/pnas.0608222104

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 - Statistical, Nonlinear, and Soft Matter Physics*, *83*(6), Article 066116. https://doi.org/10.1103/PhysRevE.83.066116

Nei, M. (1972). Genetic distance between populations. *The American Naturalist*, *106*(949), 283–292. https://doi.org/10.1086/282771

Reali, F., & Griffiths, T. L. (2010). Words as alleles: Connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1680), 429–436. https://doi.org/10.1098/rspb.2009.1513

Shibata, T. (1969). *Gengo-chirigaku no hōhō [Methods in linguistic geography]*. Chikuma Shobo.

Takahashi, T., & Ihara, Y. (2020). Quantifying the spatial pattern of dialect words spreading from a central population. *Journal of the Royal Society Interface*, *17*(168), Article 20200335. https://doi.org/10.1098/rsif.2020.0335

Yanagita, K. (1930). *Kagyūkō [A study of the snail names]*. Tōkō Shoin.