HOKKAIDO UNIVERSITY

Title	計量多次元尺度構成法の有効性と限界 ： アイヌ語を例として
Author(s)	小野, 洋平
Citation	北方言語研究, 7, 107-129
Issue Date	2017-02-15
Doc URL	https://hdl.handle.net/2115/64514
Type	departmental bulletin paper
File Information	07_ono.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北方言語研究 7: 107-129（北方言語ネットワーク編，北海道大学大学院文学研究科，2017）

［資料・研究ノート］

計量多次元尺度構成法の有効性と限界
－アイヌ語を例として－

小 野　洋 平
（統計数理研究所外来研究員）

1. はじめに

本稿は，小野(2015)の内容を大幅に改定し，統計学や数学の専門外の読者にもわかりやすい内容とし，方言研究への統計手法の適用についての幅広い理解を促進することを旨として書かれた.

小野(2015)においては紙面の都合上，用いた統計手法の数学的な詳細を省略せざるを得なかった．そのため，小野(2015)の主要な成果の一つである「平取周辺のアイヌ語諸方言が『統計学的』に見ても他のアイヌ語方言と比べ特徴的である」という知見は，未だにアイヌ語学者の間において, 十分な了解が得られたとは, 筆者には残念ながら感じられない. そこで，本稿においては，小野(2015)で用いた手法とは別の，より直感的に分かりやすい計量多次元尺度構成法を実際に適用することによってこれらの知見を再確認する. さらに, これらの分析によりアイヌ語諸方言における様々な仮説を検証し，計量多次元尺度構成法の有効性を示す．一方，本稿の後半部では，計量多次元尺度構成法の数学的な背景を説明し，その限界を示す．本稿の最終的な目的は，計量多次元尺度構成法の有効性と限界を示した上で，計量多次元尺度構成法を用いた基礎語彙統計学的研究を含めた様々な言語研究について適切な批判と再検討が行える素地を提供することである．本稿によって言語学と統計学との相互理解が深まり，学際的な研究が更に進展することを期待したい.

2. 本研究の構成と目的

本稿では 3 節で，距離行列から方言を視覚化する手段である「計量多次元尺度構成法」を表 1 の服部・知里(1960)のデータに実際に適用し，その有効性を示し，4 節で得られた結果について考察を行う．さらに，5 節では「計量多次元尺度構成法」の数学的な背景を説明する．一方 6 節では,「計量多次元尺度構成法」を含む，方言間の同根性のデータから非類似度(距離)行列を直接求める手法の問題点を論じる．7 節では，線形代数の学習の必要性を説き，8 節に結語を書く.

本研究の目的は「計量多次元尺度構成法」の有効性と限界を 3 つの点から示すことである．第一に，3 節と 4 節において，単純な非類似度(距離)行列のデータから方言間の関係を視覚化することは, 方言学上の既存の様々な仮説を検証する上では, 簡便な手段としては, なお有効であることを示す．特に，単なる非類似度の情報を多次元の軸に整理することによって，それぞれの軸の持つ意味を方言学の知見から解釈することができる点は魅力的である．一方，5 節では，計量多次元尺度構成法の持つ方法論の上の有効性と限界を示すとともに，6 節では「計量多次元尺度構成法」を含めた「方言間の同根性のデータから非類似度(距離)行列を直接求める手法」には,「データの情報を落とす」という限界があること

を示す．最終的な目的は，本論文で述べる「計量多次元尺度構成法」の有効性と限界を踏まえた上で,「計量多次元尺度構成法」を用いた基礎語彙統計学的研究を含めた様々な言語研究について,言語学者が適切な批判と再検討を行えるような素地を提供することにある．

3. 計量多次元尺度構成法の服部・知里(1960)への適用

本節では，距離データを視覚化する手法の一つである「計量多次元尺度構成法」を表 1 の服部・知里(1960)のデータに適用する．まず，表 1 は類似度行列であるため，100 から表 1 の各要素を引くことによって，非類似度行列(距離行列)に変換する．得られた非類似度行列に「計量多次元尺度構成法」を適用し(次元数は 4 に設定した)，得られた 4 次元まで座標を表 2 に示す．さらに，服部・知里(1960)より，19 地点の地図を図 1 に示す．実際の計算には R 言語(2015)の”cmdscale”コマンドを利用した．

計量多次元尺度構成法の大きな利点の一つは，表 1 の単なる(非)類似度の情報を，表 2 のような複数の軸の情報に整理することによって，それぞれの軸が担っている情報を言語学的な知見によって解釈することができる点にある．ただし，6 節で後述するように，服部・知里(1960)のように方言間の同根性に関する元のデータが存在するにも関わらず，(非)類似度行列を直接計算することは統計学の観点からは「データの情報を落とす」こととなり，限界がある点にも留意する必要がある．

表 1. アイヌ語方言間の類似度(%)．服部・知里(1960: 338)より

第 3 表	19. 内路 Na	18. ラチシイカ Ra	17. 白浦 Sh	16. 真岡 Ma	15. 多蘭泊 Ta	14. 落帆 Oc	13. 宗谷 Sô	12. 名寄 Ny	11. 旭川 As	10. 美幌 Bi	9. 訓路 Ku	8. 帯広 Ob	7. 様似 Sa	6. 新冠 Ni	5. 貫気別 Nu	4. 平取 Hi	3. 幌別 Ho	2. 長万部 Os
1. 八雲 Yakumo	72.0	70.5	71.9	70.8	70.3	73.3	84.8	86.7	89.5	87.7	88.2	87.5	88.7	89.1	88.8	90.3	91.8	96.1
2. 長万部 Oshamambe	71.3	70.0	71.3	70.0	68.9	74.5	83.2	85.2	88.6	86.7	87.9	87.0	89.8	90.0	90.5	90.2	90.7	
3. 幌別 Horobetsu	75.4	73.8	74.6	73.6	73.2	74.2	85.1	88.2	90.3	87.9	90.6	88.5	87.9	92.0	90.6	93.1		
4. 平取 Hiratori	75.7	73.8	74.3	74.1	73.4	75.3	83.5	87.2	88.7	86.9	88.0	86.7	85.8	96.1	95.6			
5. 貫気別 Nukkibetsu	72.7	70.9	72.5	71.7	70.5	74.9	82.9	85.9	87.7	83.8	85.7	84.6	85.0	95.8				
6. 新冠 Niikappu	75.4	73.3	74.1	73.5	73.1	75.3	82.2	85.2	88.4	85.8	88.0	86.2	87.1					
7. 様似 Samani	71.2	70.3	72.1	70.8	69.7	72.5	84.4	85.2	89.4	89.2	92.1	90.7						
8. 帯広 Obihiro	70.1	68.6	69.6	70.2	69.4	70.5	84.2	88.7	90.0	93.1	94.7							
9. 訓路 Kushiro	72.1	70.6	71.9	71.7	71.7	72.0	86.8	88.7	91.4	94.0								
10. 美幌 Bihoro	70.5	70.1	71.1	70.6	71.1	71.2	86.6	88.4	89.2									
11. 旭川 Asahikawa	75.1	73.4	74.2	73.9	73.6	73.4	85.4	90.8										
12. 名寄 Nayoro	73.2	72.3	73.0	72.8	73.0	73.9	87.6											
13. 宗谷 Sôya	76.8	77.5	78.3	79.4	78.8	79.7												
14. 落帆 Ochiho	92.4	89.6	91.7	91.1	88.8													
15. 多蘭泊 Tarantomari	89.8	89.0	88.5	92.6														
16. 真岡 Maoka	92.3	90.3	91.6															
17. 白浦 Shiraura	93.3	92.1																
18. ライチシカ Raichishka	90.5																	

表 2. 19 のアイヌ語方言の 4 次元空間における座標(小数点第 3 位で四捨五入)

	第 1 次元(x)	第 2 次元(y)	第 3 次元(z)	第 4 次元
X1_yakumo	-9.70	-2.91	-3.25	-0.75
X2_oshamambe	-10.20	-5.11	-5.52	0.04
X3_horobetsu	-7.10	-2.49	2.56	0.95
X4_hiratori	-6.49	-5.55	3.73	-1.14
X5_nukkibetsu	-8.20	-7.93	1.03	-3.78
X6_niikappu	-6.64	-6.48	3.09	1.07
X7_samani	-9.78	2.20	-4.61	5.50
X8_obihiro	-11.46	5.23	0.82	1.96
X9_kushiro	-9.67	4.73	0.91	2.31
X10_bihoro	-10.11	6.30	0.22	0.61
X11_asahikawa	-7.13	1.88	2.94	1.78
X12_nayoro	-7.12	4.16	2.60	-5.32
X13_souya	-0.65	6.61	-3.76	-5.71
X14_ochiho	16.06	-3.25	-5.57	-2.15
X15_tarantomari	17.89	3.85	3.11	-1.79
X16_maoka	17.89	1.94	0.62	-0.43
X17_shiraura	17.21	-0.60	-1.77	2.61
X18_raichishka	18.15	-0.08	0.48	1.40
X19_nairo	17.05	-2.54	2.36	2.83

