

C-2

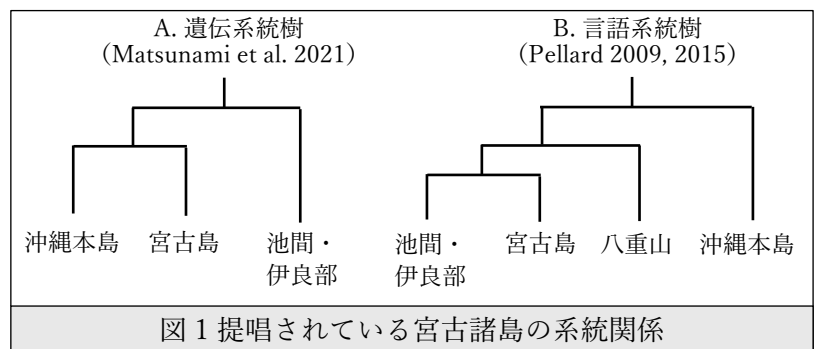
琉球列島における遺伝系統樹と言語系統樹の比較に向けて —言語系統樹の客観的作成手法を検討する—

セリック ケナン (NINJAL) ・ 麻生 玲子 (名桜大学) ・ 松波 雅俊 (琉球大学)

要旨：本発表では、琉球列島における遺伝系統樹と言語系統樹の比較に向け、言語系統樹の客観的作成手法を検討する。近年、琉球列島を対象にヒトの系統を明らかにする遺伝学の研究と言語の系統を明らかにする言語学の研究が進められているが、先行研究によって明らかにされたヒトの系統樹と言語の系統樹は一致していない。2つの系統樹の間に見られる「矛盾」を解釈する際には、言語学の研究手法が他分野の研究者にも評価可能な客観的なものであることが求められる。このような背景から、本発表ではまず系統ネットワークという手法を用いて宮古語を中心とした琉球諸語の言語系統樹を作成し提示する。その結果として、伝統的な言語学的手法で得られる結果と多少の違いはあるものの、概ね同様の結果が得られたことを述べる。最後に系統ネットワークという手法を評価し、利点と今後の課題について述べる。

1. はじめに

近年、琉球列島を対象地域に、ヒトの系統を明らかにする遺伝学の研究と言語の系統を明らかにする言語学の研究が進められており、その中で、宮古諸島を中心に



遺伝学および言語学の両方から詳細な研究結果が公表されている（遺伝学：Matsunami et. al 2021、言語学：Pellard 2009、2015、セリック 2020）。興味深いことに、これらの研究によって明らかにされたヒトの系統樹と言語の系統樹は一致していない（図 1）。具体的には、遺伝解析から宮古諸島の人々の成立には外部からの移住の影響があり、池間・伊良部島の人々が先立って分岐し、その後宮古島の人々と沖縄本島の人々が分岐したことが推定された。しかし、この結果は琉球諸語が南琉球諸語（与那国・八重山・宮古）と北琉球諸語（沖縄・奄美）に分類されるというこれまでの言語学の定説と矛盾する。このため、2つの系統樹の間に見られる「矛盾」をどのように解釈すべきかが課題である。

しかし、ヒトと言語の系統樹を比較する際には、それぞれの分野で用いられている手法が妥当な結果を導きだしているという前提が必要である。遺伝学では、世代ごとの変化（突然変異）は一定の確率で起こるので、数値化したデータに基づく定量的な推定が可能となる。そのため、方法の再現性が保証されており、信憑性の高い結果が得られていると考えられる。これに対して、伝統的な言語学では、言語の系統を反映すると思われるデータを手動で選別した上で、節約的に系統樹を作成している（ローレンス 2000、2008、Pellard 2009、2015 など）。しかし、次に述べるように、この手法は、基づく原理が妥当であっても、結果の信憑性について課題を抱えている。

