HOKKAIDO UNIVERSITY

Title	語形と規則性の複合的なデータの構築 ： アイヌ語基礎語彙の情報の最大化に向けて
Author(s)	深澤, 美香; 小野, 洋平
Citation	北方言語研究, 16, 151-173
Issue Date	2026-03-20
DOI	https://doi.org/10.14943/2115.99347
Doc URL	https://hdl.handle.net/2115/99347
Type	departmental bulletin paper
File Information	09_Fukazawa_Ono.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北方言語研究 16: 151-173（日本北方言語学会 2026）

語形と規則性の複合的なデータの構築
―アイヌ語基礎語彙の情報の最大化に向けて―

深 澤　美 香
（国立アイヌ民族博物館）
小 野　洋 平
（聖路加国際大学大学院公衆衛生学研究科）

キーワード：アイヌ語、文献学、特徴量抽出、比較言語学、基礎語彙統計学再考

1. はじめに

アイヌ語の方言研究は、言語状況や資料的価値の観点から、大部分が服部・知里（1960）の基礎語彙データに依拠してきた。Asai（1974）のアイヌ語の方言分類に関する重要な研究も、服部・知里（1960）のデータを基に加筆・修正を施し、クラスター分析を行ったものである。本研究では、Asai（1974）のアイヌ語基礎語彙のデータを、現在のアイヌ語研究の知見から適切に数値化することを試みる。具体的には、深澤・小野（2025）、Ono & Fukazawa（2025）で提案した「基礎語彙の語形のみに頼らない方言分類の手法」を拡張・発展させて、新たな方言データを構築したうえで、統計手法を用いて再分析を行う。

本稿の構成は、2 節で本研究の背景となる先行研究について説明し、3 節で本研究に直結する深澤・小野（2025）及び Ono & Fukazawa（2025）が残した課題について解説する。また、4 節では語形データと規則性データで構成される複合的なデータの構築について論じ、5 節において本研究の結果と考察について述べる。

2. 本研究の背景

２．１．基礎語彙 200（202）項目のデータと研究

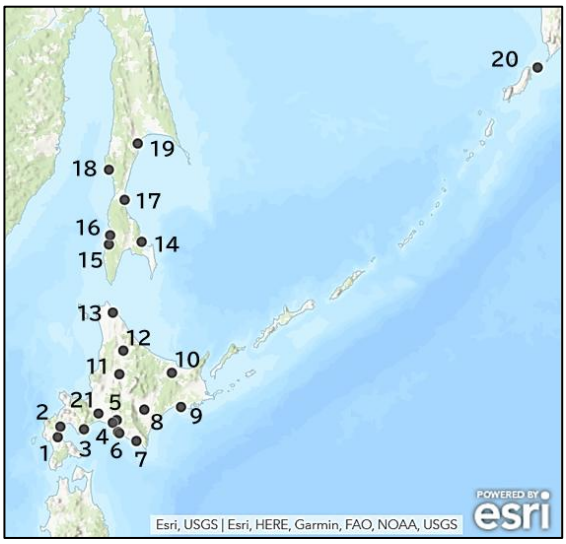
本研究の背景となる研究は、服部・知里（1960）と Asai（1974）の方言分類の研究である。服部・知里（1960）では、著者らが 1955 年にアイヌ語 19 方言と 200 項目の基礎語彙[1] を調査し、そのデータについて統計学的に分析している。また、Asai（1974）では、アイヌ語諸方言の基礎語彙をクラスター分析にかけてアイヌ語の方言分類を行っている（以下、クラスター分析にかけることを「計算」と略す）。なお、このとき Asai（1974）は、服部・知里（1960）の基礎語彙データを用い、一部の語彙や方言については文献や自身の調査データによって追加や修正を行っている。図 1 は Asai（1974）が対象にした 21 方言で、服部・知里（1960）から 2 方言（千歳方言・千島北部方言）が追加されている。

また、Asai（1974）のデータでは、服部・知里（1960）の基礎語彙 200 項目のうち 2 項目[2] がそれぞれ 2 分割されて 202 項目となっている。Asai（1974）は、202 項目中 110 項目を計

[1] アイヌ語基礎語彙 200 項目について、詳しくは、深澤・小野（2025）、深澤（2025）を参照されたい。なお、調査者・アイヌ語話者・調査日・調査経緯などは、服部・知里（1960）や服部（1964, 1999）に詳述されている。

[2] 「雨」（76. rain）と「かぐ」（187. smell）の 2 項目が名詞と動詞に分けて項目立ててある。

算に採用し、アイヌ語の方言分類を提示した。以降、Asai（1974）の方言分類の結果はさまざまな形で引用されるようになるが、Asai（1974）が計算に採用した 110 項目がどの項目なのかが同論文中に明記されておらず､結果の検証がないまま 50 年近くが過ぎることとなる。



1. 八雲、2. 長万部、3. 幌別、
4. 平取、5. 貫気別、6. 新冠、
7. 様似、8. 帯広、9. 釧路、
10. 美幌、11. 旭川、12. 名寄、
13. 宗谷、14. 落帆、15. 多蘭泊、
16. 真岡、17. 白浦、
18. ライチシカ、19. 内路、
20. 千島北部、21. 千歳

図 1：方言の地点（深澤・小野 2025）

Asai（1974）の計算に採用された 110 項目を特定し、本格的に検証を行えるようになったのが、小野・深澤（2023）による研究である。小野・深澤（2023）は、Asai（1974）の論文中に記載される語形同士の「類似」の情報を頼りに、Asai（1974）が行った類似判断の規準を復元した。そして、Asai（1974）の計算から除外された 92 項目のうち、論文中に記載のなかった 25 項目を特定した（図 2）。翌年の小野・深澤（2024）では、服部・知里（1960）と Asai（1974）の語形同士の類似判断の規準の違いが、結果的に方言分類の違いを引き起こすということを発見した。

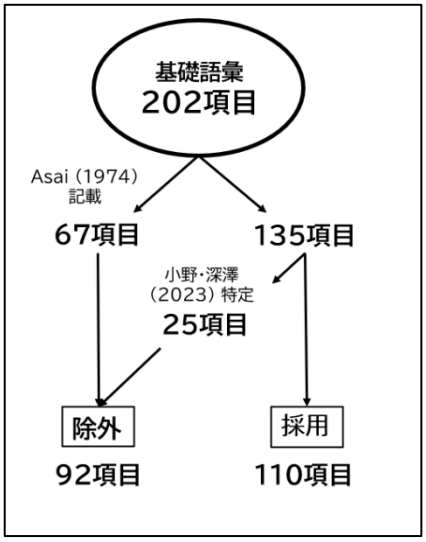


図 2：Asai (1974) で除外・採用された項目の特定の経緯（深澤・小野 2025）

そして、深澤・小野（2025）及び Ono & Fukazawa（2025）では、Asai（1974）の計算から省かれた基礎語彙 92 項目を各方言の新たなデータとして取り込む方法を検討し、基礎語彙 202 項目を最大限に活用することを試みた。具体的には、小野・深澤（2023）で特定した 25 項目は、Asai（1974）がすべての語形が「類似」しているという理由で除外した項目であったため、その類似判断の基準となる音対応を方言データとして取り込んだ。その結果、Asai（1974）のアイヌ語諸方言の分類結果に現れた方言グループ（北海道方言を北部方言、東部方言、西部方言、南部方言に分類）は、Asai（1974）が示した結果以上に弁別的な方言グループを形成することとなった（図 3）。

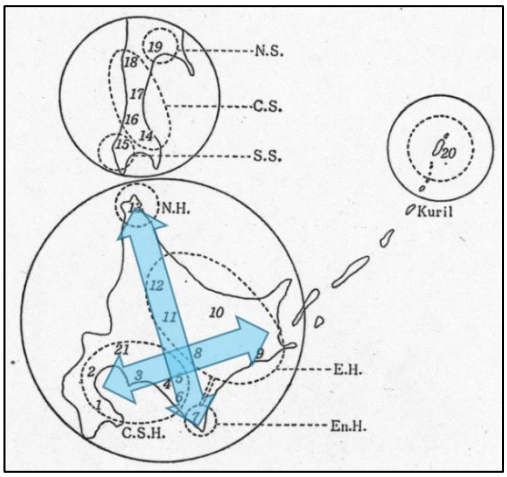


図 3：Asai (1974: 100) の方言分類と北海道方言の 4 分類（深澤・小野 2025）

なお、Asai（1974）は 202 項目中 110 項目を計算に採用したわけだが、服部・知里（1960）もまた基礎語彙 200 項目全てを採用しているわけではない。服部・知里（1960）が除外した項目は、ある方言において調査に対して疑問がある場合や話者が回答できなかった場合、調査漏れの場合、そして、Asai（1974）と同様、すべての方言において語形が類似または一致する場合などである。まとめると表 1 のようになる。

表 1 ：先行研究の基礎語彙採用項目数と特徴

先行研究	採用項目数 （全体項目数）	主な特徴
服部・知里（1960）	91（200）	基礎語彙 200 項目の語彙リストを作成
Asai（1974）	110（202）	服部・知里（1960）をベースにデータを加筆・修正し、クラスター分析を行う
深澤・小野（2025）、Ono & Fukazawa（2025）	202（202）	Asai（1974）で除外された 92 項目を中心に「類似」の情報をデータとして活用