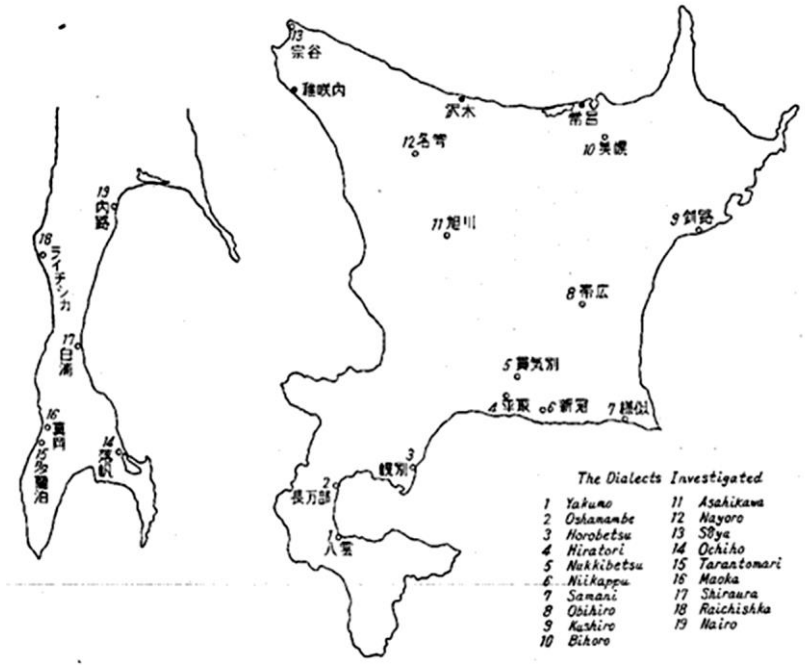

図 1. 服部・知里(1960)が調査した 19 のアイヌ語方言の地点. 服部・知里(1960: 309)

表 2 から第 1 次元(X 軸)においては，樺太方言が大きくプラスの値を取り，北海道方言が大きくマイナスの値を取り，宗谷方言が中間的な位置を占めている．さらに，北海道方言の中でも，旭川，名寄は相対的には樺太方言に近い値を取っており，地理的な関係と整合性がある．また，北海道南西部の方言の中でも，幌別，平取，貫別，新冠が相対的に樺太方言に近い値を取っていることは，中川(1996)が指摘した沙流・千歳・樺太型の表れと考えられる．第 2 次元(Y 軸)においては，樺太方言を除いて考えると，北海道の南西方言と北東方言の差が明瞭に現れている．第 3 次元(Z 軸)においては，北海道の南西方言を中心に見ると，幌別，平取，貫別，新冠と八雲，長万部との差が明瞭に現れる．さらに，北海道方言の中でも，八雲，長万部，様似，美幌，宗谷という周圏に位置する方言が，マイナスの値を取っている(ただし，美幌は 0.22)．小野(2015)では，沙流・千歳を中心とした方言がまとまることを示したが，今回の分析では逆に，周圏の方言が特徴的な値をとることが明らかになった．

また，第 4 次元においては，北海道方言の宗谷と名寄が，樺太南部の方言である落帆，多蘭泊，真岡と近い値をとっている．これは，宗谷海峡を介した地理的な近接性との関連が示唆される．

以上のように表 2 だけからでも，これらの情報は読み取れるが，表 2 の 1 次元目から 3 次元目までの座標を 3 次元空間にプロットすることにより, 直感的により解りやすい図を，図 2，図 3，図 4 に示した．

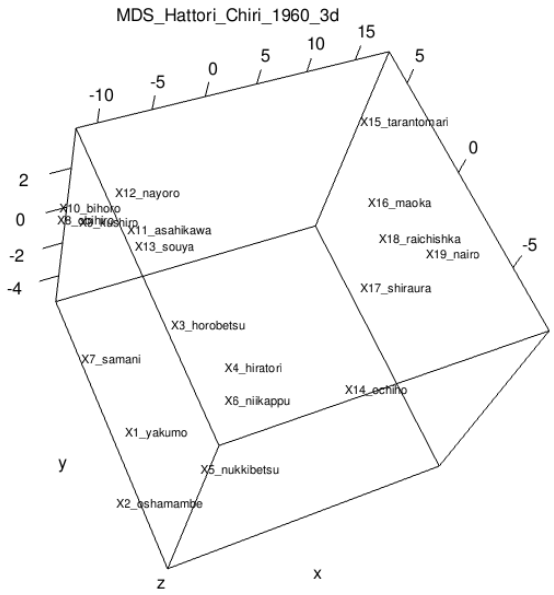

図 2．3 次元のアイヌ語方言の布置を左上から見た図

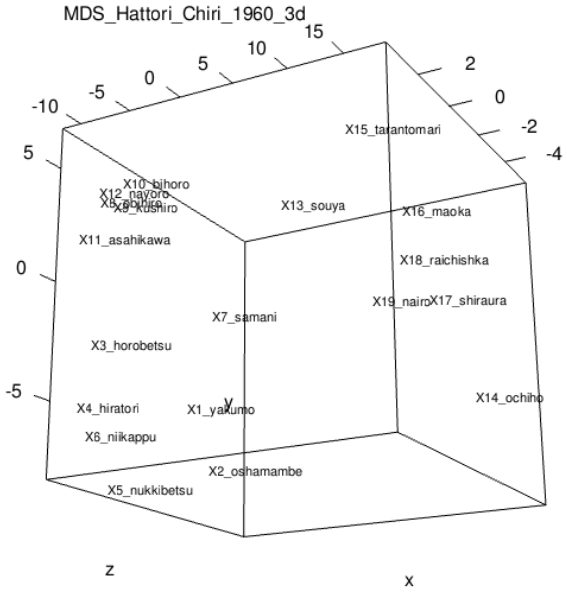

図 3. 3 次元のアイヌ語方言の布置を右上から見た図

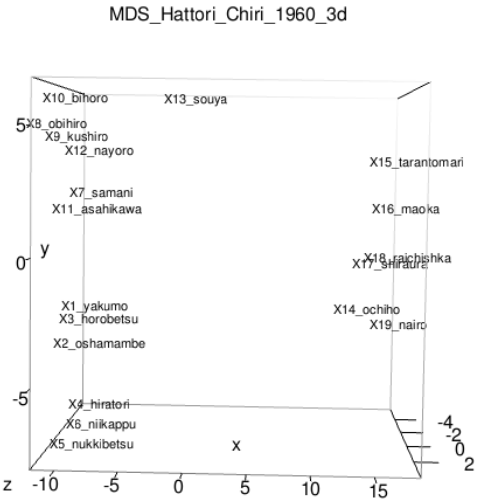

図 4. 3 次元のアイヌ語方言の布置を真上から見た図

図 2，図 3，図 4 から，北海道北東の方言が，地理的な分布と比較して，方言の類似度の観点からは非常に密集していることがわかる．また逆に，樺太方言に関しては，地理的な分布とは対照的に，方言の類似度の観点からは各方言間の差が大きいことがわかる．

4. 考察–軸の解釈

図 2，図 3 より 1)(幌別，)平取，新冠，貫別方言と他の方言との間に大きな差が認められる．この点は中川(1996)など先行研究において古くから文献学的には指摘されている．さらに，北海道南西部と北海道北東部の方言の座標を比較すると，2)北海道南西部の方言

では地域差が広く認められるのに対して，北海道北東部の方言では宗谷方言と様似方言を除けば，地域差はあまり明瞭ではなく，むしろ地理的分布の広さと対照的に，方言の類似度の点からは密集している．さらに服部・知里(1960: 340)も指摘しているように 3)(幌別，)平取，新冠，貫別方言と様似方言との間にも大きな差が認められる．

