

Los métodos de análisis cuantitativo y su aplicación

a *La flexió verbal en els dialectes catalans*

Maria Pilar Perea, Univesidad de Barcelona

Hiroto Ueda, Universidad de Tokio

1. Introducción

Antoni M. Alcover (Manacor 1862-Palma de Mallorca 1932) fue el primer dialectólogo que encuestó datos verbales con vistas al desarrollo de un proyecto sobre morfología flexiva inicialmente denominado “Estudi de la conjugació catalana”. Después de 22 años de encuestación (1906-1928), recopiló un repertorio de casi 500.000 formas verbales correspondientes a la flexión completa de 75 verbos de 149 localidades del dominio lingüístico catalán. Estos materiales, publicados por su discípulo Francesc de B. Moll, aparecieron en una publicación periódica, entre el 1929 y el 1933, con el nombre de “La flexió verbal en els dialectes catalans”. Desde 1999 estos materiales han sido completados e informatizados para mejorar su accesibilidad y consulta. En primer lugar, fueron introducidos en una base de datos, que se convirtió en el fundamento para elaborar mapas dinámicos. El año pasado se consiguió aplicar voz sintetizada a estos materiales, con lo cual se ha obtenido un atlas sonoro dedicado a la morfología verbal (Perea 2008b). Estas aplicaciones informáticas tienen carácter descriptivo: muestran gráficamente áreas dialectales, áreas subdialectales e isoglasas. Sin embargo, aun es posible explotar estos materiales desde un punto de vista interpretativo.

El objetivo de esta presentación es mostrar la necesidad de colaboración entre dialectólogos y analistas de datos cuando se trabaja con materiales muy numerosos y que sólo pueden interpretarse utilizando técnicas de análisis cuantitativo. Además del análisis dialectométrico, que ya se ha aplicado utilizando la metodología de Hans Goebel, se presentarán diversos procedimientos de análisis de respuestas múltiples de morfología verbal basados en la frecuencia, la correlación, la clasificación cluster y el análisis de componentes principales. Este tipo de análisis puede ser extrapolado a datos dialectales de otras lenguas que presenten unas características similares.

2. Características y tratamiento informático de los datos verbales

En la edición original de *La flexió verbal*, extraída de los correspondientes cuadernos de campo (figura 1.1), los 67 verbos estudiados se clasificaban según su pertenencia a una determinada conjugación y según las formas de los diferentes tiempos verbales (infinitivo, gerundio, participio, presente de indicativo, imperfecto de indicativo, pretérito, futuro, condicional, presente de subjuntivo, imperfecto de subjuntivo, e imperativo). En el desarrollo del paradigma, cada forma se relacionaba con una variante morfológica y con la transcripción fonética de la desinencia correspondiente cada persona verbal. A lado de cada variante se incluía el número de localidad o de localidades (figura 1.2) donde se registró la forma en cuestión. La presentación impresa pretendía mostrar los datos verbales de manera sintética –su publicación ocupó sólo 368 páginas–, pero el esfuerzo de condensación y de simplificación del autor hizo dificultosa la consulta de los datos e impidió explotar todas sus posibilidades.

The image shows a page from a handwritten field notebook, likely a linguistic field notebook, containing a grid of verb forms. The grid is organized into columns and rows, with each cell containing a word or a set of words representing a specific verb form and its phonetic transcription. The forms are written in various languages, including Catalan, Spanish, and others. The grid is organized into columns and rows, with each cell containing a word or a set of words representing a specific verb form and its phonetic transcription. The forms are written in various languages, including Catalan, Spanish, and others.

Fig. 1.1. Una página de un cuaderno de campo

18. — CREURE

INFINITIU

Creure: krèuro 1-31, 34-37, 39, 41-55, 57-58, 60-62, 132, 137-138, 140, 142. krèure 30, 32-33, 36, 38, 40, 58-59, 87, 137. krèure 56. krèuri 56, 75, 86, 89. krèure 63-86, 88-107, 109, 111-117. krèurer 108, 110. kròura 118-121, 123, 126-131, 133-136, 139-141, 143-147. kròure 119, 122. kròurò 121, 125. krèurò 124. krèura 132. krèure 138-139, 141. krèura 148.

