

HUMAI Anthology

いくつかの in silicio 歴史言語学に関する展望

白鳥 詩織

1. はじめに

わたくしは自分の分野の未来について思索をすることが好きで、その展望をもとに研究計画を作成するのはなかば趣味である。この趣味は将来予想される過酷な競争を生き残るために涵養してきたもので、所謂適応的選好形成であることについてはやや自覚がある。けれども、人がある種の SF を愉しむ気持ちとこの選好は近しく、これについてそんなにいやな気はしない。普段はなかなか展望を述べる場はないが、このたび、日本財団 HUMAI 事務局に「野心的な展望や思索」の提出を求められたため、公然と趣味と好きなことに関する文章を書くことができる。嬉しいことである。

わたくしの専門は言語研究であり、これから書いてゆく「展望や思索」はだいたいのところ歴史言語学と呼ばれる分野に関するものになる。日本財団 HUMAI プログラムには、*in silicio* による歴史言語学の研究手法全般を研究テーマとして掲げて採択されたので、このテーマ設定は順当である。*In silicio* とはラテン語で「シリコンの中で」を意味するフレーズであり、おそらくいまこの文章で初めて使用された言葉であり、計算機を利用した研究手法を指すためのカッコつけた表現である。生物学、ことバイオインフォマティクスの文脈ではこれではなく *in silico* と呼ぶのであるが、格変化が誤りであるため、言語学者ならばただしく専格に曲用させて *in silicio* と呼ぶべきである。言語学者は単に生物学の手法を輸入するのではなく、それを正しく理解し、消化し、自らのものとし、系統学に貢献しなければならぬと、わたくしは信ずる。わたくしの使った *in silicio* とは、この換骨奪胎の精神の象徴である。

わたくしは歴史言語学のすべてのトピックに興味があり、環境と予算と能力と寿命が許すかぎり研究してゆくつもりである。けれども、自分が論文を出すことに責任をもてる範囲での展望を話すことこそが知的徳であると信ずるから、今回この「いくつかの *in silicio* 歴史言語学に関する展望」と題した文章で議論されるのは、主として、しばしば「系統言語学」と呼ばれる (Greenhill, Heggarty & Gray 2020) ここ四半世紀ほどの間に生まれた分野の、とりわけ日本語の属する日琉語族における、ごくごく最近の展望だけになる。今回、遺憾ながら、埋め込み表現の発達によって近年最も変化が大きい計算通時意味論について述べることはない。今後研究してみたい分野だけだ。

「系統言語学 (phylolinguistics)」とは、生物系統学で発展した手法——一般にはベイズ法による系統推定——を言語学的対象に適用可能なかたちに改良して輸入するアプローチの総称である。「系統言語学」の主流の研究テーマは、言語の系統分岐の過程を推論するものである。近年はとりわけ BEAST などの系統推定ソフトウェアを使用するベイジアンアプローチが系統推定の覇権的な手法であり、世界の名だたる語族に適用されてきた。けれども、言語の系統分岐の過程を推論できることは、じつは系統学のもつ可能性のごく一部にすぎないので、ベイズ法による系統樹のみをわざわざ系統学の言語学への適用とみなすべきではないのではないかと感じる。系統樹のトポロジーを確定したり分岐年代を推定したりするのは、比較言語学の古典的なテーマにすぎないのだから、あくまでも比較言語学の下位分野であることを明示して、**計算比較言語学 (computational comparative linguistics)** と称すべきであると、わたくしは思料する。この呼称には、「系統言語学」とはちがって、比較言語学者の作成したデータなしになりたない分野であることを明示する機能がある。系統言語学の辺境には、誤りだらけのデータを利用し続け、モデルに的外れな工夫をかさねて、系統分類に対して本質的に貢献しない論文を濫造する光景がしばしば見られる。この現状は、学問として不健全である。言語の系統分岐はあくまでも比較言語学の問題であり、系統学の植民地ではないのだ。

さきほど述べた通り、わたくしが日本財団 HUMAI プログラムの助成を受けながら研究してきた研究は、*in silicio* の歴史言語学である。そして、白鳥 (2026a), 河合・白鳥 (2026) や白鳥・片岩 (2025) などその一端を披瀝してきたのは、何れも極東アジアの言語史研究、とりわけ日本語の属する語族である日琉語族の史的研究の閉塞を計算的手法によって打破するための手法群と方法論である。これらの研究が今後 2・3 年のうちに軌道に乗る研究プログラムであることについて、わたくしは確信する。こころみに、十分に実現可能な未来をあなたがたに見せてゆく。

