

# 論文の内容の要旨

## 論文題目

Pattern and Process of Language Evolution in the Japanese Islands  
(日本列島における言語進化のパターンとプロセス)

## 氏名

リー サンホン

生物と言語はどちらも垂直に伝達されるという共通点を持つ (Darwin 1871)。生物は「遺伝」を媒体に、言語は「学習・模倣」を媒体にその性質が親から子へと受け継がれる。それ故、両者においてある個体に突然変異が発生した場合、その多くは変異を起こした個体の後代に伝達され、蓄積していく。従って、生物と言語はどちらも時間の経過とともにそれぞれの進化の歴史を刻み込んだ「系統」を形成するようになり、両者の系統進化の歴史は進化生物学の統計手法を用いて再構築することが可能である。統計的手法で描き出された言語の系統進化はさらに、その話者の生物学的系統進化の有様を反映することがインド・ヨーロッパ語族 (Gray & Atkinson 2003)、バントゥー語族 (Holden 2002)、オーストロネシアン語族 (Gray et al. 2009) において報告されている。本博士論文は、この「言語の系統進化」という現象の中身を「パターン研究 (現象の記述)」と「プロセス研究 (メカニズムの究明)」という2つの軸から捉え、それぞれに対する理解を深めることを目的とした。

研究1では、地球上における言語進化の現象の記述をさらに増やすことを目指し、かつて日本列島の北部で繁栄したアイヌ語の系統進化を再構築した (パターン研究)。アイヌ語が現在に比べ遥かに豊かな多様性を保っていた時に収集された語彙データ (Hattori & Chiri

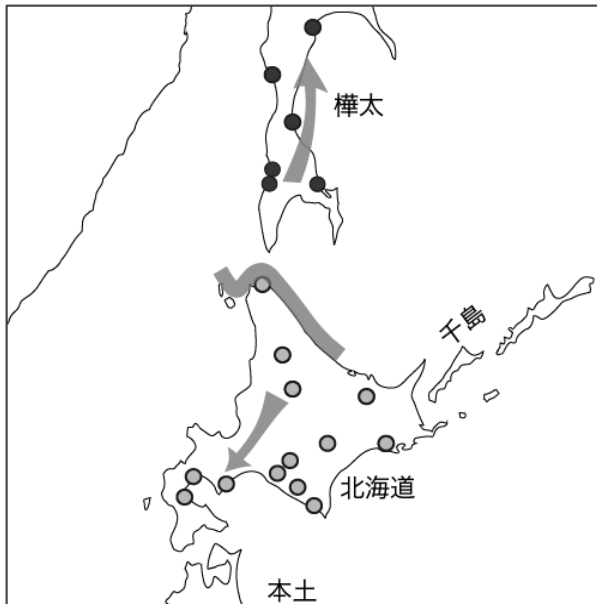


図1. 丸は分析で用いられた言語サンプル。系統地理分析の結果、アイヌ語は約1300年前に北海道の北部の海岸（灰色の曲線棒）から伝播したと推定された。

1960) と、従来の手法に比べてより一層洗練された「系統地理分析」という分析法を用い、系統進化の原点（共通祖先）が「いつ」存在したのかだけでなく、「どこ」に存在したのかをも推定した。その結果、アイヌ語は約1300年前に北海道の北部で誕生したことを示す結果が得られた(図1)。この結果を、「言語の系統進化はその話者達の系統進化の有様を反映する」という法則を踏まえた上で解釈すると、アイヌ語の系統地理進化は、アイヌの起源説として長年支持されて来た「二重構造説」(Hanihara 1991) より、近年のミトコンドリア DNA を用いた研究結果から提案

された「オホーツク人混血説」(Sato et al. 2007; 2009)と一致する。この結果の妥当性は(1) 北海道北部が伝播の原点であることに対するベイズファクターを推定し、(2) 結果のモデル依存性を検証することで確かめた。アイヌは狩猟採集民族であったことと、アイヌ語の系統進化が口バストに再構築されたことから、研究1の結果は、言語の系統進化がその話者の生物学的系統進化の有様を反映するという現象がインド・ヨーロッパ語族やオーストロネシアン語族といった農耕民族の言葉に限定されるものではなく、広範囲で観測できる普遍的な現象である可能性を示す (Lee & Hasegawa 2013)。

研究2では、パターン研究の延長線上にある作業として、日本語族から推定された系統進化とその年代の妥当性を検討した。日本語族が発生した年代は、日本方言辞典 (Hirayama 1988;1992) に基づき、北の札幌から南の与那国まで57の方言と2つの絶滅語（上代日本語と室町時代語）から抽出した語彙データを用いて、ベイズ系統樹分析を行うことで推定した。その結果、日本語族の起源（系統樹の根元）は約2200年前であるという結果が得られた (Lee & Hasegawa 2011) 。これは、現代日本の基盤となる勢力は約10000年前の縄文時代から外部の影響を受けず現在まで持続してきたということを唱える従来の学説 (Suzuki 1981) を覆し、弥生時代に日本列島に流れ込んだ農耕民族が現代日本人・日本語の原点であることを主張する学説 (Hudson 1999) を支持するものである。しかし、Lee and Hasegawa (2011)の日本語族の系統樹分析の結果に対して、これまで多くの研究者から批判がなされてきた。そこで、Lee and Hasegawa (2011)に対して頻繁に挙げられた批判である(1) 言語水平伝播の影響、(2) 年代推定のモデル依存性、(3) 年代比較

