東京外国語大学国際日本研究センター対照日本語部門第 6 回研究会
『外国語と日本語との対照言語学的研究』（2012/7/21）

広域スペイン語語彙バリエーション研究における新しい数量化の試み

－日本語計量言語地理学の方法に学ぶ－

上田博人（東京大学）

はじめに[1]

ヨーロッパ（スペイン）、南北アメリカ大陸、およびアフリカ（赤道ギニア共和国）の広大な地域で使用されるスペイン語語彙の地理的変異については多くの研究がなされているが[2]、その大部分は語彙目録の記述的研究にとどまり、計量方言学(Dialectometry)の方法を取り入れた数量的研究はきわめて少ない。残念ながらスペイン語計量言語地理学は一部の例外を除けば、言語形式と使用地域という二次元の配列（データ行列）を対象にして様々な多変量解析(Multivariate Analysis)を行う、という井上史雄氏の研究(1994, 2001, 2007)を代表とする日本語計量方言研究の水準に至っていない[3]。日本のスペイン語研究者は進んだ日本語計量言語地理学の方法を学び、それを広域スペイン語研究に応用することができる、という恵まれた環境にある。

私たちは 1993 年から継続して広域スペイン語語彙バリエーションを研究してきた（末尾の参考文献目録を参照）。以下ではこの研究計画の概要を簡単に紹介し、１つの概念（罵言）にある語彙バリエーションを例として私たち独自の数量化の方法を、これまでに学んだ日本語言語地理学研究でよく利用されている多変量解析法と比較しながら説明し、その応用可能性について考察したい。スペイン語の個々の語形のバリエーションと分布については別の機会に発表してあるので(Ueda 2005)、今回の研究会では割愛し調査・分析法について扱う。

１．資料

私たちの研究計画 VARILEX は Variación Léxica del Español en el Mundo「世界の中のスペイン語語彙バリエーション」を略した名称である[4]。これまでおよそ 20 年間にわたって毎年語彙

---

[1] 本研究は日本学術振興会の科学研究費助成の援助による研究成果の一部である（「スペイン語語彙バリエーションの総合的研究の完成」基盤研究(C), H24-27, 24520453）

[2] 次を参照：Cahuzac 1980; Chuchuy 1993; Haensch y Werner 1993; Kany 1962; Kühl de Mones 1993; Lope Blanch 1978; López Morales 1986.; Marrone 1974; Moreno de Alba 1992; Moreno Fernández1993; Rabanales 1987.

[3] 半沢(2007: 179)は「国立国語研究所による戦後の言語生活研究が統計数理研究所と共同で行われたことからも分かるとおり、日本語方言データに多変量解析を適用した研究の歴史は古く、豊富な蓄積を持っている」と述べている。

[4] 「ラテンアメリカ言語学文献学会」Asociación del Lingüística y Filología de América Latina (ALFAL)が 1993 年メキシコ・ベラクルスで開催されたとき、東京外国語大学の高垣敏博氏と私は学会本部の賛同を得て VARILEX 計画を立ち上げた(Takagaki 1993; Ueda 1994)。その後

チェックリスト法(Wood 1990)による質問票を郵送し、回答された資料を独自に開発した言語データ処理プログラムで分析し、その結果をインターネットで公開してきた。スペイン語圏諸都市に在住する研究者の協力を得て毎年 4 名のインフォーマント（39 歳以下・40 歳以上×男性・女性）から 200 ほどの質問事項の回答を送っていただき、これをコンピュータ処理する、という手順を進めてきた。現在までにおよそ 1500 の概念について調査し、次のサイトで資料を公開してきた（図 1.1）。

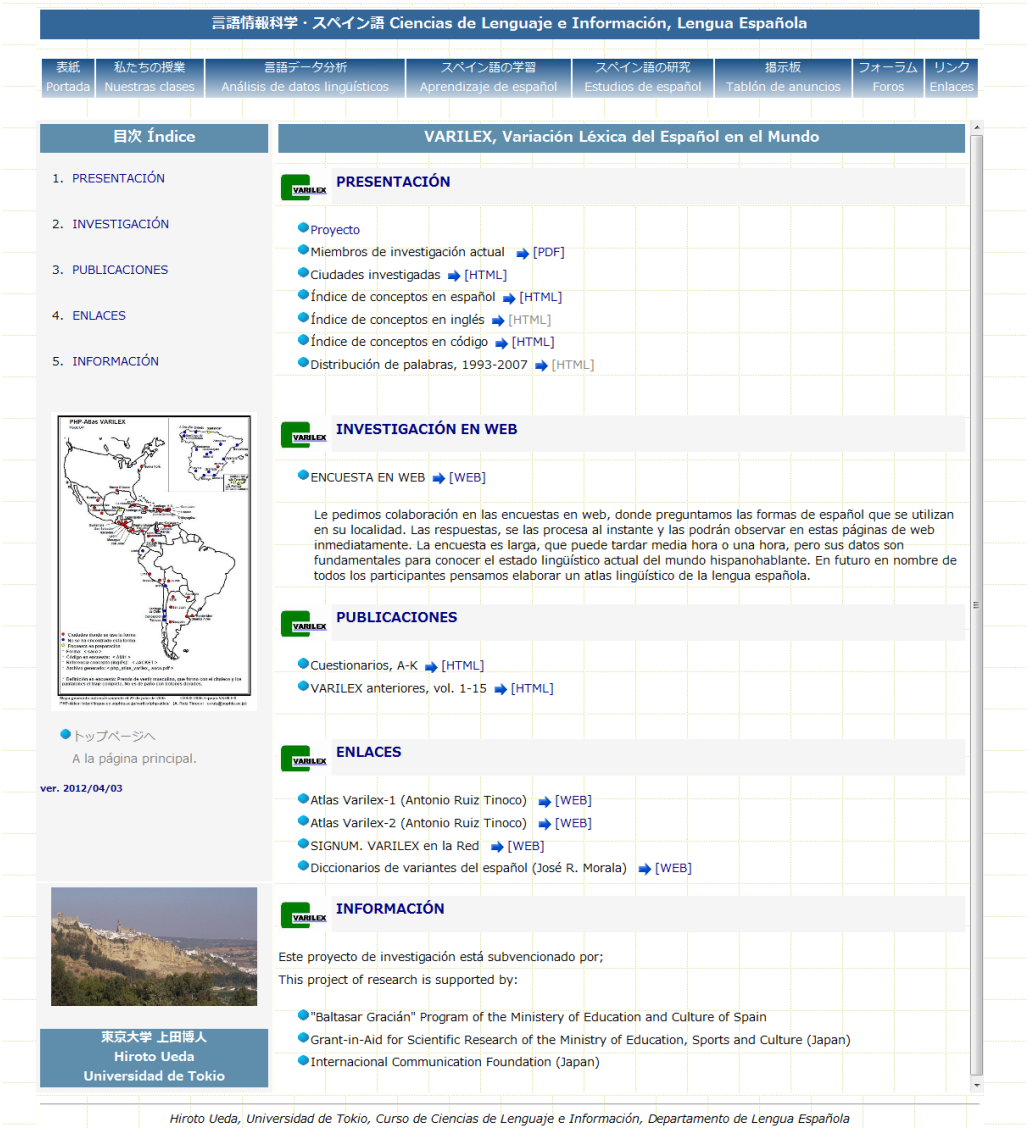


図 1.1. http://lecture.ecc.u-tokyo.ac.jp/~cueda/varilex/index.html

以下ではその中で語形の変異が最も多く観察された[D140] FOOL: Forma de insultar a una persona, refiriéndose a su falta de inteligencia.（FOOL : 頭が悪いと言って人をののしる言葉）を取り上げる。質問票を用意するにあたっては先行文献や辞書など（Carbonell: 2000, Casas: 1994, Escobar: 1986, Martín: 1974, Sanmartín Sáez: 1998, Ruiz 2001）を参考にして選択候補

Antonio Ruiz Tinoco 氏（上智大学）と青砥清一氏（神田外語大学）が参加した。

となる語彙リストを用意した[5]。実際の調査ではさらに多くの語彙を採集した[6]。

収集した資料は縦軸に語形、横軸に調査地点を配置し、二次元の行列の中で該当する回答数を載せる。これは一般のクロス集計表、Excel のピボットテーブルと同じである(図 1.2)。

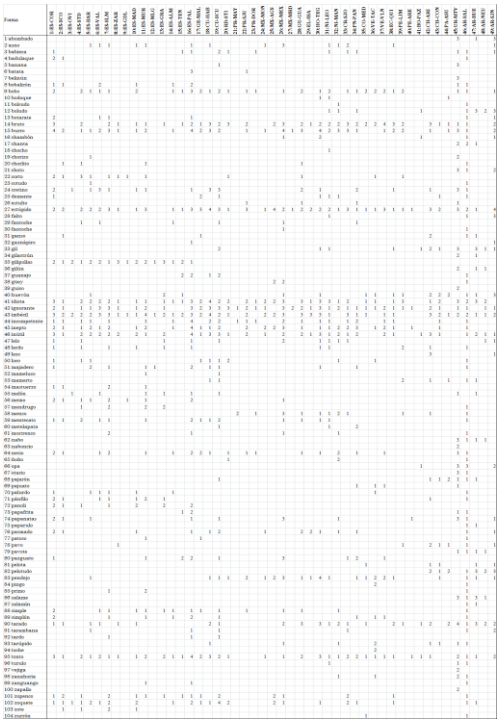


図 1.2. データ行列（地理的分布：数値 0-4）

ここで使用できるスペースの関係でデータ行列のすべてを示すことはできないが、次の図 1.3 に冒頭部分だけを拡大表示しておく[7]。

---

[5] 当初の語彙リストは次のとおりである : abodocado, abombado, alberja, alcaucil, asno, babieca, badulaque, bambaco, banana, batata, belinún, belloto, beocio, bobalicón, bobeta, bobo, bodoque, bolonio, bolsa, bolsón, bolsudo, boludo, boncha, botarate, bruto, burro, cachirulo, caspiento, caspudo, chacarón, chambón, chanta, chauchón, chocho, chorizo, chorlito, choto, cirolo, citrulo, corto, cotudo, cretino, croto, demente, estúpido, estulto, faltado, falto, fantoche, fantoche, fantoso, ganso, gaznápiro, gedeón, gil, gilastrún, gilí, gilipollas, gilún, guanajo, güey, guiso, hueva, huevón, idiota, ignorante, imbécil, incompetente, inepto, inútil, junípero, lelo, lerdo, leso, lila, loco, majadero, mamacallos, mameluco, mamerto, mapelotudo, mastuerzo, melón, memo, mendrugo, menso, mentecado, mentecato, metelapata, mochilón, mostrenco, ñoño, nabo, naboncio, necio, opa, orate, otario, pánfilo, pásula, pajarón, pajuato, palomo, palurdo, panoli, papafrita, papanatas, paparulo, pasmado, pastenaca, patoso, pavo, pavote, pazguato, pelandrún, pelota, pelotudo, pendejo, pendiolo, pingo, porro, primo, salame, salamín, sandio, sansirolé, simple, simplón, soroco, sota, tagüicho, tagüirongo, taradelli, tarado, tarambana, tardo, tarúpido, toche, tolombelo, tolongo, tonto, trolón, turulo, vejiga, vejigón, zampaboya, zanahoria, zanguango, zapallo, zopenco, zoquete, zote, zurrón.

[6] 追加された語彙リスト：infeliz, güevón, papón, pringao, impresentable, torpe, retrasado mental, cantollo, cerrojo, pollaboba, tolete, inculto, baboso, mal nacido, dundo, babas, moco, sope, limitado, sonso, mermo, badulaque, pasguato, tolete, sirguango, majarón, odioso, animal.

[7] 調査地点は次のとおりである。[ES-COR] La Coruña (España), [ES-SCO] Santiago de Compostela (España), [ES-OVI] Oviedo (España), [ES-STD] Santander (España), [ES-SLM]

Forma	1:ES-COR	2:ES-SCO	3:ES-OVI	4:ES-STD	5:ES-BAR	6:ES-VAL	7:ES-SLM	8:ES-ZAR	9:ES-GDL	10:ES-MAD	11:ES-MUR
1:abombado											
2:asno					1	1	1			1	1
3:babieca	1										
4:badulaque	2	1									

図 1.3. データ行列：冒頭部分

このクロス集計表（データ行列）はいわば言語地理データの記述のレベルを示すものである。従来のスペイン語方言学研究はこの段階で終了してあることが多いが多変量解析を応用した計量地理言語学ではこれが分析の出発点となる。

２．方法

２．１　データ行列の統合化

ここでは私たちの研究計画 VARILEX で試みているデータ行列の「統合化」について説明する。統合化とは、たとえば次の下左図のようなデータ行列の行(d-1… 5)と列(v-1…4)を並べ替えて、なるべく反応点(v)の分布を一定の位置に集中させる方法である。

Salamanca (España), [ES-ZAR] Zaragoza (España), [ES-BAR] Barcelona (España), [ES-GDL] Guadalajara (España), [ES-MAD] Madrid (España), [ES-VAL] Valencia (España), [ES-GRA] Granada (España), [ES-MLG] Málaga (España), [ES-TEN] Santa Cruz de Tenerife (España), [ES-PAL] Las Palmas de Gran Canaria (España), [GE-MAL] Malabo (Guinea Ecuatorial), [CU-HAB] La Habana (Cuba), [CU-SCU] Santiago de Cuba (Cuba), [RD-STI] Santiago (República Dominicana), [PR-SJU] San Juan (Puerto Rico), [PR-DOR] Dorado (Puerto Rico), [PR-MAY] Mayagüez (Puerto Rico), [MX-MON] Monterrey (México), [MX-AGS] Aguas Calientes (México), [MX-MEX] Ciudad de México (México), [MX-MRD] Mérida (México), [GU-GUA] Guatemala (Guatemala), [EL-SSV] San Salvador (El Salvador), [HO-TEG] Tegucigalpa (Honduras), [NI-LEO] León (Nicaragua), [NI-MAN] Managua (Nicaragua), [CR-SJO] San José (Costa Rica), [PN-PAN] Panamá (Panamá), [CO-MED] Medellín (Colombia), [VE-MED] Mérida (Venezuela), [VE-VLN] Valencia (Venezuela), [VE-TAC] Tachira (Venezuela), [EC-QUI] Quito (Ecuador), [PE-LIM] Lima (Perú), [PE-ARE] Arequipa (Perú), [BO-PAZ] La Paz (Bolivia), [CH-ARI] Arica (Chile), [CH-CON] Concepción (Chile), [PA-ASU] Asunción (Paraguay), [UR-MTV] Montevideo (Uruguay), [AR-SAL] Salta (Argentina), [AR-SJN] San Juan (Argentina), [AR-NEU] Neuquén (Argentina), [AR-BUE] Buenos Aires (Argentina).

