


広域スペイン語語彙バリエーションの調査・集計・解釈
連続する言語変異の区画化と視覚化の方法

上田博人（東京大学）

私たちは1993年からスペイン・赤道ギニア共和国・ラテンアメリカ諸国の各地で日常使用されるスペイン語語彙について現地調査や通信調査を毎年繰り返してきた。スペイン語圏の多くの研究者と連絡をとりながら、手書きの絵を添えた質問票を送付し、各地で性別と年齢に偏りのないように選ばれた4名の被調査者に、次のように予め用意した回答例の中から自分が使用する語形を選択してもらった。該当する語形が回答例にならない場合はそれを自由に書き込むスペースも用意した。これまでに調査した約2000項目の集計結果をウェブサイトで公開している。

Explicaciones

1. Subraye la(s) palabra(s) o expresión(es) si usted mismo la(s) utiliza.
Ejemplo:



[A001] JACKET: Prenda de vestir masculina, que forma con el chaleco y los pantalones el traje completo. No es de paño con botones dorados.

(1)americana, (2)capa, (3)chaleco, (4)chaqueta, (5)gabán, (6)leva, (7)paletón, (8)saco, (9)saco de terno, (10)saco de traje, (11)traje, (12)vestón.

&Otro(s) _____, #No se me ocurre.

\$Comentario:

かつてのヨーロッパの伝統的な言語地理学やスペイン語方言学の方法によれば、調査者が小さな村のお年寄りにその土地に残る言葉について多くの質問をし、その音声をノートに転写し、専門家がそれらをまとめて地図帳に書き写すという作業が行われていた。この伝統的な方法と私たちの方法を比べると次の点が対照的に異なる。比較的小さな地域とスペイン語圏全体、小村と大都市、高齢者と壮年者、面接調査と通信調査、唯一調査と継続調査、手作業の集計とコンピュータ処理、言語地図→データベース→計量分析という計量方言学の流れとデータベース→言語地図・計量分析の流れなど。どちらが優れている、ということではなく、両者は時と場所の条件や目的・対象・方法が異なるので、排他的ではなく相補的な関係にある。

私たちはウェブ上で操作できる数量化・視覚化プログラムを作成した(Ueda/Moreno)。プログラムは語形・地点の二次元配列と地図プロットを出力し、多変量解析法によって語形を地理的に連続させ、その語形間および地点間の分布係数（原点からの平均距離）をクラスター分析によって区画化し、その図式的な区画と同じ色で現実的な地図上の地

点を連携させた。従来の等語線による区画化は重複・連続する言語変異の分布では困難であった。そこで同じ区画に属する地点の緯度・経度の平均値を中心点とし、それと各語形の地点を結ぶネットワークにより分布の特徴を見る、という方法を採用した。

今回の発表では、調査項目の1番(A001)「(男性用の) 上着・ジャケット」の分布を例にして、その語形のバリエーションの21か国における地理的分布を観察し、私たちの区画化と視覚化の方法を提示したい。資料は調査の集計データ(絶対頻度)を現地の協力者に確認した検証データ VARILEX-R (1/0 データ) を使用する。

次はこの項目について各地の分布を国ごとにまとめた表である。

地点	<i>americana</i>	<i>chaqueta</i>	<i>gabán</i>	<i>saco</i>	<i>saco de terno</i>	<i>saco de traje</i>	<i>vestón</i>
1.ES		1					
2.GE	1	1					
3.CU		1		1		1	
4.RD		1		1			
5.PR		1	1	1			
6.MX				1		1	
7.GU				1			
9.HO				1			
9.EL				1			
10.NI				1			
11.CR		1		1			
12.PN				1			
13.CO		1		1			
14.VE		1		1			
15.EC		1			1		
16.PE				1	1		
17.BO				1	1		
18.CH		1					1
19.PA				1			
20.UR				1		1	
21.AR				1		1	

(1.ES: スペイン ; 2.GE: 赤道ギニア ; 3.CU: キューバ ; 4.RD: ドミニカ ; 5.PR: プエルトリコ ; 6.MX: メキシコ ; 7.GU: グアテマラ ; 8.HO: ホンジュラス ; 9.EL: エルサルバドル ; 10.NI: ニカラグア ; 11.CR: コスタリカ ; 12.PN: パナマ ; 13.CO: コロンビア ; 14.VE: ベネズエラ ; 15.EC: エクアドル ; 16.PE: ペルー ; 17.BO: ボリビア ; 18.CH: チリ ; 19.PA: パラグアイ ; 20.UR: ウルグアイ ; 21.AR: アルゼンチン)

次は頻度の高い語形(chaqueta, saco, saco de traje, saco de terno)と国名をその分布が対角線の付近に集中するようにパタン化した分布を示す(パタン分類表)。行と列の各係数は原点からのミンコフスキー距離(3乗)の平均値を使った。

