

方言の周圏分布形成条件の数理的解明： スケーリング極限における形成境界の解析的導出

Mathematical Clarification of Conditions for Circumferential Dialect Distribution: Analytic Derivation of Formation Boundaries in the Scaling Limit

八丸 世旺[†], 関 元秀[‡]
Seo Hachimaru[†], Motohide Seki[‡]

[†] 北陸先端科学技術大学院大学, [‡] 九州大学

[†] Japan Advanced Institute of Science and Technology, [‡] Kyushu University
s2620004@jaist.ac.jp

概要

方言周圏論が仮定する CLI (中心のみでの語彙革新) と UT (一方向伝達) は, 各々単独でも周圏分布形成に十分であることをベイズ繰り返し学習モデルで示した. これは両条件の同時成立を前提とする従来モデルの仮定を緩和する. さらに, 移住と新語生成がともにまれなスケーリング極限では, 定常状態方程式がリアプノフ型の線形方程式に帰着し, 言語距離が「移住率 × 集団サイズ」と新語生成パラメータのみの関数となることを代数的に証明した. UT 単独条件では, 周圏分布が形成されるこの積の閾値を解析的に導出した.
キーワード: 言語変化, 方言地理学, ベイズ学習, 繰り返し学習モデル, 周圏分布

1. はじめに

言語の地理的変異は一般に地理的近接性と相関し, 隣接集団間で類似した言語形式が共有される (De Gregorio et al., 2024). しかし一部の方言分布はこのパターンから逸脱する. 方言周圏論 (柳田, 1930) が提唱する「周圏分布」はその典型例であり, 新しい語彙が文化的中心地から外側へ拡散することで, 中心から離れるほど古い語彙が残存するという同心円状の分布パターンを説明する.

周圏分布の形成条件として, 従来二つが提案されてきた: (1) 語彙革新は文化的中心地においてのみ発生する (中心のみでの語彙革新; CLI), (2) 周辺地域は中心の威信を認識し, 中心から周辺への一方向に言語情報を受容する (一方向伝達; UT). CLI は新語が生まれる場所を中心のみに, UT は情報の流れる向きを中心から周辺への一方向に限定する仮定である. Lizana et al. (2011) はこの二条件をともに組み込んだ数理モデルで周圏分布を再現し, Takahashi & Ihara (2020) は CLI のみでも周圏分布が形成されることを示した.

一方, 実際の言語変化は必ずしもこれら二条件に従わない. たとえば「うざったい」は東京西部郊外で発生し後に中心部へ広まった (Inoue, 2010). この事例は, 語彙革新が中心以外で生じる点で CLI に反し, 情報が周辺から中心へ流れる点で UT にも反する. すなわち現実の変化は CLI・UT のいずれか, あるいは両方を破りうる. しかしいずれの先行研究も, CLI・UT が個別に充足されない場合の周圏分布形成を検証しておらず, また周圏分布が形成される具体的なパラメータ条件 (移住率 m や集団サイズ N との関係) も分析していない.

本研究では, CLI・UT を独立に検証できる拡張ベイズ繰り返し学習モデルを構築し, 二条件の有無が周圏分布形成に与える影響を数値解析と代数的証明により明らかにする. まず, CLI のみ・UT のみの各単一条件でも周圏分布が形成されうることを示す. これは, 周圏分布の形成に両条件が必須ではないことを意味し, 従来の数理モデルが前提としてきた仮定を緩和する. さらに, $m \ll 1$, $\alpha/N \ll 1$ のスケーリング極限において, 定常状態の言語距離が mN (部分集団間の移住率 × 集団サイズ) と新語生成パラメータ α の関数として記述できることを示し, UT のみ条件での周圏分布形成境界 $mN > \frac{1}{2^{n-1}}$ ($n = (M-1)/2$) を代数的に証明する.