伝統的な言語学では、同じ祖語から分岐した姉妹言語の系統関係、すなわち、姉妹言語のうち、どの言語が共通の祖語から分岐したかを明らかにするために「共有改新」と呼ばれる改新を特定するという手法を取っている。共有改新は、祖語からの分岐以降、一度だけ生じ、複数の姉妹言語に共有さ

れる改新として定義される。2つの言語が、対象の姉妹言語の歴史において一度だけ生じた改新を共有する場合、この2つの言語がこの改新を共通の祖語から継承したと解釈することは合理的である。つまり、この2つの言語は、姉妹言語の中で一つの言語群を形成しているということである（図 2-a）。

しかし、同じ改新が複数の姉妹言語に観察された場合、この改新が系統を反映させる「共有改新」であるとは限らない。なぜならば、共有改新とは異なる解釈も成り立つ可能性があるからである。一つは、ある改新が、分岐後、複数回独立して生じた並行変化である（図 2-b）。もう一つは、改新が生じた後、借用によって他の姉妹言語に伝播した場合である（図 2-c）。さらに、もっと複雑なケースなども考えられる（図 2-d）。当然のことながら、これらの場合には改新の共有が言語の系統を反映しているとは言えない。

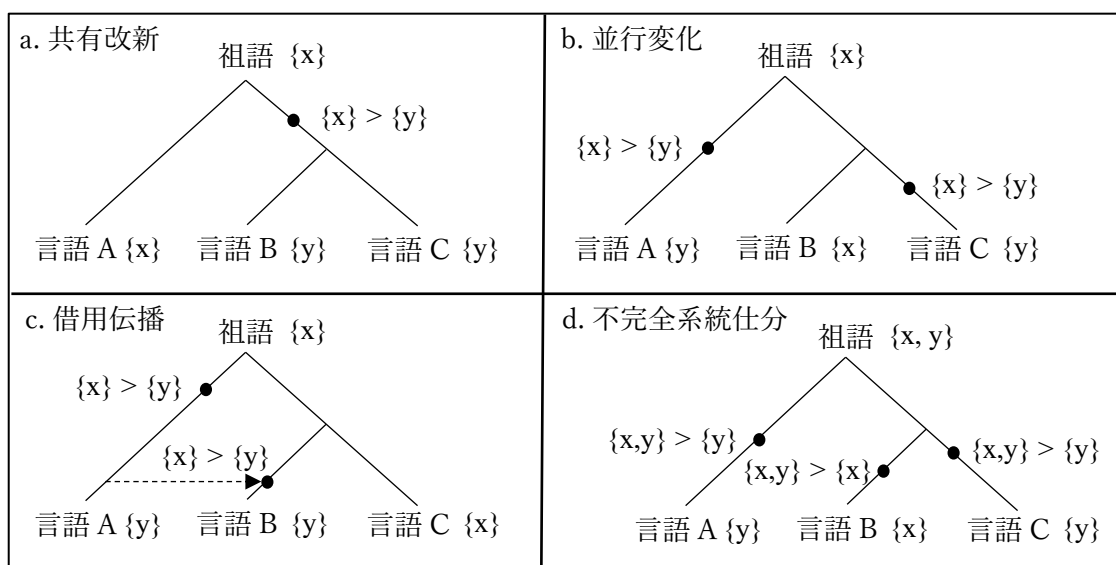


図 2 改新の共有に関する解釈（{x}：言語状態、{x} > {y}：{x}が{y}に変化する）

以上のことから、2つ以上の姉妹言語が共有している改新が共有改新であると判断するためには、その改新が独立して複数回生じた蓋然性が低く、また、借用によって伝播した蓋然性も低いことが必要である。しかし、これらの蓋然性に関わる評価基準は、歴史比較言語学の結果の蓄積に則っているとは言え、必ずしも明確に示されているわけではない。そのため、共有改新の認定が曖昧または恣意的に見えてしまう恐れがある。さらに他分野の研究者にとっては、評価基準が不透明であるため、言語学で得られた結果が妥当であるかどうか評価不能であるという問題が指摘できる。