の妥当性、(4) 系統樹そのものの解釈に関する問題などといった批判を考察し、直接的・間接的証拠を用いてそれらに答えることを試みた。その中でも「なぜ本土日本語の系統関係は琉球語のそれより曖昧（不確実）なのか」という問いに答えるために系統ネットワーク分析を行い（図2）、究極要因のレベルと至近要因のレベルで幾つかの仮説を提案した。

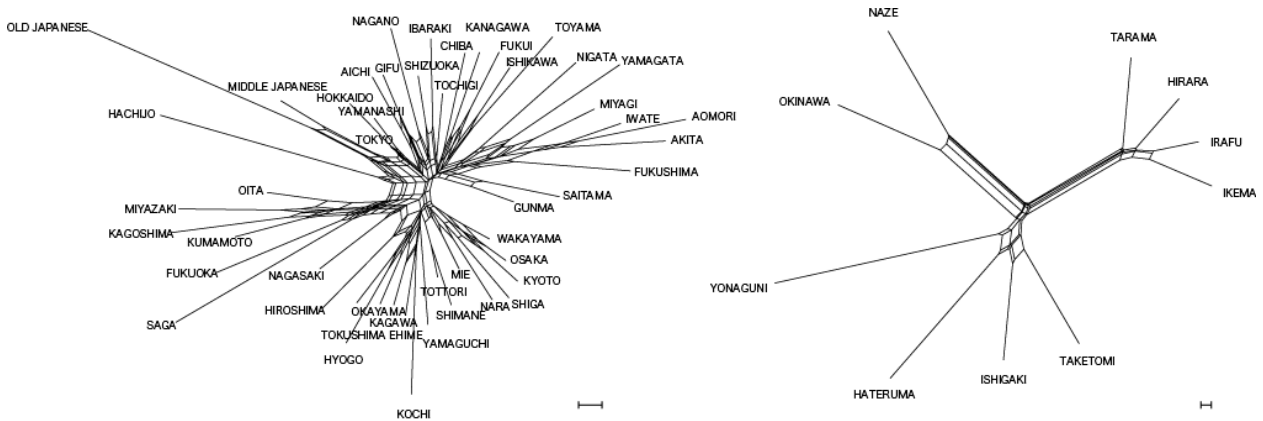


図2. NeighborNet 法で描き出した日本語族の系統ネットワーク。左は本土日本語群で、右は琉球語群。琉球語は日本語に比べて系統関係がはっきりしていて、系統の結合の不確実性（枝と枝の間を繋げる網の有無）が少ないことがわかる [Delta score (最低 0、最大 1；値が大きくなるほど不確実性が上昇) 左：0.4、右：0.2]。

研究3では、研究2の系統ネットワークから見られた日本語族内の「系統関係の不確実性の違い」を説明するメカニズムの1つとして、「地理的隔離」を検討した（プロセス研究）。生物学においては、山岳や海などの地理によって孤立した集団は独自の進化を経験し（異なる淘汰圧を受け）、はっきりした系統関係を持つようになる（種分化する）、「異所的種分化」という現象が知られている。この知見を踏まえ、また、日本語族は海によって隔離されている言葉（例：琉球列島）と陸地で繋がっている言葉（例：本州）から構成されていることから、図2のような現象の背景には「地理的隔離」という要因があるという仮説を立て、マンテル検定を用いてそれを検証した。その結果、統計的に有意な水準で、地理的に隔離された言葉の間では言語間距離が遠くなる（種分化する）傾向にあることを確認した（図3）。この結果は言語間距離の計算法を変えても、物理的距離を統制しても、ノンパラメトリック法で相関を求めても、 $R^2$  値は変動するものの、一貫して統計的に有意であることを確認した。

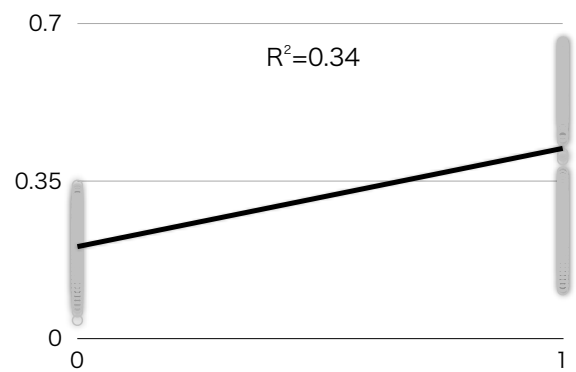


図3. 海による地理的隔離の有無 [X 軸 (0:無,1:有)] と Jaccard 法で計算した 57 の言語間の距離 [Y 軸 (最低 0、最大 1;値が大きくなるほど違う言葉になる)]の相関関係の例 (Pearson)。地理的隔離は言語間距離の分散の約 34%を説明する。