２．２．類似判断の規準と規則性

2.1 節でも述べたとおり、深澤・小野（2025）と Ono & Fukazawa（2025）では、既存の基礎語彙のデータを余すところなく最大限に活かすことで、方言分類などの分析結果にも適切な影響を及ぼすのではないかという仮説を立て、それを検証している。具体的には、Asai（1974）の計算に採用されなかった 92 項目の語形データの「類似」の情報から、13 種類の規則性データを抽出し、Asai（1974）の 110 項目の語形データに追加した。

規則性データの抽出で参考にしたのは、Asai（1974）の語形同士の類似判断の規準である。Asai（1974）は、ある項目のなかで語形同士が殆ど類似、もしくは明らかに類似している場

合、語形同士を括弧でくくって表示している。例えば、「魚」[3]（19. fish）という項目は、Asai（1974）の論文中で（ciep, cep）, ceh のように示される[4]。これにより、Asai（1974）の類似判断において ciep と cep は「類似」と判断されているということがわかる。

小野・深澤（2023）では、このような括弧で示された類似の情報を頼りに、(1）で表される類似判断の規準を得た。

（1）Asai（1974）の類似判断の規準[5]（小野・深澤 2023）

a. 母音 1 タイプ：1-1（V_1V_1：V_1）, 1-2（V_1V_1：V_1y）
b. 母音 2 タイプ：2-1（V_1y：V_1）, 2-2（V_1y：V_1i）
c. 母音 3 タイプ：（V_1V_2：V_2）
d. 母音 4 タイプ：4-1（C_1e：C_1i）, 4-2（C_1V：C_1）/4-2’（yV：y～i）[6],
4-3（C_1：C_1＋V）
e. 半母音タイプ：1-1（y：w）, 1-2（iw：uy）
f. h- タイプ：1-1（hV_1：’V_1）, 1-2（’V_1：h-＋V_1）
g. 子音タイプ：1-1（C_1C_1：C_1）, 1-2（C_1C_2：C_2）
h. r タイプ
i. 接辞タイプ
j. 所属形タイプ
k. 単複タイプ
l. 重複タイプ
m. 位置名詞タイプ

類似判断の規準のうち a.～h.のタイプは、音対応に関する方言間の語形の規則性を捉えたものである。Asai（1974）はこれらの規則性をもつ語形を「類似」とみなし、項目内の対象となる語形の異なりを計算に含めなかった。さらに、項目内のすべての語形が「類似」であった場合には、項目ごと計算から除外した。しかし、深澤・小野（2025）および Ono & Fukazawa（2025）では、この規則性こそが方言特徴のデータとなり、基礎語彙 202 項目に含まれる情報を最大限に活用することに繋がるのではないかと考えた。よって、計算から項目ごと省かれた基礎語彙 92 項目を中心として、「類似」の情報から導き出された 13 種類の規則性を各方言データとして取り入れた。（2）は、（1）の類似判断の規準のうち、除外された 92 項目を中心とした規準である。また、（3）は、（2）の類似判断の規準から構成される 13 種の規則性である。例えば、「魚」（19. fish）の語形（ciep, cep）, ceh を例にすると、括弧でくくられた ciep と cep の ie と e の音対応は、規則性 5 の V_1V_2：V_2 で表されていることになる。

[3] 項目の日本語訳は、服部・知里（1960）を参考にした。
[4] 実際には各語形の後ろに括弧付きの数字が付記されており、図 1 の方言番号に対応する。
[5] 本稿では、Asai（1974）で用いられるアイヌ語の j は y と表記する。
[6] 母音の対応関係があるわけではないため、4-2’（yV_1：y～i）から 4-2’（yV：y～i）へ表記を変更した。

(2) 除外された 92 項目の「類似」の情報を中心として抽出した類似判断の規準

a. 母音 1 タイプ：1-1（V_1V_1：V_1），1-2（V_1V_1：V_1y）
b. 母音 2 タイプ：2-1（V_1y：V_1）
c. 母音 3 タイプ：3（V_1V_2：V_2）
d. 母音 4 タイプ：4-1（C_1e：C_1i），4-2（C_1V：C_1）/4-2'（yV：y～i）
e. 半母音タイプ：1-1（y：w），1-2（iw：uy）
f. h-タイプ：1-1（hV_1：'V_1），1-2（'V_1：h-＋V_1），
g. 子音タイプ：1-3（C_1C_2：C_2C_2）
h. r タイプ　（深澤・小野 2025）

(3) 13 種の規則性（括弧内の数字は図 1 の方言番号と対応）

規則性 1：V_1V_1（14-19）：V_1（1-13, 20, 21）【母音 1-1 タイプ】
規則性 2：V_1V_1（14-19）：V_1y（1-3, 19）：V_1（1-13, 20, 21）【母音 1-1 タイプ+母音 1-2 タイプ+母音 2-1 タイプ】
規則性 3：V_1y（1, 3, 7-17, 19-21）：V_1（4-6, 18）【母音 2-1 タイプ】
規則性 4：V_1V_2（1-4, 6-12, 17-19, 21）：V_2（4, 5, 13, 20）：V_2V_2（15-19）【母音 3 タイプ+母音 1-1 タイプ】
規則性 5：V_1V_2（1-3, 21）：V_2（3-21）【母音 3 タイプ】
規則性 6：C_1e（1-16, 18-21）：C_1i（17-20）【母音 4-1 タイプ】
規則性 7：C_1V/yV（1-19, 21）：C_1/y～i（20）【母音 4-2 タイプ／母音 4-2'タイプ】
規則性 8：y（1-14, 16-21）：w（9, 10, 15）【半母音 1-1 タイプ】
規則性 9：iw（1-9, 12-17, 19 21）：uy（9, 10）【半母音 1-2 タイプ】
規則性 10：hV_1（1-9, 11, 13-21）：'V_1（8-12, 20）【h- 1-1 タイプ】
規則性 11：'V_1（1-21）：h-＋V_1（7, 10, 13）【h- 1-2 タイプ】
規則性 12：C_1C_2（1-13, 20, 21）：C_2C_2（7-11, 15, 19）：hC_2（14, 16-18）【子音 1-3 タイプ】
規則性 13：rV（14-19, 20*）：r（1-13, 20, 21）【r タイプ】　（深澤・小野 2025）

3．深澤・小野（2025）と Ono & Fukazawa（2025）で残された課題

深澤・小野（2025）と Ono & Fukazawa（2025）では、既存の基礎語彙のデータを余すところなく最大限に活かそうと試みたが、それでもなお次の 2 点が課題として残された。

① 除外された 92 項目の「類似」の情報から導き出された規則性が、採用された 110 項目のなかでは「語形」データとして扱われている点
② 採用された 110 項目では、複数の項目にわたり同様の規則性が確認できる場合であっても、「語形」データとして扱われている点

以下では、具体的な語彙例を見ながら確認していきたい。

① 除外された 92 項目の「類似」の情報から導き出された規則性が、採用された 110 項目のなかでは「語形」データとして扱われている点

（4）語彙例（●は除外された 92 項目中 25 項目に含まれる）

●83. ashes...（**uu**na：**u**na）：**uy**na → **規則性 2：V_1V_1：V_1y：V_1**

107. here...（**tee**ta：**te**yta）：teta → **規則性 2：V_1V_1：V_1y：V_1**
語形 107：（teeta：teyta）：teta

「灰」（83. ashes）の項目は、Asai（1974）の計算から除外された 92 項目中 25 項目である。Asai（1974）の論文中には、uyna という語形が uuna や una と類似すると明記されていない。しかし、小野・深澤（2023）では uyna と uuna、una という三つの語形が Asai（1974）の結果からみて類似すると判断され、項目ごと計算から除外されたのではないかと結論づけた。そして、深澤・小野（2025）は、Asai（1974）の計算から除外された「灰」（83. ashes）の情報を「規則性 2」（V_1V_1：V_1y：V_1）のデータとして取り入れた[7]。

それに対し、「ここに」（107. here）は、Asai（1974）の計算に採用された 110 項目に含まれる。Asai（1974）において teeta と teyta が同一の括弧に含められていることから、teeta と teyta は類似とされ、teta はその他の語形とは類似しないということになる。深澤・小野（2025）は、この項目も規則性 2 を抽出したデータとして扱ったが、実際この項目は 110 項目中の語形データとしても扱われており、データに含まれる情報が重複してしまったことになる。

② 採用された 110 項目では、複数の項目にわたり同様の規則性が確認できる場合であっても、「語形」データとして扱われている点

（5）語彙例の一部

61. die...ray：tay → **語形 61：ray：tay**

85. path...（ruu：ru）：tuu → **語形 85：（ruu：ru）：tuu**

100. name...（ree：re）：tee → **語形 100：（ree：re）：tee**

150. freeze...rupus：tupus → **語形 150：rupus：tupus**

Asai（1974）の計算に採用された 110 項目のうち、複数の項目にわたって語頭の rV_1 が tV_1 と対応する例が見られる。「死ぬ」（61. die）は ray と tay の対応が見られるが、これらは Asai（1974）では「類似」を表す同一の括弧には含められていない。同様に、「道」（85. path）の項目を見てみると、ruu と ru は同一の括弧に含められており、規則性 1：V_1V_1（14-19）：V_1（1-13, 20, 21）で扱われるものであるが、tuu は同一の括弧には含められていない。「名前」（100. name）や「氷る」（150. freeze）の項目にも語頭の rV_1 と tV_1 が見られる。しかし、深澤・小野（2025）では Asai（1974）が非類似として扱った音対応は規則性データとして扱わ

[7] V_1V_1 が樺太方言の長母音を表し、V_1 がそれに対する北海道方言の昇り核アクセントを持った母音を表す。また、Fukazawa（2018: 95-96）では、V_1y が V_1V_1 と V_1 の歴史的変遷過程の途中の語形を表しているのではないかと指摘した。