これらの指摘は，アイヌ語学の中における先行研究の再確認に過ぎないが，今回の計量多次元尺度構成法を用いた分析においては，2 次元(X 軸, Y 軸)までの情報から 4)中川(1996)が指摘する沙流・千歳・樺太型の分布が確認され，さらに 3 次元目(Z 軸)までの情報に注目したことによって，5)八雲，長万部，様似，美幌，宗谷からなる周圏的な分布が確認された．また，新たな知見として，4 次元目において，北海道方言の宗谷と名寄が，樺太南部の方言である落帆，多蘭泊，真岡が近い値をとっており，宗谷海峡を挟んだ地理的な近接性が示唆された．これらの結果から，2 節で述べたように，方言の同根性のデータから距離行列を直接計算し，計量多次元尺度構成法を適用する手法は，方言学における既存の仮説を検証する簡便な手法として有効であることが示された．

5. 補遺 1－計量多次元尺度構成法について[1]

5.0 用語及び定理の導入

5.0.1 ベクトル及び行列

中学，高校で扱う数が専ら 1 つの数であったのに対して，統計では数の「集まり」を扱う．その中で，「ベクトル」とは以下のようなものである(以下，ベクトルは太字で記す)．

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$$

[具体例 1] 表 2 の X1_八雲，X2_長万部，X3_幌別の X 軸の値を例に挙げると，

$$\boldsymbol{x} = \begin{pmatrix} -9.70 \\ -10.20 \\ -7.10 \end{pmatrix}$$

となる．

ただし，$x_1, x_2, x_3, \ldots, x_n$は実数，もしくは複素数とする．$\boldsymbol{x}^{\mathrm{T}}$をベクトル$\boldsymbol{x}$の「転置(transpose)」とし，

$$\boldsymbol{x}^T = (x_1, x_2, x_3, \ldots, x_n)$$

と記す．

[具体例 2] 具体例 1 の「転置(transpose)」を求めると，

---

[1] 本節における「計量多次元尺度構成法」の説明は，永田・棟近(2001)や齋藤・宿久(2009)を参考にし，筆者が独自にまとめたものである．

$$\boldsymbol{x}^T = (-9.70, -10.20, -7.10)$$

となる．

さらに，統計においては，ベクトルの「集まり」を行列として扱う(以下，行列は大文字の太字で記す)．行列とは，

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

ただし，$x_{ij}\{i = 1, \dots, n\}\{j = 1, \dots, m\}$は実数，もしくは複素数とする．

[具体例 3] 表 2 の X1_八雲，X2_長万部，X3_幌別の X 軸 Y 軸 Z 軸の値を例に挙げると，

$$\boldsymbol{X} = \begin{pmatrix} -9.70 & -2.91 & -3.25 \\ -10.20 & -5.11 & -5.52 \\ -7.10 & -2.49 & 2.56 \end{pmatrix}$$

この時，$x_{ij}\{i = 1, \dots, 3\}\{j = 1, \dots, 3\}$となり，$x_{12} = -2.91$, $x_{21} = -10.20$ となる．

ベクトル$\boldsymbol{x_1}$を

$$\boldsymbol{x_1} = \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{n1} \end{pmatrix}$$

とすると，行列$\boldsymbol{X}$は以下のように書ける．

$$\boldsymbol{X} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \dots, \boldsymbol{x_m})$$

[具体例 4] 具体例 3 においては，

$$\boldsymbol{x_1} = \begin{pmatrix} -9.70 \\ -10.20 \\ -7.10 \end{pmatrix}, \boldsymbol{x_2} = \begin{pmatrix} -2.91 \\ -5.11 \\ -2.49 \end{pmatrix}, \boldsymbol{x_3} = \begin{pmatrix} -3.25 \\ -5.52 \\ 2.56 \end{pmatrix}$$

となる．

同時に，$\boldsymbol{X^T}$を行列$\boldsymbol{X}$の「転置(transpose)」としたとき，以下のように記す．

$$\boldsymbol{X}^T = \begin{pmatrix} \boldsymbol{x_1^T} \\ \boldsymbol{x_2^T} \\ \boldsymbol{x_3^T} \\ \vdots \\ \boldsymbol{x_m^T} \end{pmatrix}$$

[具体例 5] 具体例 3 から，

$$\boldsymbol{x_1^T} = (-9.70 \quad -10.20 \quad -7.10)$$
$$\boldsymbol{x_2^T} = (-2.91 \quad -5.11 \quad -2.49)$$
$$\boldsymbol{x_3^T} = (-3.25 \quad -5.52 \quad 2.56)$$

よって,

$$\boldsymbol{X^T} = \begin{pmatrix} -9.70 & -10.20 & -7.10 \\ -2.91 & -5.11 & -2.49 \\ -3.25 & -5.52 & 2.56 \end{pmatrix}$$

となり，$\boldsymbol{X}$と比較すると,

$$\boldsymbol{X} = \begin{pmatrix} -9.70 & -2.91 & -3.25 \\ -10.20 & -5.11 & -5.52 \\ -7.10 & -2.49 & 2.56 \end{pmatrix}$$

となる.

5.0.2 ベクトルの内積の定義

今ベクトル$\boldsymbol{x_1}[\boldsymbol{x_1} = \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{n1} \end{pmatrix}]$と$\boldsymbol{y_1}[\boldsymbol{y_1} = \begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ \vdots \\ y_{n1} \end{pmatrix}]$とするとき,ベクトル$\boldsymbol{x_1}$とベクトル$\boldsymbol{y_1}$の「内積(inner product)」は以下のように定義される.

$$\langle \boldsymbol{x_1}, \boldsymbol{y_1} \rangle = \sum_{q=1}^{n} x_{q1}\, y_{q1}$$

今記法を簡略化するために，ベクトル$\boldsymbol{x}$を$\vec{\boldsymbol{x}}$と表し，$\langle \boldsymbol{x_1}, \boldsymbol{y_1} \rangle$を$\overrightarrow{\boldsymbol{x_1}} \cdot \overrightarrow{\boldsymbol{y_1}}$と表す.

[具体例 6] 具体例 4 の行列$\boldsymbol{X}$における$\boldsymbol{x_1}, \boldsymbol{x_2}$を例に挙げると,

$$\langle \boldsymbol{x_1}, \boldsymbol{x_2} \rangle = \overrightarrow{\boldsymbol{x_1}} \cdot \overrightarrow{\boldsymbol{x_2}} = \sum_{q=1}^{n} x_{q1}\, x_{q2}$$
$$= x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32}$$
$$= (-9.70)(-2.91) + (-10.20)(-5.11) + (-7.10)(2.49)$$
$$= 62.67$$

5.0.3 行列の積の定義

今行列$\boldsymbol{Z}$について,

$$\boldsymbol{Z} = \begin{pmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{pmatrix}$$

$z_{ij}\{i = 1, \ldots, n\}\{j = 1, \ldots, m\}$と表すと，行列の積$\boldsymbol{Z} = \boldsymbol{ST}$は，以下のように計算する.
行列$\boldsymbol{S}$を$s_{iq}\{i = 1, \ldots, n\}\{q = 1, \ldots, l\}$，行列$\boldsymbol{T}$を$t_{qj}\{q = 1, \ldots, l\}\{j = 1, \ldots, m\}$と表すと,

$$z_{ij} = \sum_{q=1}^{l} s_{iq}\, t_{qj}$$

[具体例 7] 具体例 5 における行列$\boldsymbol{X}$と行列$\boldsymbol{X^T}$の積$\boldsymbol{XX^T}$を計算する，

$$\boldsymbol{X} = \begin{pmatrix} -9.70 & -2.91 & -3.25 \\ -10.20 & -5.11 & -5.52 \\ -7.10 & -2.49 & 2.56 \end{pmatrix}$$

$$\boldsymbol{X^T} = \begin{pmatrix} -9.70 & -10.20 & -7.10 \\ -2.91 & -5.11 & -2.49 \\ -3.25 & -5.52 & 2.56 \end{pmatrix}$$

なので，

$$\boldsymbol{Z = XX^T} = \begin{pmatrix} (-9.70)(-9.70)+(-2.91)(-2.91)+(-3.25)(-3.25) & (-9.70)(-10.20)+(-2.91)(-5.11)+(-3.25)(-5.52) & (-9.70)(-7.10)+(-2.91)(-2.49)+(-3.25)(2.56) \\ (-10.20)(-9.70)+(-5.11)(-2.91)+(-5.52)(-3.25) & (-10.20)(-10.20)+(-5.11)(-5.11)+(-5.52)(-5.52) & (-10.20)(-7.10)+(-5.11)(-2.49)+(-5.52)(2.56) \\ (-7.10)(-9.70)+(-2.49)(-2.91)+(2.56)(-3.25) & (-7.10)(-10.20)+(-2.49)(-5.11)+(2.56)(-5.52) & (-7.10)(-7.10)+(-2.49)(-2.49)+(2.56)(2.56) \end{pmatrix}$$

$$= \begin{pmatrix} 113.1206 & 131.7501 & 67.7959 \\ 131.7501 & 160.6225 & 71.0127 \\ 67.7959 & 71.0127 & 63.1637 \end{pmatrix}$$