PARTICIPI PASSAT

Cregut: -üt 1-3, 5-148. -öt 4. — **Cres:** krés 84.

PARTICIPI PRESENT

Creient: krajén 1-2, 5, 11-13, 15-17, 22-24, 26-29, 33-41, 44-57, 59-62. krién 6. krién 63-67, 71-73, 75-76, 78-80, 82-87, 89, 91, 94. krajént 133, 141. krajént 148. — **Creguent:** krögén 1, 3-5, 8-12, 14, 16-19, 21-23, 25, 29-32, 39, 41-45, 48-49, 53, 57, 59, 142-147. kregén 68, 70, 73, 75, 81, 85, 94. kregént 101-107, 111. kregént 118, 120, 122, 124-125, 128-131. kregént 119-121, 123, 125-127, 132-141. — **Creuren:** krèuron 7, 18. — **Crevent:** krøbén 19-22, 27, 30-32, 40. — **Creent:** kræn 58. kræn 88, 90, 92-93, 95-100, 109, 112-113, 115. krènt 101, 103, 105, 107-108, 110, 114, 116-117. krènt 133, 135-136, 138, 141. — **Cregüent:** kregüén 77.

PRESENT D'INDICATIU

1.^a sg. — **Crech:** krèk 1-2, 5-14, 16-62, 87, 124, 132, 137-142. krèk 63-86, 88-117, 148. krók 118, 128-131. krók 119-123, 125-127, 133-136, 141, 143-147. — **Creni:** krépi 3, 5, 15-16. — **Cresi:** krézi 4.

2.^a sg. — **Creus:** kréus 1, 9, 11-12, 16-17, 19, 21. — **Creus:** kréus 2-3, 5-8, 10, 13-62, 87, 124, 132, 137-142. kréus 63-86, 88-117, 148. kréus 118-123, 125-131, 133-136, 141, 143-147. — **Creses:** krézas 4.

3.^a sg. — **Cren:** krép 1-62, 87, 124, 132, 137-142. krép 63-86, 88-117, 148. krép 118-123, 125-131, 133-136, 141, 143-147.

1.^a pl. — **Creym:** krajém 1-3, 5-8, 11-15, 22-24, 26-29, 31, 33-39, 41-42, 44-62. krajém 55. krajém 63-67, 69, 71-76, 78-80, 82-86, 89, 91, 94. krajém 87. krajém 148. — **Cresém:** krözém 4. — **Creguem:** krögém 9-10, 14, 16-19, 21, 30, 40, 43. krögém 68, 70, 73, 81, 85, 101-106, 108, 111. — **Crevem:** krøbém 16-17, 19-23, 25, 27, 30-32, 36, 40, 51. — **Cresevem:** krözøbém 17, 25. — **Crenem:** krèpém 77. — **Creem:** krèem 88, 90, 92-93, 95-100, 104-105, 107-110, 112-117. — **Creym:** krèjm 118-123, 125-131, 133-136, 141, 143-147. krèjm 124, 132, 137-142.

2.^a pl. — **Creyeu:** -èu 1-3, 5-8, 11-15, 22-24, 26-29, 31, 33-39, 41-42, 44-62, 87. -èu 55, 63-67, 69, 71-76, 78-80, 82-86, 89, 91, 94, 148. — **Creseu:** -èu 4. — **Cregueu:** -èu 9-10, 14, 16-19, 21, 30, 40, 43. -èu 68, 70, 73, 81, 85, 101-106, 108, 111. — **Creveu:** -èu 16-17, 19-23, 25, 27, 30-32, 36, 40, 51. — **Creseveu:** -èu 17, 25. — **Creueu:**