なかったため、これらはすべて語形データとして扱われており、rV_1 と tV_1 の異なりを示すデータの比重が大きくなってしまったことになる。

４．語形と規則性からみた新たなデータ構築

４．１．データの改善

本研究の目的は、Asai（1974）のアイヌ語の基礎語彙による方言データを適切に数値化することである。深澤・小野（2025）と Ono & Fukazawa（2025）の研究では、計算から除外された 92 項目の「類似」の情報から規則性を抽出し、それを 110 項目の語形データに追加したが、本研究では、Asai（1974）によって語形データとして扱われてきた 110 項目のなかにも音対応による規則性があることに焦点を当てる。さらに、Asai（1974）の基礎語彙 202 項目のデータを、改めて語形と規則性の二つのデータとして整理し、新たな方言データを作成する。

以下では、語彙例を通して本研究のデータの改善点について見ていく。

① 語形と規則性のデータを再整理

採用された 110 項目において、除外された 92 項目の類似の情報から導き出された規則性が見られる場合は、「語形」データとして扱わず「規則性」データに統合した。

（6）語彙例 （●は除外された 92 項目中 25 項目に含まれる）

●83. ashes...（**uu**na：**u**na）：**uy**na → **規則性 2：V_1V_1：V_1y：V_1**

107. here...（**tee**ta：**tey**ta）：**t**eta → **語形 107 を規則性 2：V_1V_1：V_1y：V_1 に統合**

② 「非類似」の情報も含めて新たな規則性データを作成

採用された 110 項目において、複数の項目にわたり同様の規則性が確認できる場合は、新たな規則性データとして語形データを統合した。

（7）語彙例の一部

61. die...ray：tay → **語形 61**
85. path...（ruu：ru）：tuu → **語形 85**
100. name...（ree：re）：tee → **語形 100**
150. freeze...rupus：tupus → **語形 150**

→ **新たな規則性 rV：tV に統合**

このように 3 節で述べた深澤・小野（2025）及び Ono & Fukazawa（2025）で残された課題を克服することで、本研究では先行研究と異なる新たなデータを構築した。比較すると表 2 のようにまとめられる。

表 2：先行研究と本研究の基礎語彙項目の採用項目数とデータの種類の比較

研究	採用項目数 （全体項目数）	採用項目から抽出した データの種類
服部・知里（1960）	91（200）	服部・知里（1960）の 91 項目の語形データ
Asai（1974）	110（202）	Asai（1974）の 110 項目の語形データ
深澤・小野（2025）、 Ono & Fukazawa（2025）	202（202）	Asai（1974）の 110 項目の語形データ +13 種類の規則性データ（主に 92 項目の「類似」の情報から抽出）
本研究	202（202）	Asai（1974）の 87 項目の語形データ +22 種類の規則性データ（202 項目の「類似」・「非類似」の情報から抽出）

４．２．語形と規則性の新たなデータ

本研究のデータの特徴については、改めて以下の二点にまとめられる。

① Asai（1974）や服部・知里（1960）が行ってきた語形間での言語学的な類似判断を、規則性というデータに変換し、方言データとして取り込んだ点
② すべての項目を対象とし、同様のデータ構成をもつ語形データを規則性データとして統合し、情報が重複しないデータを構築した点

上記の①に関わり、本研究で新たに採用した規則性データ 9 種を（8）に挙げる。また、それに関連して、Asai（1974）が行った類似判断にはなかった新たな規準を（9）に挙げる。

（8）規則性データ
規則性 14：V_1V_2（4-10, 21）：V_2（1-4, 11-13）【母音 3 タイプ】
規則性 15：C_1e（1-6, 11-21）：C_1i（7-10）【母音 4-1 タイプ】
規則性 16：yV（1-6, 11-14, 16-19）：y～i（7-11, 20, 21）【母音 4-2’タイプ】
規則性 17：C_1C_1（1-13, 15-19, 20, 21）：C_1（14, 15, 17, 19, 20）【子音 1-1 タイプ】
規則性 18：V_1p/t/k（1-13, 15, 19-21）：V_1h/s（14, 16-18）【子音 1-4 タイプ】
規則性 19：pa（1-6, 11, 12, 21）：ca（7-10, 12, 13-20）【子音 1-5 タイプ】
規則性 20：rV_1(V_1)（1-13, 14-18, 20, 21）：tV_1(V_1)（14-19）【子音 1-6 タイプ】
規則性 21：tu（1-6, 8-21）：cu（7）【子音 1-7 タイプ】
規則性 22：ara（14, 16, 17, 19）：ari（1, 8-10, 11-13, 15, 21）：aro（1-13, 20, 21）：aru（15, 18, 19, 20）【子音 1-8 タイプ】

（9）採用された 110 項目を中心に抽出した新たな類似判断の規準（a.～m.までは前掲の（1）を参照）

n. 子音タイプ：1-4（V_1p/t/k：'V_1h/s），1-5（pa：ca），1-6（rV_1(V_1)：tV_1(V_1)），1-7（tu：cu），1-8（ara：ari：aro：aru）

上記の②に関わり、本研究で情報が重複しない新たなデータとして整理したものを（10）と（11）に示す。（10）は、Asai（1974）の 110 項目のうち、語形データから規則性データに変換・統合された項目である。（11）は、規則性データへ変換・統合されたことで、厳選されることとなった語形データ 87 項目である。

（10）規則性データに統合（24 項目）

17. man, 19. fish, 21. dog, 36. feather, 39. ear, 40. eye, 41. nose, 42. mouth, 45. claw, 50. neck, 61. dye, 65. walk, 66. come, 85. path, 90. white, 93. hot, 100. name, 116. day, 117. wind, 119. sea, 146. wife, 153. cut, 156. far, 158. thick

（11）語形データ（87 項目：14. long は除外された 92 項目中 25 項目に含まれる）

1. I, 6. who, 7. what, 10. many, *14. long, 15. small, 22. louse, 23. tree, 24. seed, 25. leaf, 26. root, 31. bone, 32. grease, 37. hair, 38. head, 43. tooth, 44. tongue, 47. knee, 48. hand, 49. belly, 51. breast, 53. liver, 59. know, 62. kill, 63. swim, 64. fly, 69. stand, 70. give, 72. sun, 74. star, 75. water, 76. rain (the rain), 76'. rain (it rains), 77. stone, 78. sand, 79. earth, 81. smoke, 82. fire, 84. burn, 86. mountain, 91. black, 92. night, 94. cold, 99. dry, 101. ye, 104. how, 105. when, 106. where, 107. here, 108. there, 113. few, 114. sky, 115. day, 121. river, 122. wet, 126. back, 127. leg, 129. wing, 130. lip, 132. navel, 133. guts, 138. grass, 143. mother, 144. father, 145. husband, 147. salt, 148. ice, 149. snow, 150. freeze, 151. child, 154. wide, 155. narrow, 157. near, 159. thin, 161. heavy, 163. sharp, 164. dirty, 166. rotten, 167. smooth, 171. right, 173. rub, 174. pull, 175. push, 180. dig, 188. puke, vomit, 198. alive, 199. rope

次節からは、本研究で採用した規則性データのうち、（8）で示した新たな規則性 9 種について解説する（規則性 1〜13 の詳細については、深澤・小野（2025）を参照）。

4.2.1. 規則性 14：母音 3 タイプ（V_1V_2：V_2）

規則性 14 は、母音 3 タイプに相当する「知っている」（59. know）の項目を一つの規則性と捉えたものである。対象となるのは同項目中の eraman と eramuan という二つの語形である。eramuan を分解すると、e-ram-u-an「〜について・心・所属形・存在動詞（単数）」という 2 項動詞であり、eraman は eramuan から音縮約された語形である。沙流・千歳から北海道東部方言がひとまとまりになる方言分布を示す。

服部・知里（1960）の沙流川中流〜下流にあたる平取・福満あたりの方言では両形が見られるが、田村（1996）では沙流川中流以下および鵡川では eraman であるという報告がある。また、千歳ではもっぱら eramuan という形で用いられる（中川 1995）という報告も見られ、隣接する地域のなかでも違いが確認される。

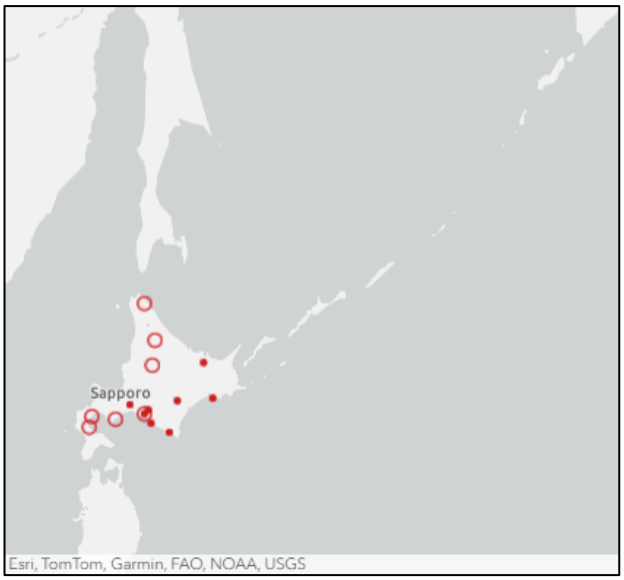


図 4：規則性 14

【規則性 14：V_1V_2（4-10, 21）：V_2（1-4, 11-13）】：

語彙例：

59. know...eramuan（4-10, 21）：eraman（1-4, 11-13）[8]