2. in silicio 歴史方言学

先述の通り、わたくしの現在の研究領域をひとことで呼ぶならば、歴史言語学のなかでもだいたいのところ「比較言語学」の範疇に収まる。ここで比較言語学とは、記録された諸言語を合理的に説明できる歴史的变化の過程を形式的にモデリングし、その仮説を主軸に言語史のさまざまな現象を説明する言語学の分野のことである。することによる特徴づけはわかりにくいので、できるところに寄せても話しておく、比較言語学とは、記録の無い言語を推定できる学問の一つであり、同一の起源を有する言語の存在を立証できる唯一の分野である。わたくしは生まれ育った極東アジアの言語史に興味があり、対象言語はこの地域から選んできた。わたくしの現在の研究領域を、主として関与する語族の名を冠して「日琉比較言語学」と呼ぶこともできる。

言語の系統とはいかなるものなのかについては、一般常識を敷衍するだけで理解できるのであるが、日常的にも学校教育でも、言語について観察する機会は少ないため、奇妙な認識をもつ人が絶えない。故に、これから言語の系統推定について話すから、言語の系統に関する議論がいかなるものなのかについて話しておく必要がある。常識から明らかであるが、われわれの母語、つまり第一言語は、われわれがある年齢までに置かれた言語環境において晒されたデータに基づいて獲得したものである。つまり、幼少期に中国語にいい晒されなかった人間が、いきなり鬼神に憑かれて中国語を話し始めることはありえない。(もちろん、鬼神に憑かれたかのごとく勉強し始めることはありえるが)。常識的に明らかであるが、行きつけの中国料理店のメニューをそらんじて中国語が話せる様になったわけではないし、漢字を書けるからといって中国語を書けるわけではない。つまり言語形式を習得することと言語体系を習得することは異なるし、文字と言語も異なる。したがって、ある年齢までのこどもの言語環境を規定する共同体(言語共同体)の構造が変化したり、破壊されたりしないかぎり、その共同体で共有される言語の根幹の部分は継承されてゆくことになる。言語は徐々に変化してゆくので、共同体が分裂したときに言語が共有されつづけることは通常なく、言語も徐々に変化し、分岐してゆく。言語の系統推論とは、過去に生じた言語共同体の分裂を間接的に推論することなのである。

日本語が属する小さな語族である日琉語族(Japonesian, 旧称 Japonic)は、初期の系統分岐の過程が不明のままである。日琉語族は九州琉球語派、東国語派と称すべき二つの巨大なグループを擁する(もしくは擁した)可能性が十分にある。つまりこの二つのグループは、過去に一つの言語共同体をなし、それが分岐することで成立したと予想される。しかし、九州琉球語派・東国語派の側系統(いとこ、はとこ)をなすはずの西日本の大部分の諸言語は、その系統分類が困難である。

白鳥(2026a)は、西日本の諸言語の系統分類の問題に体系的にアプローチするための新たな方法論を提案するものであった。白鳥(2026a)は、愛媛県を題材に、西日本の諸言語の集団遺伝学的手法を適用する方法を論じたものである。従来、西日本の諸言語の系統分類が困難であったのには、二つの理由があった。第一は、非常に古い時代に複雑な系統分岐を経験した上、長期の言語接触によって収斂してきたことである。言語は、コミュニケーションに使用されるため、通じない言葉を相互に調整することがある(応化)。詳細は略すが、ほかにもさまざまな要因により、分岐は古いのに、ずっと似た変化を経験し続けることが起きる。言語は、雪ばれの景色のごとく、降り積もったものが溶けて、もとの姿がよくわからなくなることがあるのである。西日本の諸言語はその状況にある。第二は、単純に、利用可能なデータが少ないことである。以上の二つの問題が解決されなければ、西日本の諸言語の系統分類は不可能なままである。

白鳥(2026a)は、この問題にいかにしてアプローチしたのか?