Fig. 1.2. La edición impresa de *La flexió verbal*

La introducción de los materiales en una base de datos (Perea 1999) generó un corpus de 470,255 registros, adaptados al AFI, y, superando las dificultades de visualización de la edición impresa, facilitó la realización de consultas simples y complejas y propició su cartografía automatizada (Perea 2001, 2004) (figura 3), a pesar de que *La flexión verbal* no fue concebida, en sus orígenes, como un atlas lingüístico. La mapificación automática permite generar más de 6.000 mapas, que son el resultado de multiplicar el número de formas de cada verbo (55) por el número de verbos estudiados (117). Este atlas morfológico es la suma de mapas individuales donde se visualizan fronteras dialectales que representan el inicio o el final de la utilización de una forma verbal particular o la superposición de resultados idénticos. Sin embargo, no

prevenciones posibles, una forma única que fuera la más representativa de cada localidad.

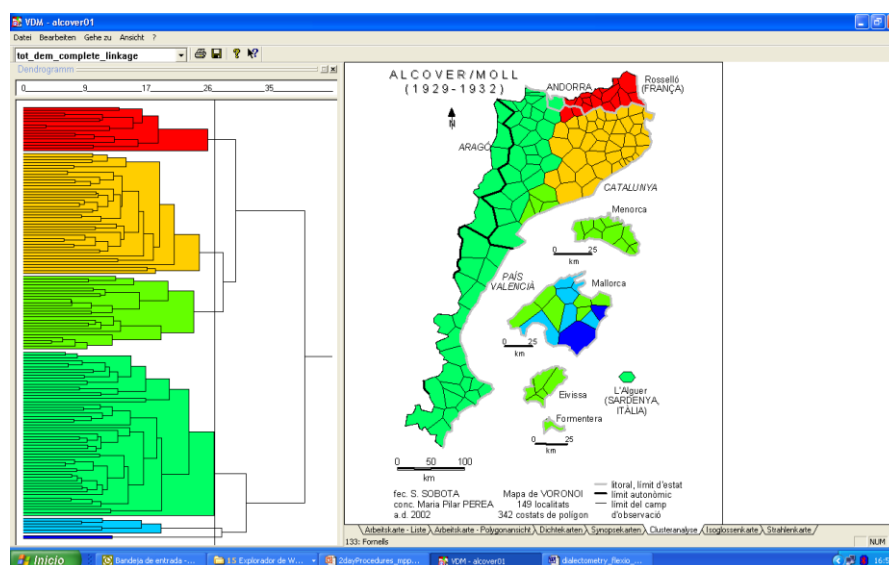


Fig. 1.4. Un mapa de cluster aplicando la metodología de Goebl

A continuación se mostrará la metodología que se ha utilizado para tratar de manera eficaz las respuestas múltiples de este corpus de morfología verbal.

2. Método

2. 1. Datos de respuestas múltiples

De la totalidad de los verbos investigados por Antoni M. Alcover, escogimos, en un primer estadio de la investigación, las formas correspondientes al presente de indicativo de 29 verbos y 128 localidades, para formar una matriz bidimensional¹. El problema a que nos enfrentamos es la presentación de numerosos casos de respuestas múltiples. Los informantes no se limitan a ofrecer una única forma para una pregunta específica, sino que pueden dar más de dos formas distintas. También puede haber casos de respuestas múltiples cuando se hacen encuestas a varias personas dentro de una