4.2.2. 規則性 15：母音 4-1 タイプ（C_1e：C_1i）

規則性 15 は、母音 4 タイプに相当する「犬」（21. dog）の項目をひとつの規則性と捉えたものである。「犬」という語彙は、seta と sita の二つの語形が確認できる。基礎語彙 202 項目のなかには類例が確認できないものの、語形が音対応に関連し、かつ方言分布も明確であったため、ここでは語形データではなく規則性データとして扱った。

アイヌ語の sita という語形が見られるのは北海道東部方言と千島北部方言に限られるが、seta はその他の北海道方言と樺太方言に広く分布する。また、金田一（1993［1937］）及び知里（2001［1952］）によれば、本州の東北地方（山形から秋田、津軽にかけて）の猟師の山言葉では、セタ seta、シェダ ʃeda、セッタ setta、ヘダ heda という語を持ち、「犬」あるいは場所によっては里言葉で「犬の肉」を意味する。本州の東北地方のアイヌ語については詳しいことがわかっていないが、これらもアイヌ語に由来すると考えられている。

[8] Asai（1974）の誤植を改めた。

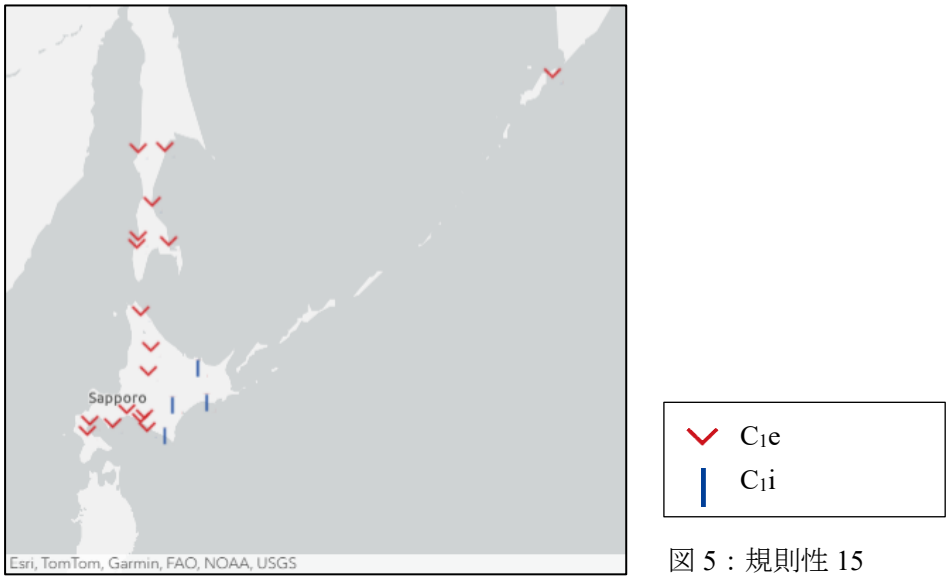


図 5：規則性 15

【規則性 15：C_1e（1-6, 11-21）：C_1i（7-10）】：

語彙例：

21. dog...**se**ta（1-6, 11-21）：**si**ta（7-10）

4.2.3. 規則性 16：母音 4-2’タイプ（yV：y ~ i）

規則性 16 は、母音 4-2’タイプ（yV：y ~ i）に属し、「男」（17. man）の項目をひとつの規則性と捉えたものである。同様の母音 4-2’タイプ（yV：y ~ i）を含む規則性 7 は千島北部方言とその他の方言という方言分布を持つが、規則性 16 は全く異なる方言分布を持つ。

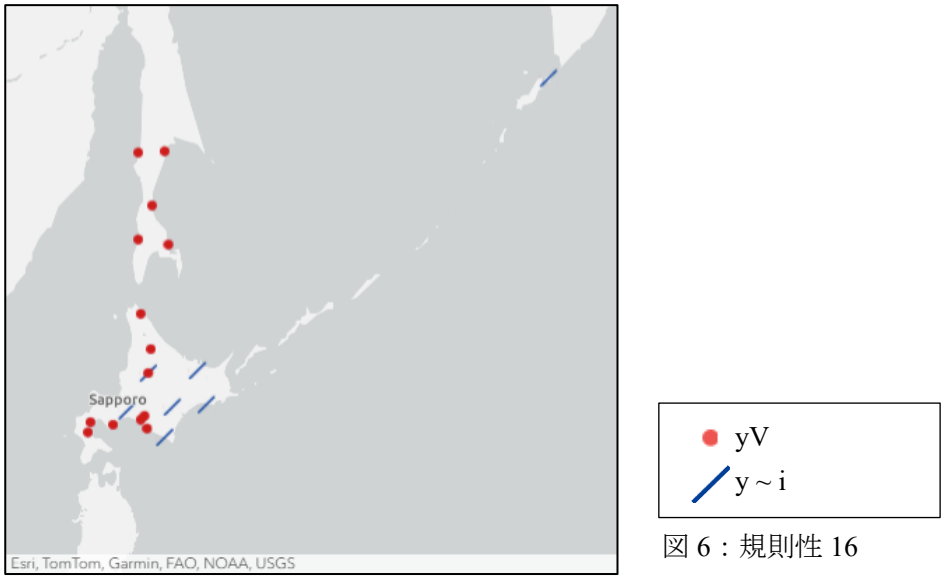


図 6：規則性 16

【規則性 16：yV（1-6, 11-14, 16-19）：y ~ i（7-11, 20, 21）】

語彙例：

17. man...okka**yo**（1-6, 11-13）, ohka**yo**（14, 16-19）：okka**y**（7-11, 20, 21）

「男」（17. man）は、okkayo や ohkayo のように後ろに o という母音が後続する yV タイプと、okkay のように y で終わる y～i タイプの二つに大きく分かれる。北海道の西部・北部方言および樺太方言では yV，北海道の東部方言および千島北部方言では y ～ i タイプとなり、地理分布上も大きく二分されるのが特徴で、歴史的にどちらが古い形なのかは明確ではない。樺太の多蘭泊方言だけ okkaw となる。なお、okkayo と ohkayo の対応は、規則性 18 で述べる。

4.2.4. 規則性 17：子音 1-1 タイプ（C_1C_1：C_1）

規則性 17 は、子音 1-1 タイプに属する。同じ子音が連続する場合に、子音が一つになる方言と語形同士で対応関係を持つ場合である。子音が一つの場合は、樺太方言の長母音に後続することや、千島北部方言の語形によく見られる。

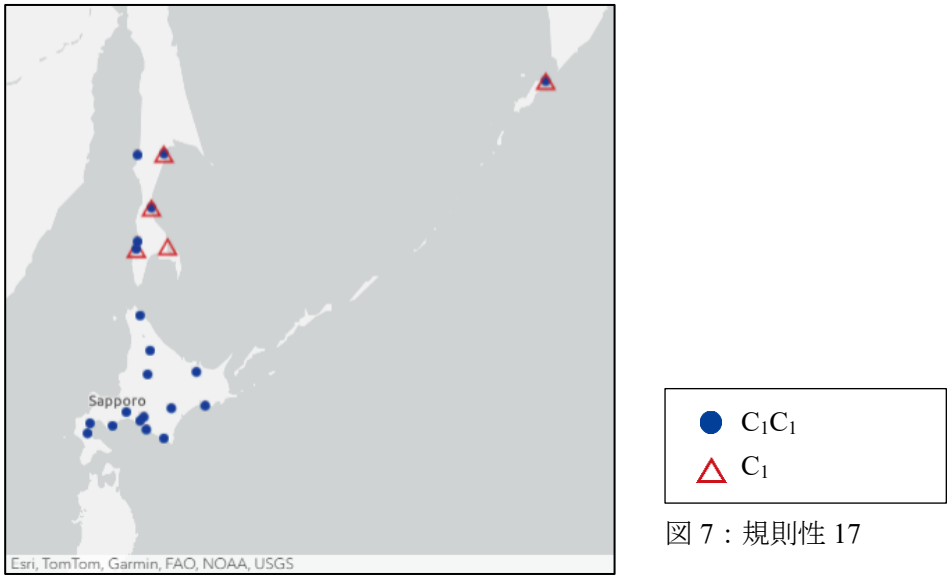


図 7：規則性 17

【規則性 17：C_1C_1（1-13, 15-19, 20, 21）：C_1（14, 15, 17, 19, 20）】：

語彙例（●は除外された 92 項目内 25 項目に含まれる）：

6. who...hu**nn**a（4-6, 21）：hu**n**at（20）

●14. long...ta**nn**e（1-12, 20, 21）, ota**nn**e（13, 15, 16, 18）：otaa**n**e（14, 17, 19）

62. kill...ro**nn**o（7-9, 21）：ro**n**a（20）

99. dry...sa**tt**ek（13, 19）：sa**t**ek（20）

158. thick...iro**nn**e（1-13, 15-19, 21）, iro**nn**’（20）：iroo**n**e（14）

164. dirty...ica**kk**ere（1-13, 19-21）：ica**k**ere（14, 15）

4.2.5. 規則性 18：子音 1-4 タイプ（V_1p/t/k：V_1h/s）

規則性 18 は、新たな子音 1-4 タイプに相当し、多くの語彙例を持つ。深澤・小野（2025）でも「非類似」の情報から例示的に抽出した規則性のデータ（論文中の通し番号で規則性 14）として取り上げたが、本研究では規則性 18 として規則性データに追加する。