となる．

5.0.4 余弦定理及び内積の性質

今三角形 ABC に関して，図 5 のような関係が成り立っているとする．

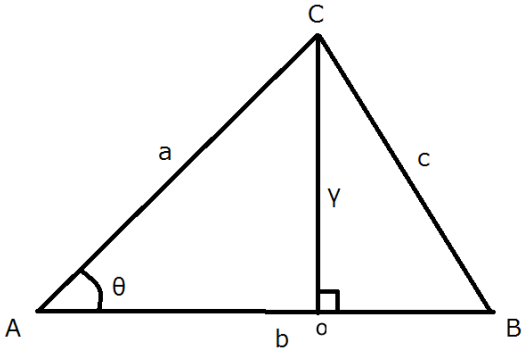

図 5. 三角形 ABC について．その 1

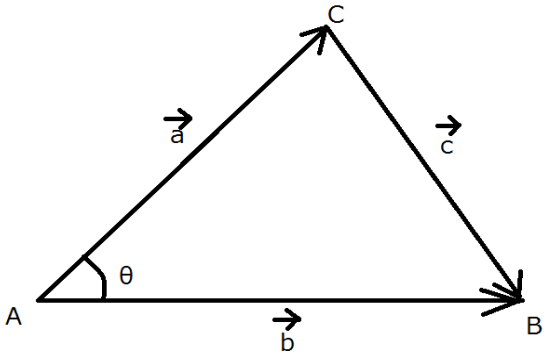

図 6. 三角形 ABC について．その 2

このとき，辺 CO の長さをγとしたとき，三平方の定理より，

$$a^2 = (a\cos\theta)^2 + \gamma^2 \cdots (1)$$

$$c^2 = (b - a\cos\theta)^2 + \gamma^2 \cdots (2)$$

(1),(2)より

$$c^2 = a^2 + b^2 - 2ab\cos\theta \cdots (3)$$

となる．一方で，三角形 ABC を図 6 のように表記を変え，$\vec{\boldsymbol{a}} \cdot \vec{\boldsymbol{a}} = |\vec{\boldsymbol{a}}|^2$と表すと，$|\vec{\boldsymbol{c}}|^2$につ

いて

$$|\vec{c}|^2 = (\vec{b} - \vec{a}) \cdot (\vec{b} - \vec{a}) = |\vec{b}|^2 + |\vec{a}|^2 - 2\vec{a} \cdot \vec{b} \cdots (4)$$

(4)における$|\vec{a}|^2, |\vec{b}|^2 |\vec{c}|^2$はそれぞれ図 5 における$a^2, b^2, c^2$に対応するので，(3)と(4)を比較することによって，

$$\vec{a} \cdot \vec{b} = ab\cos\theta = |\vec{a}||\vec{b}|\cos\theta \cdots (5)$$

という性質が示された．(5)のことを内積と呼び，この情報は計量多次元尺度構成法において方言間の類似度を表し，重要な役割を果たす．

5.0.5 エッカート・ヤング分解 (Eckart and Young 1936)

表 3. 方言 1 から方言 5 までの距離行列(非類似度行列)

	方言 1	方言 2	方言 3	方言 4	方言 5
方言 1	0	$\sqrt{5}$	2	$\sqrt{3}$	$\sqrt{2}$
方言 2	$\sqrt{5}$	0	$\sqrt{3}$	2	$\sqrt{5}$
方言 3	2	$\sqrt{3}$	0	$\sqrt{5}$	2
方言 4	$\sqrt{3}$	2	$\sqrt{5}$	0	1
方言 5	$\sqrt{2}$	$\sqrt{5}$	2	1	0

エッカート・ヤング分解については，数学的な詳細には立ち入らないが，表 3 のような「距離行列」$\boldsymbol{Z}[\boldsymbol{Z}_{ij}\{i,j=1,\ldots,n\}]$が与えられた時，ある基準のもとで最もふさわしい行列$\boldsymbol{X}$が得られ，以下のような関係式を満たす．

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}} \cdots (6)$$

この時，$\boldsymbol{Z}$の固有値，固有ベクトルを求め，Z の固有値を対角要素とする行列を$\boldsymbol{\Lambda}$，対応する固有ベクトルを列に並べた行列を$\boldsymbol{Y}$とすると，

$$\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{\Lambda}\boldsymbol{Y}^{\mathrm{T}} \cdots (7)$$

と表すことができる．これを，エッカート・ヤング分解という．この結果から，求める$\boldsymbol{X}$は，

$$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{\Lambda}^{\frac{1}{2}} \cdots (8)$$

となる．なぜならば，(6)に(8)を代入すると，

$$\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{Y}\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\boldsymbol{T}} = \boldsymbol{Y}\boldsymbol{\Lambda}^{\frac{1}{2}}\left(\boldsymbol{\Lambda}^{\frac{1}{2}}\right)^{\boldsymbol{T}}\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{Y}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Y}^{\mathrm{T}} = \boldsymbol{Y}\boldsymbol{\Lambda}\boldsymbol{Y}^{\mathrm{T}}$$

となり[2]，これは「Z の固有値を対角要素とする行列を$\boldsymbol{\Lambda}$，対応する固有ベクトルを列に並べた行列を$\boldsymbol{Y}$とする」という$\boldsymbol{\Lambda}$と$\boldsymbol{Y}$の定義を満たしている．

5.1 概要

本節では，表 3 のような方言間の距離を表した行列を「距離行列」と呼び$d_{ij}\{i,j=1,\ldots,n\}$と表す．今，分析者の手元には，d_{ij}の情報しかないものとする．分析者が，計量多次元尺度構成法を利用する目的は，d_{ij}から各方言iのm次元の座標$x_{im}\{i=1,\ldots,n\}\{m=1,\ldots,p\}$(これは，5.0.5 節の行列$\boldsymbol{X}$に対応する)を求めることにある．まず，$d_{ij}$は距離の公理を満たして

[2] 行列の固有値，固有ベクトル，転置計算の法則などは，齋藤(1966)を参照のこと．

いる必要がある．さらに，d_{ij}からx_{im}を求めるには，高校数学で学んだ余弦定理とエッカート・ヤング分解(Eckart and Young 1936) を用いる．ただし，d_{ij}から直接x_{im}を求めると，数学的に様々な答えがあり望ましくない．具体的には，求めた座標$x_{im}\{i=1,\dots,n\}\{m=1,\dots,p\}$について 1)あらゆる方向に方言間の相対的な関係を保ったまま平行移動が可能であるがX軸Y軸…Z軸の値が変わってしまう点と2)原点を中心にある角度θ回転させても方言間の相対的な関係が変わらないがX軸Y軸…Z軸の値が変わってしまう点の2点が問題となる．そこで,「二重中心化」という変換をd_{ij}に適用し，さらにエッカート・ヤング分解を適用することによって 1)の問題を解決した$x_{im}\{i=1,\dots,n\}\{m=1,\dots,p\}$が得られる．ただし，2)の問題は解決されていないため,「計量多次元尺度構成法」によって得られたp次元座標へ方言間の関係を視覚化した図は，原点を中心にある角度θ回転することが可能であるため，分析者は図の解釈の際にはこのことを留意する必要がある．

5.2 距離の公理

表3のような方言間の距離を表した行列を「距離行列」と呼びd_{ij}と表す. 表3の例では，$a_{ij}\{i,j=1,\dots,5\}$とする．今，距離行列$a_{ij}\{i,j=1,\dots,5\}$が以下の性質を満たしている時，a_{ij}は「距離の公理を満たす」という．3番目の条件は「三角不等式」と言われる．

$$\begin{cases} a_{ij} \geq 0 \\ a_{ij} = a_{ji} \\ a_{ij} + a_{jk} \geq a_{ik} \end{cases}$$

表3のa_{ij}が「距離の公理を満たす」ことは容易に確かめられる.「距離の公理を満たす」ことは,「計量多次元尺度構成法」においてq次元平面へ方言間の関係を視覚化するための必要条件である.

