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

Seo Hachimaru [1,*], Motohide Seki[1]

[1]Kyushu University, 4-9-1 Shiobaru, Minami Ward, Fukuoka 815-8540, Japan
[*]Author for correspondence (hachimaru.seo.625@s.kyushu-u.ac.jp)

short title: Necessary Conditions for Peripheral Distribution of Dialect

Abstract
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 occasional formation of a peripheral distribution of words, 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 agent-based simulations and 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 simulation

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, geographically close languages or dialects form a cluster.

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 generated exclusively at a cultural center (hereafter referred to as centralized neologism creation), and (2) people in surrounding areas perceive a "prestige of center" for a word introduced from the center side and accept it. 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.

Recently, not many but several mathematical modeling studies quantitatively evaluated the validity of verbal arguments based on the theory of peripheral distribution of dialectal forms. Lizana et al., (2011) incorporated both of the aforementioned key conditions to obtain the peripheral distribution in their simulations. Takahashi & Ihara, (2020) claimed that the realistic word distribution can be even formed only with the first condition that new words emerge solely at the center.

However, language change does not necessarily conform to the assumption that new words are created only in cultural centers. Indeed, there are documented 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 clear evidence of linguistic innovation flowing from peripheral regions into cultural centers (Inoue, 2010). Therefore, it is worthwhile to investigate whether peripheral distributions emerge even when the center does not play a special role in neologism creation.

In this study, an extended Iterated Learning Model (ILM) was developed to clarify the conditions under which peripheral distributions are formed. Previous studies have employed ILMs to explain the evolution of linguistic variants through cultural transmission processes (Kirby, 2001; Kirby et al., 2007). Although these studies have an advantage of simulating individual-level decision-making, they did not address questions concerning the geographic diffusion patterns, as they have considered an unstructured population. In contrast, the present ILM, by considering multiple subpopulations, has the capability to (analytically and mathematically) analyze how linguistic variations spread within a structured population. We specifically address the question of whether the two core assumptions of theory of peripheral distribution of dialectal forms are each independently or jointly sufficient for the formation of peripheral distributions.

Model

The Iterated Learning Model (ILM) framework simulates language transmission across generations. In this model, agents-representing subpopulation-acquire linguistic knowledge from data generated by the previous generation and subsequently produce data for the next generation. This process of intergenerational transmission, characterized by imperfect learning and production, drives language change through linguistic drift (Reali & Griffiths, 2010).

In this paper, we use Bayesian inference as individual learning (Kirby, 2001; Kirby et al., 2007). A learner receives data containing the number of linguistic forms v_k such as sounds, words, or grammatical constructions, and infers the frequency distribution of variants following Bayes' theorem (Christian P. Robert, 1994). The variables $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are K dimensional vectors ($\boldsymbol{x} = [x_1, x_2, \dots x_K]$ and $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots \theta_K]$), where K is the total number of variants. The elements x_k and θ_k denote frequency and the estimated probability, respectively, of v_k.

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

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 there is no selection of variants. Agents are neutral between variants. The distribution is determined only by parameter α and K.

$$p(\boldsymbol{\theta}) \propto \prod_{k}^{K} \theta_k^{\alpha/K-1},$$

The parameter α/K moderates the learner's preference for diversity: the larger α/K value represents a tendency to retain the greater number of variants. Since the Dirichlet distribution is the conjugate prior of a multinomial distribution, it belongs to the same family as the prior.

It can be shown that probability that an agent produces a v_k is calculated as

$$\frac{x_k + \frac{\alpha}{K}}{N + \alpha}$$

See (Reali & Griffiths, 2010) for details.
We extend this model to the extreme case of $\alpha = 0$ by taking limit as $\alpha \to 0$. In this case, the Dirichlet prior formally approaches

$$p(\boldsymbol{\theta}) \propto \prod_{k=1}^{K} \theta_k^{-1},$$

which corresponds to the Haldane prior. Under this limiting assumption, the posterior distribution is given by

$$p(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto \prod_{k=1}^{K} \theta_k^{x_k-1},$$

and the production probability for variant v_k simplifies to

$$\frac{x_k}{N}$$

Which holds even if $x_k = 0$.