北海道方言でみられる音節末の-p/-t/-k と樺太の多くの方言で見られる-h/-s の対応は、基

礎語彙 202 項目中にも非常に多く見られる。服部・知里（1960）はこの対応を「類似」と扱うが、Asai（1974）は「非類似」として扱っており、二つの論文の類似判断の規準とデータの異なりを示す最たる例の一つである。

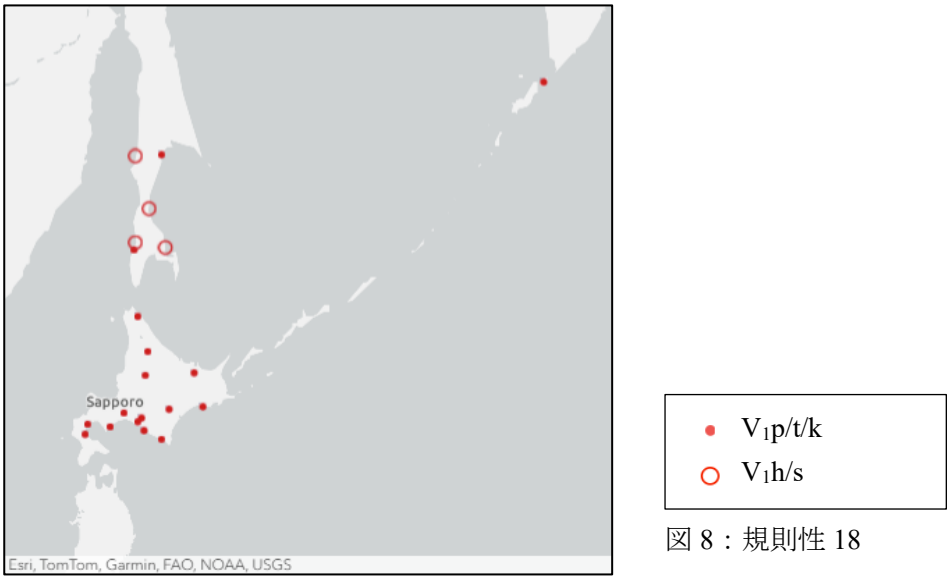


図 8：規則性 18

【規則性 18：V_1p/t/k（1-13, 15, 19-21）：V_1h/s（14, 16-18）】

語彙例の一部（〇は除外された 92 項目に含まれる）：

19. fish...ci**ep** (1-3, 21), c**ep** (3-13, 15, 19-21)：c**eh** (14, 16-18)
〇20. bird...cik**ap** (1-13, 15, 19, 21)：cik**ah** (14, 16-18)
26. root...sinr**it**（1-13, 15, 20, 21）：sinr**is**（18）
〇28. skin...k**ap**（1-9, 11-13, 15, 19-21）：k**ah**（14, 16, 18）
36. feather...r**ap**（1-13, 15, 20, 21）, t**ap**（19）：r**ah**（14, 16-18）
40. eye...s**ik**（1-13, 15, 19-21）：s**is**（14, 17, 18）[9]
43. tooth...im**ak**（7-10, 13, 15, 19, 20）：im**ah**（14, 16-18）
48. hand...t**ek**（1-13, 15, 19-21）：t**eh**（14, 16, 18）
50. neck...rek**ut**（1-13, 15, 20, 21）：rek**uh**（14, 16-18）, tek**uh**（19）
66. come...**ek** (1-13, 15, 19-21)；**eh** (14, 16-18)
72. sun...c**up**（1-13, 19-21）：c**uh**（14, 16-18）
93. hot...ses**ek**（1-13, 15, 19-21）：ses**eh**（14, 16-18）
99. dry...satt**ek**（13, 19）, sat**ek**（20）：saht**eh**（14, 16-18）
〇128. arm...t**ek**（3, 7, 9, 11-13, 15, 19, 21）：t**eh**（16-18）
129. wing...r**ap**（1, 2, 5, 7-9, 11, 13, 21）：r**ah**（14, 17）
129. wing...tekk**up**（3, 4, 6, 8-12, 15, 19, 20）：tehk**uh**（16, 18）
133. guts...tuyor**op**（6）, tuwor**op**（15）：tuyor**oh**（16-18）

[9] 樺太の真岡方言は sik-となっており、後続する要素に続く際に k が出てきていると考えられることから、sik の語形からは除外した。

○136. fruit...tu**ep**（15, 19）：tu**reh**（14, 16-18）
146. wife...m**at**（1-13, 15, 19-21）：m**ah**（14, 16-18）
148. ice...**rup**（12, 13, 15）, **tup**（19）：**ruh**（14, 16-18）
○152. dark...ekur**ok**（20）：ekur**oh**（14）
154. wide...os**ep**（13, 15, 19）：os**eh**（14, 16-18）
167. smooth...rar**ak**（1-7, 11. 13. 20. 21）. raar**ak**（15）, taar**ak**（19）：raar**ah**（14, 16-18）
○187. smell...hurahu r**ak**（15）：hurahu r**ah**（17）

4.2.6. 規則性 19：子音 1-5 タイプ（ca：pa）

規則性 19 は、新たな子音 1-5 タイプに相当し、「ロ」に関連する語彙をもつ。主に、北海道南西部の方言で pa 系、北海道北東部、千島北部、樺太の方言で ca 系の語形が確認される。ca と pa の境界にあたる方言は、旭川方言、名寄方言、さらに沙流方言より少し東に位置する静内方言であり、典型的な南西・北東型の分布を示す（ca と pa の歴史的な考察に関する詳細は、深澤（2017）を参照）。

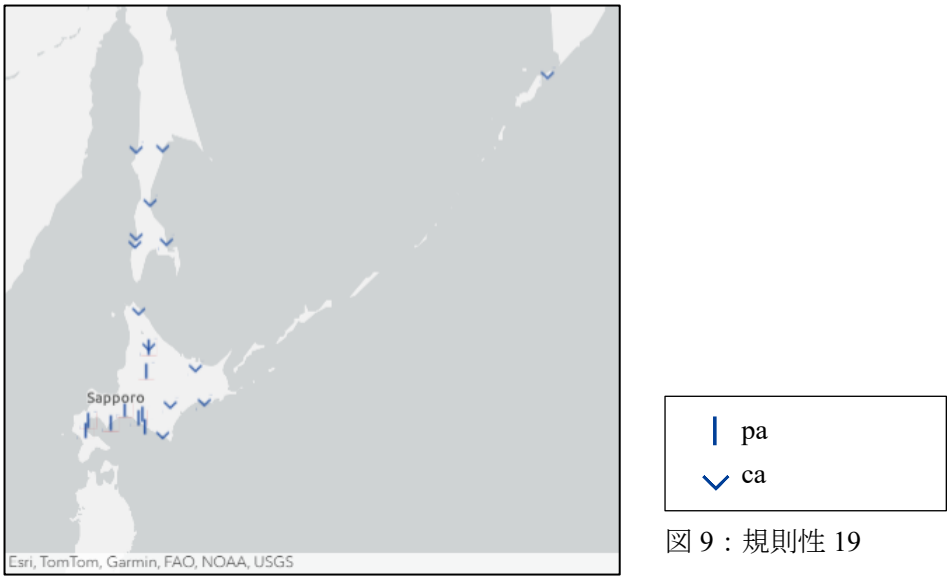


図 9：規則性 19

【規則性 19：pa（1-6, 11, 12, 21）：ca（7-10, 12, 13-20）】
語彙例：
42. mouth...**pa**ro（1-6, 11, 21）：**ca**ro（7-10, 12, 13, 20）, **ca**ra（14, 16, 17）, **ca**ru（15, 18, 19, 20）
130. lip...**pa**pus（3, 12）：**ca**pus（7-13）, **caa**pus（14-19）
130. lip...**pa**toye（4-6, 12, 21）：**ca**toy（20）

4.2.7. 規則性 20：子音 1-6 タイプ（$rV_1(V_1)$：$tV_1(V_1)$）

規則性 20 は、新たな子音 1-6 タイプに相当し、主に樺太の内路方言とその他の方言での対応関係をもつ。この規則性が見られるのは語頭の r-と t-に限られる。北海道方言で第一音節が高く発音される例外アクセントが見られる場合、樺太方言では長母音になるため、「名

前」（100. name）のように、re, ree：tee という対応関係も確認される。

語頭の r-と t-の音対応については、方言ではなく話者個人の発音によるものという可能性もあるが、「白い」（90. white）の項目のみ例外的に樺太の全ての方言で語頭の t-が確認される。Vovin（1993：16-18）は、この「白い」（90. white）の例や、内路方言の t-の例から、プロトアイヌ語では *d- だったのではないかと指摘する。例えば「名前」（100. name）は、Vovin（1993: 83）が再構したプロトアイヌ語では *dEE HL となる。しかし、Vovin 自身も、t-をもつ内路のデータがさらに得られない限り決定的な証明に至らないことを述べている。

なお、金田一（1931: 10-11）は retar のような語頭の /r/ は「t, d の間の音と r との中間音になる」と述べ、樺太の retara が detara と書かれていることは聞き違いではあるがもっともな聞き違いだと指摘した。また、知里（1942: 462）も、「樺太に於ては r が頭音に立つ時屡々破裂音の t（又は d）に紛れる」と述べている。