例えば、これまで琉球諸語の系統関係を明らかにするために提案されてきた共有改新について具体的に見てみよう。表 1 で分かるように、共有改新として、語形の変化（表 1-b,c,f,i,l-n）、文法化のプロセス（表 1-e,k）、意味変化（表 1-o）、アクセントの変化（表 1-a）、音変化（表 1-d,j）など、様々な言語変化が提案されている。後者 2 つは、特定の語彙に起きた不規則的な変化であるため、確かに独立して複数回生じた蓋然性が低いと考えられる。しかし、これに対して、語形の変化などは、ある方言で新しい語形が生じた後、借用によって他の方言に広がったという可能性が否定しきれない。

表 1 共有改新の例 (ローレンス 2000, 2008, Pellard 2009, 2015, セリック 2020)

レベル	共有改新	変化タイプ	
南琉球	a) 「幾つ」のアクセントの不規則的変化: B > A	ア変化	これに対して、伝統的な言語学で用いられる手法の恣意性を排除できるよう、広く語彙を収集した上で、遺伝学的手法を転用し言語系統樹を作成した研究も存在する (Lee & Hasegawa 2011、Saitou & Jinam 2017、斎藤 2021)。しかし、言語学者が参加していないこれらの研究は、基づ
南琉球	b) 「庭」の語形: minaka 系	語形	
南琉球	c) 「一人・独り」の語形: *tagukjaa	語形	
南琉球	d) 「額」における不規則的変化: *pitae > *putae	音変化	
南琉球	e) 「～てから (継起)」の文法的形式	文法化	
宮古語	f) 「～も (累加)」の標識: =mai	語形	
宮古語	g) 「頭」の語形: kanama ₁	語形	
宮古語	h) 「丸い」の語形: maaku	語形	
宮古語	i) 「濡れる」の語形: mmi-	語形	
宮古語	j) 「首」における音位転換: *kubi > *pugi > fug ₁	音変化	
八重山語	k) 「切れる」*kire- の可能形としての文法化	文法化	
八重山語	l) 「芽」の語形: bai	語形	
八重山語	m) 「嬉しい」の語形: sani-	語形	
八重山語	n) 「垢」の語形: gaba	語形	
八重山語	o) 「甥」における意味変化: 「甥」 > 「甥・姪」	意味変化	

いている言語データ自体に問題がある。このため、得られている結果の妥当性が疑わしく、言語学者からは全く受け入れられていない (ペラール 2018、2021)。ただし、研究方法自体が否定されているわけではないことに注意が必要である。言うまでもなく、データが適切でなければ、どのような方法を取ったとしても妥当な結果が得られるはずがない。

以上の背景を踏まえ、本発表では言語系統樹の作成について、学際的に評価可能な客観的な手法を探るため、Lee & Hasegawa (2011) などの流れを汲みながらも、その問題点を克服できるよう研究を実施する。具体的には対象地域を専門とする言語学者が作成した言語データに基づき、系統ネットワークという手法を用いて宮古語を中心として琉球諸語の言語系統樹を作成する。

