



HOKKAIDO UNIVERSITY

Title	服部・知里(1960)の統計科学的再考察 : アイヌ語方言周圏論の実証
Author(s)	小野, 洋平
Citation	北方人文研究, 8, 25-41
Issue Date	2015-03-31
Doc URL	https://hdl.handle.net/2115/58444
Type	departmental bulletin paper
File Information	08_02_ono.pdf



服部・知里(1960)の統計科学的再考察 —アイヌ語方言圏論の実証—

小野洋平

総合研究大学院大学複合科学研究科統計科学専攻

1. はじめに

服部・知里(1960)は昭和 30 年から 31 年にかけて、北海道 13 地点、樺太 6 地点の計 19 地点のアイヌ語の方言について、大規模な調査を行った結果である。

日本における、基礎語彙統計学的研究の嚆矢であり、また彼らの調査の後、多くのアイヌ語の話者がなくなっていることから、アイヌ語の資料の保存という観点からも記念碑的な論文である。

服部・知里(1960)では、得られたデータの類似度などを定義し、方言同士の関係について推測を行っているが、残念ながら、計算機の性能などに関する当時の時代的制約上、適切な統計処理を行える環境になかった。

さらに、服部・知里(1960)のデータに基づき、Asai(1974)はアイヌ語の方言に関するクラスター分析をおこなった。しかし、Asai(1974)は統計分析を行える環境にあったが、データの扱い方が統計学的には問題があるように思われる。

さらに、近年の Lee and Hasegawa(2013)の研究でも、服部・知里(1960)のデータをもとに、ベイズ法を用いた系統樹推定を行っているが、やはりデータの扱い方が適切でないように思われる。

本稿では、服部・知里のデータを適切に扱い統計分析を行い、言語地理学的な観点からより妥当な結果が得られることを示す。特に、樺太の 6 つの方言についてより妥当の結果が得られた。

さらに、服部・知里(1960)のデータを統計学的に妥当に処理したデータを、著者が独自に提案する MCANeighbor-Net を適用することによって、今までのクラスター分析では見落とされていたアイヌ語における方言圏論的構造を明らかにすることに成功した。

今後は、本稿で提案した方法の有効性と、適用範囲の限界について調べていきたい。

2. 先行研究

2.1 先行研究の概要

2.1.1 服部・知里(1960)

服部・知里(1960)は Swadesh(1955)の基礎語彙 200 語をもとに独自の言語学的な知見をもって実際に調査する語を選択し、19 の地点の語彙の同根性について、表 1 のような基準ののっとり、データを作っている。

この表から、服部・知里(1960)は以下のような手続きで、方言間の類似度を求めている。

1. + は 1 とカウントする
2. ± は + の部分を 0.5 と、- の部分を 0.5 と分解する

3. -は1とカウントする

4. それ以外の記号は無視する

そして、方言間の語彙に関して、+, ±, -の個数の合計に対して、+と±の+(0.5とカウントする)の割合(%)を方言の類似度として用いる。

架空例: 八雲と長万部について

+が60個、±が30個、-が10個の場合。

分子=60×1+30×0.5=75

分母=70×1+30×1+10×1=110

よって八雲と長万部との類似度は75/110=68.2%

表1: 服部・知里(1960)での語彙の同根性に関する基準。服部・知里(1960: 309)より作成

+	語根が互に対応し合うもの
—	語根が互に対応しないもの
±	比較さるべき形態素が、一方または両方の方言に二つ(以上)あり、そのうち一つどうしの関係が+で、他方の関係が—であると認められる場合。被調査者が、問題の二つ(以上)の単語の使用法の違いについて詳しく且つ正確な報告ができ、その結果意義素が正確に記述されている場合には、そのうちどちらかを採るべきか明らかなのが普通だが、主としてこれらの条件が欠けたために、ただ二つ(以上)の単語が並べ報告されている場合には、±という記号をつけるという処置をとらざるを得なかった。
○	対応関係の不明な場合、または比較のためにいずれの形態素をとるべきか、不明の場合。
?	調査結果に疑問のある場合
•	一方の方言に、該当の単語のない場合、被調査者が答えられなかった場合
()	調査漏れ

このような計算を各方言間に行うと、表2が得られる。さらに、表2の値は類似度であるため、100から表2の値をひいたものを非類似度(方言間の距離)と定義し、クラスター分析(4章参照)にかけた結果が図1(クラスター分析には最長距離法を使用)である。

図1の結果と、服部・知里(1960)の表2の解釈とを比較すると、

1. 「北海道方言とカラフト方言との間には大きな断層がある」という服部・知里(1960)の見解は図1でも支持された
2. しかし、「宗谷方言は、他の北海道方言から比較的遠く、且つカラフト方言に最も

3. さらに、樺太西海岸にある多蘭泊(15)と真岡(16)がまとまり、樺太東海岸の落帆(14)、白浦(17)と内路(19)と樺太西海岸の北にあるライチシカ(19)がまとまるのは、樺太が南北に山脈が走り、交通も海岸沿いを中心としていることから、地理的に疑問が残る。地理的には、西海岸の 15,16,18 と東海岸の 14,17,19 がまとまるべきである。この 3 点が挙げられる。