Lv	v-1	v-2	v-3	v-4
d-1	v	v		
d-2			v	
d-3		v		
d-4			v	v
d-5	v	v	v	

→

Lv	v-2	v-1	v-3	v-4
d-3	v			
d-1	v	v		
d-5	v	v	v	
d-2			v	
d-4			v	v

統合化にはさまざまな方法が考えられる。次は Cahuzac (1980)のラテンアメリカスペイン語「農夫」の語形分布資料を使って各種の統合化を行った結果である。次がデータ行列である。

語形	AR	BO	CH	CO	CR	CU	EC	EL	GU	HO	MX	NI	PA	PE	PN	PR	RD	UR	VE
1 cacahuero				v															v
2 cafetalista						v					v					v			
3 camilucho	v												v					v	
4 campero	v												v					v	
5 camperuso				v															v
6 campirano				v	v			v	v	v		v			v				v
7 campiruso					v			v	v	v		v			v				
8 campista					v			v	v	v	v	v			v	v			v
9 campusano	v														v			v	
10 campuso					v			v	v	v		v							
11 colono																v	v		
12 comparsa	v												v					v	
13 conuquero				v		v										v	v		v
14 coquero		v					v							v					
15 chagrero				v			v												
16 changador	v												v					v	
17 chilero					v			v	v	v	v	v			v				
18 chuncano	v												v					v	
19 enmaniguado						v										v	v		
20 estanciero	v												v					v	
21 gaucho	v	v											v					v	
22 guajiro						v											v		
23 guanaco					v			v	v	v		v			v				
24 guaso	v	v	v			v	v							v					
25 huasicama				v			v												
26 huertero	v		v								v			v					
27 hulero					v			v	v	v	v	v			v				
28 invernador	v		v										v	v				v	
29 jibaro																v	v		
30 lampero	v	v												v					
31 lanudo				v			v												v
32 llanero				v															v
33 macanero					v						v								
34 manuto															v		v		
35 montero						v											v		
36 montubio				v		v	v							v			v		
37 paisano							v							v			v		
38 pajuerano						v	v							v					
39 partidario	v	v																v	
40 payazo				v															v
41 piona	v												v					v	
42 ranchero						v					v				v		v		
43 rondin		v																	
44 sabanero				v															v
45 veguero															v		v		
46 viñatero	v		v										v	v				v	
47 yanacón	v	v	v											v					

図 2.1a. データ行列

データ行列では縦軸に語形を、横軸に国名コードをそれぞれアルファベット順に並べてある。このデータ行列に各種の統合分析を適用すると、次のようにさまざまな分布パタンが提示される。

語形	PA	UR	AR	BO	CH	GU	HO	NI	EL	CR	PE	PN	EC	MX	CO	VE	CU	RD	PR
16 changador	v		v	v															
12 comparsa	v		v	v															
41 piona	v		v	v															
18 chuncano	v		v	v															
20 estanciero	v		v	v															
3 camilucho	v		v	v															
4 campero	v		v	v															
21 gaucho	v		v	v	v														
39 partidario			v	v	v														
43 rondin					v														
10 campuso							v	v	v	v	v								
46 viñatero	v		v	v		v							v						
28 invernador	v		v	v		v							v						
30 lampero				v	v								v						
9 campusano			v	v										v					
47 yanacón				v	v	v							v						
23 guanaco							v	v	v	v	v			v					
7 campiruso							v	v	v	v	v			v					
17 chilero							v	v	v	v	v			v					
27 hulero							v	v	v	v	v			v					
26 huertero				v		v							v						
14 coquero					v								v		v				
6 campirano							v	v	v	v	v			v		v	v		
33 macanero											v				v				
24 guaso				v	v	v							v					v	
8 campista							v	v	v	v	v			v		v			v
38 pajuerano													v					v	
15 chagrero														v					
25 huasicama														v					
37 paisano													v						v
31 lanudo														v		v			
36 montubio													v					v	
42 ranchero														v					
44 sabanero															v	v			
40 payazo															v	v			
1 cacahuero															v	v			
5 camperuso															v	v			
32 llanero															v	v			
45 veguero													v						v
34 manuto													v						v
2 cafetalista														v			v		v
13 conuquero															v	v	v	v	v
22 guajiro																	v	v	
35 montero																	v	v	
19 enmaniguado																	v	v	v
11 colono																		v	v
29 jibaro																		v	v

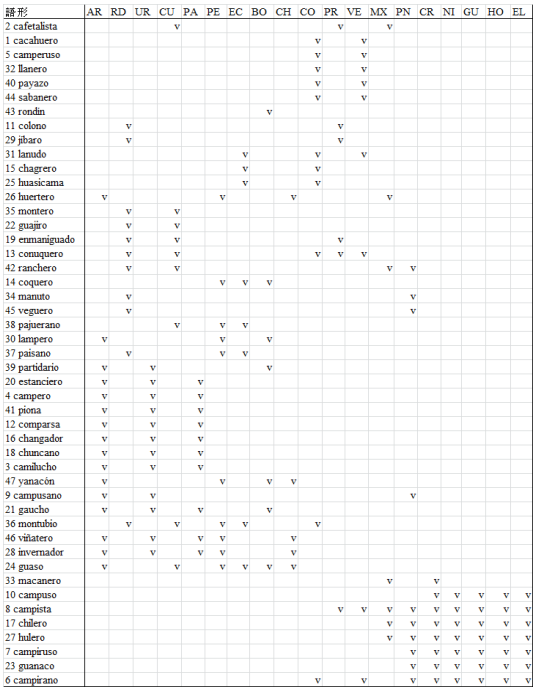
図 2.1b. 原点距離統合分析

語形	AR	UR	PA	CH	PE	EC	CO	VE	PR	RD	CU	MX	CR	EL	GU	HO	NI	PN	BO
1 cacahuero							v	v											
5 camperuso							v	v											
32 llanero							v	v											
40 payazo							v	v											
44 sabanero							v	v											
31 lanudo						v	v	v											
15 chagrero						v	v												
25 huasicama						v	v												
36 montubio					v	v	v			v	v								
37 paisano						v	v			v									
14 coquero					v	v													v
24 guaso	v				v	v					v								v
47 yanacón	v				v	v													v
30 lampero	v					v													v
39 partidario	v	v																	v
21 gaucho	v	v	v																v
3 camilucho	v	v	v																
4 campero	v	v	v																
12 comparsa	v	v	v																
16 changador	v	v	v																
18 chuncano	v	v	v																
20 estanciero	v	v	v																
41 piona	v	v	v																
28 invernador	v	v	v	v	v														
46 viñatero	v	v	v	v	v														
26 huertero	v			v	v														
33 macanero												v	v						
17 chilero												v	v	v	v	v	v	v	
27 hulero												v	v	v	v	v	v	v	
7 campiruso													v	v	v	v	v	v	
23 guanaco													v	v	v	v	v	v	
10 campuso													v	v	v	v	v		
6 campirano								v	v				v	v	v	v	v	v	
8 campista									v			v	v	v	v	v	v	v	
2 cafetalista									v		v								
19 enmaniguado									v	v	v								
11 colono									v	v									
29 jibaro										v									
13 conuquero								v	v	v	v								
22 guajiro										v	v								
35 montero										v	v								
42 ranchero										v	v	v						v	
34 manuto										v								v	
45 veguero										v								v	
9 campusano	v	v																v	
38 pajuerano						v	v					v							
43 rondin																			v

図 2.1c. 隣接距離統合分析

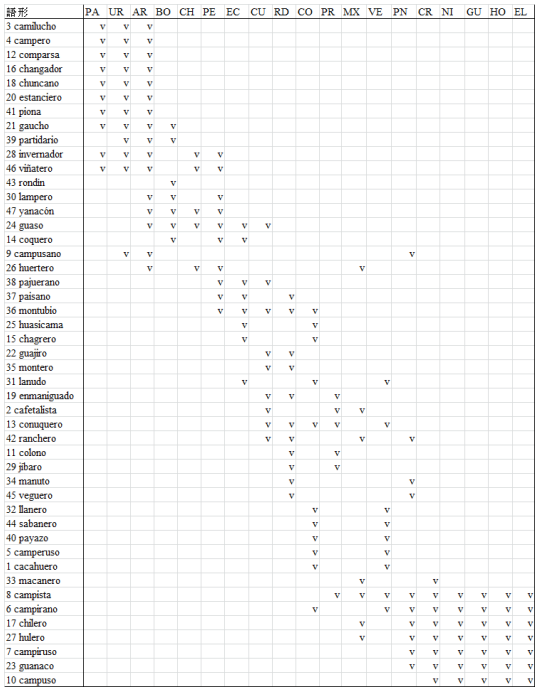
語形	AR	BO	PA	UR	EC	PE	CH	RD	CO	EL	GU	HO	NI	CR	PN	MX	PR	VE	CU
6 campirano									v	v	v	v	v	v	v			v	
7 campiruso										v	v	v	v	v	v				
23 guanaco										v	v	v	v	v	v				
33 macanero															v	v			
27 hulero										v	v	v	v	v	v	v			
17 chilero										v	v	v	v	v	v	v			
8 campista										v	v	v	v	v	v	v	v	v	
10 campuso										v	v	v	v	v	v				
1 cacahuero									v									v	
32 llanero									v									v	
5 camperuso									v									v	
40 payazo									v									v	
44 sabanero									v									v	
31 lanudo						v			v									v	
2 cafetalista																v	v		v
15 chagrero						v			v										
25 huasicama						v			v										
9 campusano	v			v											v				
18 chuncano	v		v	v															
41 piona	v		v	v															
20 estanciero	v		v	v															
3 camilucho	v		v	v															
12 comparsa	v		v	v															
16 changador	v		v	v															
4 campero	v		v	v															
21 gaucho	v	v	v	v															
39 partidario	v		v	v															
13 conuquero									v	v							v	v	v
34 manuto									v						v				
45 veguero									v						v				
28 invernador	v			v	v		v	v											
46 viñatero	v			v	v		v	v											
24 guaso	v		v	v		v	v	v									v		
11 colono									v									v	
29 jibaro									v									v	
19 enmaniguado									v									v	v
38 pajuerano						v	v												v
14 coquero			v			v	v												

図 2.1d. 主成分統合分析



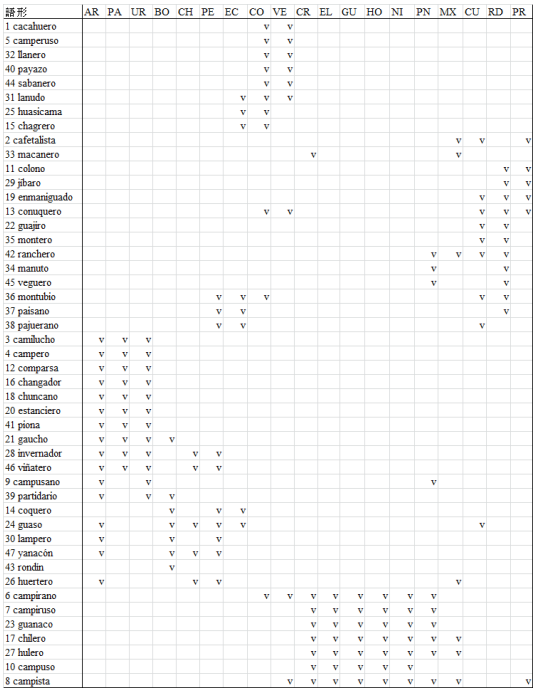
語形	AR	RD	UR	CU	PA	PE	EC	BO	CH	CO	PR	VE	MX	PN	CR	NI	GU	HO	EL
2 cafetalista				v							v		v						
1 cacahuero										v		v							
5 camperuso										v		v							
32 llanero										v		v							
40 payazo										v		v							
44 sabanero										v		v							
43 rondin								v											
11 colono		v									v								
29 jibaro		v									v								
31 lanudo							v			v		v							
15 chagrero							v			v									
25 huasicama							v			v									
26 huertero	v					v			v					v					
35 montero		v		v															
22 guajiro		v		v															
19 enmaniguado		v		v							v								
13 conuquero		v		v						v	v	v							
42 ranchero		v		v										v	v				
14 coquero						v	v	v											
34 manuto		v													v				
45 veguero		v													v				
38 pajuerano				v		v	v												
30 lampero	v					v		v											
37 paisano		v				v	v												
39 partidario	v		v					v											
20 estanciero	v		v		v														
4 campero	v		v		v														
41 piona	v		v		v														
12 comparsa	v		v		v														
16 changador	v		v		v														
18 chuncano	v		v		v														
3 camilucho	v		v		v														
47 yanacón	v					v		v	v										
9 campusano	v		v											v					
21 gaucho	v		v		v			v											
36 montubio		v		v		v	v			v									
46 viñatero	v		v		v	v			v										
28 invernador	v		v		v	v			v										
24 guaso	v			v		v	v	v	v										
33 macanero													v		v				
10 campuso															v	v	v	v	v
8 campista											v	v	v	v	v	v	v	v	v
17 chilero													v	v	v	v	v	v	v
27 hulero													v	v	v	v	v	v	v
7 campiruso														v	v	v	v	v	v
23 guanaco														v	v	v	v	v	v
6 campirano										v		v		v	v	v	v	v	v