¹ Los verbos escogidos son: *anar, beure, bullir, cantar, caure, conèixer, córrer, cosir, coure, creure, deure, dir, dormir, entendre, estar, fer, fugir, mereixer, morir, nàixer, obrir, partir, perdre, poder, prendre, riure, sentir, tenir, valer*; en total 29 verbos. Estos verbos son completos en cuanto a las respuestas ofrecidas en todas las localidades. Excluimos las localidades siguientes, clasificadas según su pertenencia a una determinada área dialectal, por no ofrecer respuestas en algunas formas verbales: pirenaico oriental: 1 [Canet de Rosselló], 4 [Formiguera], 7 [Portè], 8 [Montlluís], 10 [Angostrina], 13 [Sallagosa], 23 [Sant Joan de les Abadesses]; catalán oriental: 41 [Vic]; catalán occidental: 66 [Boí], 67 [Andorra la Vella], 68 [Llavors], 70 [la Torre de Cabdella], 81 [Artesa de Segre], 85 [Cervera], 94 [Calaceit]; balear: 128 [Sineu], 132 [Manacor], 137 [Ferreries], 144 [Corona], 145 [Sant Jordi de ses Salines], 146 [Jesús]. Hechas estas excepciones quedan 128 localidades en total.

misma localidad. La figura siguiente demuestra en color rojo los casos de respuestas múltiples en una distribución bidimensional: 29 verbos por 128 localidades:

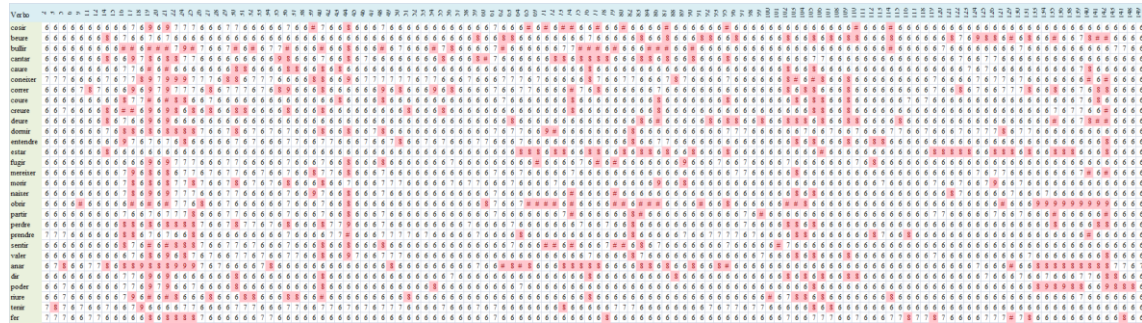


Fig. 2.1. Distribución de verbos por localidades, respuestas múltiples

2. 2. Grado de confluencia

La práctica usual en la dialectometría es tratar todas las formas igualmente sin distinción de las preguntas que se hacen a los encuestados. Por ejemplo en la Figura 2.2.a, tenemos dos formas coincidentes entre L2 y L3, lo mismo que entre L2 y L5. Los datos, sin embargo, se clasifican no solamente por las formas sino también por las categorías gramaticales a que pertenecen las formas. Las tres formas *vai*, *vaig*, *vau* pertenecen al verbo *anar* en primera persona singular, mientras que *anau* y *aneu* al mismo verbo pero en segunda persona plural. Deberíamos tener en cuenta esta estructura a la hora de calcular el grado de similitud entre las localidades².

Forma	L2	L3	L5	L14
anar1-vai	+	+	+	+
anar1-vaig		+	+	
anar1-vau			+	
anar5-anau				+
anar5-aneu	+	+	+	+

Fig. 2.2.a. Respuestas múltiples

[Conflu]	L2	L3	L5	L14
L2	1.000			
L3	0.854	1.000		
L5	0.789	0.908	1.000	
L14	0.854	0.707	0.642	1.000

Fig. 2.2.b. Grados de confluencia

Nuestra propuesta es evaluar el grado de similitud en términos del Grado de confluencia (G.C): la cantidad relativa de confluencia de las reacciones. Supongamos un caso en que la Localidad-1 ofrece dos formas y la L-2, igualmente dos y ninguno de estas formas coinciden entre ellas (Figura 2.2.c. izquierda).

² El coeficiente de correlación phi y otros índices que tratan los datos cualitativos toman en consideración los casos de no coincidencias. No obstante, no se considera en su cálculo la forma de agrupación en categorías. Con relación a los varios modos de calcular la similitud en los datos cualitativos, véase Ueda (1995).