We extended the aforementioned model to scenarios where each generation consists of multiple agents. In this model, each agent probabilistically selects a parent agent and samples a piece of data from the data pool generated by that parent agent for its learning process. We assume that the data pool generated by the parent agents is sufficiently large, so the data received by each agent in the subsequent generation can be considered independently generated. The probability that the i-th agent in generation t receives v_k depends only on the number of times v_k received by each agent of generation $t-1$, denoted as $x_{k(t-1)}^{(j)}$.

$$\sum_{j}^{M} W_{ij} \frac{x_{k(t-1)}^{(j)} + \frac{\alpha^{(j)}}{K}}{N^{(j)} + \alpha^{(j)}},$$

where W_{ij} is the probability that i-th agent adopts a data element produced by j-th agent of the parent generation. In other words, W_{ij} represents the degree of connection between agents.

Insert Figure 1 here

Figure 1. Each agent updates its hypothesis by receiving data generated by the previous generation

with a weighted probability and then generates data for the next generation.

Consider the case where every mutation produces a variant that has never been observed in the population. In this case, the number of variants is potentially infinite, which yields the use of an infinite dimensional Dirichlet distribution for the prior distribution. The agents generate already existing variant k with probability $x_k/(N+\alpha)$ under the x_k. It also generates a completely novel word with probability $\alpha/(N+\alpha)$. As in the case of finite variants, errors generate all variants equally, but no longer generate existing variants because there are a finite number of variants that already exist, whereas there are an infinite number of variants that do not yet exist.

The probability of receiving an already produced variant is given by

$$\sum_{j}^{M} W_{ij} \frac{x_{k(t-1)}^{(j)}}{N_j + \alpha_j}$$

the probability of receiving a novel variant is given by

$$\sum_{j}^{M} W_{ij} \frac{\alpha^{(i)}}{N^{(i)} + \alpha^{(i)}}$$

We confirmed that our model can reproduce the peripheral distribution predicted by theory of peripheral distribution of dialectal forms, which explains how linguistic changes diffuse from cultural centers to peripheral areas, under certain parameter settings. According to this theory, new linguistic features emerge in influential cultural centers and gradually spread to surrounding regions due to the prestige associated with these centers (Yanagita, 1930). Consequently, newer linguistic forms cluster near the center, while older forms persist in the outer areas, resulting in a pattern of concentric circles.

To examine the effects of specific conditions on this distribution, we conducted four sets of simulations, each with and without the following two conditions:
centralized neologism creation: New words are generated exclusively at a cultural center.
center prestige: People in surrounding areas perceive a "prestige of center" for a word introduced from the center side and accept it.

We performed simulations on a one-dimensional lattice, where agents are arranged in a line and can only interact with their immediate neighbors. A designated cultural center was established along this line, and we analyzed how the above conditions affect the distribution of linguistic features.

In the scenario without the center prestige (condition 2), each agent i receives input from both its left neighbor $i-1$ and right neighbor $i+1$, with appropriate adjustments at the boundaries. The interaction weight matrix is defined as follows:

$$W_{ij} = \begin{cases} 1 - \frac{m}{2}, if\ i = j\ and\ (i = 1\ or\ i = n) \\ 1 - m, \quad if\ i = j\ and\ 1 < i < n \\ \frac{m}{2}, \quad if\ |i - j| = 1 \\ 0, \quad ohterwise \end{cases}$$

When the center prestige is applied, the matrix W becomes asymmetric. Specifically, we define the cultural center at position

$$c = \frac{M+1}{2}$$

where M is restricted to odd numbers to ensure that a unique center exists. The weight matrix is redefined as:

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

We define the linguistic distance between agents as the Manhattan distance, calculated based on the frequency of the data each agent receives. In addition, a peripheral distribution is said to be realized when at least one agent has a partner agent that is geographically more distant and linguistically closer to it than the center agent.

Results

Our analysis revealed that the most pronounced peripheral distribution of linguistic features emerged when two key factors were simultaneously applied: the centralized neologism creation in the cultural center and the center prestige (Figure 2, bottom right). Interestingly, we also observed peripheral distributions when only one of these two conditions was present, as illustrated in Figure 2. This finding suggests that the mechanisms underlying the formation of peripheral distribution may be simpler than previously thought. Distribution is qualitatively different for each condition. The top left panel of Figure 2 is the case where none of the two conditions are present, and the linguistic distance is simply greater the farther away the group is. The result shows that the model setup is sufficient to represent the peripheral distribution.

Insert Figure 2 here

Figure 2. Heatmap showing the time-averaged linguistic distance between subpopulations during the simulation. Four conditions were examined based on the presence or absence of two factors: the centralized neologism creation and the center prestige. For all simulations, parameter values were set to M =15, m =0.01, and $N^{(i)}$ =100 for all i. For the parameter α, when the centralized neologism creation was enabled, only the central subpopulation $i = (n+1)/2$ was assigned $\alpha_i = 0.1$ while all other subpopulations were set to 0; when it was disabled, all subpopulations had $\alpha_i = 0$. Data were time-averaged over the final 90% of 1,000,000 generations and ensemble-averaged over 200 independent runs.