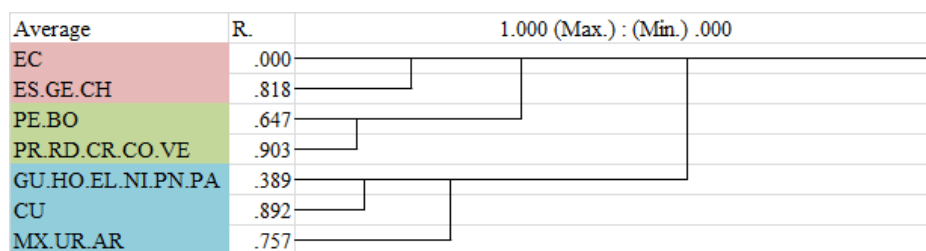
Dst.	s. de terno	chaqueta	saco	s. de traje	Row	Xn
EC	1	1			EC	1.651
ES.GE.CH		1			ES.GE.CH	2.000
PE.BO	1		1		PE.BO	2.410
PR.RD.CR.CO.VE		1	1		PR.RD.CR.CO.VE	2.596
GU.HO.EL.NI.PN.PA			1		GU.HO.EL.NI.PN.PA	3.000
CU		1	1	1	CU	3.208
MX.UR.AR			1	1	MX.UT.AR	3.570

Column	s. de terno	chaqueta	saco	s. de traje
Yp	2.410	4.165	5.372	6.538

斜めの矢印が示すような対角化された分布は語形と地点の連関を明示し、各語形と各地点の数値(Xn, Yp)は分布的距離を連続的に示している。これを離散的に区画化するために、はじめにプログラムですべての地点の原点からの距離(Xn)の組み合わせについて係数の差の絶対値を計算し、次の対照行列を作成した。

Dis.	EC	ES.GE.CH	PE.BO	PR.RD.CR.CO.VE	GU.HO.EL.NI.PN.PA	CU	MX.UR.AR
EC	.000	.349	.759	.945	1.349	1.557	1.919
ES.GE.CH	.349	.000	.410	.596	1.000	1.208	1.570
PE.BO	.759	.410	.000	.186	.590	.797	1.160
PR.RD.CR.CO.VE	.945	.596	.186	.000	.404	.611	.974
GU.HO.EL.NI.PN.PA	1.349	1.000	.590	.404	.000	.208	.570
CU	1.557	1.208	.797	.611	.208	.000	.362
MX.UT.AR	1.919	1.570	1.160	.974	.570	.362	.000

これに基づいてクラスター分析を行うと次の樹状図が得られる。



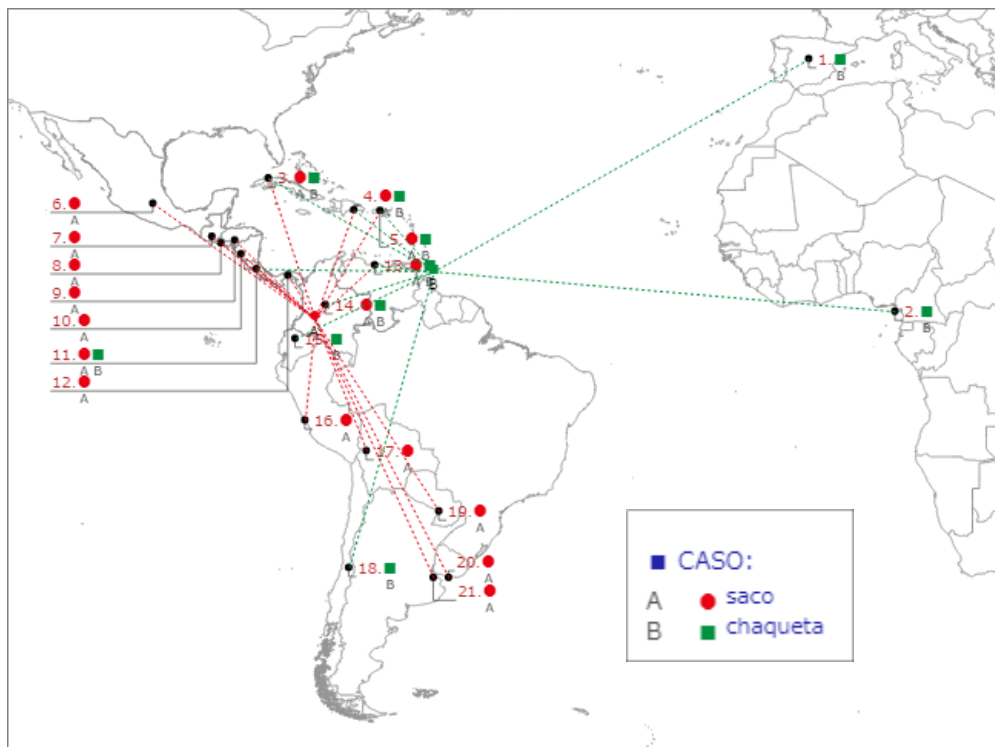
同様にして、語形の原点からの距離(Yp)クラスター分析を行う。

Average	R.	1.000 (Max.) : (Min.) .000
saco de terno	.000	
chaqueta	.575	
saco	.354	
saco de traje	.717	

次に地点と語形のクラスター分析の結果を先のパタン分類表に反映させる。

Dst.a	saco de terno	chaqueta	saco	saco de traje
EC	1	1		
ES.GE.CH		1		
PE.BO	1		1	
PR.RD.CR.CO.VE		1	1	
GU.HO.EL.NI.PN.PA			1	
CU		1	1	1
MX.UR.AR			1	1