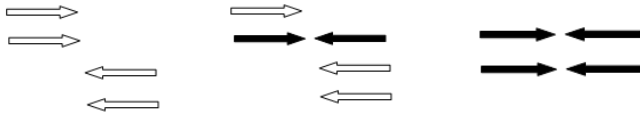


Fig. 2.2.c. Varios tipos de confluencia: c. nula, c. parcial, c. completa

Naturalmente este caso debería presentar un G.C. cero de similitud. El caso contrario es la coincidencia total, que cobraría el G.C. 1 en la escala normalizada entre cero y uno (derecha). La Figura central de 2.2.c representa un caso de confluencia parcial. Para la L-1 es un medio de confluencia, una entre las dos, mientras que para la

L-2, un tercio, una entre las tres. El G.C. entre estas dos localidades sería: $\sqrt{\left(\frac{1}{2} \times \frac{1}{3}\right)} = 0.408$. Por esta operación se obtiene la totalidad de la confluencia relativa. El signo de raíz que se pone encima es para bajar un grado de la elevación debida a la multiplicación que se hace entre las dos localidades. Si hubiera dos confluencias, el G.C.

sería: $\sqrt{\left(\frac{2}{2} \times \frac{2}{3}\right)} = 0.816$.

La fórmula del G.C. de una sola categoría que aquí proponemos es tan sencilla como la siguiente:

$$\text{Grado de confluencia (G.C.)} = \sqrt{\frac{a^2}{(a+b)(a+c)}} = \frac{a}{\sqrt{(a+b)(a+c)}}$$

donde a es número de confluencias entre las dos localidades; b es número de las formas lingüísticas no confluyentes en la L-1, c es número de las formas no confluyentes en la L-2. Así $(a + b)$ representa el número de todas las formas de L-1; $(a + c)$, el de L-2.

La fórmula del G.C. de todas las categorías se obtiene por la suma de G.C. dividida por el número de categorías:

$$\text{Grado de confluencia media (G.C.M.)} = \frac{\sum \frac{a}{\sqrt{(a+b)(a+c)}}}{n}$$

donde n es número de categorías.

Volviendo al ejemplo presentado en la Figura 2.2.a., podemos calcular como sigue:

$$G. C. M. (L2:L3) = \frac{\sqrt{\left(\frac{1}{1} \times \frac{1}{2}\right)} + \sqrt{\left(\frac{1}{1} \times \frac{1}{1}\right)}}{2} = 0.854$$

$$G. C. M. (L3:L5) = \frac{\sqrt{\left(\frac{2}{2} \times \frac{2}{3}\right)} + \sqrt{\left(\frac{1}{1} \times \frac{1}{1}\right)}}{2} = 0.908$$

Nuestro dato concreto ofrece una matriz de G.C.M de la manera siguiente (parte inicial):

[Conflu]	2	3	5	6	9	11	12	14	15	16	17	18	19	20
2	1.000													
3	0.929	1.000												
5	0.922	0.883	0.994											
6	0.920	0.873	0.905	1.000										
9	0.846	0.855	0.848	0.805	1.000									
11	0.850	0.840	0.907	0.857	0.869	1.000								
12	0.861	0.851	0.918	0.869	0.857	0.939	1.000							
14	0.754	0.706	0.814	0.790	0.744	0.770	0.781	1.000						
15	0.794	0.781	0.836	0.810	0.822	0.870	0.859	0.819	1.000					
16	0.753	0.728	0.814	0.760	0.751	0.797	0.808	0.847	0.827	1.000				
17	0.711	0.674	0.764	0.727	0.700	0.745	0.757	0.812	0.758	0.887	1.000			
18	0.690	0.634	0.713	0.689	0.687	0.675	0.687	0.889	0.721	0.782	0.752	1.000		
19	0.749	0.720	0.786	0.744	0.741	0.775	0.786	0.847	0.792	0.901	0.889	0.806	0.994	
20	0.757	0.715	0.823	0.790	0.744	0.789	0.800	0.877	0.831	0.903	0.870	0.801	0.876	1.000