図 2.1e. 因子統合分析



語形	PA	UR	AR	BO	CH	PE	EC	CU	RD	CO	PR	MX	VE	PN	CR	NI	GU	HO	EL
3 camilucho	v	v	v																
4 campero	v	v	v																
12 comparsa	v	v	v																
16 changador	v	v	v																
18 chuncano	v	v	v																
20 estanciero	v	v	v																
41 piona	v	v	v																
21 gaucho	v	v	v	v															
39 partidario		v	v	v															
28 invernador	v	v	v			v	v												
46 viñatero	v	v	v			v	v												
43 rondin					v														
30 lampero				v	v		v												
47 yanacón				v	v	v	v												
24 guaso				v	v	v	v	v	v										
14 coquero					v		v	v											
9 campusano		v	v												v				
26 huertero				v		v	v						v						
38 pajuerano							v	v	v										
37 paisano							v	v		v									
36 montubio							v	v	v	v	v								
25 huasicama								v		v									
15 chagrero								v		v									
22 guajiro									v	v									
35 montero									v	v									
31 lanudo								v					v						
19 enmaniguado									v	v	v								
2 cafetalista									v		v	v							
13 conuquero									v	v	v	v		v					
42 ranchero									v	v		v			v				
11 colono										v	v								
29 jibaro										v	v								
34 manuto										v					v				
45 veguero										v					v				
32 llanero											v		v						
44 sabanero											v		v						
40 payazo											v		v						
5 camperuso											v		v						
1 cacahuero											v		v						
33 macanero												v			v				
8 campista											v	v	v	v	v	v	v	v	v
6 campirano											v		v		v	v	v	v	v
17 chilero												v		v	v	v	v	v	v
27 hulero												v		v	v	v	v	v	v
7 campiruso														v	v	v	v	v	v
23 guanaco														v	v	v	v	v	v
10 campuso															v	v	v	v	v

図 2.1f. 数量化Ⅲ類統合分析



語形	AR	PA	UR	BO	CH	PE	EC	CO	VE	CR	EL	GU	HO	NI	PN	MX	CU	RD	PR
1 cacahuero								v	v										
5 camperuso								v	v										
32 llanero								v	v										
40 payazo								v	v										
44 sabanero								v	v										
31 lanudo							v	v	v										
25 huasicama							v	v											
15 chagrero							v	v											
2 cafetalista																v	v		v
33 macanero										v						v			
11 colono																		v	v
29 jibaro																		v	v
19 enmaniguado																	v	v	v
13 conuquero								v	v								v	v	v
22 guajiro																	v	v	
35 montero																	v	v	
42 ranchero															v	v	v	v	
34 manuto															v			v	
45 veguero															v			v	
36 montubio							v	v	v								v	v	
37 paisano							v	v										v	
38 pajuerano							v	v									v		
3 camilucho		v	v	v															
4 campero		v	v	v															
12 comparsa		v	v	v															
16 changador		v	v	v															
18 chuncano		v	v	v															
20 estanciero		v	v	v															
41 piona		v	v	v															
21 gaucho		v	v	v	v														
28 invernador		v	v	v		v	v												
46 viñatero		v	v	v		v	v												
9 campusano		v		v											v				
39 partidario		v		v	v														
14 coquero					v		v	v											
24 guaso		v			v	v	v	v									v		
30 lampero		v			v		v												
47 yanacón		v			v	v	v												
43 rondin					v														
26 huertero		v				v	v									v			
6 campirano								v	v	v	v	v	v	v	v				
7 campiruso										v	v	v	v	v	v				
23 guanaco										v	v	v	v	v	v				
17 chilero										v	v	v	v	v	v	v			
27 hulero										v	v	v	v	v	v	v			
10 campuso										v	v	v	v	v					
8 campista									v	v	v	v	v	v	v				v

図 2.1g. クラスター統合分析

統合分析の応用例として Kawasaki (2012)を紹介する。この研究では発行日が記されている多くの中世スペイン語公証文書の言語特徴を原点距離法を使って統合し、その分布パタンをもとに発行日が記されていない文書の年代を推定している。

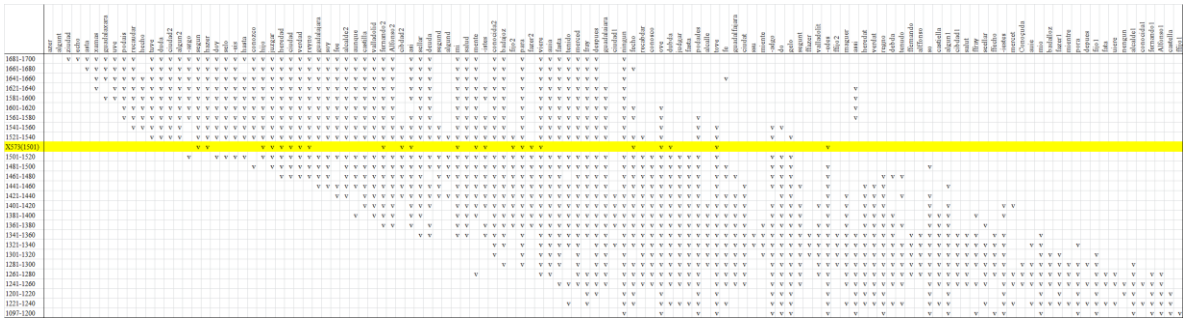


図 2.1h. 中世スペイン語公証文書の年代推定

上の図では縦軸に年代を入れ固定し横軸（言語特徴）を統合化している。この黄色の部分の横行が年代不詳の文献である。これを含めて全体を統合化すると、この行が一定の年代に位置づけられるので、その年代を推定することができる。そのためには適切な言語特徴（年代差を示す、頻度が高い、地域差が少ない、など）を選択し、実験を繰り返さなければならない。

２．２　多次元空間距離による統合化

林知己夫が開発した「数量化Ⅲ類」という多変量解析法では、先のようなデータ行列を縦軸と横軸に与えた一定の数値（以下では統一して「参照値」と呼ぶことにする）をもとに並べ替え、データが二次元行列の対角線の近くに集まるようにする、つまり、データの分布の相関係数が最大になるような参照値を求め、それをもとに並べかえる（これを「パタン化」とよぶ：林 ・樋口・駒沢 1970; 駒澤・橋口 1988)。そのために与える縦軸と横軸の参照値を求める際に線形代数の方法を応用するが、一方、上田(1993) が考案した「原点平均距離法」は文系の学生にとって難解な線形代数を使わない簡便な方法で並べ替えのための参照値を求める[8]。大きなデータの分析結果は数量化Ⅲ類とは異なるが、それでもおおよそのパタン化が達成できる。

たとえば下左図はデータ行列の例であるが、これの縦軸（d-1, 2, …, 5）と横軸（v-1, 2, 3, 4）を並べ替えて下右図のようにパタン化することができる。並べ替えの基準として使う値は反応点の位置情報によって得られる。

Lv	v-1	v-2	v-3	v-4
d-1	v	v		
d-2			v	
d-3		v		
d-4			v	v
d-5	v	v	v	

→

Lv	v-1	v-2	v-3	v-4
d-1	v	v		
d-3		v		
d-5	v	v	v	
d-2			v	
d-4			v	v

このように統合化すると、右図の行に関しては[d-1, 3, 5]と[d-2, 4]がそれぞれ統合化され、列に関しては[v-1, 2]と[v-3, 4]がそれぞれ統合化されていることがわかる。ここで「統合化」(integration)とは反応の分布が互いに近接し、全体で一定の傾向を示すことを意味する。分布の

[8] この方法は Bertin (1977)の手作業による方法を数量化したものである。

相関を高くする、つまり分布図の対角線の近辺に集中させる「パタン化」は統合化の一種である。そのためには、はじめに各行の反応点の原点からの距離の平均を次のようにして計算する。たとえば d-1 は v-1 と v-2 に反応しているので、$1^2 + 2^2$ を計算し、その平均をとって根を開く（下ではルートの記号√を使う代わりに 1/2 を乗数とする）。これはいわゆる多元空間内のユークリッド距離の平均の計算である。

d-1: $[(1^2 + 2^2) / 2]^{1/2}$ = 1.581 (...1)
d-2: $[(3^2) / 1]^{1/2}$ = 3.000 (...4)
d-3: $[(2^2) / 1]^{1/2}$ = 2.000 (...2)
d-4: $[(3^2 + 4^2) / 2]^{1/2}$ = 3.535 (...5)
d-5: $[(1^2 + 2^2 + 3^2) / 3]^{1/2}$ = 2.160 (...3)

この数値（原点平均距離）を基準にして昇順(上の計算式で...で示した)で並べ替えると次のようになる。

Lv	v-1	v-2	v-3	v-4
d-1	v	v		
d-3		v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Lv	係数
d-1	1.581
d-3	2.000
d-5	2.160
d-2	3.000
d-4	3.536

簡単だがこれで一応のパタン化ができている。この場合横軸 v-1 …4 を距離の計算の基準として使っているので、横軸を「外的基準」にしたパタン化と呼ぶことにする。つまり、たとえば、地理的分布が南北や東西、または街道に沿った地点の配置であれば、それを外的基準にすることができる。その基準にしたがって語形を見ると、d-1, 3, 5, 2, 4 という語形の配置が地点の配置に沿っている、と解釈できる。

しかし、広大なスペイン語圏のような対象を扱うときは、地点が必ずしも線上に並ぶことはなく、少なくとも東西・南北の二次元の分布を考えなければならない。さらに、都市と周辺、街道のネットワーク、文化圏、大陸・半島・島嶼部、海岸部と山間部など多くのパラメータが考えられるので、地点の連続線は複雑になる[9]。これを地点と語形の二次元の統合された分布にまとめるには、語形の並べ替えだけでなく地点の並べ替えも必要である。そこで、今度は地点を示す各縦列の原点からの距離を計算する。たとえば地点 v-1 は縦列の 1 番目の語形(d-1)と 3 番目の語形(d-5)に反応しているので、その原点平均距離は次の第 1 式のようになる。以下の地点についても同様である。

v-1: $[(1^2 + 3^{2s}) / 2]^{1/2}$ = 2.236 (...2)

[9] このような多くの変数を同時に扱うには、それぞれの特徴を変数とした多変量解析が有効である。しかし、ここで扱っている原点平均距離法は複雑な様相を示す地点（と語形）を統合化した一元的な線に配置することを目的としている。

v-2: $[(1^2 + 2^2 + 3^2) / 3]^{1/2}$ = 2.160 (...1)
v-3: $[(3^2 + 4^2 + 5^2) / 3]^{1/2}$ = 4.082 (...3)
v-4: $[(5^2) / 1]^{1/2}$ = 5.000 (...4)

この数値によれば v-1 と v-2 が位置を交代しなければならない。その結果が次図である。

Lv	v-2	v-1	v-3	v-4
d-1	v	v		
d-3	v			
d-5	v	v	v	
d-2			v	
d-4			v	v

Lv	係数
d-1	1.581
d-3	1.000
d-5	2.160
d-2	3.000
d-4	3.536

Lv	v-2	v-1	v-3	v-4
係数	2.160	2.236	4.082	5.000

これで第 1 回目の縦と横の並べ替えが終わるが、この段階で再び各横行の原点からの平均距離を計算すると次のようになる。

d-1: $[(1^2 + 2^2) / 2]^{1/2}$ = 1.581 (...2)
d-3: $[(1^2) / 1]^{1/2}$ = 1.000 (...1)
d-5: $[(1^2 + 2^2 + 3^2) / 3]^{1/2}$ = 2.160 (...3)
d-2: $[(3^2) / 1]^{1/2}$ = 3.000 (...4)
d-4: $[(3^2 + 4^2) / 2]^{1/2}$ = 3.535 (...5)

これを見ると、d-1 と d-3 を交替しなければならないことがわかる。そのように並べ替えたのが次の図である。

Lv	v-2	v-1	v-3	v-4
d-3	v			
d-1	v	v		
d-5	v	v	v	
d-2			v	
d-4			v	v

Lv	係数
d-3	1.000
d-1	1.581
d-5	2.160
d-2	3.000
d-4	3.536

Lv	v-2	v-1	v-3	v-4
係数	2.160	2.550	4.082	5.000

さらに各縦列の原点からの平均距離を計算すると次のようになる。

v-2: $[(1^2 + 2^2 + 3^2) / 3]^{1/2}$ = 2.160 (...1)

v-1: $[(2^2 + 3^2) / 2]^{1/2}$ = 2.550 (...2)

v-3: $[(3^2 +4^2 + 5^2) / 3]^{1/2}$ = 4.082 (...3)

v-4: $[(5^2) / 1]^{1/2}$ = 5 (...4)

これで横行も縦列も正しく昇順に並んだので分布パタンは収束したことになる。原点平均距離法で分布がパタン化される理由は、それぞれの行または列の反応点が示す距離の総合値が近いものの位置を近くに寄せ集め、さらにパタンの集合が行列の各地にばらばらに生まれるのではなく[10]、距離の総合値を大小順に並べ替えることによって、全体の推移にグラデーションができるからである。その操作を繰り返すことによって、よりよいパタン化が達成される。大きなデータ行列では繰り返し回数が増えるので数値処理のプログラミングが必要である[11]。

次の図 2.1a は先のデータ行列(図 1.2)の周縁部に縦軸と横軸の原点平均距離係数を与え、グラデーション処理を加えたものである。データ行列は統合されていないので原点平均距離係数はまちまちの値を示している[12]。図 2.1b はデータ行列をパタン化した結果を示している。パタン化した図では縦と横の青色のグラデーションが示すように原点平均距離係数が昇順に並んでいる。そこで、横軸の地点、縦軸の語形、そして左上から右下に徐々に変化する分布パタンの三者に統合して同じ解釈を与えることができる。仮に地点が、おおよそ北→南の並びを示しているならば、語形もおおよそ北→南の配置になり、頻度の分布も左上から右下に向かっておおよそ北→南の流れを示していることになる。以下に、データ行列と比較した原点距離統合分析の結果とそれぞれの地図上の値を示す[13]。

[10] 後述するように、クラスター分析を使った統合化は各地に分布の集合を作る。

[11] ここで採用した平均ユークリッド距離で計算することで基本的なパタン化でできるが、同距離・異分布という問題を回避するために、距離 2 乗和の平均（の 2 乗根）ではなく 3 乗和の平均（の 3 乗根）を求める方法（Minkowsky の距離）を使うことが多い。なお、原点平均距離法によるパタン化はデータ行列の初期状態の違いによって、異なる状態で収束することが多い。これは数量化Ⅲ類による厳密な方法にはないことである。

[12] なお、このデータ行列ではセルの値が先の例のような質的データではなく、0 – 4 の間の整数をとる量的データであるが、距離の計算は同様に可能である。詳細は次のサイトを参照されたい。
http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/index.html