ただし, 1) あらゆる方向に方言間の相対的な関係を保ったまま平行移動が可能であるがX軸Y軸…Z軸の値が変わってしまう点と 2) 原点を中心にある角度θ回転させても方言間の相対的な関係が変わらないけれどもX軸Y軸…Z軸の値が変わってしまう点の2点が問題である．両者がなぜ問題であるかをそれぞれ図7，図8に例示する．

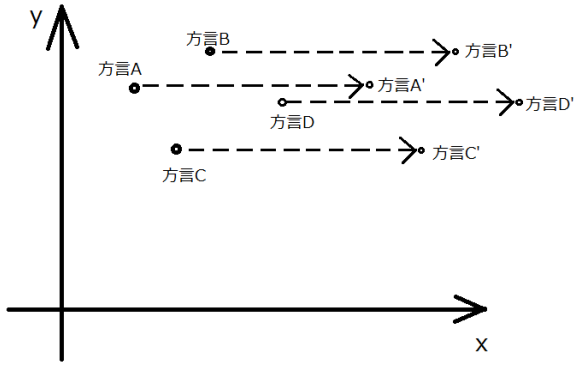

図7. あらゆる方向に方言ABCD間の相対的な関係を保ったまま，平行移動が可能であるがX軸Y軸の値が変わる.

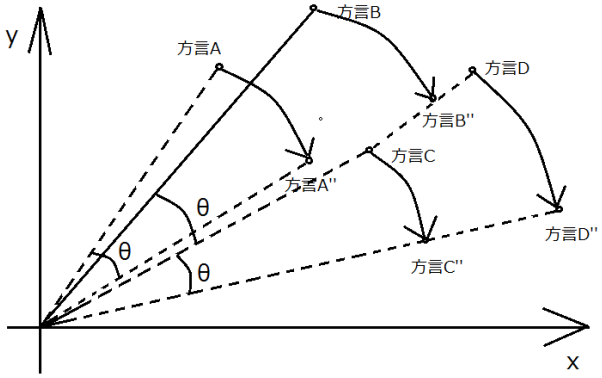

図 8. 原点を中心にある角度 θ 回転させても方言 ABCD 間の相対的な関係が変わらないが X 軸 Y 軸の値が変わる.

図 9, 図 10 ともに求めたい方言 ABCD 間の相対的な関係が変わらないが，X 軸 Y 軸…Z 軸の値が変わってしまうことの原因になっている．そこで，本節で紹介する「計量多次元尺度構成法」においては，距離行列を$d_{ij}\{i,j=1,\dots,n\}$とし，求めたいi番目の方言のm次元目の座標を$x_{im}\{i=1,\dots,n\}\{m=1,\dots,p\}$と表した時に，方言の各次元の座標の平均が 0 になるという制限をつけることで，1) あらゆる方向に方言間の相対的な関係を保ったまま平行移動が可能であるが X 軸 Y 軸…Z 軸の値が変わってしまうという問題を解決する．直感的には，各方言の座標の「重心」を原点に置くことに等しく，数学的には，各方言の各次元の座標について，

$$\sum_{i=1}^{n} x_{im} = 0, \{m=1,\dots,p\} \cdots (9)$$

という制限をつけることに等しい[3].

5.3 計量多次元尺度構成法の導出

我々の目標は，距離行列d_{ij}から i番目の方言のm次元目の座標を$x_{im}\{i=1,\dots,n\}\{m=1,\dots,p\}$を求めることである．方言$i,j,k\{i,j,k=1,\dots,n\}$は距離の公理を満たしていることから以下の余弦定理が成立する．

$$d_{ij}^2 = d_{ik}^2 + d_{jk}^2 - 2d_{ik}d_{jk}\cos\theta \cdots (10)$$

しかし，式(10)は 1)あらゆる方向に方言間の相対的な関係を保ったまま平行移動が可能であるが X 軸 Y 軸…Z 軸の値が変わってしまう点と 2)原点を中心にある角度 θ 回転させても方言間の相対的な関係が変わらないけれども X 軸 Y 軸…Z 軸の値が変わってしまう点の 2 点の問題を抱えている．

そこで，前節において，$x_{im}\{i=1,\dots,n\}\{m=1,\dots,p\}$に，

[3] 実は，この操作によって，距離行列にふくまれている「誤差」が取り除かれる．「重心」を原点にせず方言 1 や方言 2 などを中心すると，誤差の影響で，求めた方言の相対的な位置が方言 1 を中心とした場合と方言 2 を中心とした場合で異なってしまう．詳しくは，齋藤・宿久(2009: 44-45)を参照.

$$\sum_{i=1}^{n} x_{im} = 0, \{m = 1, \dots, p\}$$

とし，原点 0 に諸方言の重心を置いたことによって，式(10)のkをoに置き変えると，

$$d_{ij}^2 = d_{io}^2 + d_{jo}^2 - 2d_{io}d_{jo}\cos\theta \cdots (11)$$

さらに，式(11)の右辺第 3 項を中心に展開すると，

$$d_{io}d_{jo}\cos\theta = \frac{1}{2}\left(d_{io}^2 + d_{jo}^2 - d_{ij}^2\right) \cdots (12)$$

ここで，式(12)の左辺に注目すると，式(5)より，

$$d_{io}d_{jo}\cos\theta = \langle \overrightarrow{\boldsymbol{d_{io}}}, \overrightarrow{\boldsymbol{d_{jo}}} \rangle = \overrightarrow{\boldsymbol{d_{io}}} \cdot \overrightarrow{\boldsymbol{d_{jo}}}$$

となり，$d_{io}d_{jo}\cos\theta$は内積であることがわかる．

よって，「計量多次元尺度構成法」においては，エッカート・ヤング分解を用いて，

$$\begin{aligned} z_{ij} &= d_{io}d_{jo}\cos\theta \\ &= \frac{1}{2}\left(d_{io}^2 + d_{jo}^2 - d_{ij}^2\right) \\ &= \langle \overrightarrow{\boldsymbol{d_{io}}}, \overrightarrow{\boldsymbol{d_{jo}}} \rangle \\ &= \overrightarrow{\boldsymbol{d_{io}}} \cdot \overrightarrow{\boldsymbol{d_{jo}}} \\ &= \overrightarrow{\boldsymbol{x_{io}}} \cdot \overrightarrow{\boldsymbol{x_{jo}}} \\ &= \sum_{m=1}^{p} x_{im}x_{jm} \cdots (13) \end{aligned}$$

を満たすようなi番目の方言のm次元目の座標$x_{im}\{i = 1, \dots, n\}\{m = 1, \dots, p\}$を求める．

今，$d_{io}d_{jo}\cos\theta$の値に注目するのは，図 11 のように，各方言は座標の重心を原点 O においているため，任意の回転 θ ’に対して，内積の値は$d_{io}d_{jo}\cos\theta = d_{i'o}d_{j'o}\cos\theta$となり「不変量」となるためである．

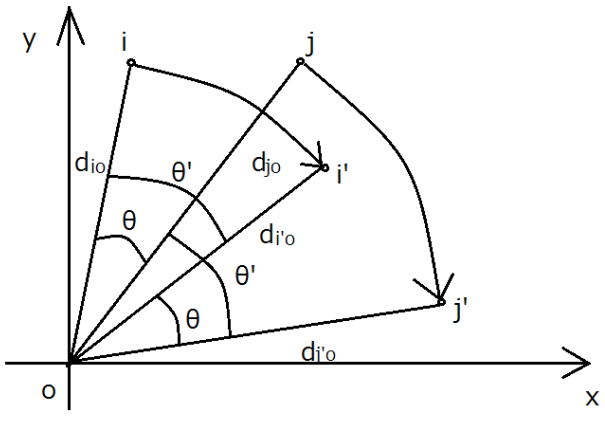

図 9. 各方言の座標の「重心」を原点に置いた図

我々の手元には，今，「距離行列」$d_{ij}\{i, j = 1, \dots, n\}$の情報しかなく，$d_{io}\{i = 1, \dots, n\}$の情報はない．よって，以下では，(9)の情報を用いて式(12)を「距離行列」$d_{ij}\{i, j = 1, \dots, n\}$の

情報で表すことを目標とする[4].