[illegible]

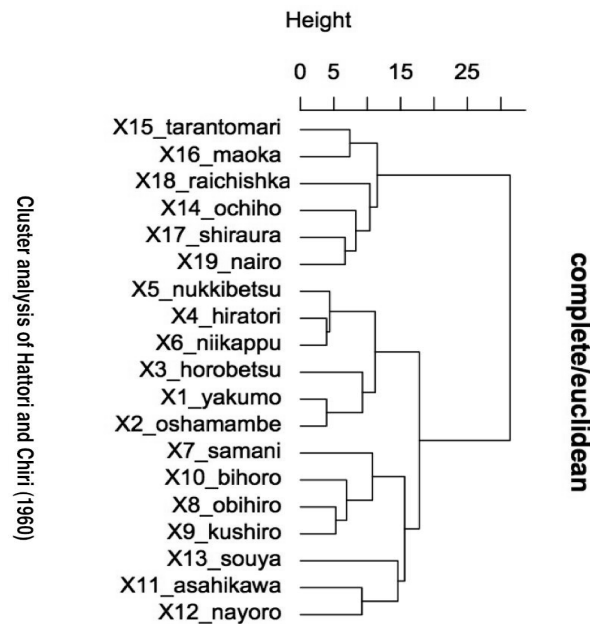


図1: 表2の類似度から非類似度を計算し、最長距離法に基づき、クラスター分析を行った結果。

2.1.2 Asai (1974)

Asai(1974)の研究では、服部・知里(1960)のデータに千島(Kuril)と千歳(Chitose)のデータを追加して分析している。

Asai(1974)では、服部・知里(1960)とは方言の類似度の定義がことなる。服部(1960)では「±」の「+」に関しては0.5の重み付けがなされ、普通の「+」については1.0の重み付けがなされていたが、Asai(1974)では、どちらも1.0の重み付けがなされている、それ以外の点は服部・知里(1960)と同じである。

Asai(1974)の手法と同じ方法でデータを、クラスター分析にかけた結果が図2である。Asai(1974)から転用した図3を並べてみると以下のようなことがいえる。

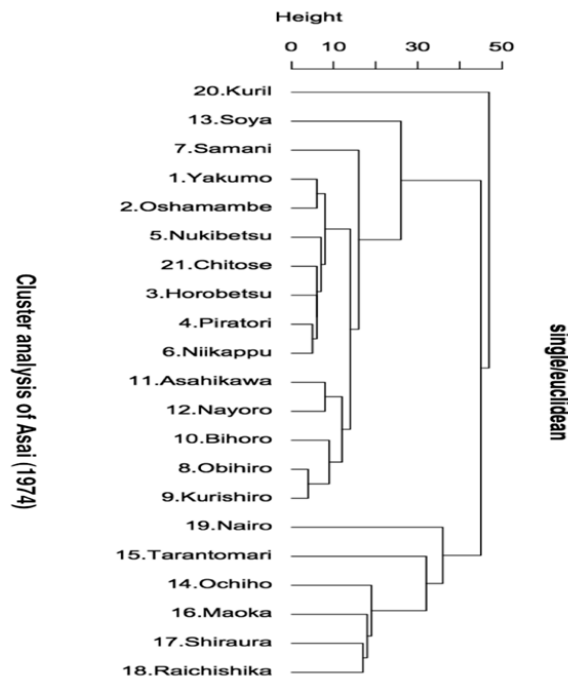


図 2: Asai (1974) の手法でデータをクラスター分析にかけた結果。クラスター分析には Asai (1974) と同じ最短距離法を使った。北海道方言が北東と南西にわかれている。宗谷方言が北海道方言の中ではもっとも樺太方言に近くなっている。

1. 服部・知里(1960)の結果である図 1 と同じように、北海道方言と樺太方言との間には大きな隔たりがあることがわかる。
2. 服部・知里(1960)の結果と一貫している点として、北海道方言は大きく北東方言と南西方言に分かれることが確認できる。
3. 図 1 からは支持されなかったが、服部・知里(1960)の考察にある「宗谷方言は、他の北海道方言から比較的遠く、且つカラフト方言に最も近い北海道方言である」ということが図 2 からは支持された。
4. しかし、Asai(1974)では、多蘭泊(15)を南サハリン方言、落帆(14)、真岡(16)、白浦(17)、ライチシカ(18)を中央サハリン方言、内路(19)を北サハリン方言と考えているが、これもサハリンには南北に山脈があることなどから地理的に疑問が残る。