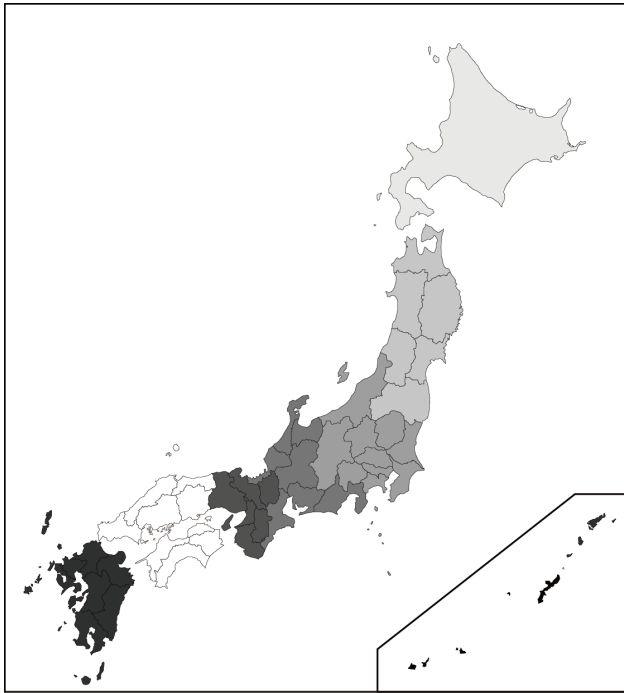


図4. ゲノムと言語間の集団分化指数を比較する際に使われた集団の区画。グラデーションが明るい方から北海道、東北、関東甲信越、東海北陸、近畿、九州、沖縄。中国四国はゲノム解析に含まれていないため比較から排除した。

その上、Nettle (1999) が報告した「話者数の減少による言語進化の高速化」という現象の影響をあまり受けていないということをも確認した。言語進化において異所的種分化のようなプロセスを確認したのはこの研究が初めてであり、同じ日本語族の中でも琉球語は系統関係が鮮明なのに本土日本語はそうではない理由は、それぞれがおかれた地理にあるという進化生物学的メカニズムを確認したことにその意義がある。

更に、言語と話者の進化的関係を調べるために、日本列島の7003人から得られた全ゲノム SNP 分析 (Yamaguchi-Kabata et al. 2008) による集団分化指数 ( $F_{ST}$ ) と本研究の言語間距離から得られた集団分化指数 ( $\Phi_{ST}$ ) を比較した。図4の集団区画に従い、各集団内の SNP と言語それぞれの分散値を計算した後、ペアワイズ比較を行い、その結果を持って集団間の遺伝的・言語的構造を比較したところ、両者は強い相関関係を持ち、言語距離の計算法を変えても、物理的距離を統制しても、ノンパラメトリック法で相関を求めても、一貫して正の相関があることを確認した。これは、日本列島における遺伝的構造は言語的構造と類似していることを示唆すると同時に、この地域における言語とその話者の共進化は現在まで続いていた可能性を示す結果である。

上記の一連の研究は、進化の歴史という大きな研究対象を、現象の記述 (アイヌ語と日本語族の時空間上の系統進化) を行ってその有様を十分理解した上で背景にあるメカニズムの究明 (地理的な孤立が日本語族進化に与える影響) に取り組む、というシステムティックな枠組みを採用しており、これまで散在したエビデンスをまとめ、総括する研究パラダイムの1つの先例となると思われる。本研究は、地理的には日本列島という地球上のごく一部に限定された研究ではありつつも、これまで定量化が難しいと思われていたヒトの言語という現象を定量的なアプローチで捉えているため、同様のデータ処理方法や分析方法を持って他の地域で検証・再現することが安易であり、幅広い分野の研究者のための踏み台を提供すると思われる。

上記の一連の研究は、進化の歴史という大きな研究対象を、現象の記述 (アイヌ語と日本語族の時空間上の系統進化) を行ってその有様を十分理解した上で背景にあるメカニズムの究明 (地理的な孤立が日本語族進化に与える影響) に取り組む、というシステムティックな枠組みを採用しており、これまで散在したエビデンスをまとめ、総括する研究パラダイムの1つの先例となると思われる。本研究は、地理的には日本列島という地球上のごく一部に限定された研究ではありつつも、これまで定量化が難しいと思われていたヒトの言語という現象を定量的なアプローチで捉えているため、同様のデータ処理方法や分析方法を持って他の地域で検証・再現することが安易であり、幅広い分野の研究者のための踏み台を提供すると思われる。