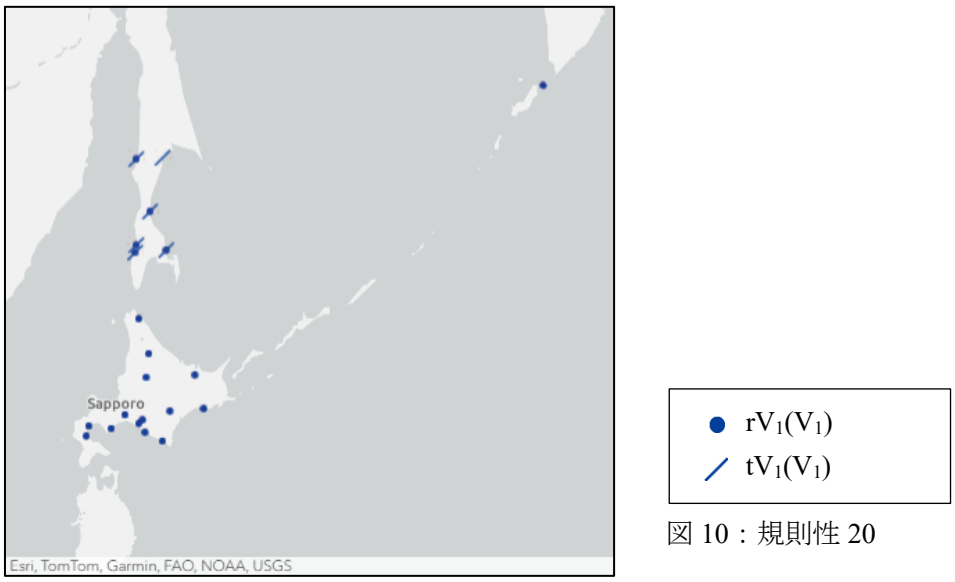


図 10：規則性 20

【規則性 20：$rV_1(V_1)$（1-13, 14-18, 20, 21）：$tV_1(V_1)$（14-19）】

語彙例（〇は除外された 92 項目に含まれる）：

22. louse...**ra**si（14, 16-18）：**ta**si（19）

〇28. skin...**ru**s（10, 15-17）：**tu**s（19）

36. feather...**ra**p（1-13, 15, 20, 21）, **ra**h（14, 16-18）：**ta**p（19）

50. neck...**re**kut（1-13, 15, 20, 21）, **re**kuh（14, 16-18）：**te**kuh（19）

61. dye...**ra**y（1-18, 20, 21）：**ta**y（19）

62. kill...**ra**yke（1-6, 10-16, 21）, **ra**yki（17, 18）：**ta**yki（19）

85. path...**ru**（1-13, 20, 21）, **ruu**（14-18）：**tuu**（19）

90. white...**re**tar（1-13, 20, 21）：**te**tara（14-19）

100. name...**re**（1-13, 20, 21）, **ree**（14-18）：**tee**（19）

117. wind...**re**ra（1-13, 20, 21）, **ree**ra（14-18）：**tee**ra（19）

148. ice...**ru**p（12, 13, 15）, **ru**h（14, 16-18）：**tu**p（19）

150. freeze...**ru**pus（1-14, 16-18, 20, 21）：**tu**pus（19）

167. smooth...**ra**rak（1-7, 11, 13, 20, 21）, **raa**rak（15）, **raa**rah（14, 16-18）: **taa**rak（19）

4.2.8. 規則性 21：子音 1-6 タイプ（tu：cu）

規則性 21 は、新たな子音 1-6 タイプに属する。この規則性は、あらゆる方言で tu となる部分が、北海道様似方言では cu と記録されるものに相当する。小松（2004）は、様似方言話者の一人が tu が cu になる傾向を指摘し、「ツ」か「チュ」と発音されたものではないかと述べている。方言特徴なのか話者や特定の地域の特徴なのかは、もう少し注意深く様似方言の事例、あるいはその他の方言の音声特徴を見ていく必要があると思われる。

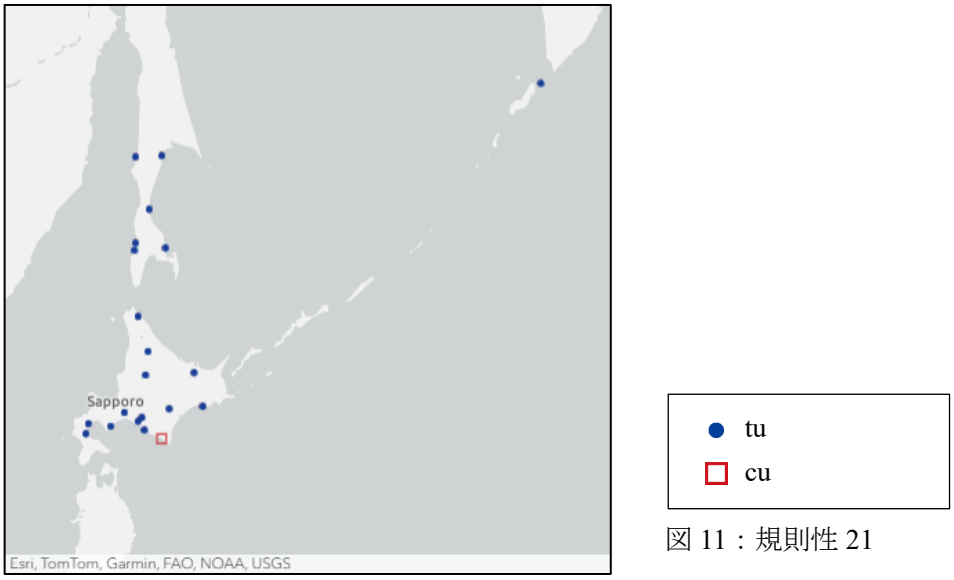


図 11：規則性 21

【規則性 21：tu（1-6, 8-21）：cu（7）】

語彙例：

41. nose...e**tu**（1-6, 8-13, 18, 20, 21）：e**cu**（7）

119. sea...a**tu**y（1-6, 8-19, 21）, a**tu**yka（20）：a**cu**y（7）

126. back...se**tu**r（1-6, 8-13, 21）, se**tu**ru (14, 16-20), se**tu**ri (15)：se**cu**r（7）

133. guts...**tu**y（1-3, 5, 8-13, 20, 21）：**cu**y（7）

153. cut...**tu**ye（1-6, 8-14, 16-19, 21）, **tu**y（20）, **tu**we（15）：**cu**ye（7）

156. far...**tu**yma（1-6, 8-21）：**cu**yma（7）

175. push...opu**tu**ye（1-6, 8-11, 21）：opu**cu**ye（7）

199. rope...**tu**s（1-6. 8-13, 15, 16, 20, 21）：**cu**s（7）

4.2.9. 規則性 22：子音 1-8 タイプ（ar：arV）

規則性 22 は、新たな子音 1-8 タイプに相当する。この形式は文法的なものと音韻的なものを両方含んでいることに注意が必要である。

アイヌ語北海道方言は名詞の概念形に母音を後続させることで所属形をつくることができる。北海道方言では、「口」（42. mouth）という意味をもつ概念形の par や car に、-o を後続させると、所属形の「～の口」を形成する。一方、樺太方言で見られる caru や cara は概

念形か所属形かが区別されない形である。樺太方言では、北海道方言の語末の r が rV に対応し、いつも母音が後続するためである。この音対応は、(2) で示した類似判断の規準のうち「h. r タイプ」に相当する。

規則性 22 は、北海道方言での概念形が ar で終わる形式を基本とした場合、北海道方言の所属形や樺太方言の対応する形でどの母音が後続するかを調べたものである。結果として、地理的に大まかな傾向がみられたことから、ここでは規則性データとして抽出した。

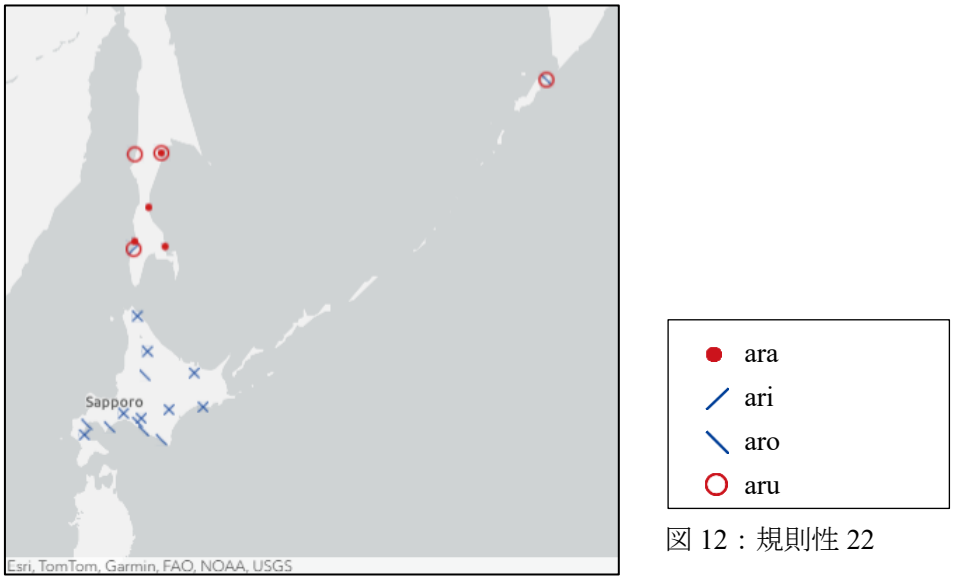


図 12：規則性 22

【規則性 22：ara（14, 16, 17, 19）：ari（1, 8-10, 11-13, 15, 21）：aro（1-13, 20, 21）：aru（15, 18, 19, 20）】