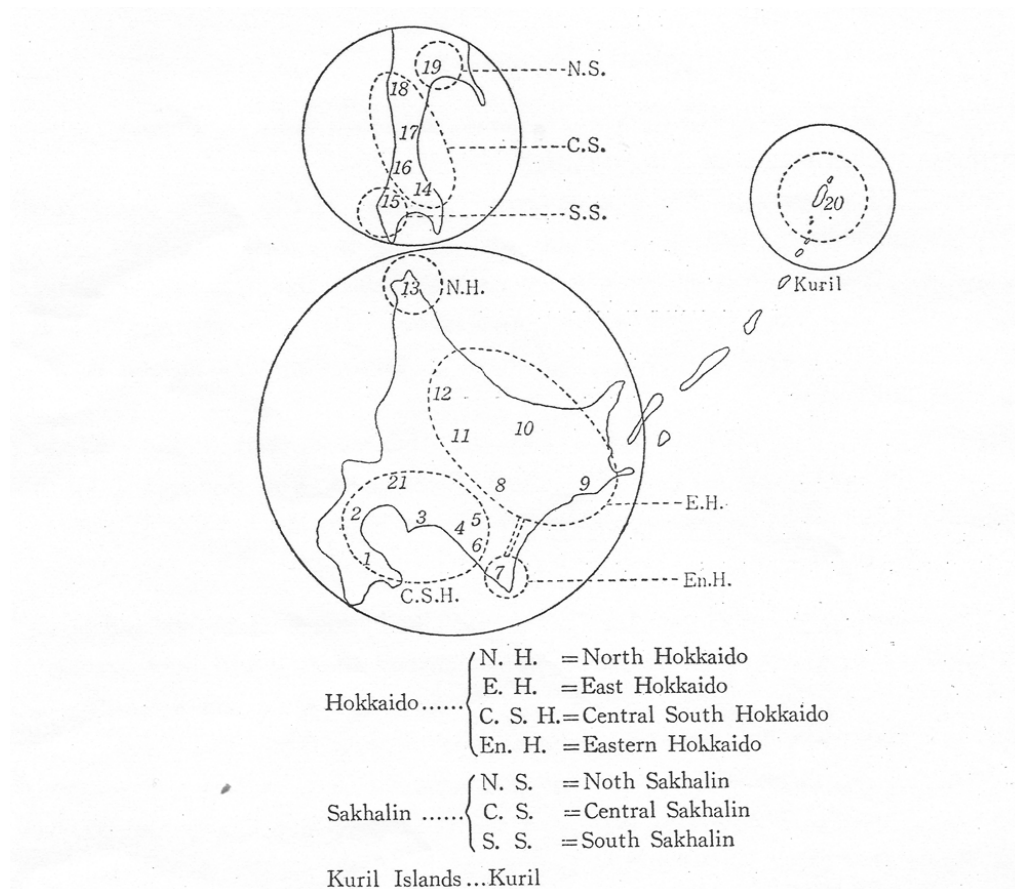


図3: Asai (1974)のクラスター分析の結果を地図上に示したもの。Asai (1974: 100)

2.1.3 Lee and Hasegawa (2013)

服部・知里(1960)のデータに基づき、ベイズ法による系統樹推定法を行ったのが、Lee and Hasegawa である¹。ベイズ法による系統樹推定法については詳細についてはここでは立ち入らないが、Lemmon and Lemmon(2008)や Drummond, Suchard, Xie and Raumbaut(2012)が参考になる。

Lee and Hasegawa(2013)の結果を図4に示す。

1. Asai (1974)と同様に、北海道方言に関する分類は概ねよい
2. ただし、サハリン方言の分類は図1の結果と同様であるが、樺太西海岸にある多蘭泊(15)と真岡(16)がまとまり、樺太東海岸の落帆(14)、白浦(17)と内路(19)と樺太西海岸の北にあるライチシカ(19)がまとまるのは、樺太が南北に山脈が走り、交通も海岸沿いを中心としていることから疑問が残る。Lee and Hasegawa(2013)の結果は、距離的に離れた方言同士の関係が密であるとしている点で興味深い、これを説明する言語外要因については何も触れるところがなく、説得力に欠ける。

¹ 系統樹推定法とクラスター分析法は異なった考えに基づいた別の統計手法であり、解釈には注意が必要であることを留意する必要がある。

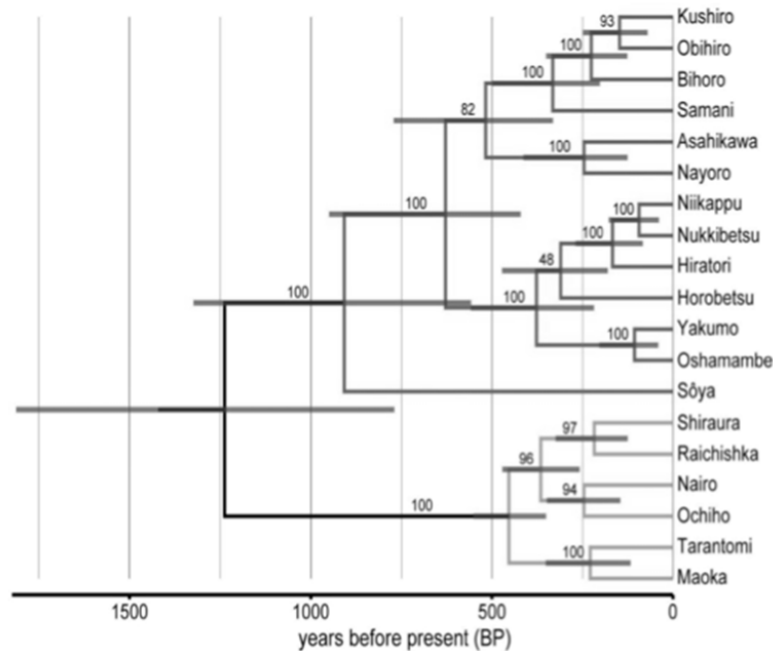


図 4: Lee and Hasegawa (2013) のベイズ法による系統樹推定の結果。Lee and Hasegawa (2013: 3)