Fig. 2.2.d. Grados de confluencia (parte inicial)

3. Interpretación

3.1. Evidencias cuantitativas

El resultado de cálculos sin más como el de la sección anterior es abstracto y difícil de evaluar, sin apoyo de otras evidencias. En nuestro caso concreto podríamos y deberíamos ofrecer las evidencias siguientes:

- | | |
|---|---------------------|
| { | Evidencia intuitiva |
| | Evidencia teórica |
| | Evidencia práctica |
- | | |
|---|----------------------------|
| { | Evidencia práctica externa |
| | Evidencia práctica interna |

Intuitivamente, las cifras de 0.408, 0.816, 0.854, 0.908 en los casos tratados en la sección anterior y otras cifras presentadas en la Figura 2.2.b. serán convincentes a ojos del investigador. Las fórmulas presentadas de G.C. y G.C.M resultan siempre cero en caso de la(s) confluencia(s) nula(s) y 1 en caso de la(s) confluencia(s) completa(s), lo

cual nos convence sin problema. El G.C. del caso de la confluencia parcial de la Figura 2.2.c., un 0.408, puede parecer alto por tratarse de un caso entre los 2 (L-1) + 3 (L-2) miembros. No obstante la cifra es razonable por ser una confluencia de una mitad por parte de L-1, y un tercio por parte de L-2, entre las cuales se sitúa el G.C. (0.408).

En el plano teórico la ecuación de G.C. es derivable matemáticamente a partir del Coeficiente de correlación phi:

$$\text{Coeficiente de correlación phi (C.C.P.)} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$$

donde d es número de formas lingüísticas que no se presentan ni en L-1 ni en L-2. En teoría la cantidad de d se supone infinita, de modo que el C.C.P. resulta ser equivalente a G.C.³:

$$\begin{aligned} \text{C.C.P. (} d > \infty \text{)} &= \frac{(ad - bc) / d}{\sqrt{(a+b)(a+c)(d+b)(d+c) / d^2}} \\ &= \frac{(ad / d - bc / d)}{\sqrt{(a+b)(a+c)(d / d + b / d)(d / d + c / d)}} \\ &= \frac{a}{\sqrt{(a+b)(a+c)}} = \text{G.C.} \end{aligned}$$

Las evidencias prácticas externas en forma de cartografía sirven de apoyo para demostrar la validez del análisis. El siguiente gráfico demuestra la distribución geográfica de puntos investigados con clasificación hecha por el método de cluster, a partir de la matriz simétrica de la Figura 2.2.d.

³ Sobre el uso del coeficiente phi y el de su versión modificada, véanse Kroeber (1937, 1969), Ellegard (1959), Herdan (1964).