The tendency for peripheral areas to retain older words was found to be driven solely by the centralized neologism creation and not by the prestige of the center alone (Figure 2). However, the distribution is different when only the centralized neologism creation is applied and when both two conditions are applied. With the center prestige, the age of words increases linearly with distance from the center.

Insert Figure 3 here

Figure 3. Expected age of variants (i.e., the number of generations since the variant first emerged) across subpopulations under four different conditions. The blue solid line represents the time-averaged simulation results (with the colored regions indicating the variance, computed as the variance of the means from 200 independent runs), while the red dashed line indicates the analytical solution. For all simulations, parameter values were set to $M = 15$, $m = 0.01$, and $N^{(i)} = 100$ for all i. For the parameter α, when the centralized neologism creation was enabled, only the central subpopulation $i = (n+1)/2$ was assigned $\alpha_i = 0.1$ while all other subpopulations were set to 0; when it was disabled, all subpopulations had $\alpha_i = 0$. Data were time-averaged over the final 90% of 1,000,000 generations and ensemble-averaged over 200 independent runs.

The expected value of oldness can be obtained analytically. Derivation is performed below. Let $a_i(t)$ be the expected age of the variants possessed by the i-th subpopulation at time t.

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

where $\mu_i = \alpha_i \,/\, (N_i + \alpha_i)$. Let us introduce the state vector $\boldsymbol{a} = (a_1, a_2, \dots, a_M)^T$ and denote by $\mathbf{1}$ the vector with all entries equal to one. Define the diagonal matrix

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

In this notation, the stationary condition becomes

$$\boldsymbol{a} = W \,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})(\boldsymbol{a} + \mathbf{1})$$

Rearranging the terms, we obtain

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

where I is the identity matrix. Assuming that the matrix $I - W\,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})$ is invertible, which is typically guaranteed if the spectral radius of $W\,\mathrm{diag}(\mathbf{1} - \boldsymbol{\mu})$ is less than one, the solution for $\boldsymbol{a}$ is given by

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

Consider the case where there is not a condition of the centralized neologism creation. In this case μ and $a(t)$ are constant across all subpopulations, and it is clear that $a(t+1)$ is equal for all subpopulations. Therefore, under the condition $a_i(0) = a_j(0)\ (\forall i, j)$, we have $a_i(t) = a_j(t)\ (\forall i, j, t)$.

In the steady state, we can assume $a_i(t+1) = a_i(t) = a$ for all i. This leads to
$a = (a+1)(1-\mu)$
Solving for a:

$$a = \frac{1-\mu}{\mu}$$

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

Discussion

This study investigated the dynamics of language change using an extended Iterated Learning Model (ILM) that incorporates multiple agents and network structures. 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 only examined the formation of peripheral distribution under the assumption that new words are generated exclusively at the center, our model extends this by evaluating whether such patterns can also emerge without this constraint.

The present study defined and calculated linguistic distance between subpopulations as a key measure to define a peripheral distribution. Specifically, we evaluated the expected word distance, which can be regarded as a measure of overall linguistic distance. A previous study (Takahashi & Ihara, 2020), on the other hand, used subpopulation mean age to define peripheral distribution. Both measures yielded the same conclusion that a peripheral distribution is achieved when the condition of the centralized neologism creation holds.

Furthermore, the present model can examine the situation in which the centralized neologism creation, which was inevitably built-in in the model of Takahashi & Ihara (2020), does not hold. Our findings suggest that even when linguistic innovation occurs at multiple locations, peripheral distribution patterns can emerge under certain conditions, particularly when transmission mechanisms emphasize prestige or other forms of differential adoption. This finding challenges the traditional view that both factors are essential for the peripheral distribution, suggesting instead that multiple mechanisms can independently produce similar macroscopic language patterns.

While one of the two conditions is sufficient to generate peripheral distribution as discussed above, only the centralized neologism creation can explain the retention of older linguistic forms in peripheral regions. This was not only observed in our numerical simulations but also confirmed analytically: When all subpopulations generate new words equally, they exhibit the same expected age of variants is indifferent among subpopulations. On the other hand, when innovation is restricted to the center, peripheral subpopulations predominantly acquire words diffused from the center, resulting in the accumulation of older forms. The prestige of cultural center influences this process by insulating the center from surrounding subpopulations, preserving a lower expected age similar to the uniform innovation scenario. In contrast, without the prestige, new words spread less efficiently, increasing the overall expected age of variants. It follows that we can distinguish whether there is a centralization bias in the rate of neologism creation by examining the age distribution of linguistic forms.