2.2 先行研究の問題点

服部・知里(1960)、Asai (1974)も Lee and Hasegawa (2013)もサハリン方言に関して、地理的妥当性のある結果を得ることができなかった。

これには、服部・知里(1960)の「o, ?, •」という所謂、同根性が「わからなかった」というカテゴリーを分析では無視していることが問題ではないかと考えた。

つまり、調査時点で語根の対応が「わからない」ほど、A 方言の語彙 C と B 方言の語彙 C' について「o, ?, •」であることはそれだけ A 方言と B 方言とが「遠い」という情報を表していると考えられる。

「o, ?, •」を無視することは、有用な情報をあえて無視していることとなっており、「o, ?, •」を一つのカテゴリーと捉え分析に加えることが必要である。

3. 目的

服部・知里(1960)における +, ±, −, [o, ?, •] の 4 カテゴリーを統計的に適切に扱い、アイヌ語の方言の分類を行うことを目的とする。

ただし、統計科学的な知見からは、+, ±, −, [o, ?, •] を同時に扱うことは「尺度水準」の問題が生じる。

なぜならば、もし +, ±, − だけを扱えばよいのなら、

1. + > ± > − という順序関係は認めることができるだろう(順序尺度)
2. さらには、服部・知里(1960)のように + に 1.0 の重み付け、± の + に 0.5 の重み付けをすることで加減乗除まで仮定することも可能かもしれない(間隔尺度)

しかし、 $+, \pm, -, [o, ?, \bullet]$ を同時に扱おうとすると、

1. $+, \pm, - \rightarrow [o, ?, \bullet]$ という最低限の順序尺度すら成立しない
2. このような時には、犬(1)、猫(2)、猿(3)といった便宜的に値を付けるだけの「名義尺度」しか成立しない

名義尺度のデータを分析する手段の一つが、林の数量化3類である(Hayashi's Quantification Method III)(Hayashi, 1951)。

4. 方法(林の数量化3類とクラスター分析)

林の数量化3類は、数学的に同等な方法が世界各地で「再発見」されており、さまざまな呼称がある。ヨーロッパでは多重対応分析(Multiple Correspondence Analysis)(Benzécri, 1973)、カナダのトロント大学の西里静彦によって双対尺度法(Dual Scaling)(Nishisato, 1980)、オランダのグループ Gifi によって等質性分析(Homogeneity Analysis)(Gifi, 1990)などと呼ばれている。

数量化3類の役割は、主に2つある。

第1に、名義尺度のデータにある基準によった適切な「数値」を与えることである。第2に、適切な数値を与えられたデータの構造の中から「核」となる部分の情報を効率よく抽出することである。

本研究では、数量化3類によって抽出されたデータがある基準に従ってグルーピングする。この方法をクラスター分析という。クラスター分析とは、最も素朴な理解としては、「似たもの同士をまとめていく」ということである。

クラスター分析には様々な手法があるが、詳細については、Everitt, Landau, Leese and Stahl (2011)が参考になる。

データの分析に際して、このデータは語彙×地域×地域の3相データであるため、19の地域に関する語彙×地域の2相データに分割し、このデータに対して数量化3類をかけた。データの固有値のプロットを図5に示す。足立・村上(2011)に従い、データの固有値のプロットの減少が急に緩やかになる前までの次元を採用した。それぞれの19の地域に関するデータに数量化3類を適用し、固有値のプロットから次元を選択肢、選択した次元までの座標をもとに19のデータそれぞれでユークリッド距離に基づいた距離行列を計算した。