第一の、系統分類の困難さについて、白鳥(2026a)は、西日本の諸言語をめぐる系統分類を急ぐのではなく、さしあたって内的構造を解明すべきであるとした。非常に初期の言語共同体の分岐を推論するためには、日琉語族の非常に初期の言語(早期言語)の言語体系を復元しなければならない。早期言語の復元は将来的に可能な見込みがあるが、現段階ではまだ困難であるから、その問題の手前の、比較的最近に起きた言語共同体の分岐と融合を精密に議論することで、将来の系統分類のための準備を行ってゆくべきであるとしたのである。そしてその手法として、詳細の説明は省くが、集団遺伝学的手法を利用してゆくことを論じ

た。この種の手法は、言語の分岐が樹状構造をなすか（その言語が分裂してから間もないものか）を議論したり、言語のもつ形質の類似性を論じたりすることができるからである。

第二の問題について、白鳥（2026a）は、従来、省みられてこなかったデータを使用することで、データの不足を補ってゆくことができる」と指摘した。言語の系統推定には、通常、基礎語彙表と呼ばれるあらかじめ定められた 100 語とか 200 語のリストの語彙を使用する。これは「目」だとか「私」みたいな、頻度が高くなかなか借用されない語彙のリストである。しかし、日琉語族の基礎語彙の記述は、その系統的多様性を十分に被覆できる規模のものが存在しないから、標準的な手法の適用は困難である。英語から日本語への借用語を観察してみてもいい。すると、明らかに頻度の高い意味領域（技術など）と、低い意味領域（身体など）があることがわかるし、所謂「助詞」などの文法構造の根幹に関する語よりも名詞の方がはるかに借用されやすいことなどもわかるはずである。この現象を論ずるための言語類型論の概念が、借用可能性である。借用可能性の低い形質は、水平伝播に対する頑健性を有する。つまり、祖先から継承されたものであり、言語系統を反映する蓋然性が大きい。系統推定は言語接触の影響を統制しておこなってゆく必要があるため、基礎語彙表には借用可能性の低いものの中から単語が選ばれてきた。同様に、借用可能性の低い形質を集めることで、非標準的なデータセットに基づいた内的構造の解析をおこなってゆくことができるのである。

白鳥（2026a）の推定により、愛媛県の諸言語は単一のグループではなく、明らかに複数のグループからなることがわかった。愛媛県の諸言語のグループ自体は、前世紀にも議論されてきたものであったが、従来の研究は、職人芸的に認定された少数の特徴の有無に基づくため反証が困難で、それらが系統的にどんな意味をもつグループなのか不明瞭であった。曖昧な「区画」を脱し、多数の特徴に基づいた再現可能な計算的手法によって地域集団を検出できるようになったことの意義は大きい。今後、前世紀には困難だった内的構造の分析により、その形成過程を推定してゆくことができる。白鳥（2026a）の研究プログラムは、類似した状況にある西日本の諸言語に適用してゆくことができる。今後、この方法論によって西日本の諸言語の内的構造が精密に解明されれば、存在が予想される日本語の未知の姉妹言語の境界線が劃定されてゆく。つまり、日本語の未知の「新種」が発見され、日琉語族の歴史に関する理解が深まることが期待される。そして実のところ、これは単なる期待に留まるものではなく、わたくしがここ 7 年ほど取り組んできた研究のなかで、着実に裏付けられつつある事実でもある（cf. 白鳥 2026b など）。日本諸語には、愛媛県の諸言語や周防大島語をはじめとする、10 年前には省みられもしなかった姉妹言語が含まれることが明らかになりつつある。これらの姉妹言語の研究が進めば、日琉語族の早期言語の復元の精度が（おそらく飛躍的に）高まり、西日本の諸言語の系統分類の問題は、徐々に、確実に解決してゆく。

方言連続体の内的構造の集団遺伝学的手法による解析は、系統言語学の新たな可能性を切り拓くものでもある。他の語族においても、今後、この方法論の拡張によって、方言連続体の内的構造が精密に理解されてゆくことが十分にありうる。

3. in silicio 歴史文字言語学

前節では系統言語学のうち言語系統について論ずるものを扱ってきた。系統言語学は、一般に、言語の系統推定を扱ってきた分野であり、言語学の内部にある、言語系統以外の系統学的な構造を持つ対象について無関心である。しかし、系統学的手法の射程は、言語の系統推定にとどまるものではなく、今後、さまざまな問題に適用されてゆくはずである。その実例が、わたくしが取り組んできた文字体系の分類に関する問題である。