2. 方法

2.1. 概要

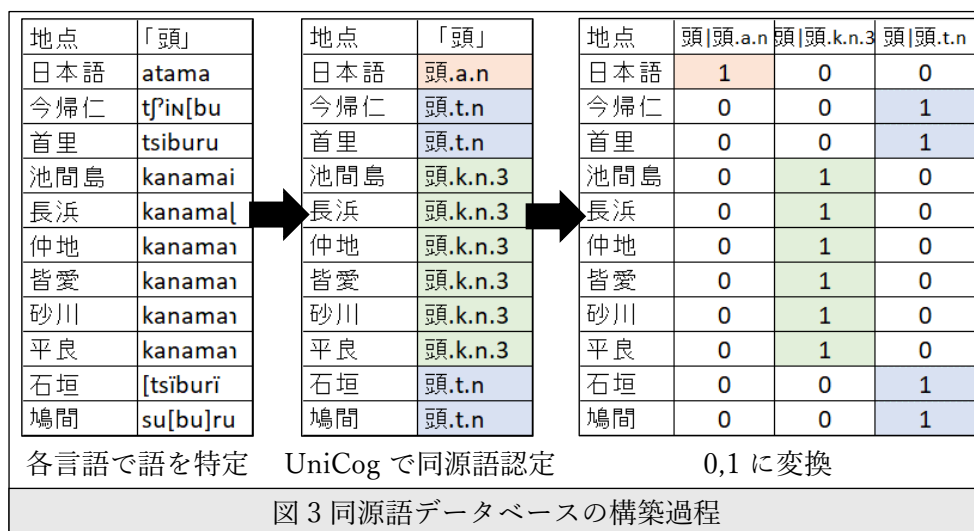
琉球諸語の系統関係を客観的な手法で検討するために、印欧諸語の系統関係について採用された定量的手法 (Heggarty et al. 2023 など) に則る。すなわち、借用されにくい基礎的な概念のリストを基に、対象言語の「同源語データベース」を構築した上で、各概念に対する同源語の有無をバイナリデータ 0,1 で数値化し、このデータに基づき、系統ネットワーク (2.3 節参照) を作成する。

2.2. 対象言語と同源語データベースの構築

対象地点は琉球諸語 11 地点 (沖縄語 2 地点、宮古語 7 地点、八重山語 2 地点) と日本語である。使用する基礎的な概念のリストは、トマ・ペラール氏が日琉諸語の系統関係を探るために用意した 217 の基礎概念 (「頭」「笑う」「大きい」など) のリスト (未公開) である。ただし、概念の曖昧さや対象言語への翻訳の難しさなどにより、統一的なデータが得られなかった 14 の概念 (「細い」「一杯」「叩

く」など)を除いたため、本発表の分析は203の基礎概念に基づく。

同源語データベースの構築過程は図3の通りである。第一に、本稿の末尾にある語彙資料を参照し、各概念に対する対象言語の語形を検索した。同じ概念に対して複数の回答を可能としたが、同じ概念に対して複数の語形が見つかる場合、語形の選定にあたり、次の①から③に当てはまる語形は除外した。すなわち、①明らかに借用語である語形(石垣四箇の *ujantfu*「鼠」)、②明らかに有標である語形(例:皆愛の無標の *kanamaŋ*「頭」に対する有標の *kanamarja*「(動物の、または罵り言葉としての)頭」)、③日常語ではない、特定の言語レジスターに属する語形(例:文語である首里の *sura*「空」)である。この結果、2913語が得られた。第二に、琉球諸語の同源語をアノテーションするための枠組



である UniCog (Celik et al. 2024) を使用して、得られた語に対して同源性を判断し、同源語 ID を付与した。語彙データは 475 個の同源語に該当した。第三に、各概念に対する同源語の有無をバイナリデータに変換した。この結果、498 個のバイナリ

変数が得られた。

2.3. 系統ネットワーク

系統による変化(1節で言う「共有改新」と独立して複数回生じた並行変化などが混在する言語データの関係性は、単純な二分岐に基づく系統樹を用いて正確に再現するのは難しい。そこで、本研究では、系統樹では得ることができない言語の並行変化などを考慮できる系統ネットワークを用いて解析結果を可視化した。系統ネットワークの描画には、SplitsTree (version4.19.2)に実装されている Neighbor-Net 法を用いた (Bryant & Moulton 2004、Huson & Bryant 2006)。前節で述べたように、琉球列島 11 地点と日本語 1 地点から収集した 498 個の概念に対する同源語の有無を 0,1 のバイナリデータに変換後、距離行列を計算し、各地点間の距離を推定した。距離行列の算出には、同一概念の多重変化や変化のバイアスは考慮せず、単純に地点間の距離を計算する p-distance に基づいた “Uncorrected_P” を使用した (Nei and Kumar 2000)。リサンプリングによる統計解析法であるブートストラップ検定を 1000 回実施し、分岐の正確性を推定した。