[13] 地図化には埼玉大学の谷謙二作成の地理情報支援システム MANDARA を使用した。
http://ktgis.net/mandara/

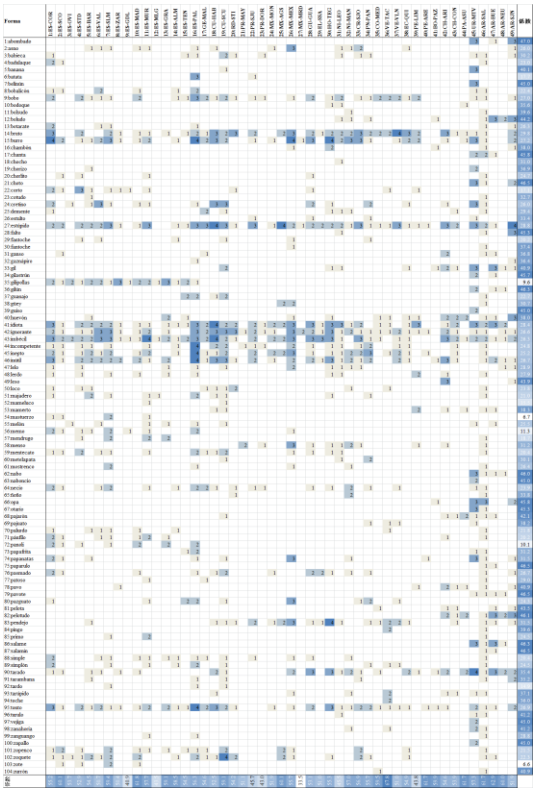


図 2.2a. データ行列

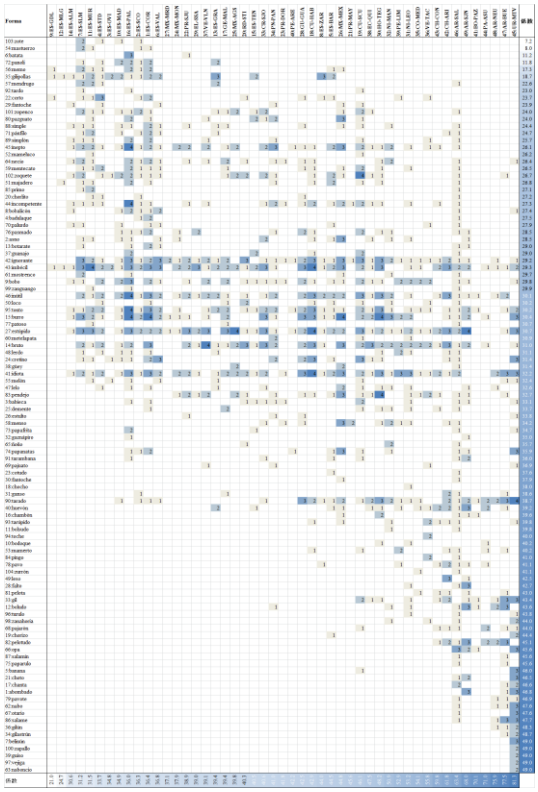


図 2.2b.原点距離統合分析

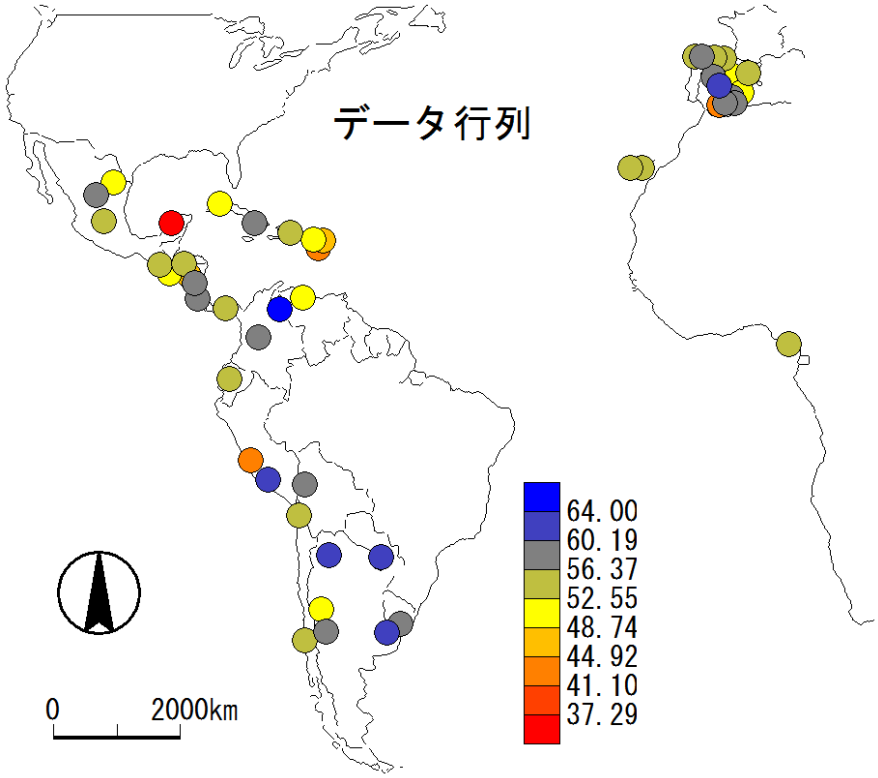


図 2.2c 原点距離地図：データ行列

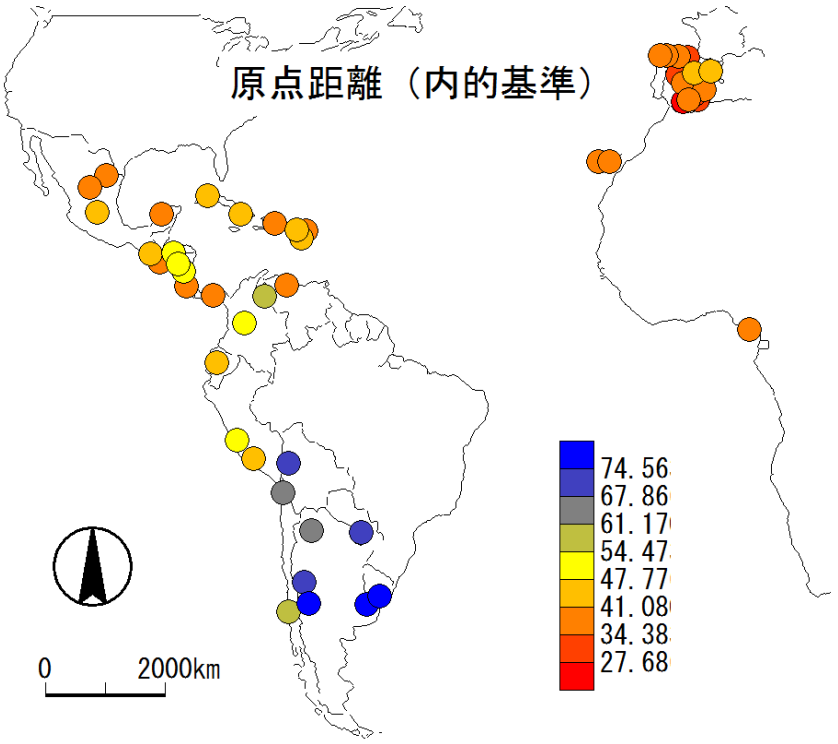


図 2.2d 原点距離地図（内的基準）

原点距離法はデータ行列の行全体、または列全体が作る多次元空間内の距離を計算して、その結果に基づいて行と列の並べ替えを行っている。その統合化によって反応点は対角線に近い位置に集中する。一方、次に見る「隣接距離統合分析」では、行（または）列どうしの反応点の差の自乗を全部足して、どちらかに反応のあるケースの数で割り、その根を求める。その数が一番小さい行（または列）を隣に置く、という操作を全体の行（または列）について行う。つまり、それぞれの行に一番近い行を選んで、次々に並べ替える、という手順になる。列についても同様である。その結果は次のような分布を示す。

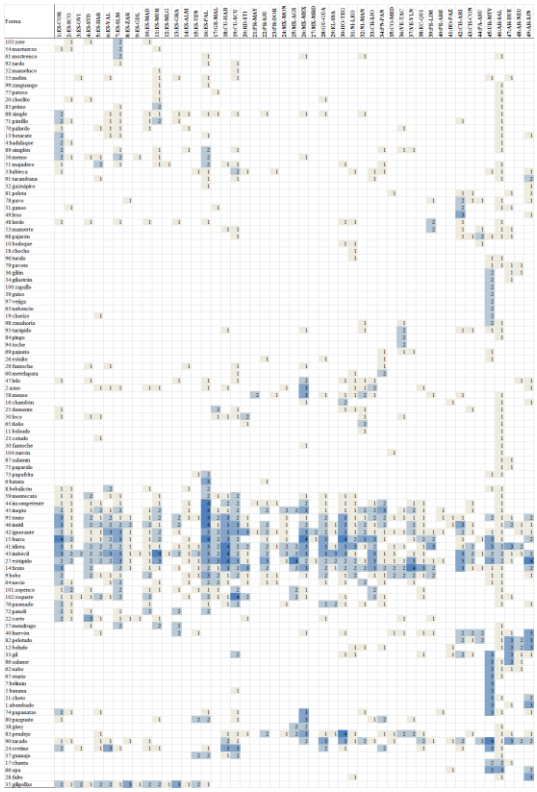


図 2.2e 隣接距離統合分析

隣接距離統合分析による統合化行列は高い相関係数を示すことはないが、次々に近い行データ（または列データ）を連続させるので、反応点の一定の集団を生む働きがある。しかし、この集中化は隣接するデータだけの情報によるものであるために、「鎖効果」chain effect を招きやすい。つまり、A > B > C > D という連続において、A > B, B > C, C > D のそれぞれについてはたしかに連続性が認められるが、A > D に至るときには大きく変わってしまうことがあったり、逆に A > D が近接することがあったりする。

２．３．多変量解析による統合化

次にデータ行列ではなく、相関係数などの関係を示す行列（対照行列:Coefficient of Correlation Matrix）の統合化を考えてみたい（安本・本多 1977: 52-53）。次は先の質的データ(P1)の相関行列（下左図）とその統合化の結果である（下右図）。右図でより強い対角化が見られる。

P1	v-1	v-2	v-3	v-4
v-1	1.000	0.667	-0.167	-0.408
v-2	0.667	1.000	-0.667	-0.612
v-3	-0.167	-0.667	1.000	0.408
v-4	-0.408	-0.612	0.408	1.000

P1	v-2	v-1	v-3	v-4
v-2	1.000	0.667	-0.667	-0.612
v-1	0.667	1.000	-0.167	-0.408
v-3	-0.667	-0.167	1.000	0.408
v-4	-0.612	-0.408	0.408	1.000

同様に個体の相関係数表を統合化する。

P1	d-1	d-2	d-3	d-4	d-5
d-1	1.000	-.577	.577	-1.000	.577
d-2	-.577	1.000	-.333	.577	.333
d-3	.577	-.333	1.000	-.577	.333
d-4	-1.000	.577	-.577	1.000	-.577
d-5	.577	.333	.333	-.577	1.000

P1	d-4	d-3	d-1	d-2	d-5
d-4	1.000	-.577	-1.000	.577	-.577
d-3	-.577	1.000	.577	-.333	.333
d-1	-1.000	.577	1.000	-.577	.577
d-2	.577	-.333	-.577	1.000	.333
d-5	-.577	.333	.577	.333	1.000

このように変数についても個体についてもそれぞれの相関係数行列を統合させ、その結果得られる両軸の並びに基づいて、改めてデータ行列を並べ替えると次のようになる。この統合化のパタン化の結果はあまりよくないが、反応点(v)を隣接させる効果が表れている。

P1	v-2	v-1	v-3	v-4
d-4			v	v
d-3	v			
d-1	v	v		
d-2			v	
d-5	v	v	v	

次がデータ行列を関係係数行列（相関係数行列）で統合化した結果である。分布が中央に集中していることがわかる。また一定のパタン化がなされている(図 2.3a)。

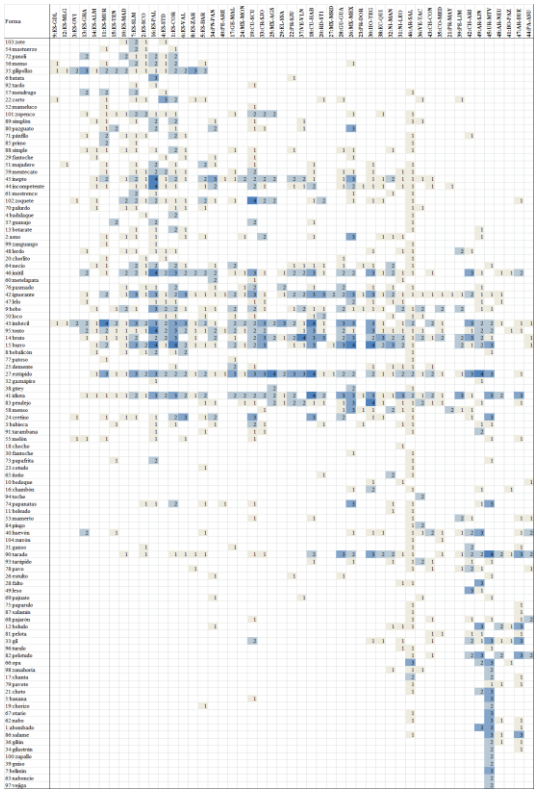


図 2.3a 関係係数統合

また、統合分析の縦軸と横軸の係数として主成分分析(Principal Component Analysis: Wood

et al. 1986, 273-290)で求める負荷と得点を使うことができる。反応点(v)が行列の中心部に集まっている(図 2.3b)。同様に、因子分析(Factor Analysis: Rietveld and van Hout 1993: 251-295; Wood et al. 1986, 290-295)の出力の因子と得点を統合分析の縦軸と横軸の係数にすると、データ行列は次のように統合化される(図 2.3c)[14]。因子の数値が近いものが寄せ集まるので反応点が互いに隣接するようになる。