The assumption that linguistic innovation occurs at a single location and diffuses outward is echoed in a traditional geolinguistic hypothesis known as the Principle of Adjacent Distribution (Shibata, 1969). This principle posits that when geographically adjacent communities A–B–C each use a distinct word a–b–c, the historical relationship among the forms likely follows the geographic order: either the words were innovated in this chronological order a, b, and c, or in the reverse order c, b, and a. As the assumption above leads to peripheral distribution of dialect, it might be thought that the Principle of Adjacent Distribution can be applied to any population in which a peripheral distribution is observed. However, the results in the present study clearly indicated that is not true, the peripheral distribution is achieved without the centralized neologism creation, which results in a gradient of word age.

Further studies using the present model or an updated model are needed. For example, our simulations modeled only a unidirectional flow of linguistic data and assumed the centralized neologism creation as the key differentiator between cultural centers and surrounding regions. However, 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 variables and validate our findings with historical linguistic data and real-world case studies to refine the proposed framework. In addition, leveraging the model's capability to represent language change in arbitrary network structures presents a promising avenue for exploring language evolution within complex social relationships, such as those observed in online communities where linguistic variations emerge and spread in unique ways.

Acknowledgments

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

Author contribution

SH developed the model concept and analyzed the model, and MS supervised. SH and MS wrote the manuscript.

Data accessibility & program code

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

References

Bromham, L. (2025). Population Size and Language Change: An Evolutionary Perspective. *Annu. Rev. Linguist. 2025*, *11*, 183–208. https://doi.org/10.1146/annurev-linguistics-031422

Christian P. Robert. (1994). *The Bayesian choice : a decision-theoretic motivation* (1st ed.). Springer. https://doi.org/10.1007/978-1-4757-4314-2

De Gregorio, J., Toral, R., & Sánchez, D. (2024). Exploring language relations through syntactic distances and geographic proximity. *EPJ Data Science*, *13*(1). 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. In *Dialectologia* (Vol. 5).

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*, *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). https://doi.org/10.1103/PhysRevE.83.066116

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 hoho (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). https://doi.org/10.1098/rsif.2020.0335

Yanagita, K. (1930). *Kagyuko*. Toko Shoin.

Figure 1.

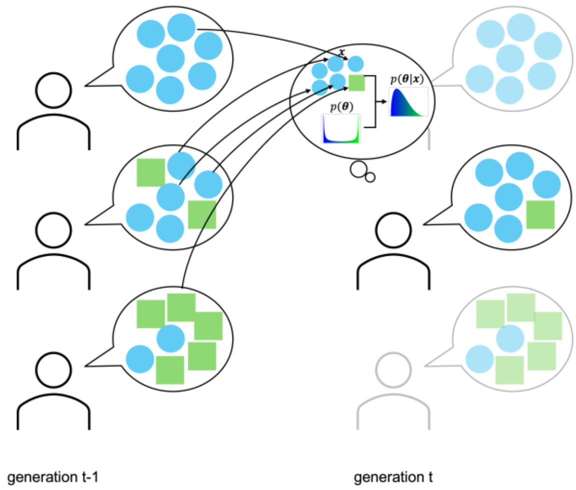

Figure 2.

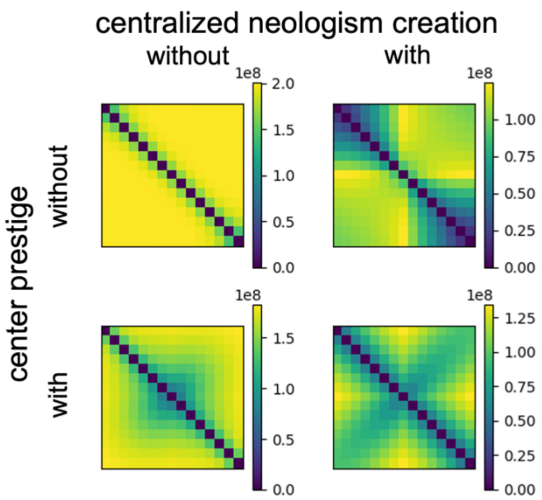

Figure 3.