語彙例：

○3. we...anokay ut**ara**（14, 16, 19）：kuut**ari**（9, 12）, cokayut**ari**（13）, ciut**ari**（8, 9, 12, 21）, anokay ut**ari**（13, 15）, anut**ari**（8-10, 12, 21）, aut**ari**（1）

39. ear...kis**ara**（14, 16, 17. 19）：kis**aru**（15, 18）

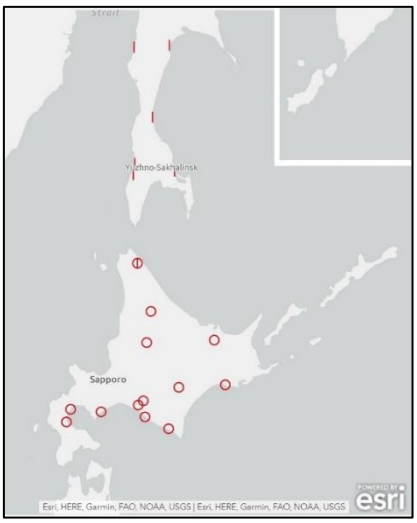
42. mouth...c**ara**（14, 16, 17）：p**aro**（1-6, 11, 21）：c**aro**（7-10, 12, 13, 20）：c**aru**（15, 18, 19, 20）

５．本研究の結果と考察

本研究の統計学的考察と結果については Ono & Fukazawa（2026）で述べるため、本稿では規則性の情報をデータに入れることで大きく値が変化した宗谷方言に着目し、本研究による新たなデータによって何が見えるのかということを述べたい。

宗谷方言は、北海道方言と樺太方言の境界の方言と考えられる。例えば、「火（82. fire）」の項目について、宗谷方言は北海道方言で見られる ape という語形、そして樺太方言で見られる unci という語形の二つを持っている（図 13）。このように、宗谷方言は、北海道と樺太の両方の語形を持つということもあれば、語彙項目によっては北海道方言の語形のみを持つことや、樺太方言の語形のみを持つこともある（境界の方言の詳細は深澤・小野（2024）を参照。）

図 14 は、本研究のデータに対して改良した等質性分析にかけ、方言間の関係性を 3 次元空間で可視化したもので、それを 2 次元目からみた図である（等質性分析については、Ono & Fukazawa（2026）で詳しく述べる）。2 次元目から見た場合の配置で特徴的なのは、宗谷よりもマイナス（左）側に樺太方言が位置し、宗谷よりもプラス（右）側に北海道方言が位置することである。



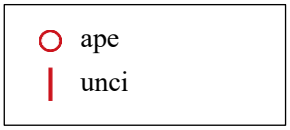


図 13：「火（82. fire）」の分布図
（深澤・小野 2024）

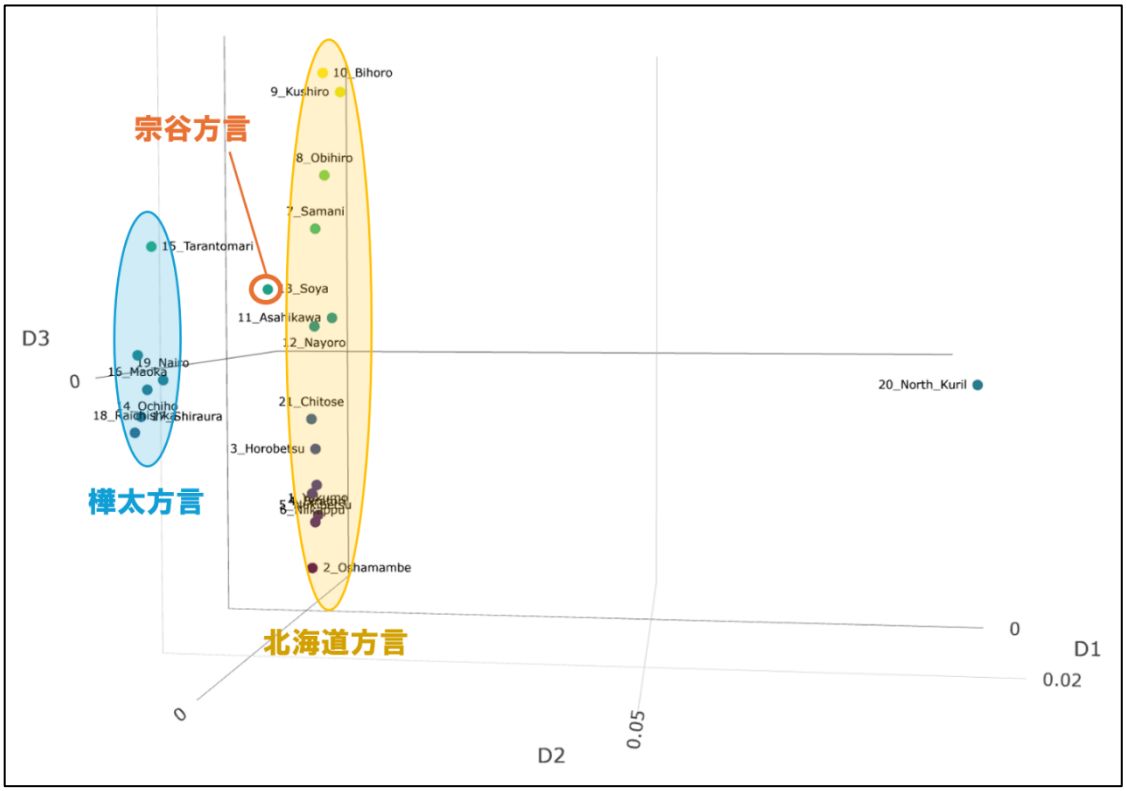


図 14：宗谷方言とその他の方言の関係（3 次元空間上で 2 次元目から見た場合）

図 15 は、語形データと規則性データのどの数値が、各次元の方言配置に影響を与えているかを捉えたグラフである。各項目と規則性について宗谷方言を 2 次元目からみたものであり、青色は語形のみのデータの数値、赤色は語形データに規則性データを加えた際の数値を示す。よって、青色の数値と赤色の数値を比べることにより、どのように 2 次元目で配置の変化があったかということを考察することができる。

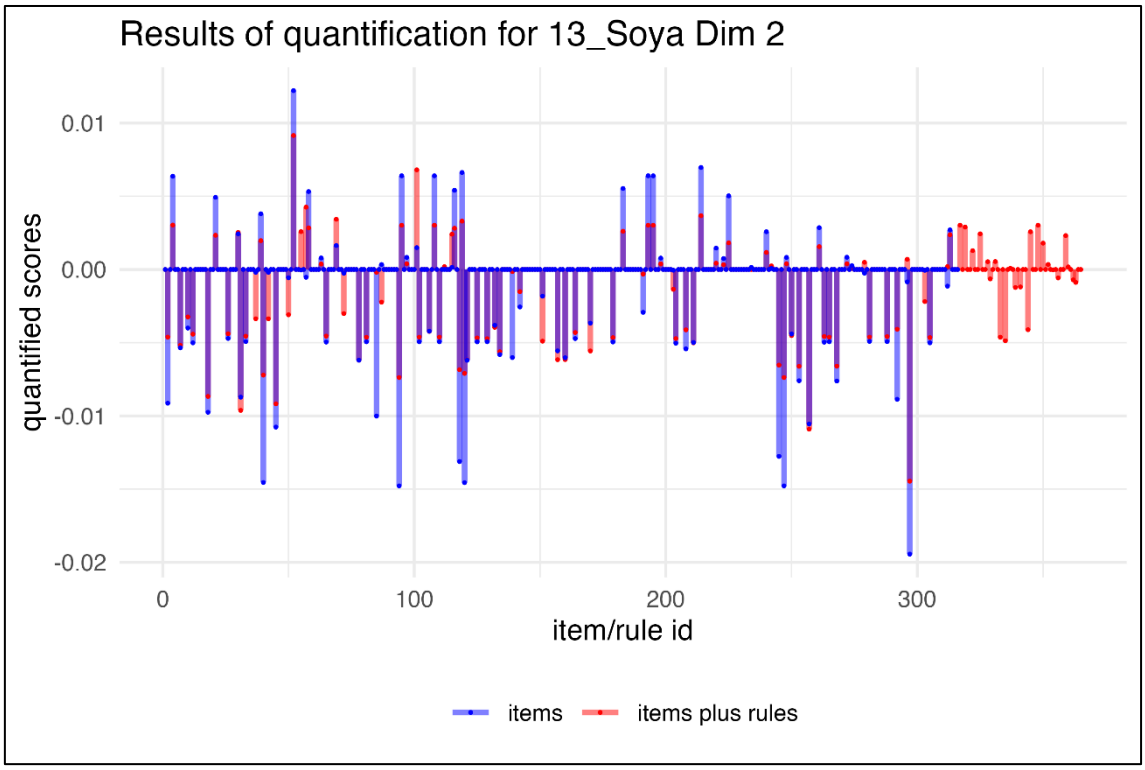
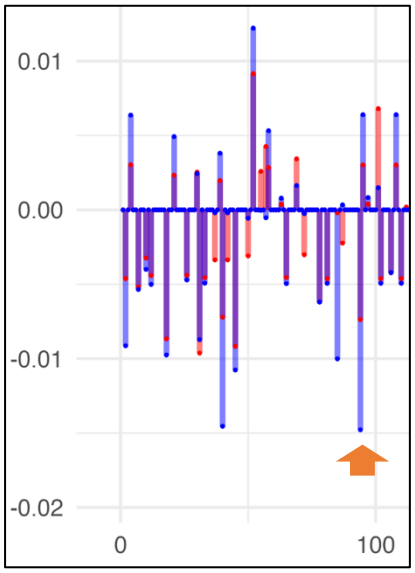