2. モデル

個体の学習と再生産は Real & Griffiths (2010) のベイズ繰り返し学習モデル (ILM) に基づく. 本研究はこれを, $M = 7$ 個の部分集団が直線状 (1 次元鎖) に並んだ空間構造に拡張する. これは文化的中心から外側へ一次元的に延びる地理を理想化したものである. 部分集団を $i \in \{1, \dots, M\}$ で添字づけし, 中央の部分集団 $c = (M+1)/2$ ($M = 7$ では $c = 4$) を文化的中心地とする. 各部分集団は集団サイズ N をもち, 隣接す

る部分集団 $i \pm 1$ とのみ移住を通じて交流する（最近接結合）；両端の部分集団（ $i = 1, M$ ）は片側にのみ隣接する。

各世代の学習者は、前世代が産出した語から N 語を変異形の標本として観測し、その頻度分布を推論して次世代へ語を産出する。事前分布を対称ディリクレ分布とし、変異形の総数を無限と仮定すると、部分集団 i の学習者が既存の変異 k を一つの観測として受け取る確率は

$$\sum_j W_{ij} \frac{x_k^{(j)}}{N + \alpha^{(j)}}$$

で与えられ、未観測の新たな変異を受け取る確率は $\sum_j W_{ij} \mu_j$ ($\mu_j = \alpha^{(j)} / (N + \alpha^{(j)})$) である。ここで $x_k^{(j)}$ は前世代の部分集団 j で変異 k が観測された回数 ($\sum_k x_k^{(j)} = N$)、 W_{ij} は学習者が部分集団 j の産出から観測を得る確率（部分集団間の変異形の移住に相当）、 $\alpha^{(i)}$ は部分集団 i のディリクレ事前分布の濃度パラメータであり新語生成率 μ_i を定める（ $\alpha^{(i)}$ が大きいほど新語が生まれやすい）。

CLI の有無は $\alpha^{(i)}$ の設定で制御する：CLI ありでは $\alpha^{(c)} = \alpha$ 、それ以外の部分集団では $\alpha^{(i)} = 0$ ；CLI なしでは全部分集団で $\alpha^{(i)} = \alpha$ 。UT の有無は、隣接部分集団間を結合する重み行列 $W = I + mL$ の構造で制御する（ m ：移住率、 L ：交流グラフの離散ラプラシアン）。UT なしでは L を対称な離散ラプラシアンとし、各部分集団は両隣と等方的に双方向で交流する；UT ありでは L を中心から周辺への有向グラフに対応する非対称行列とし、各非中心部分集団は中心側の隣接部分集団からのみ移住を受け入れる（中心は流入を受けない）。

言語距離は Nei の遺伝距離 $D_{ij} = -\ln(J_{ij}/\sqrt{J_{ii}J_{jj}})$ で定義する（ J_{ij} ：部分集団 i と j の変異形頻度ベクトルの内積の期待値）。周囲分布は、非中心部分集団 i について、地理的に中心を越えた反対側の部分集団 j との距離が中心との距離より小さい（ $D_{ij} < D_{ic}$ ）場合であるとする。

3. 結果

数値計算：片方の条件のみでも周囲分布は形成される

4 条件（CLI・UT 各有無の組み合わせ）を数値計算で調べた（図 1）。CLI も UT もない場合、いかなるパラメータでも周囲分布は形成されない。両条件が存在する場合、探索した全パラメータ領域で周囲分布が出現する。重要なのは単一条件のケースである：CLI のみ、あるいは UT のみでも、特定のパラメータ領域で周囲分布が形成される（表 1）。また、UT のみと CLI

のみとは異なるパラメータ条件下で周囲分布が形成され、二条件は mN に対して逆の依存性を示す：UT のみでは mN が大きいほど、CLI のみでは mN が小さいほど形成されやすい。