3. 結果

以上のように、大きな系統区分については先行研究との完全な一致が見られるものの、宮古語の内部の系統関係については、結果に異同も観察される。一方では、水納島方言を含む多良間諸方言が、宮古祖語から最初に分岐したとする先行研究の仮説（Pellard 2009、セリック 2020）と矛盾しない結果（水納を除いた宮古語諸方言が 68.3%の確率で言語群を成す）が得られた。しかしその一方で、池間諸方言と伊良部諸方言が宮古語の中で一つの言語群を成すとする従来の仮説とは異なり、系統ネットワークでは、伊良部仲地が池間島および長浜と一つの言語群を成さず、宮古語の他の方言（平良、皆愛、砂川）と同じ言語群に分類されるという結果が得られた。

4.2. 宮古語の系統的位置について

宮古語は図 4 で見られるように非常に明確な形で一つの言語群としてまとまるだけではなく、宮古の語群を特徴付ける枝（分岐点④）が長い。系統ネットワークでは枝の長さが言語間の距離を表すので、宮古語は他の言語から「遠い」ということになる。これは、宮古語諸方言が多くの共有改新を共有するという事に起因する可能性がある。しかし、本手法では距離を計算する際に保持と改新を区別していないので、枝の長さを正しく解釈するためには、関連する変数を個別に確認しておく必要がある。

宮古語の枝が長くなっている原因について確認すると、宮古語の全地点で値が 1、その他の地点で 0 となる変数が 12 個、その逆の場合の変数は 1 個（「歌 | 歌.u.n」）見つかった。前者 12 個の変数を見ると、確かに宮古語の共有改新である可能性が高い変数（「血 | akats₁」「投げる | tivv-」「灰 | karapa₁」「良い | dzo:-」「頭 | kanama₁」「歌 | a:gu」「小さい | imi-」※語形は平良）が多く含まれている（他の変数については、地点が少ないことによる見かけ上の共有改新である）。つまり、本発表の手法では、保持と改新を区別していないにも関わらず、少なくとも宮古語に関しては、共有改新を共有する地点が正しくまとめられるという結果になった。

4.3. まとめ:本発表の手法の評価

以上、伝統的な言語学の結果と本発表で得られた結果の整合性について見てきたが、結論として、伝統的な言語学の結果に照らし合わせても妥当な結果が得られており、本手法が有効であるということが確認できた。本発表で用いた手法は、次の 3 点で有用である。第一に、方法が透明で再現性が高い点である。このことにより、他分野の研究者であっても結果を客観的に評価できる。第二に、言語間の距離が枝の長さで定量的に表現されるため、伝統的な言語学の手法では得られない情報を提供できる点である。第三に、宮古語について見てきたように、同源語データベースの構築と系統ネットワークによる可視化を通じて、共有改新の発見を容易にする点である。

一方で課題も指摘できる。すなわち、第一に、どのような基礎的概念が使用されるべきか、第二に、各概念に対して対象言語の語形をどのように選定すべきか、第三に、バイナリ変数が完全に独立しているか、の 3 点について検討を深めていく必要がある。いずれにせよ、今回の客観的な手法でも言語系統樹と遺伝系統樹との間に矛盾があることを確認できたため、その課題に取り掛かることが望まれる。

語彙資料 日本語（首都圏）：第二著者内省／沖縄語今帰仁：仲宗根政善 (1983)『沖縄今帰仁方言辞典 今帰仁方言の研究・語彙篇』角川書店／沖縄語首里：国立国語研究所 (1963)『沖縄語辞典』大蔵省印刷局、宮古語池間島：平山輝男(1983)『琉球宮古諸島方言基礎語彙の総合的研究』桜楓社、上里武・吉浜朝榮 (2010)『池間島の方言と口語集』私家版／宮古語伊良部長浜：平山(1983)／宮古語伊良部仲地：富浜定吉 (2013)『宮古伊良部方言辞典』沖縄タイムス社／宮古語平良：渡真利章良 (2019)『宮古の方言集～平良の方言を中心とした言葉～』ボーダーインク／宮古語皆愛・砂川：第一著者調査ノート／宮古語水納：セリック ケナン・大浦辰夫 (2022)『みんなふつ語彙集』国立国語研究所／八重山石垣：宮城信勇 (2003)『石垣方言辞典』沖縄タイムス社／八重山語鳩間：加土工真市 (2020)『鳩間方言辞典』国立国語研究所。