日本書紀は漢文で書かれたわけであるが、当時の日本語による歌謡が挿入される。そんな日本書紀歌謡には、その文字の用法によって大きく α 群と β 群の別があることが知られる。 α 群の用字法は唐代長安の音韻体系にちかく、中国語話者が編纂を担ったと推定される（森 1991）。この説は、ここ四半世紀の日本書紀研究における基礎的な出発点となる仮説であり、多大な影響を及ぼし続けてきたものであるが、定性的な観察に基づいて発見法的に帰納された分類に過ぎなかった。

白鳥(2026c)は、日本書紀歌謡を計算的手法によって分類することを試みたはじめての研究である。白鳥(2026c)の提案手法は、通説の α 群・ β 群の境界線を完全に検出するとともに、通説の二分法に問題があることを指摘することに成功した。ここで採用した手法は、自然言語処理における潜在ディリクレ配分法に類似した集団遺伝学の手法である STRUCTURE (Pritchard, Stephen & Donnelly 2000) の一種で、ベイジアンアプローチによる音節の漢字頻度分布のクラスタリングである。つまり、この手法は系統学のひとつの応用であるといえる。白鳥(2026c)の手法は音節ごとの漢字選択の傾向性をみるものであるが、20世紀になされたほとんどの万葉仮名研究は音節ごとの漢字選択の傾向の記述に終始するものであるため、この手法の完成をもって、古典的な万葉仮名研究の手法をほぼ完全に計算機のなかで再現することに成功することになる。現段階ではモデル選択の方法などに若干の改善すべき点があるが、ちかく、上代日本語文字論にはささやかな計算論的転回が到来する。上代文献の全体にこの手法を適用し、その内的構造を解明することが望まれる——いな、受動態で他人事の様に書くべきではあるまい。このさきわたくしは第一線でこれをおこなってゆくべきである。こんな研究をしたら勘違されることが増えてしまったけれど、わたくしの専門は上代日本語の音韻論的再建でもあり、この研究は未来の精密な音声学的再建と言語史の解明にとっての不可欠の基礎となる。

また、河合・白鳥(2026)は、平安時代の漢文訓読のために使用された片仮名の祖先の多様性を、以上の研究と同様に集団遺伝学で発達した計算的手法の応用によって解明しようとした試みである。今回は利用できたデータが少なく、あまりうまくゆかなかったが、こちらも、文字と遺伝子の間に同型性を見出す方針自体に誤りはない。従来の定性的な観察を超えた歴史研究の精密化には着実な第一歩を踏み出すことができたといえる。中古日本語研究に閉込められがちな片仮名を、上代日本語研究の一環としておこなってゆくわたくしが言っても、あなたはまだ信じてくれないかもしれないが、わたくしはこれからも一貫した問題関心のなかで仮名を研究してゆくつもりである。

4. さいごに

この文章では、著者がここ数ヶ月行ってきた研究について概観してきた。系統言語学は、文字の系統分類や、方言学の最も古典的な問題である方言の形成過程などへの適用可能性がまだまだ広がる、発展中の分野である。冒頭に述べた、計算比較言語学としての系統言語学を超えた系統言語学の可能性を感じていただけたなら、幸甚である。

引用文献

- Greenhill, Simon J.; Paul Heggarty & Russel D. Gray (2020) "Bayesian phylolinguistics." In: Richard D. Janda, Brian D. Joseph & Barbara S. Vance (eds.) *The Handbook of historical linguistics, vol. II*. Hoboken: Wiley.
- 河合翼・白鳥詩織(2026)「造字原則に基づいた片仮名の系統分類の試み」人文科学とコンピュータ第140回研究発表会。
- 白鳥詩織・片岩拓也(2025)「n-gram モデルの KL 情報量を用いた万葉集訓字主体表記巻の仮説探索」第20回言語処理若手シンポジウム。
- 白鳥詩織(2026a)「借用耐性のある形質による系統ネットワーク—愛媛県の言語はいくつあったか?—」人文科学とコンピュータ第140回研究発表会。
- 白鳥詩織(2026b)「比較言語学的に面白いかもしれない日本語「方言」」言語学フェス2026。
- 白鳥詩織(2026c)「集団遺伝学的アプローチによる日本書紀歌謡の教師なし分類の試み— α 群・ β 群を超えて」第五回計算社会科学大会。
- 森博達(1991)『古代の音韻と日本書紀の成立』。東京:大修館書店。
- Pritchard, Jonathan K.; Matthew Stephens, & Peter Donnelly (2000). Inference of Population Structure Using Multilocus Genotype

Data. *Genetics*, **155** (2): 945–959.