今，地点iと地点jの内積z_{ij}は 5.0.2 節より，式(13)を展開すると，

$$z_{ij} = \sum_{m=1}^{p} x_{im}x_{jm} = \frac{1}{2}\left(d_{io}^2 + d_{jo}^2 - d_{ij}^2\right) \cdots (14)$$

となる．ここで，式(9)の条件を使うと，

$$\sum_{i=1}^{n} x_{im} = 0, \{m = 1, \dots, p\}$$

より，

$$\sum_{i=1}^{n} z_{ij} = \sum_{i=1}^{n}\sum_{m=1}^{p} x_{im}x_{jm} = \sum_{m=1}^{p}\sum_{i=1}^{n} x_{im}x_{jm}\ [5] = \sum_{m=1}^{p}\left(\sum_{i=1}^{n} x_{im}\right)x_{jm} = 0 \cdots (15)$$

$$\sum_{j=1}^{n} z_{ij} = \sum_{j=1}^{n}\sum_{m=1}^{p} x_{im}x_{jm} = \sum_{m=1}^{p}\sum_{j=1}^{n} x_{im}x_{jm} = \sum_{m=1}^{p} x_{im}\left(\sum_{j=1}^{n} x_{jm}\right) = 0 \cdots (16)$$

式(15)(16)を式(14)に適用すると，

$$\sum_{i=1}^{n} z_{ij} = \frac{1}{2}\sum_{i=1}^{n}\left(d_{io}^2 + d_{jo}^2 - d_{ij}^2\right) = \frac{1}{2}\left(\sum_{i=1}^{n} d_{io}^2 + nd_{jo}^2 - \sum_{i=1}^{n} d_{ij}^2\right) = 0 \cdots (17)$$

$$\sum_{j=1}^{n} z_{ij} = \frac{1}{2}\sum_{j=1}^{n}\left(d_{io}^2 + d_{jo}^2 - d_{ij}^2\right) = \frac{1}{2}\left(nd_{io}^2 + \sum_{j=1}^{n} d_{jo}^2 - \sum_{j=1}^{n} d_{ij}^2\right) = 0 \cdots (18)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij} = \frac{1}{2}\sum_{i=1}^{n}\left(nd_{io}^2 + \sum_{j=1}^{n} d_{jo}^2 + \sum_{j=1}^{n} d_{ij}^2\right) = \frac{1}{2}\left(n\sum_{i=1}^{n} d_{io}^2 + n\sum_{j=1}^{n} d_{jo}^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2\right) = 0 \cdots (19)$$

式(17)と式(18)を足しあわせると，

$$\frac{1}{2}\left(\sum_{i=1}^{n} d_{io}^2 + nd_{jo}^2 - \sum_{i=1}^{n} d_{ij}^2\right) + \frac{1}{2}\left(nd_{io}^2 + \sum_{j=1}^{n} d_{jo}^2 - \sum_{j=1}^{n} d_{ij}^2\right) = 0$$

$$\Rightarrow d_{io}^2 + d_{jo}^2 = \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 + \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} d_{io}^2 + \sum_{j=1}^{n} d_{jo}^2\right) \cdots (20)$$

[4] 以下の式展開は永田・棟近(2001: 228-229)に従った.

[5] 和の計算は足す順番を変えても(交換しても)同じ結果になる.

となる．今，我々の式展開の目的は，$d_{io}^2 + d_{jo}^2$をd_{ij}だけの式で表すことであった．よって，式(18)における第 3 項をd_{ij}だけで表すことができればよい．今，式(19)は，

$$\frac{1}{2}\left(n\sum_{i=1}^{n} d_{io}^2 + n\sum_{j=1}^{n} d_{jo}^2 - \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right) = 0$$

なので，これを展開すると，

$$\frac{1}{n}\left(\sum_{i=1}^{n} d_{io}^2 + \sum_{j=1}^{n} d_{jo}^2 \right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right) \cdots (21)$$

となり，式(21)を式(20)の第 3 項に代入することで，

$$d_{io}^2 + d_{jo}^2 = \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 + \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right) \cdots (22)$$

となる．さらに，式(12)に式(22)を代入することによって，

$$z_{ij} = d_{io} d_{jo} \cos\theta = \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 + \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right) - d_{ij}^2 \right) \cdots (23)$$

となり，式(12)を「距離行列」$d_{ij}\{i, j = 1, \ldots, n\}$の情報で表すことに成功した[6]．

式(23)のz_{ij}は「距離行列」d_{ij}よって表されているため計算可能であり，z_{ij}にエッカート・ヤング分解を適用することにより(5.0.5 節)，求めたいi番目の方言のm次元目の座標を$x_{im}\{i = 1, \ldots, n\}\{m = 1, \ldots, p\}$(これが，5.0.5 節の行列$\boldsymbol{X}$に対応する)が求まる．以上が計量多次元尺度構成法の概要である[7]．

5.4 計量多次元尺度構成法の補足

ここで，式(21)による「二重中心化」によって，5.0.5 節のエッカート・ヤング分解から得られた行列$\boldsymbol{X}$(i番目の方言のm次元目の座標を$x_{im}\{i = 1, \ldots, n\}\{m = 1, \ldots, p\}$とする行列)が 1)あらゆる方向に方言間の相対的な関係を保ったまま平行移動が可能でない点と 2)原点を中心にある角度 θ 回転させても方言間の相対的な関係が変わらないけれども X 軸 Y 軸…Z 軸の値が変わってしまう点の 2 点が数学的にどのような性質をもたらすか補足する．

今，任意の平行移動を$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1m} \\ \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \cdots & \varepsilon_{nm} \end{pmatrix}$,ただし$\varepsilon_{1i} = \varepsilon_{2i} = \cdots = \varepsilon_{ni}\{i = 1, \ldots, m\}$.

とすると，任意の平行移動がもし可能であるならばエッカート・ヤング分解から得られた行列$\boldsymbol{X}$ついて，新たに$\boldsymbol{X_1}$として，$\boldsymbol{X_1} = \boldsymbol{Y}\boldsymbol{\Lambda}^{\frac{1}{2}} + \boldsymbol{\varepsilon}$を考えることができる．

しかし，新たに得られた$\boldsymbol{X_1}$は，式(6)より，

---

[6] 式(21)の変換を「二重中心化(double centering)」と呼ぶ．

[7] 厳密には，5.0.5 節におけるエッカート・ヤング分解の際に，固有値が負の固有ベクトルは，$\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{\Lambda}^{\frac{1}{2}}$を計算するとき，$\boldsymbol{\Lambda}^{\frac{1}{2}}$が複素数になってしまう．そのため，固有値が正の上位 p 次元の固有ベクトルの値を用いる．

$$\boldsymbol{X_1 X_1^T} = (\boldsymbol{Y\Lambda^{\frac{1}{2}}} + \boldsymbol{\varepsilon})\left(\boldsymbol{Y\Lambda^{\frac{1}{2}}} + \boldsymbol{\varepsilon}\right)^{\boldsymbol{T}}$$

$$= (\boldsymbol{Y\Lambda^{\frac{1}{2}}} + \boldsymbol{\varepsilon})\left(\left(\boldsymbol{Y\Lambda^{\frac{1}{2}}}\right)^{\boldsymbol{T}} + \boldsymbol{\varepsilon^T}\right)$$

$$= (\boldsymbol{Y\Lambda^{\frac{1}{2}}} + \boldsymbol{\varepsilon})\left(\left(\boldsymbol{\Lambda^{\frac{1}{2}}}\right)^{\boldsymbol{T}} \boldsymbol{Y}^{\mathrm{T}} + \boldsymbol{\varepsilon^T}\right)$$

$$= \boldsymbol{Y\Lambda^{\frac{1}{2}}}\left(\boldsymbol{\Lambda^{\frac{1}{2}}}\right)^{\boldsymbol{T}} \boldsymbol{Y}^{\mathrm{T}} + \boldsymbol{Y\Lambda^{\frac{1}{2}}\varepsilon^T} + \boldsymbol{\varepsilon}\left(\boldsymbol{\Lambda^{\frac{1}{2}}}\right)^{\boldsymbol{T}} \boldsymbol{Y}^{\mathrm{T}} + \boldsymbol{\varepsilon\varepsilon^T}$$

$$= \boldsymbol{Z} + \boldsymbol{Y\Lambda^{\frac{1}{2}}\varepsilon^T} + \boldsymbol{\varepsilon\Lambda^{\frac{1}{2}}Y}^{\mathrm{T}} + \boldsymbol{\varepsilon\varepsilon^T}$$

となり，元の距離行列$\boldsymbol{Z}$とは違った行列から$\boldsymbol{X_1}$は求めることができる．元の距離行列$\boldsymbol{Z}$とエッカート・ヤング分解から得られた行列$\boldsymbol{X}$との間に対応づけが全くできない．数学的に望ましくない性質であるため，二重中心化によって「原点を重心とする」ことで，これを不可能とした．