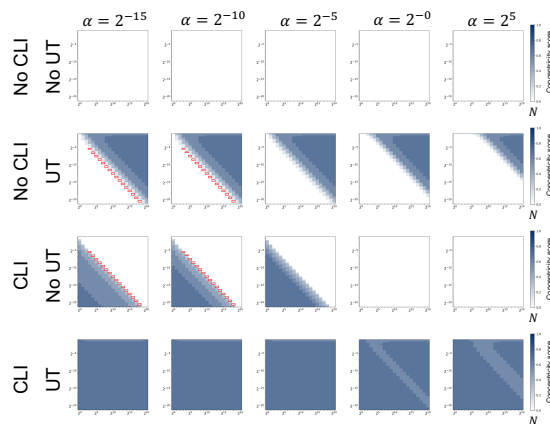


図 1 周囲分布出現のパラメータ空間（ $M = 7$ ）。行は 4 条件、列は α の値に対応。軸は m （部分集団間の移住率、対数スケール）と N （集団サイズ、対数スケール）。着色は周囲分布の程度を示す。 α が小さい領域で UT のみ・CLI のみの境界が $mN = \text{const}$ の直線として現れ、二条件は mN に対して逆の依存性を持つ。

表 1 4 条件における周囲分布形成のまとめ

条件	周囲分布が現れる領域
CLI なし・UT なし	形成されない
UT のみ	mN が大きい領域
CLI のみ	mN が小さい領域
CLI × UT	全領域

境界の直線構造：スケーリング極限での解析的説明

パラメータ空間を系統的に調べると、 α が小さい領域で単一条件の形成境界が対数-対数プロット上で傾き -1 の直線として現れる。この構造は、 $m \ll 1$, $\alpha/N \ll 1$, $mN = O(1)$ というスケーリング極限で解析的に説明できる。

このスケーリング極限において J の差分方程式を一次展開すると、定常状態がリアプノフ型方程式

$$(mN)(LJ + JL^T) - (\Lambda J + J\Lambda) + (I - J_{\text{diag}}) = \mathbf{0} \quad (1)$$

に帰着する（ L ：ラプラシアン行列、 $\Lambda = \text{diag}(\alpha^{(i)})$ ）。この方程式に含まれるパラメータは mN と α のみであるため、この極限においては定常解 J^* が mN と α のみに依存し、周囲分布の境界は $mN = f(\alpha)$ の曲線（対数-対数プロット上の傾き -1 の直線）として帰着する。

さらに UT のみ条件では、 L が中心→周辺の有向グラフであるため、式 (1) を中心成分から再帰的に解くことができる。伝播減衰率 $\beta = mN/(mN + 2\alpha)$ ，交差伝播係数 $\gamma = mN/(mN + \alpha)$ ，および $n = (M - 1)/2$ を用いると J_{ij} の厳密解が得られ、 $\alpha \rightarrow 0$ 極限で周囲分布条件を解くと

$$mN > \frac{1}{2^n - 2} \quad (2)$$

が導かれる。数値計算で観察された直線境界が、代数的に $mN = \frac{1}{2^n - 2}$ という定数に帰着することが厳密に証明された。

CLI と UT の異なる距離パターンと語彙年齢

CLI と UT は異なるメカニズムで周囲分布を生成するため、言語距離パターンにも質的な差異が現れる (図 2)。CLI のみでは、中心を超えた後に距離が再び増加する (中心周辺で語彙が固定されるため反対側の周辺と類似する)。UT のみでは、中心を超えた後も距離が低いまま維持される (周辺が中心の語彙状態を時間遅れで再現するため)。

語彙年齢 (変異形が最初に出現してからの世代数) の空間分布も CLI と UT で質的に異なる (図 3)。CLI がない場合 ($\mu_i = \mu$ 一様)，定常状態では全部分集団で期待年齢が等しく $a = (1 - \mu)/\mu$ となる。CLI がある場合は周辺ほど古い語彙が蓄積し非線形な年齢勾配が現れる一方、UT の追加により線形勾配になる。すなわち、周辺部の語彙年齢の増大は CLI のみによって駆動され、UT のみでは生じない。