最終的には、19のデータから得られた19の距離行列を平均した距離行列をクラスター分析にかけることとした。

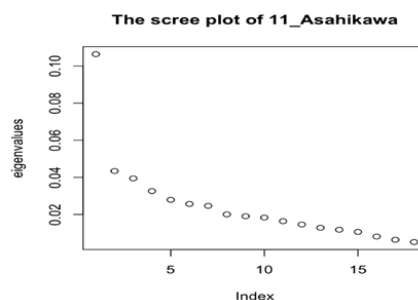


図5: 旭川のデータの固有値のプロット。3次元目までのデータを採用することにした。

5. 結果

クラスター分析の結果、図 6 を得た。

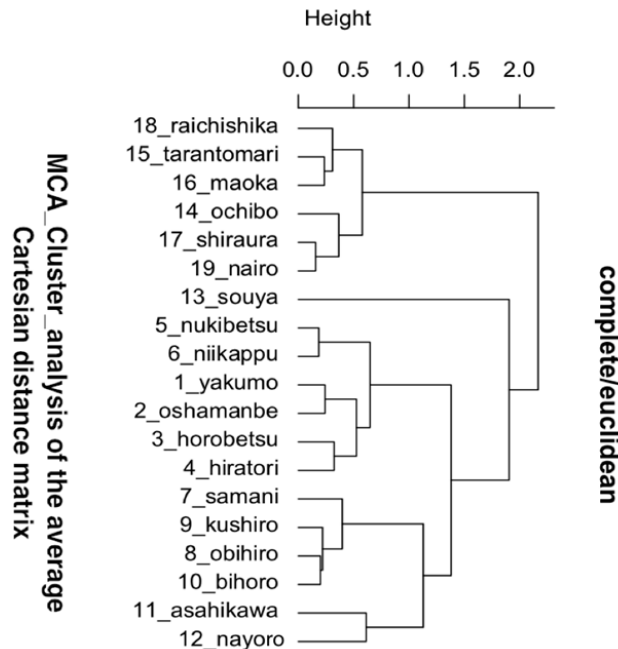


図 6: 19 のデータから得られた 19 の距離行列を平均した距離行列をクラスター分析にかけた結果 (クラスタリングには最長距離法を用いた)。宗谷が北海道方言の中で最もサハリン方言に近く、北海道方言は北東と南西に分かれる。さらに、サハリン方言が西海岸と東海岸で分かれている。これは言語地理学的な知見と一致する。

1. 北海道の結果は Asai (1974) と一貫して、宗谷が北海道方言の中では最も樺太方言に近く、北海道方言の中では北東方言と南西方言にわかれるという傾向がある。
2. サハリン方言が、西海岸の方言 (多蘭泊 (15)、真岡 (16)、ライチシカ (18)) と東海岸の方言 (落帆 (14)、白浦 (17)、内路 (19)) とに分かれ、樺太には南北に山脈が走り、交通は海岸沿いを基本としていたという地理学的な知見と整合している。

サハリン方言が、なぜ西海岸と東海岸で分かれることが妥当かということに関しては、Ono (2014) において文化人類学的、文献学的観点から詳細な議論がなされているので、参照されたい。本稿の目的は、本章でクラスター分析によって言語地理学的に妥当な分類をあたえたアイヌ語方言の距離行列に、Neighbor-NetAnalysis を適用することで、アイヌ語における方言周囲論的構造を統計科学的に考察することであるため、先に進むこととする。

6. Neighbor-NetAnalysis による分析

6.1 Neighbor-NetAnalysis とは

点 A、B、C、D が $A(0,2), B(0,0), C(1,0), D(1,2)$ のような位置にあったとする。伝統的

なクラスター分析の考え方では、AとD、BとCとが近いため、それらがまずまとまる。最後にAとDのまとまりと、BとCとのまとまりが、まとまる。結果、図7のようなクラスターが描かれる。

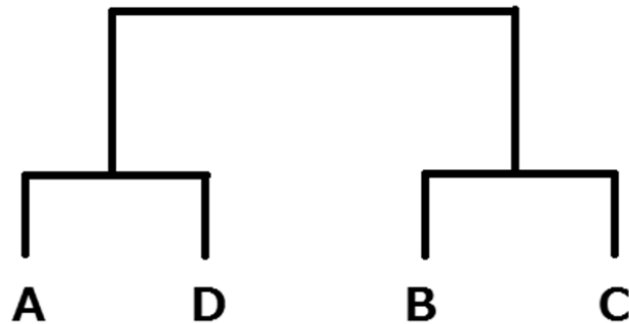


図7: A, B, C, Dに関する「強い」構造。AとDとがまとまり、BとCとがまとまり、最後に、二つのまとまりがまとまる。

しかし、実際には、AとBもAとDほどではないが、近い。またCとDもBとCほどではないが近い。よって、AとBとがまとまり、CとDとがまとまり、最後に、AとBのまとまりと、CとDとのまとまりが、まとまる図8のようなクラスターも考えることができる。

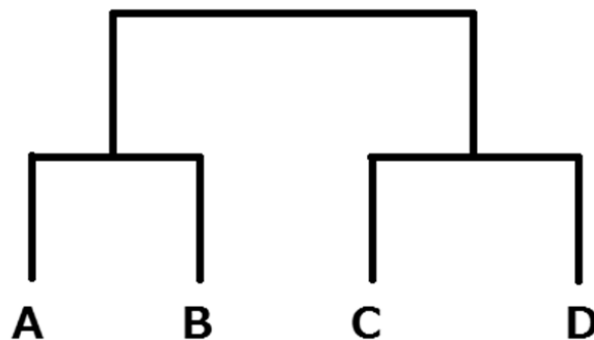


図8: A, B, C, Dに関する「弱い」構造。AとBとがまとまり、CとDとがまとまり、最後に、二つのまとまりがまとまる。

しかし、伝統的なクラスター分析では、「強い」構造である図7が選択され、「強い」構造である図7と「弱い」構造である図8を「同時に」表現することはできない。

つまり、クラスター分析で描かれる構造というのは、データの持つ「最も強い」構造なのである。

そこで、Neighbor-NetAnalysisでは、点A、B、C、Dの関係を図9のように表現する。四角形をたてに沿って切れば、AとBのまとまりと、CとDとのまとまり(先述した「弱い」構造)が表現され、横に切ればAとDのまとまりと、BとCとのまとまり(先述した「強い」構造)が表現される。四角形の辺の長さが近ければ近いほど、辺の両端の関