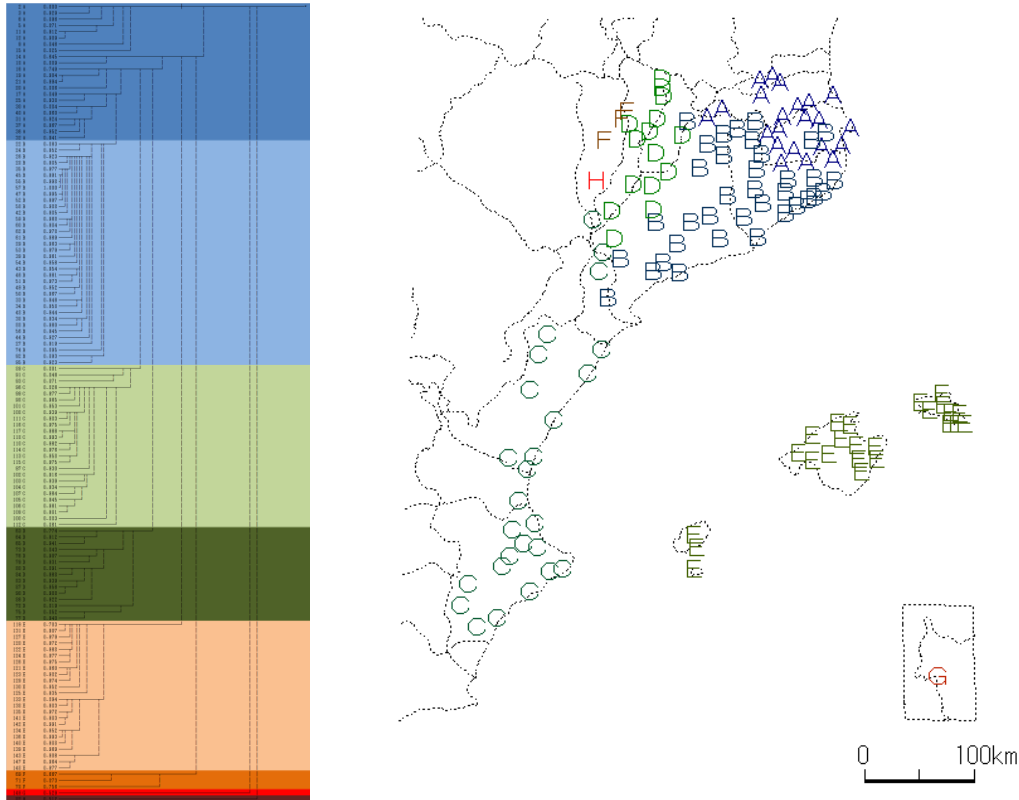


Fig. 2.3.a Análisis de cluster y distribución geolingüística⁴

Las congruencias de los puntos que se observan dentro del mismo grupo y la continuidad geográfica que se aprecia entre los distintos grupos puede ser una evidencia externa que asegura la robustez del cálculo de G.C.

Dentro de la distribución misma se puede ofrecer otra evidencia de carácter interno. Ahora el análisis de cluster no se aplica solo a las localidades, sino también a las formas lingüísticas, con el propósito de adquirir un alto grado de conglomeración en la distribución de reacciones en el gráfico bidimensional⁵. Véase la Figura 2.3.b.

⁴ El mapa fue dibujado por medio del sistema Mandara, elaborado por el profesor Kenji Tani de la Universidad de Saitama (Japón). Su sitio es: <http://ktgis.net/mandara/>

⁵ Para el método de cluster patronizado, véase Ueda (en prensa).