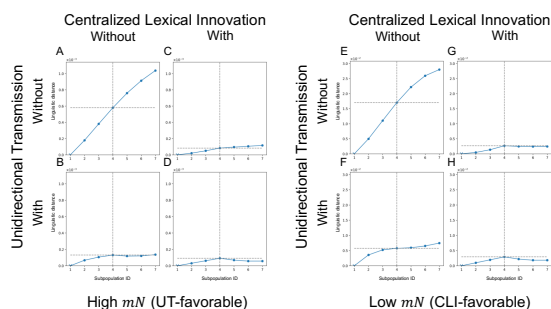


図 2 異なるパラメータ条件下での言語距離パターン ($M = 7$)。 (A–D) UT のみで周囲分布を生じるパラメータ条件； (E–H) CLI のみで周囲分布を生じるパラメータ条件。部分集団 1 からの言語距離を他の部分集団に対してプロットした。垂直破線は中心 ($c = 4$)，水平破線は部分集団 1 と中心との距離を示す。CLI のみ (C, G) では中心を超えた後に距離が再び増加し，UT のみ (B, F) では低いまま維持される。

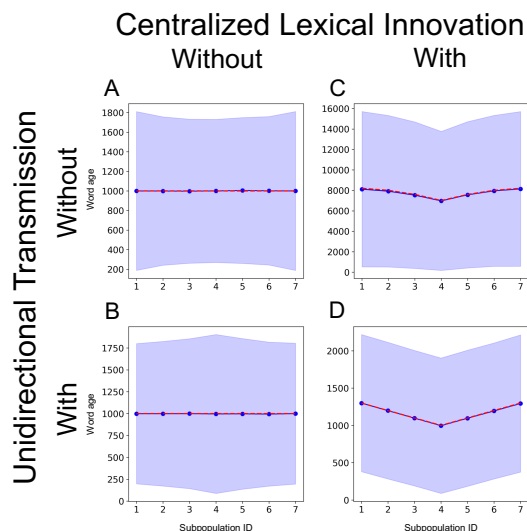


図 3 4 条件における各部分集団の語彙年齢期待値 ($M = 7$, $m = 0.01$, $N = 100$, $\alpha = 0.1$)。青実線はシミュレーション結果 (着色領域は標準誤差)，赤破線は解析解。CLI なし条件 (A, B) では全部分集団で期待年齢が同様。CLI あり条件 (C, D) では周辺ほど古い語彙が蓄積し，UT の追加 (D) により線形勾配になる。

4. 考察

本研究は、数値計算で観察された境界の直線構造を解析的に説明することに成功した。 $m \ll 1$, $\alpha/N \ll 1$ のスケーリング極限でリアプノフ型方程式 (1) に帰着し、その解 J^* が mN と α のみに依存するため、境界は対数-対数プロット上で直線状に現れる。ただしこれはあくまで当該スケーリング極限での帰結であり、一般的な主張ではない。方言の周囲分布は地域間移動が限られた古い時代に形成されたと考えられ、 $m \ll 1$ および $\alpha/N \ll 1$ (新語生成率が小さい) はいずれも妥当な仮定である。

UT のみの境界 $mN > \frac{1}{2^n - 2}$ (式 2) は $\alpha \rightarrow 0$ 極限での値であり、この極限で境界は新語生成パラメータに依存しなくなる。 α が大きいほど、UT のみで周囲分布が形成されるのに必要な mN の閾値は高くなる。すなわち、 α が小さいほど、より広い mN 領域で周囲分布が形成される。

CLI と UT は異なる言語距離パターンと語彙年齢分布を生成するため、実方言データに観察されるこれらの指標を分析することで、観察された周囲分布の形成に CLI・UT のいずれが関与しているかを推定できる可能性がある。今後の課題として、CLI のみ条件における境界 $f_{CLI}(\alpha)$ の代数的導出、現代の複雑な交流

ネットワークをもつ集団での言語変化への応用，および実方言データへの診断指標の適用が挙げられる．

文 献

- 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>.
- Inoue, F. (2010). Real and apparent time clues to the speed of dialect diffusion. *Dialectologia*, 5.
- 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. <https://doi.org/10.1103/PhysRevE.83.066116>.
- Real, 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*, 277 (1680), 429–436. <https://doi.org/10.1098/rspb.2009.1513>.
- 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), 20200335. <https://doi.org/10.1098/rsif.2020.0335>.
- 柳田 国男 (1930). 蝸牛考. 刀江書院.