係は強いと考えてよい。逆に、辺の長さが長ければ長いほど、辺の両端の関係は弱いといえる。

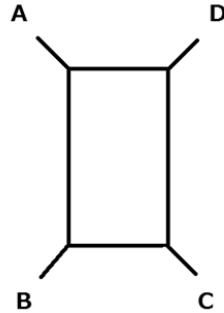


図 9: A, B, C, D に Neighbor-NetAnalysis を適用した結果。横に切ることによって、図 7 の A, B, C, D に関する「強い」構造を表現し縦に切ることによって「弱い」構造が表現できる。

本稿では、第 5 章の結果に用いた距離行列を、Neighbor-NetAnalysis にかける。今まで、服部・知里(1960)のデータに基づいた研究はいずれも、データの「もっとも強い構造」が取り出されるクラスター分析を用いていた。しかし、Neighbor-NetAnalysis を適用することによって、クラスター分析によって隠されていた「弱い」構造が明らかになると期待できる。本稿では、服部・知里(1960)のようなカテゴリカルデータ(名義尺度)に対して Neighbor-NetAnalysis を適用する新たな方法を MCANeighbor-Net と命名する。

Neighbor-NetAnalysis の詳細については、Huson and Bryant (2006)が参考になる。

6.2 MCANeighbor-Net を適用した結果と考察

第 5 章において、クラスター分析を適用した距離行列に、Neighbor-NetAnalysis を適用した。結果、図 11 を得た。また、参考に図 12 を載せた。

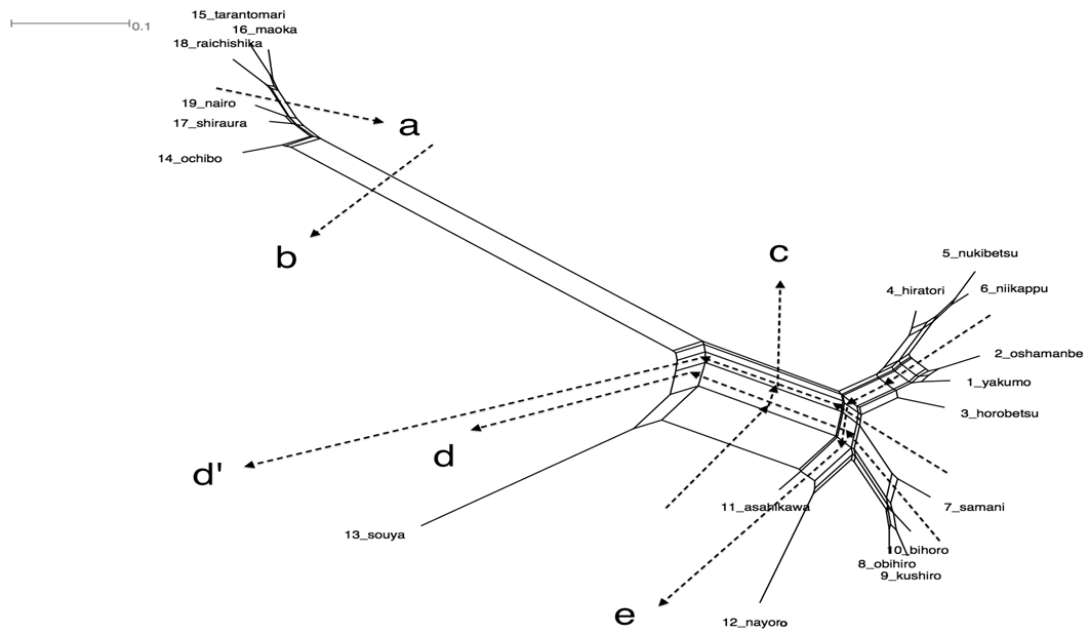


図 11: 数量化 3 類の結果得られた 3 次元目までの座標に基づいた 19 の距離行列を平均した距離行列に Neighbor-NetAnalysis を適用した結果