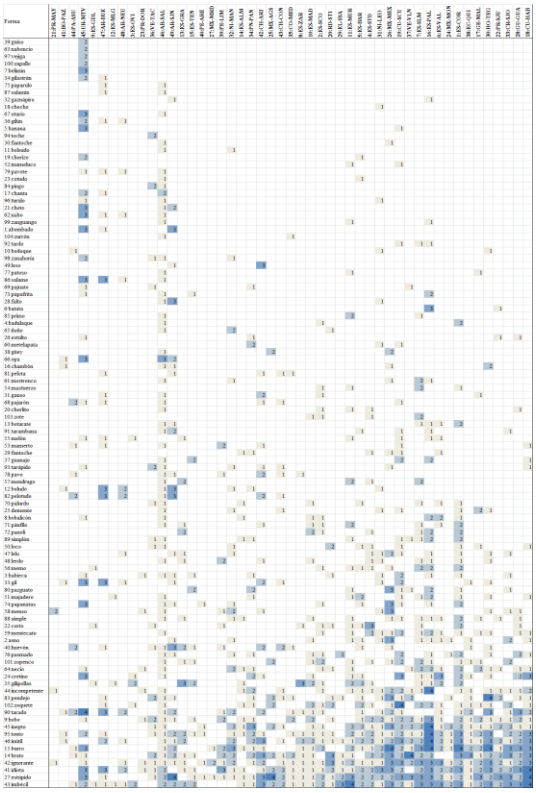


図 2.3b 主成分統合

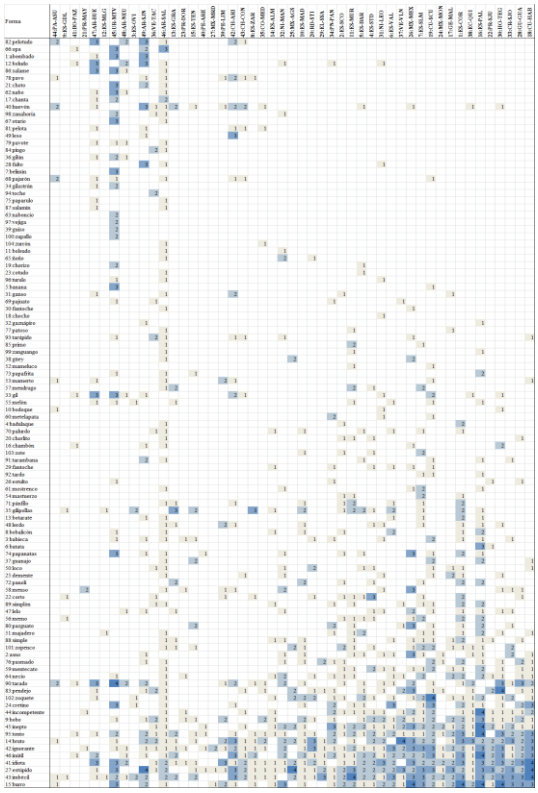


図 2.3c 因子統合

次に、統合分析の縦軸と横軸の係数として数量化Ⅲ類分析で求められる負荷と得点を使う。数量化Ⅲ類の本来の目的は分布パタンの相関係数を最大化することにあるので、当然もっともすぐれた対角化（パタン化）が得られる(図 2.3d)。一方、ここで興味深いのはクラスター分析による統合化である。横軸の変数をクラスター分析し、その並びに連番をつけて統合分析の係数とし、縦軸でも同様に係数を作り、これらの係数を使ってデータ行列を統合化させると次のような結果になる(図 2.3e)。クラスター化は必ずしもパタン化を保証しないが、反応点を各所に集中させる働きがあるので、言語地理学の観点からの集中的観察を可能にする（Perea and Ueda, 2011）。

[14] ここでは Direct Varimax 法を使った。芝(1975: 90-103)を参照。

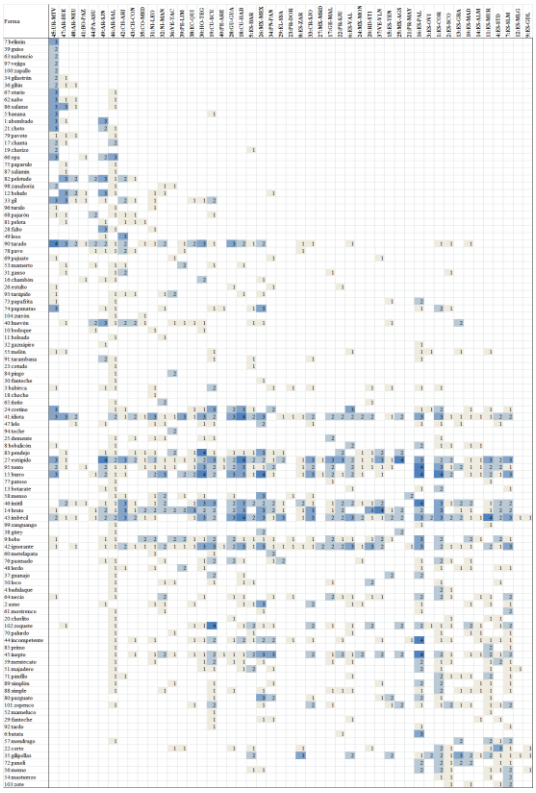


図 2.3d 数量化Ⅲ類統合

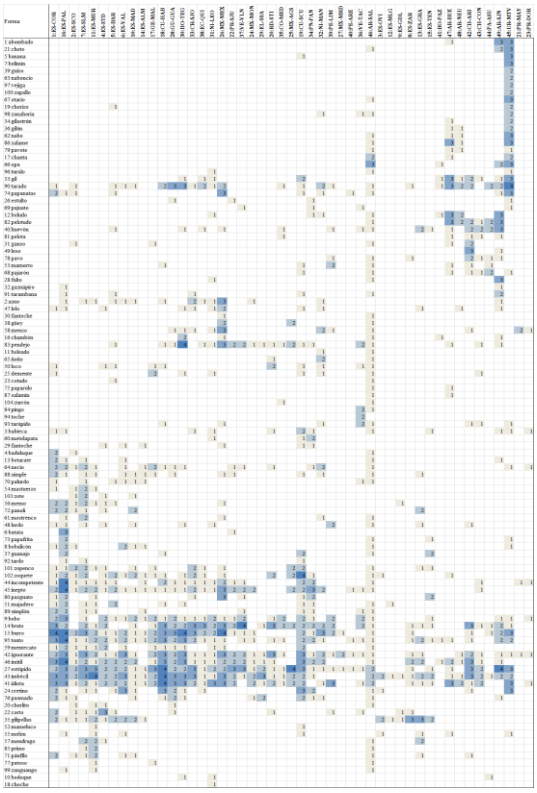


図 2.3e クラスター統合

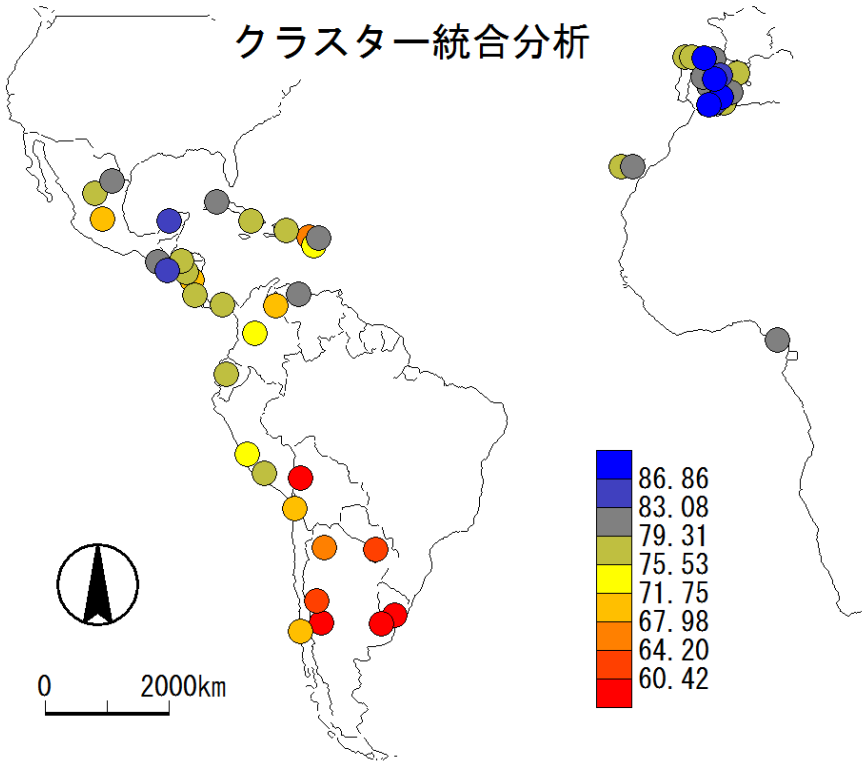


図 2.3f クラスター統合地図

３．考察

３．１．データ行列の補充

私たちの VARILEX 計画では各地点で 4 名に質問しているが、同一地点の回答が必ずしも同じでなるとは限らない[15]。そこで個別の語彙の個別の分布を見るのではなく語彙全体の分布の傾向を観察するという方法を使っている。一般に、数量分析にはデータ行列を固定したものとして分析し一定の分析結果を結論として提示する方法と、同じデータ行列にさまざまな方法を実験的に適用し、その解釈を仮説として提示する方法がある。前者の方法を使って各種の集計表、相関行列、言語地図の作成がなされ、後者の方法では各種の多変量解析が試みられている。おおまかには前者は「記述的方法」、後者は「解釈的方法」と呼ぶことができるだろう。私たちの研究計画では先述した資料の性質（不統一性）から記述的方法がとれない。その限界性を認めた上で解釈的な方法を採用している。

統一した資料の確定的記述ができていないのに、その解釈を試みるのは無謀ではないかと思われるかもしれない。たしかに VARILEX の資料については、たとえば Madrid で使われていないはずの語が反応数 1 を記録している、または逆に、Madrid で使われているはずなのに 4 人の回答者の誰もマークしていない、というケースもある。データ分布表や言語地図を提示すると、しばしば現地の人から、その語形が実際に使われている、という報告を受けることがある。言語地図で語形の分布を提示することは、それが言語現象という一律には扱えない複雑な実態であるために、たとえば天気図で各地の気圧を提示すること以上に困難なのである。しかし、私たちの研究の目的は地域差を明示するような辞書の編纂や語彙目録を作成することにあるのではなく、語彙の地域分布の全体的傾向を調べることにあるので、個別の例外はあまり問題にしない。むしろ、天気図の等圧線のような大勢を提示することが目的である。等圧線が気圧の地理的分布を精密に区分するのではなく気圧の一定のグラデーションを便宜的に示しているのと同様である。実際に、語彙バリエーションの分布も旧来の方法による精密な「等語線」(isogloss)やその束(bundle)を設定することは困難である[16]。

また、たとえば falto という語がアルゼンチンの 4 都市において、それぞれ 1, 0, 0, 3 という頻度を記録しているが、その絶対数そのものは重視しない。たまたま回答者が個人的にこの語を使わない、ということなのかもしれない。また回答時に見逃したというケースもありうる[17]。数値そのものの意味は自然科学で扱うデータがもつような意味ではなく、むしろ大まかに全体的な傾

---

15 言語地図作製を目的とする言語地理学の方法では、各地点で 1 名の話者から聴取するのがふつうであるが、スペインの言語地理学を率いた Alvar は、各地で唯一のインフォーマントに加え副次的に農業や建築などの専門語彙を複数の住民から聴取した、と述べている（1973: 151-155)。一方、日本の言語地理学で考案された「グロットグラム」では地点の軸と年齢の軸の中で語形の分布を見る（井上 1994; 2001; 真田 2007)。VARILEX 計画では各地で男性と女性・39 歳以下と 40 歳以上の組み合わせで 4 名の回答者に質問した。

16 等語線については Coseriu (1975, 5.7.1; 1984, 62-65)、グロータース（1976: 114-5), Chambers and Trudgill (1998: 103)を参照。

17 私たちの計画ではそのような個人的な事情や事故を防ぐために複数の話者（4 名）に問い合わせている。

向をつかむための手段にすぎない。よって私たちはアルゼンチンのどの都市で頻度が 1 であり、どの都市でその 3 倍の頻度を記録したか、ということにはあまり関心がない。むしろ、falto がアルゼンチンの 2 都市で頻度の多寡はどうであれ観察されたこと自体に関心がある。

次の図は原点距離法によって統合化された分布全体の中での falto の位置（下左図）と該当部の拡大図（下右図）である。Haensch y Werner (1993: s.v.)は falto がアルゼンチン中央部の口語で使われると述べている。一方、Asociación de Academias de la Lengua Española (2010)には記録がない。私たちの調査ではニカラグアの 1 都市でも記録された。このように語彙の分布については調査ごとに結果が異なるので確定的な結果を示すことが困難である。そこで、大まかに falto が基本的にアルゼンチンにおいて優勢で、一部ニカラグアでも使われる可能性がある、と言えるだろう。ここで注目したいのはこれらの地域では全体の分布傾向を見ると統合化されていて、falto はたまたまこの調査では 47:AR-BUE, 48:AR-NEU に反応していないが、やはりこの地域の特徴として統合されているということである。

そこで、47:AR-BUE, 48:AR-NEU のゼロ回答はその地域に falto が使われていない、ということではなくて、この調査では欠測値であったか、または、たまたま回答者が見逃した可能性が高い。そこで、統合化された地域での言語特徴の一定の等質性を考慮して、それぞれのセルの左右 2 つの隣接値の平均で補充する、という方法が考えられる。その結果が次の図 3.1b である。ここでは PA-ASU に 2, UR-MTV に 3 という補充値が加わっている[18]。

[18] 補充は 1 回だけでなく可能な限り繰り返される。ここでははじめに 48:AR-NEU について隣接値を含めた[0, 0, 3]という分布から平均値の 1 で補充し[0, 1, 3]という分布を作り、さらに 47:AR-BUE について[1, 0, 1]という分布から平均値 0.66 を四捨五入した値 1 で補充し[1, 1, 1]という分布を得ている。