図 15：各次元に影響を与えている語形データと規則性データの数値
（宗谷方言での 2 次元目からみた可視化）

具体的に二つの項目の事例を見ていきたい。まず一つ目は「立つ」（No. 69）である。図 16 で矢印で示された部分のグラフを見ると語形データのみの数値よりも、語形データに規則性データを加えた数値のほうがプラスに変化している。「立つ」（No. 69）は、宗谷以外の北海道方言と北千島方言では as、宗谷方言と樺太方言で etaras という語形を持つ。つまり、宗谷方言は樺太方言と同じ語形を持っていることになるが、規則性データを入れることで数値がプラスに変化したということは、北海道方言が配置される方向に変化があったということになる。

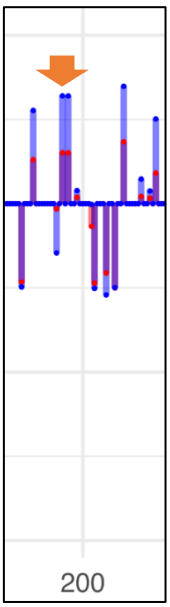
二つ目の事例は「川」（No. 121）である。図 17 の矢印で示された部分のグラフを見ると語形データのみの数値よりも、語形データに規則性データを加えた数値のほうがマイナスに変化している。「川」（No. 121）は、宗谷を含む北海道方言と北千島方言では pet、樺太方言では nay という語形を持つ。つまり、宗谷方言は北海道方言と同じ語形を持っていることになるが、規則性データを入れることで数値がマイナスに変化したということは、樺太方言が配置される方向に変化があったということになる。



「立つ」（69. stand）
「語形」データのみ：-0.01478122
「語形＋規則性」データ：-0.00737815

語形（番号は図 1 の方言番号と対応、宗谷方言は 13）：
as (1-12, 20, 21)
etaras (13-19)

図 16：「立つ」（69. stand）：宗谷方言の変化（2 次元目）



「川」（121. river）
「語形」データのみ：0.006492286
「語形＋規則性」データ：0.003026545

語形（番号は図 1 の方言番号と対応、宗谷方言は 13）：
pet (1-13, 20, 21)
nay (14-19)

図 17：「川」（121. river）：宗谷方言の変化（2 次元目）

以上のことから、宗谷方言は語形データに規則性データを入れることで値が変化し、北海道方言らしさと樺太方言らしさの両方を兼ね備えた境界の方言の特徴を持つようになったと言える。宗谷方言は、北海道方言のなかで地理的に最も樺太に近い。改めて語形と規則性の複合的なデータが示す図 14 を見ると、宗谷方言が北海道と樺太の中間的な位置に配置されている。宗谷方言が規則性データに影響を受けたことで、北海道方言と樺太方言の二つのグループを結び付け、これらの方言グループの系統関係を表しているかのようにも見える[10]。

地理言語学では、語形の地図と音韻の地図がそれぞれ分けて作成されるが、方言境界など量的に方言特徴を探る際には語形地図の重ね合わせが主となっている。また、比較言語学は

[10] 語形データのみの場合と、語形データと規則性データの複合的なデータの場合によって見られるアイヌ語方言の関係性の異なりについては、Ono & Fukazawa（2026）で詳しく述べるので、そちらを参照されたい。

音対応から言語や方言の歴史的な関係性を見るが、地理言語学のように語彙項目自体の歴史には目を向けないこともあり、統計手法に用いられるような量的な可視化はされにくい。Fukazawa（2025）では、地理言語学において語形地図と音韻地図を組み合わせた関係地図を作成することを提案したが、質的に捉えることはできても量的に捉えることが難しかった。本研究では語形と規則性という複合的なデータによって基礎語彙の情報を捉えなおし、統計分析にかけることでデータを量的に捉えることを試みたものである。これによって、規則性と語形の両方の情報が方言間の歴史にどのように影響を与えているのかなど、新たな考察の可能性も広がったと言えるが、歴史的な解釈はより慎重に議論すべきところであるため今後の課題としたい。

謝辞

本研究は、JSPS 科研費「アイヌ語の地理言語学・統計学的研究：近世アイヌ語資料の方言分析」（JP25K04114、代表者：深澤美香）の助成を受けたものである。日本北方言語学会第 8 回大会の口頭発表における質疑および本稿の 2 名の査読者には、有益な助言や誤りの訂正など細部にわたってコメントを頂いた。ここに記して御礼申し上げる。

参考文献

小野洋平・深澤美香（2023）「アイヌ語諸方言の語形の類似に関する基礎データの復元：論文に書ききれなかった研究者の判断・思考に迫る」『北方言語研究』13: 213-246.

小野洋平・深澤美香（2024）「比較不可能だったアイヌ語方言分類 : 統計的方言分類を類似判断の点から再考する」『アイヌ・先住民研究』4: 93-126.

金田一京助（1931）「アイヌユーカラ語法摘要」『アイヌ叙事詩ユーカラの研究』2. 東洋文庫.

金田一京助（1993 [1937]）「山間のアイヌ語」『山岳』33(1). 再収録 :『金田一京助全集』6: 255-260. 三省堂.

小松和弘（2004）『アイヌ語様似方言辞典』改訂版. 北海道ウタリ協会様似支部様似アイヌ語教室.

田村すず子（1996）『アイヌ語沙流方言辞典』草風館.

知里真志保（1941）「アイヌ語法研究：樺太方言を中心として」（再録 :（1973［1993］）『知里真志保著作集』3: 455-594. 平凡社.）

知里真志保（2001 [1952]）「マタギの山言葉とアイヌ語」. 再収録 :『アイヌ語考』3: 1-3. ゆまに書房.

中川裕（1995）『アイヌ語千歳方言辞典』草風館.

服部四郎（編）（1964）『アイヌ語方言辞典』岩波書店.

服部四郎（1999）『日本語の系統』岩波文庫.

服部四郎・知里真志保（1960）「アイヌ語諸方言の基礎語彙統計学的研究」『民族學研究』24(4): 307-342.

深澤美香（2017）『加賀家文書におけるアイヌ語の文献学的研究』千葉大学.

深澤美香（2025）「〔資料紹介〕国立アイヌ民族博物館所蔵知里真志保記入のアイヌ語基礎語彙調査表」『国立アイヌ民族博物館研究紀要』3：136–158.
深澤美香・小野洋平（2024）「方言境界再考：アイヌ語の方言境界を例として」『北方言語研究』14: 155-176.
深澤美香・小野洋平（2025）「基礎語彙による方言分類の諸問題：Asai（1974）のアイヌ語方言データを最大限活用するために」『北方言語研究』15: 141-163.
Alexander, Vovin (1997) *A Reconstruction of Proto-Ainu*. Leiden, New York, Köln: E.J. Brill.
Asai, Toru (1974) Classification of dialects: Cluster analysis of Ainu dialects. *Bulletin of the Institute for the Study of North Eurasian Cultures, Hokkaido University* 8: 45-136.
Fukazawa, Mika (2025) Extraction of regularities and geographical patterns from the basic vocabulary of the Ainu language, In. Jalaluddin Nor Hashimah, Suzuki Hiroyuki & Mitsuaki Endo (eds.) (2025) *Proceedings of the sixth International Conference on Asian Geolinguistics*. 62-80.
Ono, Yohei and Mika Fukazawa (2025) Statistical approaches to the quantification of regularities in languages and dialects: An exercise in Ainu dialects of Asai (1974). *Northern language studies,* 15: 179-202.
Ono, Yohei and Mika Fukazawa (2026) From Classification to History: Examining New Methodologies in Lexicostatistics Using the Data in Ainu Dialects. *Northern language studies,* 16: 121-150.

Building Composite Data on Lexical Forms and Regularities: Extracting Maximum Information from the Basic Vocabulary of Ainu Dialects

Mika FUKAZAWA
（National Ainu Museum）
Yohei ONO
（St. Luke's International University）

Keywords: Ainu language, philology, feature engineering, comparative linguistics, beyond lexicostatistics

This study aims to appropriately quantify the basic vocabulary of Ainu dialects using linguistic and statistical methods. Previous statistical studies of basic vocabulary have relied on manual cognacy judgments conducted by linguistic or statistical researchers, who applied their own criteria to the lexical forms within each vocabulary item. Thus, the statistical results depend on their judgments based on grammatical, lexical, morphological, or phonological differences. Representative studies, such as those by Hattori & Chiri (1960) and Asai (1974), concluded with different results of grouping of Ainu dialects (Ono & Fukazawa 2024).

In this study, data preparation for statistical research proceeded as follows: Multiple judgments and criteria were set within a single vocabulary item, enabling the construction of two datasets: one of lexical forms (item data) and one of phonological rules (regularity data). This approach was proposed and developed by Fukazawa & Ono (2025) and Ono & Fukazawa (2025). Furthermore, subtracting the phonological regularity data from the lexical item data avoided the duplication of linguistic information.

This study attempts to maximize the available information on basic vocabulary by reorganizing dialect data into two feature sets, and generating new data. Lexical information was treated as composite data consisting of lexical forms and phonological regularities and was examined through statistical analysis. This approach may also open new avenues, such as how the interaction between lexical forms and phonological rules imforms both the historical development and description accounts of dialects. Ono & Fukazawa (2026) presented the results of applying statistical analysis to the data produced in this study.

（ふかざわ・みか　mk.fukazawa@hotmail.co.jp）
（おの・ようへい　mathematical.humanities@hotmail.com）