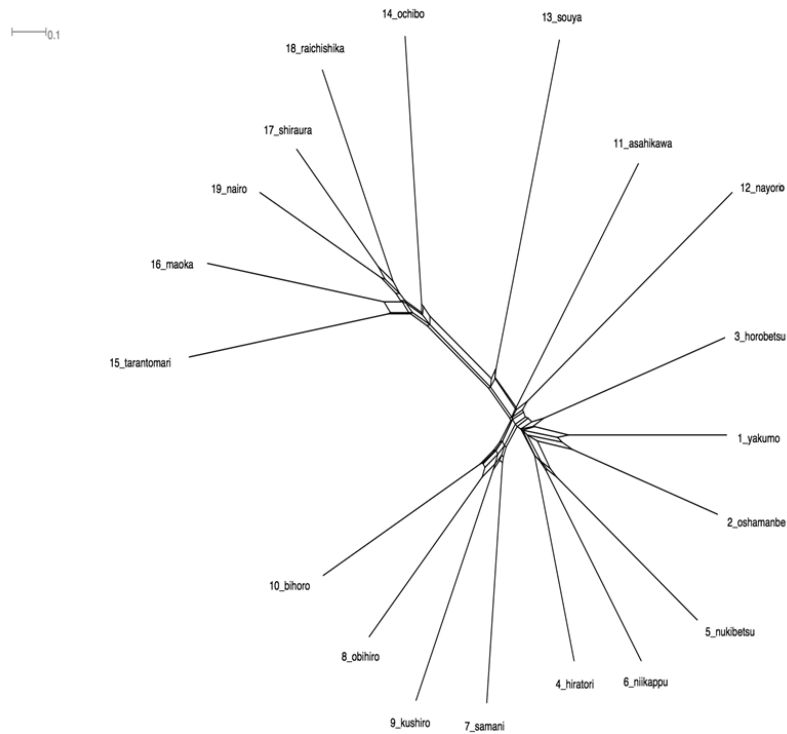


図 12: 数量化 3 類の結果得られた全次元 (18 次元) までの座標に基づいた 19 の距離行列を平均した距離行列に Neighbor-NetAnalysis を適用した結果

図 12 をみれば分かるように、次元数が多くなると、Neighbor-Net の結果における枝が長くなっていることが分かる。このことは、距離行列の共通の部分の情報を取り出し、データにおける本質的な類似の部分を示す Neighbor-Net の平行四辺形の部分が少なくなっていることを意味し、6.2.1 以下で行われる「どこで図を切るべきか」という議論がしにくい。

また、図 11 と比較して図 12 において「枝が長くなっている」ことは、図 11 に対して、図 12 の方が、距離行列がよりノイズを取り込んでいると考えられる。

数量化 3 類の結果、どこまでの次元の座標を用いるかということは、「最終的な解決」は本稿では提案できないが、図 11 と図 12 を比較した結果、本稿では 3 次元目までの結果に基づいた図 11 を中心に図の解釈を行うこととした。

図 11 は a, b, c, d, d', e の線で分割し解釈することが可能である。

6.2.1 a, b による解釈

a, b の線で、図 11 を分割することにより、まず第 1 に北海道方言と樺太方言の断絶が非常に大きいという従来の知見が支持される。また、樺太のアイヌ語の方言が樺太西岸方言(15, 16, 18)と樺太東岸方言(14, 17, 19)とに分かれていることは、第 5 章や、Ono (2014)の結果を支持している。さらに、(14, 17, 19)がより北海道方言に近いことから、樺太のアイヌ語においては西岸方言より東岸方言の方が、北海道方言に近いことが示唆される。この知見については、語彙だけではなく文法など総合的な面から今後研究がなされることが期待される。

6.2.2 c による解釈

c の線で、図 11 を分割することにより、宗谷方言は北海道方言の中でも最も樺太方言に近いことが示される。これは、服部・知里(1960)でも示唆されたことであり、その後の Asai(1974)などの研究も支持する結果である。

6.2.3 b と d, d' による解釈

b と d の線で、図 11 を分割することにより、北海道方言は北東のグループと南西のグループに合致する。この知見は、服部(1964)の分類や、Asai(1974)のクラスター分析の結果と合致する。先行研究を検討した第 2 章の図 1、図 3、図 4 とともに概ね一致しているため、北海道方言を北東のグループと南西のグループに分けることは服部・知里(1960)のデータの「もっとも強い」構造と考えられる。ただし、b と d の線で図 11 を分割することは、様似方言(7)を南西のグループに分類することになる。Asai (1974)などにおいて北東のグループにわけていることは、b と d' の線で分割することに相当する。様似方言をどのように考えるかについては、今後の研究を待ちたい。

6.2.4 c と e による解釈

c と e の線で、図 11 を分割することによって、c で北海道方言と樺太方言とが分かれる。さらに、e では北海道方言は、平取(4)、貫別(5)、新冠(6)、旭川(11)がまとめられ、北海道南西の八雲(1)、長万部(2)、幌別(3)と、北海道北東の様似(7)、帯広(8)、釧路(9)、

美幌(10)、名寄(12)が同じグループにはいる。これを地理的に見ると平取(4)、貫別(5)、新冠(6)、旭川(11)が中心を、八雲(1)、長万部(2)、幌別(3)、様似(7)、帯広(8)、釧路(9)、美幌(10)、名寄(12)が外輪を構成していることがわかる。これは、柳田国男が『蝸牛考』で述べた「方言圏論」的考えと合致する(柳田、1930)。アイヌ語が沙流、平取を中心とした方言圏論的構造をとることは、中川(1996)においても指摘されている。

アイヌ語における方言圏論的構造は、今まで文献学的な言及はあったが、統計科学的に示されたことはなかった。服部・知里(1960)や Asai(1974)、Lee and Hasegawa(2013)などにおいても、示されたのはアイヌ語の方言を北東と南西にわけた「もっとも強い」構造である。アイヌ語の方言圏論的構造が、服部・知里(1960)のデータでは「弱い」構造であったことがその原因と考えられる。

MCANeighbor-Net によって、弱い構造が可視化されたことによって、本稿では、アイヌ語の方言圏論的構造が示された。ただし、旭川が「中心」にくるかについては、今後の議論を待ちたい。