図 3.1a 補充前

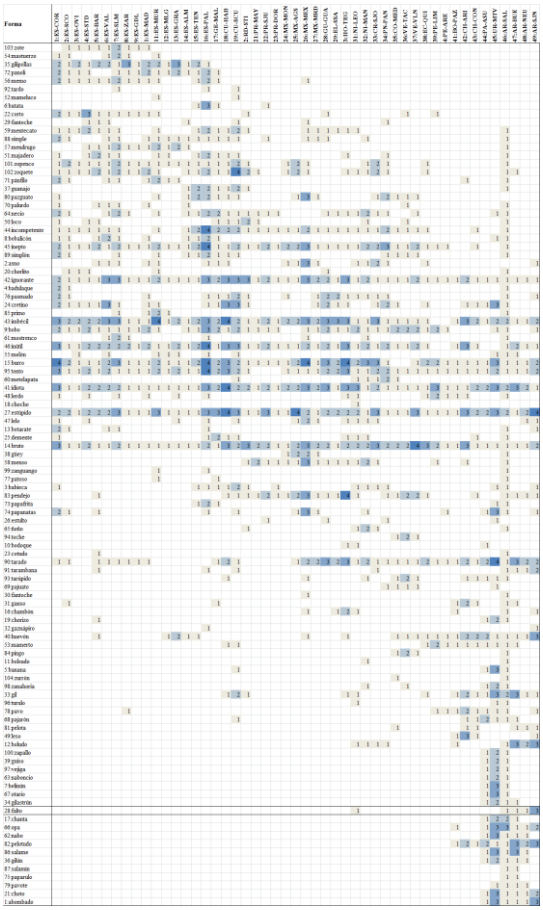


図 3.1b 補充後

45:UR-MTV	46:AR-SAL	47:AR-BUE	48:AR-NEU	49:AR-SJN
補充前	1			3
補充後	1	1	1	3

このように調査から得られたデータ行列を統合化し、内的基準から得られた地理的配列を考慮して欠測値（と思われる値）を統合隣接値によって補完して調整するという方法は、「資料を変換するという手順が入るために危険である」、「そのようなデータは信頼できない」、さらには「データを改竄している」という批判を受けるかもしれない。たしかに私たちは、言語資料の分析において採集されていない数値を他の数値（統合隣接値）で補完する、という方法を寡聞にして知らない。調査によって得られた数値は神聖視されるほどに重い意味をもっているからである。

しかし、調査で採集された原データ（採集データ）と、統合化補完処理をした調整データのどちらが言語の現実に近いか、と問い直してみると、経験的には後者（調整データ）である。また、複数の他の資料を比較すると、やはり調整データのほうが信頼性が高い。これは、「そもそも研究計画の方法（郵送法・選択法：「はじめに」を参照）に問題があって、綿密な面接法であれば信頼できるデータが得られたはずである」という反論も当然予想される。しかし、面接法を行った調査結果であっても、その発表時に、やはり、「私の村では～という言葉も使われています」という反応をよく見ることがある。つまり、絶対の真理というものは存在しないのであって、す

べて実施された調査の性質に依存するのである。そして、それぞれの方法に長所と短所があり、一律にその優劣を決定できない。私たちの今後の研究計画では、他の研究成果も参照しながら、原データに調整データを付して提示し、資料に絶対的な価値を認めるのではなく、むしろそれを比較し相対化する方法を開発していきたい。

このように VARILEX ではデータ行列が補完されたり変形されたりしている。ここで説明したようにそれぞれに理論的・実際的理由があるのだが、その根拠が研究の目的や資料の用途によって一律ではない。また、データ補完の実際的な適用においても資料の性質・分析の目的によって方法が異なる。たとえば欠測値（と見なす値）の補完において、[D-140] FOOL のデータ行列では統合化した横軸（地点）の 2 個の隣接値だけを参照し、縦軸（語形）の隣接値は参照していない。これは一般に地域の連続性は認められるにしても、語形間の連続性は認められないからである。仮に縦軸が語尾-s の脱落の割合(%：10 段階）であれば、縦軸と横軸両方の 4 個の隣接値の平均値で補完することも考えられるであろう。また、たとえば地点と音韻特徴からなる二元的配列の分析では、地点だけでなく音韻特徴の連続性も観察されることが多い。アンダルシア方言での子音連続 /s/＋/b, d, g/ において、/sb/ > [ɸ], /sd/ > [θ]が記録される地点では sg > [x]の出現も予想される(Ueda 1993)。調査ではそれぞれの地点で独立して調査票を用意するので、これら 3 つの音韻変化が必ずしも一致しないことがあるが、その場合地点と音韻特徴の隣接地を参照してデータを補完することが可能である。

３．２．データ行列の変形

一般の計量方言学の方法によれば、その分析データは既存の言語地図に基づくことが多い[19]。言語地図からデータ行列が作成され、それに相関分析、クラスター分析、主成分分析、因子分析、数量化Ⅲ類などのさまざまな多変量解析を適用される。相関分析によって得られた相関行列（対照行列）やクラスター分析によって得られた樹形図（デンドログラム）は一定の結論を導く一元的な解釈を提供する（Ueda 1995)。一方、主成分分析、因子分析、数量化Ⅲ類などの多変量解析法はデータ行列の変数の数だけ因子数が存在するため、その因子ごとに多元的な解釈を可能にする(Ueda 2008a)。また、重要な因子（I 軸と II 軸）の重さを取り出し、それを平面に配置することによって、変数間または個体間の関係を解釈することも可能である。日本の計量言語地理学の分野ではこのような多変量解析の高度な技術が駆使されている（井上 2001)。

私たちの研究計画では変数間または個体間の関係を解釈することとは別に、個体と変数からなるデータ行列（補完調整データ行列）そのものを多変量解析が提示する参照値をもとに変形し、原データ行列や調整データ行列では見つけることができなかった新しい諸相・視点を探究する。私たちの原点平均距離による統合化は数量化理論Ⅲ類と類似して、データ行列に強い相関を生み出す（井上 2001: 20; 本稿 2.1.を参照)。また、相関行列を含む関係係数行列分析、主成分分析、因子分析が提示する変数と個体の係数による統合化はデータ行列内の反応点を集中させる効力がある。さらに、隣接距離法や変数と個体のクラスター分析が提示するそれぞれの順序は、行列の各地に反応点の集中域を形成する（Perea and Ueda, 2011)。次は、各手法による統合分析の

[19] 参照：Goebl 1996, 1998, 2007; 市井 1993; 河西・真田 1982; Kletzschmar and Schneider 1996; 沢木 2002.

結果を評価する指数を示している[20]。

統合指数	連番平均距離	参照平均距離	連番相関係数	参照相関係数	平均隣接係数	標準隣接係数
データ行列	2.121	0.574	0.010	0.098	1.177	0.420
原点距離統合	2.707	1.517	0.570	0.616	1.209	0.449
連続隣接統合	2.028	1.549	0.079	0.000	1.564	0.539
関係係数統合	2.620	0.856	0.563	0.595	1.311	0.477
主成分統合	2.283	0.457	0.361	0.258	1.409	0.500
因子統合	2.718	0.112	0.444	0.297	1.393	0.492
数量化Ⅲ類統合	2.711	0.004	0.552	0.666	1.297	0.473
クラスター統合	1.289	0.310	-0.394	-0.404	1.630	0.567

図 3.2a 統合指数の比較

「連番平均距離」はすべての反応点どうしのユークリッド距離をセルの行と列の連番から計算し、それぞれの値を考慮に入れた値である。これによればクラスター統合による変形行列がもっとも反応点どうしの距離を短縮している、という結果を示している。一方、セルの行と列の連番ではなく、変形の際に与えられる縦軸（語形）と横軸（地点）の値から「参照平均距離」を計算すると、数量化Ⅲ類が距離を最小にしている。同様に、変形されたデータ行列の相関係数を計算すると、「連番相関係数」は原点距離統合が最大値を示し、「参照相関係数」は数量化Ⅲ類が最大値を示している。主成分分析と因子分析による統合化データ行列にはあまり相関がない。クラスター統合はわずかに逆相関を示しているが、クラスター分析はそもそも相関の上昇を目的にしないからである。接合の度合いを示す「平均隣接係数」と「標準隣接係数」は、どちらもクラスター統合で最大値を示している。それに続くのは連続隣接統合である。

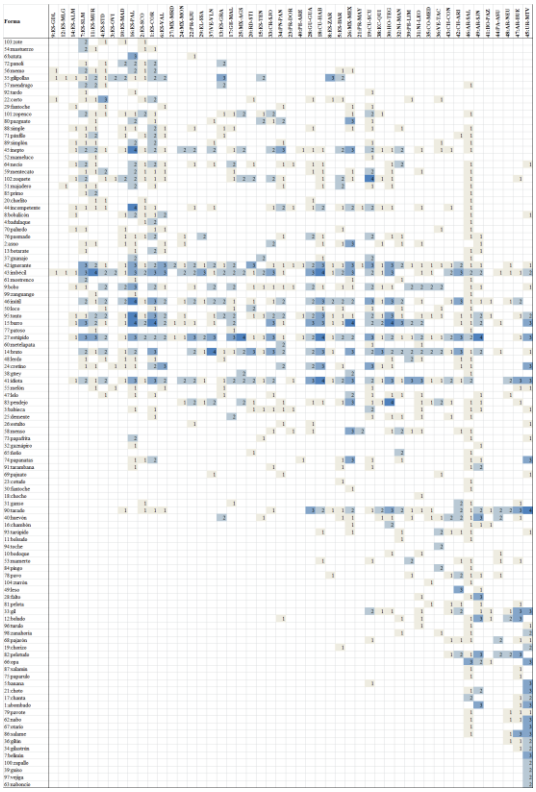
このように、それぞれの多変量解析の手法には特徴があり、変形データ行列の優劣を一概に決定できない。むしろ研究・分析の目的に応じて方法を適宜選択すべきである。たとえば、反応点をなるべく寄せ集める必要があるときは、集中点が複数でよいならばクラスター分析や連続隣接統合が適しているが、一点に集中させる必要があるときは、関係係数統合、主成分統合、因子統合がよい。反応点がデータ行列の対角線に集まると都合がよいならば、数量化Ⅲ類または原点距離統合を使うべきである。この場合、縦軸と横軸の並びに意味があるので、それぞれの軸の統一した解釈が興味深い。原点距離統合は唯一の解しか示さないが、数量化Ⅲ類は複数の解を提示するので、行列の固有値の大きなものを 2 つ選んで変数間または個体間の関係を二次元の平面で観察することができる。行と列の流れを別々に観察するには原点距離統合が適している。

アンケート調査で記入された質問票を集計して作成されるデータ行列は基本的な記述統計（平均値、分散、順位、率など）から高度な多変量解析に至るまで多様な手法で分析することができる。そこでは、一般にデータ行列の縦方向と横方向の順番を変えて配置を変形することはしない。しかし、私たちの研究計画ではデータ行列の配置をさまざまな技法によって変形する。変形してもデータの配置が変わるだけで、その本質的価値に変化はない[21]。本質的に同じデータであっても、その提示の仕方が変わることによって、初めは気づかなかった意味が見えてくることがある。このようなデータ行列の変形は私たちに新しい視点を示唆するものである。

---

20 詳細は末尾に載せた言語データ分析プログラム集 NUMEROS のウェブページを参照。
21 それぞれの分布でクラメア係数を算出すると、どれも同じ値を示す。

ここで原点平均距離法と数量化Ⅲ類による統合化の結果を再掲して比較しよう。どちらの方法でもその統合化の結果には全体的に左上から右下に向かう分布の流れが観察される。



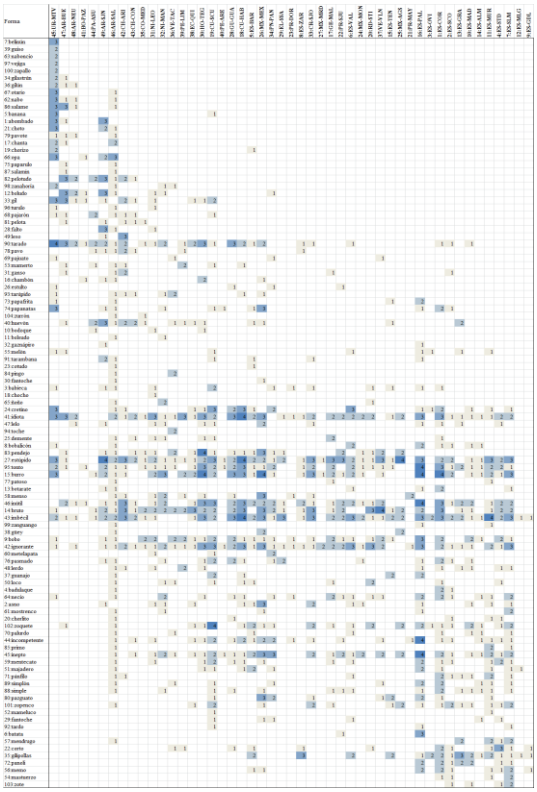


図 3.2a 原点平均距離法　　図 3.2a 数量化Ⅲ類

先述のように数量化Ⅲ類によるパタン化は理論的に最大の相関係数を獲得するが、一方、原点平均距離法は実際的にその近似値を示すだけに過ぎない。また、原点平均距離法は数量化Ⅲ類のように複数の固有値に対応する変数（または個体）のそれぞれの軸を提示すること（井上 2001: 3-25）もないので、平面や空間で変数間の関係を観察することもできない。一方、原点平均距離法は簡便であるだけでなく、内的基準と外的基準のどちらも選択することができる、という利点もある。

データ行列を分析するとき一般によく行われるのは、はじめに地点を行政区画などに従って、たとえば東地域と西地域に分割し、それぞれの地域の言語特徴を記述する、という手順である。このような方法を「前範疇化」precategorization と呼ぶことができるだろう。しかし行政区画は必ずしも言語特徴ととくに強い関係を示すとは限らないので、大まかには分析できても、たとえば東地域の地点に西地域の言語特徴が現れるという例外が多く発生することがある。一方、数量化Ⅲ類や内的基準による原点平均距離法では、はじめに地点や語形を分類するのではなく、データ行列の分布を分析し、その後で地点や語形を分類する、という「後範疇化」postcategorization と呼べるような方法をとる[22]。データ行列そのものから後範疇化を行うことにより、よりよく語形と地点の分布を記述し理解することが可能になる。さらに、後範疇化を経た変形データ行列を

[22] 井上史雄氏（私信）によれば、これは、これまでの多変量解析法の適用者が「外的基準を使わずにデータそのものに語らせる、またはデータの内部構造を読み取る」などの表現で効果を説明していたことに相当する。

改めて原点平均距離法で地点を外的基準にして、つまり前範疇化して、再度分析することも可能である。この場合、初めに前範疇化した分析とは当然その分析の結果と性質が異なる(Ueda 1993)。