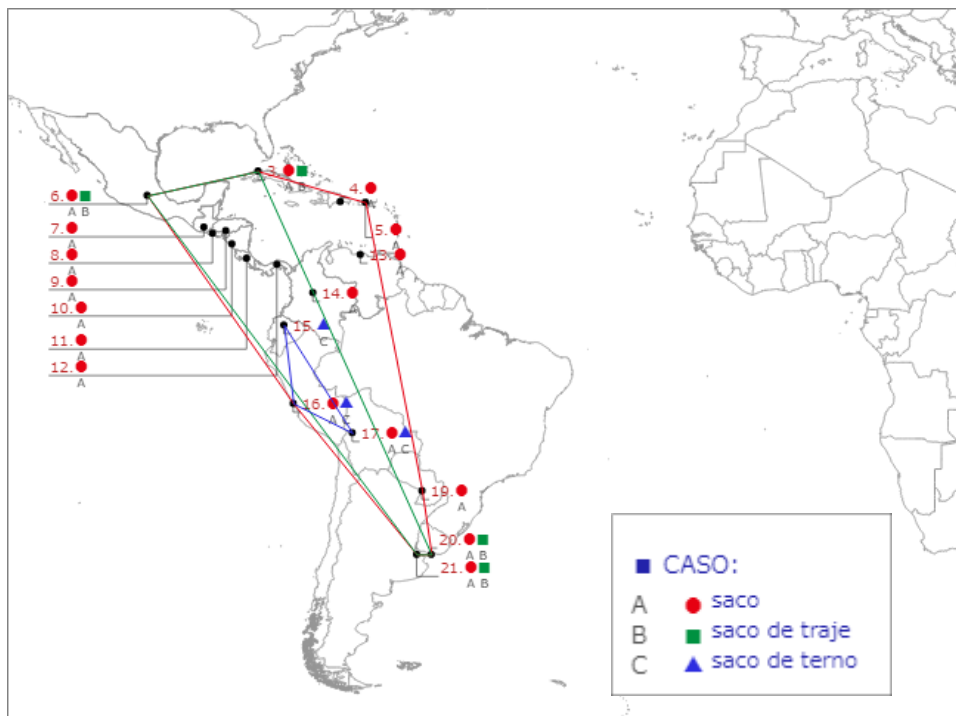
このパタン・クラスター表から、(1) saco : chaqueta と、(2) saco : saco de traje : saco de terno の分布を取り出して考察したい。次は(1)のネットワーク図である。



(1) この図を見ると、A: saco は西側の MX, 中米に、B: chaqueta は東側の ES, GE, CH に、そして A+B は中央のカリブ海諸国に分布しているので、A:A+B:B という地

理的連続が確認できる。一般に、A, B, A+B という分布の年代的变化は $A > A+B > B$, または (同じことであるが) $B > A+B > A$ という順番が普通であり, $*A > B > A+B$ ($*B > A, A+B$) や, $*A+B > A > B$ ($*A+B > B > A$), $*A+B > A=B$ (A, B が同時) という順番は普通ではない。従来の語源研究では, ラテン語起源の *saco* (<lat. SACCUS) が「上着」の意味で 14 世紀からスペインで使われていたが, 19 世紀の初めにフランス語 *jaquette* が *chaqueta* としてスペイン語化した(初出 1804), と説明されている (Corominas / Pascual, 1983)。大規模コーパス(RAE. CORDE)ではさらに古い用法が見つかる。*saco* はスペイン(ES), 赤道ギニア(GE), チリ(CH)では現在使われていない。よって $A: \text{saco} > A+B: \text{saco}+\text{chaqueta} > B: \text{chaqueta}$ という歴史的变化が確認できた。

(2) 先のパタン・クラスター表を見ると *saco* の拡張形 *saco de terno* (EC, PE, BO) と *saco de traje* (CU, MX, UR, AR) の分布は重なっていない。これを地図で視覚化する。



この図は各語形の出現位置の全体を包むようにして描いた多角形の図である。よって、多角形線で包まれた面の中に実は例外も含まれてしまう。しかし、それがかえって好都合なときもある。ここでは地理的に $A: \text{saco} \supset B: \text{saco de traje} \supset C: \text{saco de terno}$ という包含関係が視覚化されている。一般に $A \cdot B \cdot A$ という包含的分布から (例外も多いが) $A > B$ という継時的変化が想定される。ここでは $A > B > C$ という二重の変化が想定されるが、このことは先行研究や RAE コーパス(CORDE)で検証できなかった。

Corominas, J. / Pascual, J. A. (1983) *Diccionario crítico etimológico castellano e hispánico*. Madrid, Gredos. // RAE. CORDE. Real Academia Española. *Corpus Diacrónico del Español (CORDE)* <http://corpus.rae.es/> [2019/9/29] // Ueda, H. / Moreno, A. LYNEAL: <https://lecture.ecc.u-tokyo.ac.jp/~cueda/lyneal/varilex.htm>