また、e の線だけで図 11 を分割することは、中川(1996)で指摘されている「沙流・千歳・樺太型」に対応することも注目される。

c によって樺太方言と北海道方言がわかれ、さらに e によって、北海道方言の中に方言圏論的構造(沙流千歳グループとその他)が存在することが、データの統計分析により今回初めて明確になったといえる。このことは、直感的にはこれまでもあるいは気づかれていたと思われる、沙流方言とそれと関係の深い方言群がアイヌ語諸方言で持つ位置づけの大きさを客観的に示したものである。

7. 結語と今後の課題

本研究では、服部・知里(1960)のデータにおける「わからない」というカテゴリーを先行研究のように無視するのではなく、1 つの情報をもった独立したカテゴリーとして扱うことによって、樺太のアイヌ語の方言が、言語地理学的に妥当な分類となったことを示した。

さらに、その際に用いられた距離行列を「もっとも強い」構造のみを導く「クラスター分析」ではなく、「弱い」構造も把握することができる Neighbor-NetAnalysis を適用することによって、今まで文献上は言及されてきたアイヌ語における方言圏論的構造を、統計学的に示すことができた。

今後は、日本語のなどその他の言語のデータにおいても MCANeighbor-Net を適用し、その可能性と限界を探っていきたい。

統計手法は世の中にあまた存在しているが、本稿で筆者が示したように、もっとも肝心なのは、そのデータの性質をよく観察した上で、「どのような尺度水準でデータを扱うべきか」ということを十分に考慮することである。その上で、本稿で提案した MCANeighbor-Net のように、より柔軟な方法を用いることによって、今までの統計解析では死角になっていたことが、明らかになることがある。

統計科学者と言語学者との相補的な研究が、進むことを期待したい。

【参考文献】

足立 浩平・村上 隆

2010 「非計量多変量解析法」 朝倉書店.

Asai Toru

1974 Classification of dialects: Cluster analysis of Ainu dialects. *Bulletin of the Institute for the Study of North Eurasian Culture*. 8: 45-136.

Benzécri, Jean-Paul

1973 *L'Analyse des Données. Volume II. L'Analyse des Correspondances*. Dunod.

Drummond, Alexei, Suchard, Marc, Xie, Dong and Rambaut Andrew.

2012 Bayesian phylogenetics with BEAUti and BEAST 1.7. *Molecular Biology and Evolution*, 29(8): 1-4.

Everitt, Brian, Landau, Sabine, Leese, Morven and Stahl Daniel.

2011 *Cluster Analysis*. Wiley.

Gifi, Albert

1990 *Nonlinear multivariate analysis*. Wiley.

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」 『季刊民族学研究』 24(4): 307-342.

服部 四郎編

1964 「アイヌ語方言辞典」 岩波書店.

Hayashi, Chikio

1951 On the prediction of phenomena from qualitative data and the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics*, 3(1): 69-98. Springer.

Huson, Daniel and Bryant, David

2006 Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2): 254-267. Oxford.

Lee, Sean and Hasegawa Toshikazu

2013 Evolution of the Ainu Language in Space and Time. *PloS one*, 8(4), e62243.

Lemmon, Alan and Lemmon Emily Moriarty

2008 A likelihood framework for estimating phylogeographic history on a continuous landscape. *System Biology*, 57: 544-561.

中川 裕

1996 「言語地理学によるアイヌ語の史的研究」 『北海道立アイヌ民族文化研究センター研究紀要』 2: 1-17.

Nishisato, Shizuhiko

1980 *Analysis of categorical data: Dual scaling and its applications*. University of Toronto Press.

Ono, Yohei

2014 How to Handle “Don’t know” Judgments in Lexicostatistical Survey: An

Exercise in Ainu Dialectology. *Language Dynamics and Change*. Brill.
(Under Review)

Swadesh, Morris

1955 Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*. 21: 121-137.

柳田 国男

1930 「蝸牛考」 刀江書院.

Statistical Reanalysis about the Classification of Ainu Dialects On the Data of Hattori and Chiri(1960)

Yohei ONO

The Graduate University for Advanced Studies

Abstract:

The main objective of this paper is to reconsider the classification of 19 Ainu dialects, based on the data of Hattori and Chiri(1960). The results of 3 previous studies using Hattori and Chiri(1960)'s data are inconsistent with the geolinguistical facts in southern Sakhalin. Handling the “Don't know” Judgment about the cognacy between dialects as an independent category, 6 Ainu dialects in southern Sakhalin classify into east and west coast successfully. Utilizing the new statistical methodology, which the present author calls “MCANeighbor-Net”, to the data of Hattori and Chiri(1960), the present author confirms not only the traditional classification of Ainu dialects into northeast and southwest but also another classification, which forms “Center versus Periphery” structure at the center of Hiratori or Saru region and was indicated many times philologically in Ainu dialectology but not statistically. This fact suggests that MCANeighbor-Net can extract the weak relationship or structure efficiently, which the traditional statistical analysis, like cluster analysis, tends to overlook. For future research, dealing the data appropriately in terms of measurement, utilizing the flexible statistical method to it and combining the linguistic and philological knowledge and statistical approach, those complementary studies will matter.