前範疇化による分析は一定のクロス集計を提示するので、基本的に分析は一回で終了する。うまく分析できないときは別の範疇（データのグループ）を作り直し再びクロス集計をすることもあるが、それも前範疇化を繰り返しているにすぎない。また、そのようなグループの作り直しに分析者の恣意的な操作が入り込む余地がある。つまり、分析が良い結果を生まないとき良い結果を出すまで分析者が様々な分類を試みることになる。このようにして得られた「良い」結果は分析者が都合よくまとめたデータということになるだろう。一方、ここで取り上げている後範疇化による方法は純粋に内的基準に基づくので、そこに分析者の恣意的な判断が入り込むことがない。さらに実際的に重要なのは、はじめから分析者の判断で前範疇化するよりも、データ行列の内的構造から得られる後範疇化の方が、すぐれた相関・分類を提示するということである。広域スペイン語語彙バリエーションのケースで言えば、はじめから（アプリオリに）スペインとラテンアメリカ、またはさらに区分して６地域区分、または国別の区分で比較分析するのではなく、すべての（未分類の）地点における語形の分布をそのまま分析し、パタン化した分類から、後で（アポステリオリ）範疇化・分類をするほうが例外も少なく、分類そのものの根拠もデータ行列そのものから明示することができる。前範疇化による方法ははじめから外的基準を使うので、内的な根拠を示すことが困難である。

一般に分類がどのようなものであれその根拠を示すことが困難であることは、「分類」という問題に特有の循環論から理解できる。たとえば、一定の地域の東部と西部の言語特徴を分析するとしよう。このとき、アプリオリに地域を限定しないとすれば、東部（または西部）地域を地理的に画定するときの根拠は東部（または西部）地域で記録された一定の言語特徴がある地域ということになるだろう。そして、東部（または西部）地域の言語特徴を示すには、東部（または西部）地域で記録された一定の言語特徴の集合を列挙することになる。これでは、「言語的観点から東部地域はどのように確定されるか」という問いに「東部地域の言語特徴がある地域である」と答え、一方「それでは、東部地域の言語特徴とは何か」という問いに「東部地域に記録される言語特徴である」と答えていることになる。このように、何らかの外的基準を設定しないかぎり、地域と言語特徴のそれぞれの定義（確定）が循環する。この循環論の解消のためには、あらかじめ東部と西部を地理的に（外的基準によって）画定しておき、それぞれの言語特徴を記述すればよい、という方策がとられている。しかし、このような方法は先に述べたように分類に恣意性が混入する恐れがある。

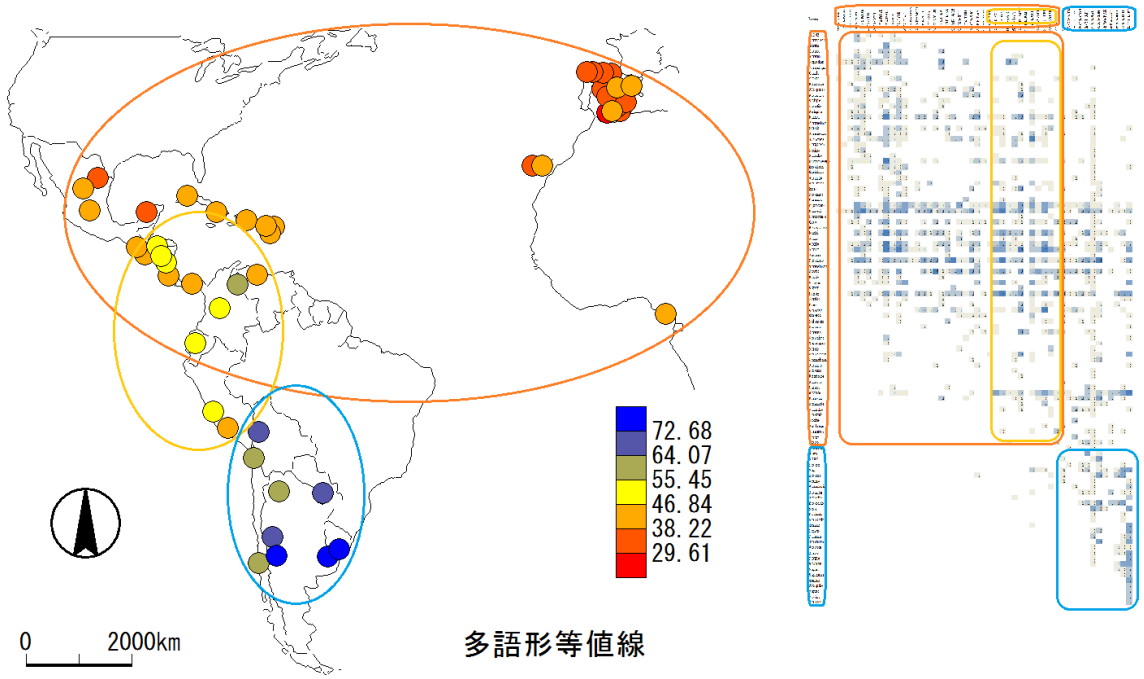
私たちのVARILEX研究計画では(Ávila et al. 2003)、総合的な語彙バリエーションの観察から、スペイン・赤道ギニア共和国→カリブ海諸国→メキシコ→中米諸国→南米北部諸国（コロンビア・ベネズエラ）→アンデス諸国（エクアドル・ペルー・ボリビア）→チリ→ラプラタ諸国（パラグアイ・ウルグアイ・アルゼンチン）という地点の連続性を見た[23]。そこで、はじめに内的基

[23] この連続は語彙バリエーションのデータ行列に基づくもので、とくにスペインとラテンアメリカという対比や、北から南へという地理的な配置に基づくものではない。結果的にそのような配列になったことは興味深い。これには植民地時代にスペイン語使用圏が拡大したという歴史地理的な背景があると思われるので、そのような言語外的な基準で分類するならば歴史地理言語分析になる。しかし、ここでも方法論的に前範疇化と後範疇化の区別をしておくとよいだろう。

準としてデータ分析の分布から地点の配置を求め、次にそれを外的基準にして個別の語彙のバリエーションを提示する、という方法を提案したい。

３．２．多語形等値線

先述のように（「はじめに」）言語地理学では個別の語彙によって「等語線」を追究する。また、複数の語彙の地理的な分布から「等語線」の「束」を設定する。しかし、ここで扱うスペイン語の罵言のように非常に多くの語彙がある場合には、その束は錯綜し、語形の等語線またはその束を選択するための先験的基準がないかぎりどのような線を描けばよいか、決めるのは困難である。このような問題には多変量解析を応用して、先験的な基準ではなく、データ行列全体から導かれる内的な基準による総合的な等値線を設定することができる。次の図は、内的基準を用いた原点平均距離法による、いわば「多語形等値線」(multilexical isogloss)を描いたものである。



多語形等値線

このように、スペイン語の罵言の地域バリエーションを示すデータ行列を内的基準によって統合化すると、とくに右下に配置される一定の語形がボリビア、チリ、ラプラタ諸国（パラグアイ、ウルグアイ、アルゼンチン）に集中していることがわかる[24]。地域内のとくに南東部に高い数値が観察される。一方、その他の地域は比較的均一であるが、それでもスペイン・アフリカ・カリブ海諸国・メキシコが一群をなし、中米・ベネズエラ・コロンビア・エクアドル・ペルーが南部地域への移行部になっていることがわかる[25]。

---

[24] 33:gil, 12:boludo, 96:turulo, 98:zanahoria, 68:pajarón, 19:chorizo, 82:pelotudo, 66:opa, 87:salamín, 75:paparulo, 5:banana, 21:choto, 17:chanta, 1:abombado, 79:pavote, 62:nabo, 67:otario, 86:salame, 36:gilún, 34:gilastrún, 7:belinún, 100:zapallo, 39:guiso, 97:vejiga, 63:naboncio.

[25] 一般にボリビアはエクアドル・ペルーとともにアンデス諸国を形成するのだが、ここではむしろラプラタ諸国と同じグループになっている。

４．結語

日本語の罵言と同様に（松本 1996）、スペイン語の罵言の語彙バリエーションも非常に多い[26]。現在の広域スペイン語の歴史はスペインの新大陸およびアフリカの植民地時代に遡るが、その歴史はおよそ 500 年間で日本語地域の歴史と比べると短い。この短期間にスペイン語圏各地で実に多くの語彙が生まれたのである。そこには日本の方言周圏論や語形伝播の各種のモデル（松本 1996; Lizana et al. 2011）では説明できない複雑さがある。

地点・地域ごとに複雑な諸相を見せるデータ行列を分析するには多変量解析が有効である。しかし、先述したように、スペイン語計量言語地理学の研究者は一般に多変量解析を使わない。一部ではクラスター分析のアプリケーションを適用しているが、日本の研究者に見られるような多元的な解釈を行うことは稀である。その理由を探ってみると、線形代数などの数学的手法に慣れていない文献学・言語学研究者が多変量解析の理論を正確に理解できないことにあるようだ。たとえ既成の統計パッケージで分析しても、それが出力する数値行列やグラフの数学的な導出過程が不明なので研究成果として示せない、ということである。数理の理論に関わる質問をすると「統計学についてはよくわからない」、または「私は統計学者ではなく言語学者として統計学を応用した」という答えが返されることがある。しかし、数理の理論的基盤を知らないでそれを応用することができるのだろうか。

幸い日本では文系でも大学の数学を履修すると線形代数の基礎が含まれることが多い。そして文系・理系を問わず多くの分野で多変量解析が利用され、その入門書から専門書に至るまで多くの参考書が出版されている[27]。ウェブにも多くの情報が載せられている。そして日本の計量的方言研究は高い成果を上げてきた（半沢 2007）。私たちの研究計画でもこれまで積極的に多変量解析を応用し、拙いものであるが自らプログラムを作成し試行錯誤の実験を繰り返しながら少しずつ適用の可能性を探ってきた。自らが収集したデータを自らが開発したプログラムで分析するという方法は能率が悪いことがある。自分でデータを収集しなくても先行研究や言語地図からデータを作成することができるし、分析プログラムは各種のパッケージが開発されている。しかし、データにしてもプログラムにしても既成のものを使うと、その構成や性質がブラックボックスになる恐れがある。説明を求められても「…を使用した」という答えしかできない。スペイン語言語地理学研究においてそのような例が多いのは残念なことである。私たち日本のスペイン語研究グループはそのような依存状態から脱却し、独自のデータとメソッドを開拓し、日本語計量言語地理学の水準に近づきたいと願っている。本稿はその経過報告の一部である。

＊謝辞

この研究をまとめるにあたっては井上史雄先生に多くのご示唆とご教示をいただきました。私

[26] 南北アメリカ大陸のスペイン語の特徴語彙を調査した Asociación de Academias de la Lengua Española (201:2241-2)は 413 語を記録している。これにはスペインのスペイン語の特徴語彙は含まれないので全体の数はさらに拡大するはずである。

[27] 次を参照：足立 (2005), Anderberg (1973), Hartigan (1975), Horst (1965); 井上(1998), 井上・広川(2000), 石村(1995), 河口(1978) 三野(2001), 奥村(1986), Rosemburg (1989), 芝(1975), 白井(2009), 竹内・柳井(1972), 安田・海野(1977)。

は先生から直接教育を受ける機会には恵まれませんでしたが、東京外国語大学に奉職した 1980 年代に先生とご一緒に電算機室でパンチカード入力とラインプリンター出力の作業を繰り返しながら、折々計量言語地理学に関する多くのことを教えていただきました。その上、ご著書やご論文をいただき多くのことを学びました。言語地理学の国際学会にも誘われ、英語で交換される興味深い議論のなかで先生の世界的な研究レベルの高さを拝見いたしました。また、Google Maps と Google Insights を使って個々の単語の地理的分布を世界地図の形で出力された先生は(井上 2011, 2012)、私信で「英語やスペイン語のように地表上で広く使われている言語の世界地図は興味深い」と述べられています。井上先生のいつものご指導とご厚意に深く感謝申し上げます。

参考文献

Abad de Santill'an, Diego. (1991) *Diccionario de argentinismos de ayer y de hoy.* Buenos Aires, Tipográfica Editora Argentina.

足立堅一(2005)『多変量解析入門：線形代数から多変量解析へ』篠原出版新社.

Alvar, Manuel. (1973) *Estructuralismo, geografía lingüística y dialectología actual.* Madrid, Gredos.

Anderberg, Michael R. (1973) *Cluster analysis for applications.* New York, Academic Press. 西田英朗・佐藤嗣二他訳(1988)『クラスター分析とその応用』内田老鶴圃.

Ávila, R. Samper, J. A. y Ueda, H. (2003) *Pautas y pistas en el análisis del léxico hispano(americano).* Iberoamericana Vervuert, 278pp.

Asociación de Acedemias de la Lengua Española. (2010). *Diccionario de americanismos.* Madrid, Santillana.

Bertin, Jacques. (1977) *La graphique et le traitement graphique de l'information.* Paris: Flammarion. 森田喬訳『図の記号学』平凡社, 1982. Antonio Muñoz Carrión (tr.) *La gráfica y el tratamiento gráfico de la información.* Madrid, Taurus, 1977

Cahuzac, Philippe. (1980) “La división del español de América en zonas dialectales. Solución etnolingüística o semántico-dialectal”, *Lingüística Española Actual*, 10, pp. 385-461.

Carbonell Basset, Delfín. (2000) *Gran diccionario del Argot*, Barcelona, Larousse.

Casas Gómez, Miguel. (1994), “Marcas diatópicas en el léxico eufemístico- disfemístico”, en G. Wotjack y K. Zimmermann (eds) Unidad y variación léxicas del español de América, pp.133-184.

Chambers, J. K. and Trudgill, Peter. (1998) *Dialectology.* Second edition. Cambridge University Press.

Chuchuy, Claudio; Hlavacka de Bouzo, Laura. (1993) *Nuevo diccionario de americanismos. Tomo II. Argentinismos.* (Dirigido por G. Haensch y R. Werner) Santafé de Bogotá: Instituto Caro y Cuervo.

Coseriu, Eugenio. (1975) *Die Sprachgeographie*. Tubingen : G. Narr. 柴田武・W. グロータース共訳『言語地理学入門』三修社 1984.

Escobar, Raúl Tomás. (1986) *Diccionario del hampa y del delito*. Buenos Aires, Editorial Universidad.