参考文献 Bryant D. & Moulton V. (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol.* 21: 255-265／セリック ケナン (2020)「南琉球宮古語史」未刊行博士論文、京都大学／Celik, Kenan, Kohei Nakazawa, Reiko Aso (2024) UniCog: A Framework Proposal for the Dynamic Compilation of Comparative Data for the Reconstruction of proto-Ryukyuan. *NINJAL Research Papers.* 26: 69-97／Heggarty, Paul, Cormac Anderson, Matthew Scarborough, Benedict King, Remco Bouckaert, Lechosław Jocz, Martin Joachim Kümmel, Thomas Jügel, Britta Irlinger, Roland Pooth et al. (2023) Language trees with sampled ancestors support a hybrid model for the origin of Indo-European languages. *SCIENCE.* 381(6656)／Huson DH & Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23: 254-267／Lee, Sean & Hasegawa, Toshikazu (2011) Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. *Proceedings of the Royal Society B: Biological Sciences* 278(1725): 3662-3669／Matsunami, Masatoshi, Koganebuchi K., Imamura M., Ishida H., Kimura R, Maeda S. (2021) Fine-scale genetic structure and demographic history in the Miyako Islands of the Ryukyu Archipelago. *Mol Biol and Evol.* 38: 2045-2056／Nei M. & Kumar S, (2000) *Molecular Evolution and Phylogenetics.* Oxford University Press, New York／Pellard, Thomas (2009) *Ōgami: Éléments de description d'un parler du sud des Ryūkyū.* Unpublished doctoral dissertation, Ecole des Hautes Etudes en Sciences Sociales (EHESS)／Pellard, Thomas (2015) The linguistic archeology of the Ryukyu Islands. In H. Patrick, M. Shinsho, & S. Michinori (Eds.), *Handbook of the Ryukyuan languages: History, structure, and use.* De Gruyter Mouton 13-37／ペラール トマ (2018)「日琉諸語の系統分類と分岐について」シンポジウム「フィールドと文献から見る日琉諸語の系統と歴史」発表資料. 国立国語研究所 (2018年12月22～23日)／ペラール トマ (2021)「日琉諸語の系統分類と分岐について」林由華・衣畑智秀・木部暢子 (編)『フィールドと文献からみる日琉諸語の系統と歴史』開拓社 2-16／ローレンス・ウエイン (2000)「八重山方言の区画について」石垣繁 (編)『宮良當壯記念論集』宮良當壯生誕百年記念事業期成会 547-559／ローレンス・ウエイン (2008)「与那国方言の系統的位置」『琉球の方言』32: 59-67／Saitou, Naruya & Jinam, T. A. (2017) Language diversity of the Japanese archipelago and its relationship with human dna diversity. *Man in India* 95(4): 205-228／斎藤成也 (2020)「ゲノムデータと言語データの関係」長田俊樹 (編)『日本語「起源」論の歴史と展望 日本語の起源はどのように論じられてきたか』三省堂 253-267.

謝辞 本研究は JSPS 科研費・挑戦的研究 (萌芽)「遺伝系統樹と言語系統樹の矛盾を検討する」(23K17503)、特別研究員奨励費「日琉祖語で再建される語の語形成の解明」(22KF0370)、若手研究「琉球諸語を対象とした効率的かつ大規模な言語資料収集・蓄積方法に関するメタ研究」(23K12167)の助成金を受けています。概念語彙リストをご提供下さったトマ・ペラール氏にお礼申し上げます。