さらに，今原点を中心にある角度θ回転させる行列を$\boldsymbol{\Theta}$と表すと，原点を中心にある角度θ回転することがもし可能であるならばエッカート・ヤング分解から得られた行列$\boldsymbol{X}$ついて新たに$\boldsymbol{X_1}$として，

$$\boldsymbol{X_1} = \boldsymbol{X\Theta}$$

を考えることができる．新たに得られた$\boldsymbol{X_1}$は，式(6)より，

$$\boldsymbol{X_1 X_1^T} = \boldsymbol{X\Theta}(\boldsymbol{X\Theta})^{\boldsymbol{T}}$$

$$= \boldsymbol{X\Theta\Theta^T X^T}$$[8]

$$= \boldsymbol{XX^T}$$

$$= \boldsymbol{Z}$$

となり，任意の直交回転が可能だとしても新たに得られた$\boldsymbol{X_1}$も距離行列$\boldsymbol{Z}$から得られたものであり，$\boldsymbol{Z}$は二重中心化をしているため，原点を中心として$\boldsymbol{X}$をある角度θ回転しても，各方言の相対的な位置関係は変わらない．

5.5 まとめ−計量多次元尺度構成法の有効性と限界

ここでは，5 節における計量多次元尺度構成法の議論をまとめ，その利点と限界を述べる．

計量多次元尺度構成法の利点としては，3 節でも述べたように，単なる非類似度の情報を多次元の軸の情報に整理することができ，その軸を言語学的な知見に基づいて解釈することができる点にある．

一方,限界としては,原点を中心としているため原点からの値の大小には意味があるが,任意の直交回転が可能であるため軸からの角度の情報には意味がない点にある．この点に関しては,「非計量多次元尺度構成法」においては，軸からの角度に意味を持たせようという研究が，近年見られる(Miyadera, 2015).

[8] $\boldsymbol{\Theta\Theta^T} = \boldsymbol{I}$となり，これは，同じ大きさの正方行列$\boldsymbol{A}$に対して，$\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$となる．

6. 補遺 2–計量多次元尺度構成法の限界について：先行研究(服部・知里 1960)の問題点

服部・知里(1960)は昭和 30 年から昭和 31 年にかけて，北海道 13 地点，樺太 6 地点の計 19 地点における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．当時，アイヌ語は消滅の危機に瀕していると識者の間では認識されており，実際彼らの調査の後に多くのアイヌ語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記念碑的な研究である．調査する語彙は，Swadesh(1955)の基礎語彙 200 語をもとに彼ら独自の言語学的な知見を持って選定され，日本における基礎語彙統計学的研究の嚆矢といえる．服部・知里(1960)は 19 地点の語彙の同根性について，表 4 のような基準にのっとり，データを作成している．

表 4. 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

表 4 から，服部・知里は以下のような手続きで，方言間の類似度を求め，表 1(3 節)を得ている．

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
5. A 方言と B 方言の類似度を，方言間の語彙に関して＋，±，－の合計に対して，＋と±の＋(0.5 とカウントする)の割合(%)とする．

架空例: 名寄と宗谷について
＋が 50 個，±が 30 個，－が 20 個，{〇, ?, ・, ( )}が 10 個の場合
分子=50×1＋30×0.5=65
分母=50×1＋30×1＋20×1=100
よって，名寄と宗谷の類似度は 65/100=65%

3 節では，表 1 の類似度のデータに対して多次元尺度構成法を適用し，方言間のデータの視覚化を行った．しかしここでは，そもそも方言間の同根性のデータについて，服部・知里(1960)のように直接類似度を求めてしまうことには, 小野(2015)で指摘した点とは別の観点から，方法論上重大な問題が含まれていることを指摘したい.

今，言語に方言 1~5 があり，それらの言語には語 I1 と語 I2 があり，語 I1 には 4 つの語形 W11，W12，W13，W14 が，語 I2 には 3 つの語形 W21，W22，W23 があり，ある方言がある語形を持っている場合には 1 を持っていない場合には 0 を記入するものとする．架空例として，表 5，表 6，表 7 を載せる．例えば，表 5 の方言 1 は語 I1 に関しては語形 W11 と語形 W14 を語 I2 については語形 W22 を持っている.

表 5. 方言 1～5 に関する架空のデータ 1

語	I1				I2		
語形	W11	W12	W13	W14	W21	W22	W23
方言 1	1	0	0	1	0	1	0
方言 2	0	1	1	0	1	1	0
方言 3	0	0	1	1	1	0	0
方言 4	1	0	1	0	0	1	1
方言 5	1	0	1	1	0	1	1

表 6. 方言 1～5 に関する架空のデータ 2

語	I1				I2		
語形	W11	W12	W13	W14	W21	W22	W23
方言 1	0	0	0	1	0	1	1
方言 2	1	1	1	0	1	1	1
方言 3	1	0	1	1	1	0	1
方言 4	0	0	1	0	0	1	0
方言 5	0	0	1	1	0	1	0

表 7. 方言 1〜5 に関する架空のデータ 3

語	I1				I2		
語形	W11	W12	W13	W14	W21	W22	W23
方言 1	0	1	1	1	1	1	1
方言 2	1	0	0	0	0	1	1
方言 3	1	1	0	1	0	0	1
方言 4	0	1	0	0	1	1	0
方言 5	0	1	0	1	1	1	0

例えば表 5 において，方言 1 と方言 2 との距離(非類似度)は，

$d_{12} = \sqrt{(1-0)^2 + (0-1)^2 + (0-1)^2 + (1-0)^2 + (0-1)^2 + (1-1)^2 + (0-0)^2} = \sqrt{5}$

となる(ここではユークリッド距離を用いた).

注目するべきは，表 5，表 6，表 7 のそれぞれの方言間の距離行列(非類似度行列)は表 8 に帰着できる点である(できれば読者は手計算で確認していただきたい).

表 8. 表 5，表 6，表 7 の距離行列(非類似度行列)

	方言 1	方言 2	方言 3	方言 4	方言 5
方言 1	0	$\sqrt{5}$	2	$\sqrt{3}$	$\sqrt{2}$
方言 2	$\sqrt{5}$	0	$\sqrt{3}$	2	$\sqrt{5}$
方言 3	2	$\sqrt{3}$	0	$\sqrt{5}$	2
方言 4	$\sqrt{3}$	2	$\sqrt{5}$	0	1
方言 5	$\sqrt{2}$	$\sqrt{5}$	2	1	0

この状況を図示すると図 10，図 11 のようになる.

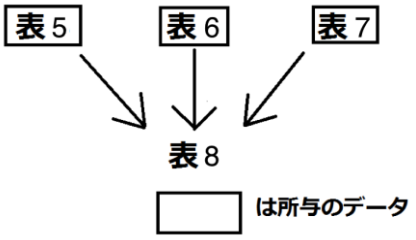

図 10. 表 5，表 6，表 7 の方言データから表 8 の距離行列を求める過程を図示したもの.

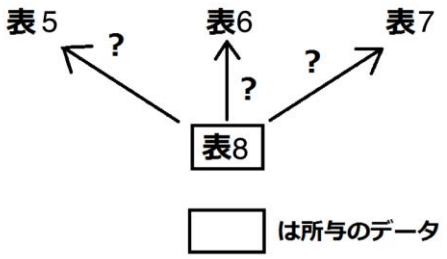

図 11. 表 8 の距離行列から元の方言データを求める過程を図示したもの．表 8 の情報だけでは，元のデータが表 5, 表 6, 表 7 のどれかは分からない.

図 10 から, 表 5, 表 6, 表 7 のデータはいずれも表 8 に帰着することは明らかであるが, 図 11 のように表 8 のデータだけからは元のデータが表 5，表 6，表 7 のどれかであるかを特定することはできない．このように，距離行列(表 8)を与えられても元の方言間における同根性のデータを完全に再現(復元)できないことを「元のデータに対して距離行列は情報が落ちている」と呼ぶこととする．

服部・知里(1960)におけるデータの扱い方に関して，筆者が問題と感じる点は，(方言間の同根性に関する)元のデータが存在するにも関わらず，わざわざ「情報の落ちた」(非)類似度行列を直接計算し，方言間の関係を考察している点にある．「情報の落ちた」(非)類似度行列を使っていることで，データの中の重要な情報が(非)類似度行列に反映されていない可能性は十分にある[9]．　計量多次元尺度構成法もこの点に限界があると言えよう．

統計学に関わるものはできるだけ「データの情報を落とさず」に分析することを意識しているが, その姿勢の背景にはこのような現象があることを読者には知っていただきたい．事実，筆者は小野(2015)において，服部・知里(1960)の方言間の同根性に関する元のデータを分析することによって，アイヌ語の方言に関して，興味深い分析結果を示している．