Goebl, Hans (1996) "La convergence ente les fragmentations géo-génétique de l'Italie du Nord", *Revue de Linguistique Romane*, t. 60, pp. 25-49.

_____. (1998) "On the nature of tension in dialectal networks: A proposal for interdisciplinary discussion", *Systems. New Paradigms for the Human Sciences*, ed. by G. Altamann and W. K. Koch, Berlin,

Walter de Gruyter, pp. 549-571.

___. (2007) "Dialectometry: theoretical prerequisites, practical problems, and concrete applications (mainly with examples drawn from the *Atlas linsguistique de la France*, 1902-1910", 第14回国立国語研究所国際シンポジウム『世界の言語地理学』*Proceedings of the 14th NIJL International Symposium*, pp. 65-74.

林知己夫・樋口伊佐夫・駒澤勉 (1970)『情報処理と統計数理』産業図書.

半沢康 (2007)「方言を量る方法」『シリーズ方言学４．方言学の技法』岩波書店, pp. 179-216.

Hartigan, J. A. (1975) *Clustering Algorithms*. New York. John Wiley & Sons.

Haensch, Günther; Werner, Reinhold. (1993) *Nuevo diccionario de americanismos. Tomo II. Argentiismos.* Santafé de Bogotá: Instituto Caro y Cuervo.

Horst, Paul. (1965) *Factor Analysis of Data Matrices*. Holt, Rinehart and Winston. 柏木繁男・芝祐順・池田央・柳井晴夫訳『コンピュータによる因子分析法』科学技術出版社, 1978.

市井外喜子 (1993)『方言と計量分析』新典社.

池田央 (1976)『統計的方法 I 基礎』新曜社.

井上史雄 (1992)「社会言語学と方言文法」『日本語学』11-6, 94-105.

_____. (1994)『方言学の新地平』明治書院.

_____. (2001)『計量的方言区画』明治書院.

_____. (2007)『変わる方言　動く標準語』筑摩書房.

_____. (2011)「Google 言語地理学入門」『明海日本語』16, 43-52

_____. (2012)「日本語世界進出のグーグル言語地理学：グーグルインサイトにみる外行語総合分布」『明海日本語』17, **-**.

Inoue, Fumio. (1988) "Dialect Image and New Dialect Forms", *Area and Culture Studies*, Tokyo University of Foreign Studies, 38: 13-23.

_____. (1996) "Computational Dialectology", *Area and Culture Studies*, Tokyo University of Foreign Studies, 52: 67-102; 53: 115-134.

井上勝雄(1998)『パソコンで学ぶ多変量解析の考え方』筑波出版会.

井上勝雄・広川美津雄(2000)『エクセルで学ぶ多変量解析の作り方』筑波出版会.

Kany, Charles E. 1962. *Semántica hispanoamericana*. Madrid: Aguilar.

河口至商 (1978)『多変量解析入門 I, II』森北出版.

河西秀早子・真田信治(1982)「『日本言語地図』による標準語形の地理的分布」『日本語研究』5, 34-47.

Kawasaki, Yoshifumi. (2012) "Datación estadística de los textos medievales sin fecha: Análisis", *Encuentro de investigadores de los textos medievales españoles*, Madrid, CSIC.

駒澤勉・橋口捷久 (1988)『パソコン数量化分析』朝倉書店.

Kletzschmar, William. A. and Schneider, Edgar W. (1996) *Introduction to Quantitative Analysis of Linguistic Suvey Data*. Thousando Oaks. SAGE Publications.

Kühl de Mones, Úrsula. (1993) *Nuevo diccionario de americanismos. Tomo III. Nuevo diccionario de uruguayismos*. Santafé de Bogotá: Instituto Caro y Cuervo.

Lizana, Ludvig; Mitarai, Namiko; Kim, Sneppen (2011). "Modelling the Spatial Dynamics of Culture

Spreading in the Presence of Cultural Strongholds" *Phyical Review.* E 83, 066116. (http://arxiv.org/pdf/1101.3998v1.pdf)

Marrone, Nila G. (1974) "Investigaciones sobre variaciones léxicas en el mundo hispano", *The Bilingual Review; La revista bilingüe*, 1, pp.152- 158.

Martín, Jaime (1974), *Diccionario de expresiones malsonantes del español. Léxico descriptivo,* Madrid, Ediciones Istmo, 2ª ed.

松本修 (1996).『全国アホ・バカ分布考：はるかなる言葉の旅路』新潮文庫.

三野大來(2001)『統計解析のための線形代数』共立出版.

Moreno de Alba, José G. (1992) *Diferencias léxicas entre España y América.* Madrid: Mapfre.

Moreno Fernández, Francisco. (1993) "Las áreas dialectales del español americano. Historia de un problema", en Moreno Fernández, F. (ed.) *La división dialectal del español de América.* Alcalá de Henares: Univ. de Alcalá de Henares, pp.10-38.

奥村晴彦(1986)『パソコンによるデータ解析入門．数理とプログラミング実習』技術評論社.

Perea, Maria-Pilar and Ueda, Hiroto. (2011). "Applying quantitative analysis techniques to *La flexió verbal en els dialectes catalans*", *Dialectologia et Geolinguistica, Journal of the International Society for Dialectology and Geolinguistics,* vol. 18, pp. 99-114.

Rietveld, Toni and van Hout, Roeland. (1993) *Statistical Techniques for the Study of Language and Language Behavior*. Berlin, Mounton de Gruyter.

Rosemburg, Ch. H. (1989) *Cluster analysis for researchers.* Robert E. Krieger Publishing Company, Inc. Malabar, Florida. 西田英朗・佐藤嗣二訳『実例クラスター分析』内田老鶴圃(1992).

Ruiz, Ciriaco. (2001) *Diccionario ejemplificado de argot*, Barcelona, Península.

Ruiz Tinoco, Antonio. (1999) "El Proyecto VARILEX en Internet. Base de datos compartida de variación léxica", *Varilex*, 7, pp. 50-60.

真田信治 (2007)「日本で編み出された"グロットグラム"」第 14 回国立国語研究所国際シンポジウム『世界の言語地理学』*Proceedings of the 14th NIJL International Symposium*, pp. 19-28.

Sanmartín Sáez, Julia (1998) *Diccionario de argot*. Madrid, Espasa.

沢木幹栄 (2002)「方言地図データの活用；GAJ のデータによる地点のクラスター分析」馬瀬良雄（監修）『方言地理学の課題』明治書院, pp. 432-444..

芝祐順(1975)『行動科学における相関分析法』東京大学出版会.

白井豊(2009)『Excel と VBA による実用数値解析入門』ゆたか創造舎.

竹内啓・柳井晴夫(1972)『多変量解析の基礎』東洋経済新報社.

Takagaki, Toshihiro. (1993) "Hacia la descripción del español contemporáneo de las grandes ciudades del mundo hispánico", *Lingüística Hispánica*, 16, 65-86.

Ueda, Hiroto. (1993) "División dialectal de Andalucía: Análisis computacional", *Actas del Tercer Congreso de Hispanistas de Asia*, Asociación Asiática de Hispanistas, Tokio, pp.407-419.

_____. (1994) "Banco de datos léxico del español. Un proyecto internacional de investigación", *Verba* (Univ. de Santiago de Compostela), 21, pp.397-416.

_____. (1995) "Zonificación del español. Palabras y cosas de la vida urbana", *Lingüística (ALFAL)*, 7,

pp.43-86.

_____. (1996a) "Variación léxica del español urbano. Vestuario y equipo", *Publicaciones del Departamento de Idiomas Extranjeros, Facultad de Artes y Ciencias,* Universidad de Tokio, 43/4, pp.99-144.

_____. (1996b) "Estudio de la variación léxica del español. Métodos de investigación", *Homenaje al profesor Makoto Hara. Trabajos reunidos con motivo de la jubilación universitaria.* Tokio, pp.341-375.

_____. (2000) "Distribución de palabras variables. España y América. Léxico de transporte". en *Estudios de Lingüística Hispánica, Homenaje a María Vaquero, Universidad de Puerto Rico,* pp.637-655.

_____. (2005) "Léxico de la blasfemia: Análisis por patronización", Josefina Prado Aragonés y María Victoria Galloso Camacho (eds.) *Diccionario, léxico y cultura*. Universidad de Huelva, España, pp. 233-245.

_____. (2008a) "Análisis dialectométrico del léxico variable español: Interpretación taxonómica de resultados", en *El español de América, Actas del VI Congreso Internacional de "El español de América" (Tordesillas, Valladolid, 25-29 de octubre 2005),* Valladolid, pp. 813-822. Instituto Interuniversitario de Estudios de Iberoamérica y Portugal, Universidad de Valladolid.

_____. (2008b) "Resultados y proyectos en las investigaciones sobre variación léxica del español". *Actas de XV Congreso de la Asociación de Lingüística y Filología de América Latina, Edición corregida y aumentada.* ISBN 978-9974-8002-6-7. Montevideo, 2008/8/18-21. 24p.

Wood, Gordon R. (1990) "Using a Printed Vocabulary Checklist", in *Computer Methods in Dialectology,* ed. by W. A. Kretzschmar Jr., E. W. Schneider, E. Johnson, An American Diaclect Society Centennial Publication, University of Georgia, pp. 6-16.

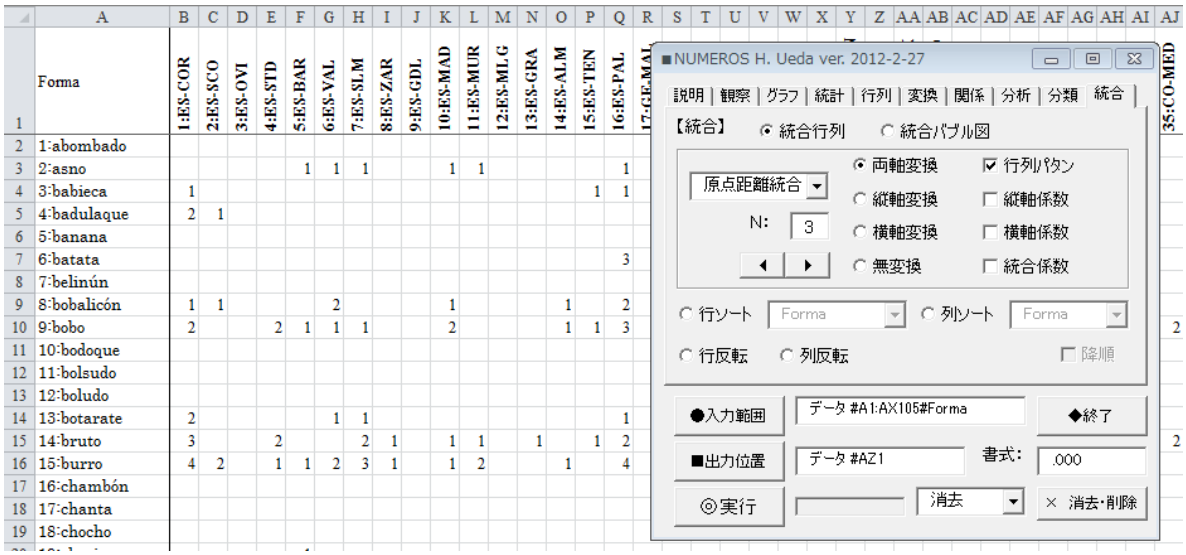
Woods, Anthony; Fletcher, Paul and Hughes Arthur (1986) *Statistis in Language Studies*. Cambridge, Cambridge University Press.

安田三郎・海野道朗(1977)『社会統計学』(改訂 2 版) 丸善.

安本美典・本多正久 (1981)『現代数学レクチャーズ D-2 因子分析法』培風館.

＊補足

本研究では ExcelVBA による言語データ分析プログラム集 NUMEROS（図 4）を使用した。
http://lecture.ecc.u-tokyo.ac.jp/~cueda/gengo/index.html



言語データ多変量分析プログラム NUMEROS.xlsm

次は原点距離統合のサブルーチンである。Snp はデータ行列、Vn は縦列の参照値ベクトル、Hp は横列の参照値ベクトルを示す配列である。

```
Sub ◆原点距離統合(Snp, Vn, Hp)
  Dim c#, d#, n&, p&, h&, i&, j&, bolV As Boolean, bolH As Boolean

  n = UBound(Snp, 1): p = UBound(Snp, 2)

  For h = 1 To 100 '100 回の繰り返しで終了
    bolV = bT: bolH = bT '配列変化のフラグ

    For i = 1 To n   '行の距離を計算
      c = 0: d = 0 '反応数と距離を初期化

      For j = 1 To p
        c = c + Snp(Vn(i, 0), Hp(j, 0)) '反応の総和
        d = d + Snp(Vn(i, 0), Hp(j, 0)) * j ^ Val(Fn.txtIntN) '距離
      Next

      If c = 0 Then
        d = 0 'DIV/0 を回避
      Else
        d = Abs(d / c) ^ (1 / Val(Fn.txtIntN)) * IIf(d > 0, 1, -1) 'N 乗根＊負記号
      End If
```

```
      If d <> Vn(i, 2) Then Vn(i, 2) = d: bolV = bF '距離：フラグ
    Next

    If Fn.optIntV Or Fn.optIntA Then Vn = SortM(Vn, 2, bT) '配列をソート

    For i = 1 To p   '列の距離を計算
      c = 0: d = 0   '反応数と距離を初期化

      For j = 1 To n
        c = c + Snp(Vn(j, 0), Hp(i, 0)) '反応の総和
        d = d + Snp(Vn(j, 0), Hp(i, 0)) * j ^ Val(Fn.txtIntN) '距離
      Next j

      If c = 0 Then
        d = 0 'DIV/0 を回避
      Else
        d = Abs(d / c) ^ (1 / Val(Fn.txtIntN)) * IIf(d > 0, 1, -1) 'N 乗根＊負記号
      End If

      If d <> Hp(i, 2) Then Hp(i, 2) = d: bolH = bF '距離：フラグ
    Next i

    If Fn.optIntH Or Fn.optIntA Then Hp = SortM(Hp, 2, bT) '配列をソート

    If Fn.optIntV Or Fn.optIntH Then Exit For '縦軸 or 横軸ならば終了
    If bolV And bolH Then Exit For '両軸の配列に変化がなければ終了

    Fn.ProgressBar.Value = h 'プログレスバー
  Next
End Sub
```