

補遺

以下では、本稿で行なった対応分析の計算過程について、具体例を通して説明する。

表 5 は、「歯 (43. tooth)」と「何 (7. what)」という基礎語彙項目に関する 19 地点のアイヌ語の語形一覧である。

	「歯 (43. tooth)」	「何 (7. what)」
八雲 (1)	mimak	nep
長万部 (2)	nimak	nep
幌別 (3)	nimak	nep, hemanta
平取 (4)	nimak	hemanta
貫気別 (5)	mimak	hinta
新冠 (6)	mimak	hemanta
様似 (7)	imak	nep
帯広 (8)	imak	nep
釧路 (9)	imak	nep
美幌 (10)	imak	nep
旭川 (11)	nimak	nep
名寄 (12)	nimak	nep
宗谷 (13)	imak	hemata, nep
落帆 (14)	imah	hemata
多蘭泊 (15)	imak	hemata
真岡 (16)	imah	hemata
白浦 (17)	imah	hemata
ライチシカ (18)	imah	hemata
内路 (19)	imak	hemata

表 5 : 「歯 (43. tooth)」と「何 (7. what)」の語形 (服部・知里 (1960) より作成)

まず、表 5 について各語形の有無を 0, 1 に変換した表 6 を得る。例えば、本研究のように旭川方言を中心として各方言との類似度について対応分析により可視化する場合、表 6 は表 7 のように変換される。旭川方言は、「歯 (43. tooth)」に *nimak*、「何 (7. what)」に *nep* という語形を持つ。表 7 では *nimak* と *nep* 以外の語形のデータは全て 0 となり、旭川方言以外で *nimak* と *nep* を持つ方言のデータが 1 として残る。

次に、表 7 を具体的な距離とともに対応分析によって可視化した方法の概略を述べる。まず、表 7 の 0, 1 データのまま方言の距離を直接計算した場合、方言間の距離にバイアス (偏り) が生じることを確認しよう。表 7 で *nimak* の語形を有するのは、長万部、幌別、平取、旭川、名寄の 5 方言、*nep* の語形を有するのは、八雲、長万部、幌別、様似、帯広、釧路、美幌、旭川、名寄、宗谷の 10 方言である。よって、*nimak* を持つ方言数より *nep* を持つ方言数が合計すると多い。この合計の方言数の違いは、方言間の距離の違いを生じさせるが、対応分析はこの距離の補正も目指したものである。例えば、平取方言は旭川方言と同じ

表7：旭川方言との同一語形に関する各方言の語形の有無を0,1データへ変換した結果
(1は旭川の語形が存在し、0は存在しないことに対応)。

そこで、対応分析では、表7を表8に変換する。表8では各行のデータを各行の合計で割っている。この変換によって、語形の合計数の違いによる方言の類似度の異なりに対するバイアスが補正される。

	mimak	nimak	imak	imah	hemata	hemanta	hinta	nep	合計
八雲 (1)	0	0	0	0	0	0	0	1	1
長万部 (2)	0	0.5	0	0	0	0	0	0.5	1
幌別 (3)	0	0.5	0	0	0	0	0	0.5	1
平取 (4)	0	1	0	0	0	0	0	0	1
真気別 (5)	0	0	0	0	0	0	0	0	0
新冠 (6)	0	0	0	0	0	0	0	0	0
様似 (7)	0	0	0	0	0	0	0	1	1
帯広 (8)	0	0	0	0	0	0	0	1	1
釧路 (9)	0	0	0	0	0	0	0	1	1
美幌 (10)	0	0	0	0	0	0	0	1	1
旭川 (11)	0	0.5	0	0	0	0	0	0.5	1
名寄 (12)	0	0.5	0	0	0	0	0	0.5	1
宗谷 (13)	0	0	0	0	0	0	0	1	1
落帆 (14)	0	0	0	0	0	0	0	0	0
多蘭泊 (15)	0	0	0	0	0	0	0	0	0
真岡 (16)	0	0	0	0	0	0	0	0	0
白浦 (17)	0	0	0	0	0	0	0	0	0
ライチシカ (18)	0	0	0	0	0	0	0	0	0
内路 (19)	0	0	0	0	0	0	0	0	0
平均	0	0.158	0	0	0	0	0	0.421	

表8：表7を対応分析に用いる数値に変換した結果
(各行の合計が1となる比率に変換し、小数点第4位で四捨五入した。
「平均」の列は、変換した各列の平均値に対応)

さらに、変換したデータの各列での変換値も計算される。例えば、nimakの列では、 $(0 + 0.5 + 0.5 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0.5 + 0.5 + 0 + 0 + 0 + 0 + 0 + 0) / 19 = 3 / 19 = 0.158$ となる。旭川方言を中心としたデータに対応分析を適用した時、例えば、幌別方言と平取方言の距離は、表8をもとに以下のような計算によって得られる。

$$\text{幌別方言と平取方言の距離} = \sqrt{(0.5 - 1)^2 / 0.158 + (0.5 - 0)^2 / 0.421} = 1.475$$

幌別方言と平取方言を直線的に測ったユークリッド距離 ($\sqrt{(0.5 - 1)^2 + (0.5 - 0)^2}$) に関して、 $(0.5 - 1)^2$ の計算に対応するnimakの列は、 $(0.5 - 0)^2$ の計算に対応するnepの列と比べてその語形を有する方言数が少ない。よって、nimakの列の平均値0.158の逆数をかけることで、 $(0.5 - 1)^2$ での距離(違い)に大きな重みづけをし、当該語形を持つ方言数の違いを補正する。

一方で、 $(0.5 - 0)^2$ の計算に対応するnepの列は、 $(0.5 - 1)^2$ の計算に対応するnimakの

列と比べ、語形を有する方言数が多い。よって、nep の列の平均値0.421の逆数をかけることで、 $(0.5 - 0)^2$ での距離（違い）に小さな重みづけをすることで、当該語形を持つ方言数の違いを補正する。

このようにして、対応分析では、中心となる方言と類似する語形の合計数や、当該語形を有する方言数の違いなどによる方言間の距離のバイアス（偏り）を補正する。また、このような計算を適用することで、対応分析による各方言間の距離が計算される。

さらに、本研究では、例えば、旭川を中心としたデータに対応分析を適用した場合、図の中心には、各方言が有する旭川の語形の特徴を平均的に有する仮想的な方言が配置されると述べた。例えば、幌別と各方言が有する旭川の語形の特徴を平均的に有する仮想的な方言は以下のように計算される。

$$\begin{aligned} & \text{幌別と各方言が有する旭川の語形の特徴を平均的に有する仮想的な方言} \\ & = \sqrt{(0.5 - 0.158)^2 / 0.158 + (0.5 - 0.421)^2 / 0.421} = 0.870 \end{aligned}$$

つまり、表8の「平均」の列にある数値は、各方言における旭川方言と類似する語形データの特徴を平均的に有する仮想的な方言の数値であり、各方言間の距離を求めた方法によって、仮想的な方言との距離が計算される。対応分析では、この計算された距離をもとにして、方言間の関係を図8～図14のような「地図」のような形にプロットする。