しかし，Black (1973)や Dyen, Kruskal and Black (1992)など過去における基礎語彙統計学的研究では服部・知里(1960)のように(非)類似度の情報から方言(言語)間の関係を視覚化しており，また，実際の論文では語の同根性に関するデータではなく，それらから計算された(非)類似度の情報のみを載せているものも多い．また，それらの研究の多くが，方言(言語)間の関係を視覚化のために，(非)類似度の情報に対して多次元尺度構成法を適用している．そのような過去の研究について適切な批判や再検討を行うためにも，3 節で述べたようなそれらの手法の持つ利点や限界だけではなく，本節で扱ったような，データの扱いにおける限界を認識する必要がある．

7. 補遺 3－線形代数の必要性について

最後に，線形代数(線型代数)を学ぶことの必要性について説きたい．線形代数は様々な学問の基礎をなしており，統計学もその上に成り立っている．しかし，言語学者の方と意見を交える際に，統計学徒としてなるべく話を砕いてみるものの，なかなか線形代数の基礎の話までは手が回らず，言語学者の方にどこか煙に巻かれたような印象を持たせてしまっている可能性がある．本稿も，例に漏れず，紙面の都合上，線形代数の基礎的な話は足早に終わらせてしまった．

我々が，データ分析を行う際には扱うデータの形式は通常の表 9 のような「行列」の形をしている．

[9] もっともデータの本質的(重要)な情報と，ノイズの情報とをアプリオリに区別することは難しく，本稿のように，方言間の同根性のデータから直接計算した(非)類似度行列が，重要な情報を反映していないとは必ずしも言えない．

表 9. 都道府県別，年齢階級別人口（平成 26 年）（総務省統計局［2016］より作成）

都道府県	総数	0～14 歳	15～19	20～24	25～34	35～44	45～54	55～64	65 歳以上
北海道	5,400	621	238	249	554	739	693	787	1,519
青森	1,321	155	64	53	123	168	172	204	383
岩手	1,284	156	61	53	121	160	160	194	380
宮城	2,328	297	111	124	276	330	291	327	573
秋田	1,037	112	46	35	88	123	127	168	339
山形	1,131	139	54	44	107	138	137	173	338
福島	1,935	241	99	84	190	244	243	298	537
茨城	2,919	376	142	137	312	415	368	416	754
栃木	1,980	257	94	90	216	286	252	288	498
群馬	1,976	258	99	90	196	283	251	270	529
埼玉	7,239	929	347	380	833	1,128	967	918	1,737
千葉	6,197	778	286	307	694	944	819	798	1,571
：	：	：	：	：	：	：	：	：	：
沖縄	1,421	249	80	77	169	203	180	192	270

この時，表 9 は 47 県，9 変数の行列$b_{ij}\{i=1,\dots,47\}\{j=1,\dots,9\}$と表すことができる「行列」を扱う際には，私達が今まで扱っていた実数の四則演算とは異なった性質があるので注意しなければならない．

例えば，実数においては，

$ab=0$ ならば$a=0$ もしくは $b=0$

であるが，「行列」においてはこの性質が成立しない．例えば，

$\boldsymbol{A}=\begin{pmatrix}1/2 & -1/2\\ 1/2 & -1/2\end{pmatrix}, \boldsymbol{B}=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$の時，$\boldsymbol{AB}=\begin{pmatrix}1/2-1/2 & 1/2-1/2\\ 1/2-1/2 & 1/2-1/2\end{pmatrix}$[10] $=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$

よって，$\boldsymbol{AB}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$ ならば $\boldsymbol{A}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$ もしくは $\boldsymbol{B}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$とはならない．

また，実数における$ab=ba$ という性質も「行列」では成立しない．

例えば，$\boldsymbol{BA}=\begin{pmatrix}1/2+1/2 & -1/2-1/2\\ 1/2+1/2 & -1/2-1/2\end{pmatrix}=\begin{pmatrix}1 & -1\\ 1 & -1\end{pmatrix}$ であるが，$\boldsymbol{AB}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$であった．

よって，$\boldsymbol{AB}\neq\boldsymbol{BA}$となる．

このように我々が行列を扱う場合には，実数での四則演算とは異なった性質を扱わなければならない．そして，そのような行列の性質を体系化した学問が「線形代数」である．

以上，最後に簡単であるが，線形代数を学ぶ必要性を説明した．

8. 結語

本論文では，前半部において，「計量多次元尺度構成法」を実際に服部・知里(1960)のデ

[10] 行列の積の計算については 6.0.3 節を参照

ータに適用した．結果，基礎語彙統計学的研究における「計量多次元尺度構成法」の有用性を例示した．特に，単なる非類似度の情報を多次元の軸の情報に整理することは方言間の関係の考察において有用であることが示唆された．このことから，距離行列から方言間の関係性を視覚化するという統計分析は，既存の方言学における仮説を検証する簡便な手法としては有効であると考えられる．本論文の後半においては，最初に，「計量多次元尺度構成法」の数学的な背景をできるだけ簡明に説明し，「計量多次元尺度構成法」の方法論上の有効性と限界を示した．次に，データの扱い方の点から，服部・知里(1960)の問題点を指摘した．統計分析の際には如何に「データの情報を落とさずに分析するか」が重要であることを述べ，(「計量多次元尺度構成法」を含めた)オリジナルのデータから直接距離行列を求める方法は統計学的には望ましくないことを指摘した．本論文によって，「計量多次元尺度構成法」の有効性と限界を提示することによって，「計量多次元尺度構成法」を用いた過去の基礎語彙統計学研究を中心とした様々な言語研究を批判，再検討する素地を与えた．本論文によって，言語学データへの統計手法の適用について言語学の中において理解が深まり，言語学者と統計学者の相補的かつ学際的な研究が進展することを望む．

参考文献

Black, Paul (1973) Multidimensional Scaling Applied to Linguistic Relationships. *Cahiers de l'Institut de Linguistique de Louvain* 3: n5-6.

Dyen, Isidore, Joseph B. Kruskal and Black Paul (1992) An Indoeuropean classification: A lexicostatistical experiment. *Transactions of the American Philosophical Society* 82(5): iii-132.

Eckart, Carl, and Young Gale (1936) The approximation of one matrix by another of lower rank. *Psychometrika* 1(3): 211-218.

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」『季刊民族學研究』24(4): 307-342.

Miyadera, Takayuki (2015) An examination of regionality in a configuration of smallest space analysis using Loevinger's homogeneity coefficient. *Quality & Quantity*, 49(3): 1203-1218.

永田靖・棟近雅彦 (2001) 『多変量解析法入門』東京: サイエンス社

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.

小野洋平 (2015) 「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」『北方人文研究』8: 25-41.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. (Accessed on 2016-09-01)

齋藤正彦 (1966) 『線型代数入門』東京: 東京大学出版会

齋藤堯幸・宿久洋 (2009) 『関連性データの解析法－多次元尺度構成法とクラスター分析－』東京: 共立出版

総務省統計局 (2016) 『日本の統計 2016』(Accessed on 2016-10-31)

Swadesh, Morris (1955) Toward greater accuracy in lexicostatistic dating. *International Journal of American Linguistics.* 21: 121-137.

Possibilities and Limitations on Multidimensional Scaling: an Exercise in Ainu

Yohei ONO
(Visiting Research Fellow, The Institute of Statistical Mathematics)

The main objective of this paper is to clarify possibilities and limitations on Multidimensional Scaling (MDS) in linguistics, taking the (dis-) similarity data in Hattori and Chiri (1960) as an example. The present author first shows the advantages of MDS, which can transform the dissimilarity data into multidimensional coordinate systems, and we can interpret those plots with the linguistic knowledge. However, MDS is an approach, which cannot be recommended from statistical point of view, in which the dissimilarity or distance matrix is calculated directly from the cognacy data. In this paper, the author then demonstrates that such method can omit the core information from the original data. Indeed, Ono (2015) found out the interesting insight on Ainu dialect with the statistically recommended approach using the original data of Hattori and Chiri (1960). However, the way the author, as statistician, does not think as desirable is one approach that can be easily used for non-specialist of statistics to test the hypothesis about the relationship of languages or dialects and it is useful for linguists to understand how MDS works in the frame of statistics because some previous lexicostatistical researches, which applied MDS to cognacy data, should be reconsidered from this point of view. Therefore, demonstrating the possibilities on MDS, the present author introduces (metric) Multidimensional Scaling, applies it to the (dis-) similarity data in Hattori and Chiri (1960) and then reconfirms statistically one of the characteristics in Ainu on dialects in and around Hiratori, which was indicated many time philologically but first verified statistically in Ono (2015). Through this introduction about possibilities and limitations on MDS, the author desires to promote the fruitful reexamination of linguistic researches using MDS especially in lexicostatistics, the better understanding on the application of statistics to linguistic data among linguists and complementary studies among linguists and statisticians.

（おの・ようへい　metrics.languages@gmail.com）