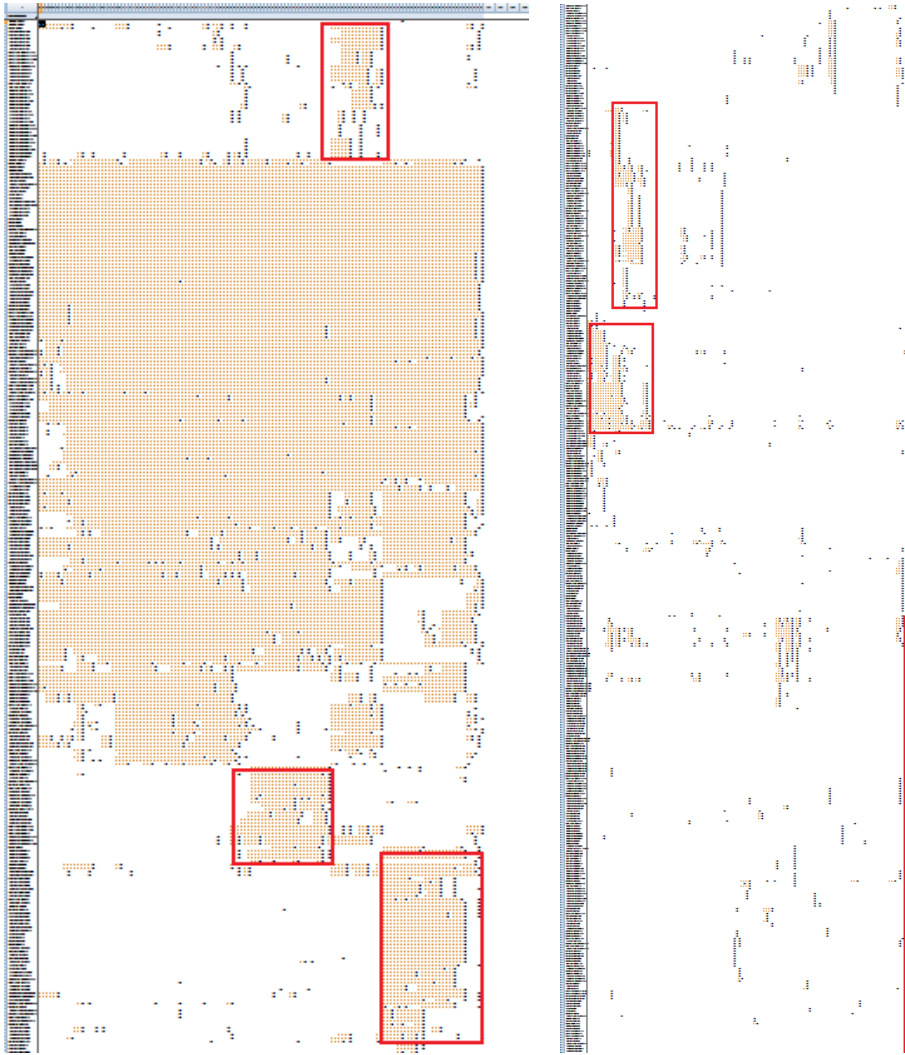


Fig. 2.3.b Análisis de cluster patronizado

En el eje vertical se colocan las formas lingüísticas y en el eje horizontal, las localidades. Tanto el orden de las formas lingüísticas como el de las localidades son debidos a los resultados del análisis cluster. Lo curioso de esta figura es que suele ofrecer unas conglomeraciones peculiares gracias a la agrupación que ofrece el análisis de cluster tanto de las localidades como de las formas lingüísticas. Esto es natural puesto que el propósito del análisis de cluster es una taxonomía numérica presentada en forma de agrupación de miembros, agrupación que se consigue por el índice razonable de la confluencia⁶.

⁶ También el lenguaje “R” tiene la función de hacer lo parecido, denominado “Heat map”. El paquete gratuito del Proyecto “R”, elaborado por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland (Nueva Zelanda) se encuentra en: <http://www.r-project.org/>.

Nos llama la atención por presentar unas conglomeraciones peculiares casi exclusivas, las partes marcadas por líneas rojas, de las que comentaremos en la sección siguiente tanto desde el punto de vista lingüístico como geográfico.

3.2. Evidencias cualitativas

a) De tipo geográfico

En el análisis de clúster se muestra la distribución de las localidades estudiadas clasificadas en 8 grupos (A, B, C, D, F, G y H), los cuales se corresponden parcialmente con las seis agrupaciones dialectales en que se suele dividir el dominio lingüístico catalán: rosellonés [pirenaico oriental, en terminología de Alcover] (A), catalán oriental (B), valenciano (C), catalán occidental (D), balear (E) y la localidad del Alguer en Cerdeña (G), y que el mismo Alcover ya constató. A estos seis grupos se añade la subdivisión que aparece en el extremo occidental, que coincide en general con la franja de Aragón, área de transición que se halla entre el territorio catalán y el aragonés: las localidades de la franja de Aragón – Vilaller, Bonansa y Benavarri – (F) y al sur aisladamente la localidad de Tamarit (H).

El análisis de cluster corrobora, pues, la división dialectal que el mismo Alcover propuso y añade dos nuevas subáreas que coinciden con un área de transición, que, aunque se tiene en cuenta en los estudios sobre la variación lingüística, nunca ha servido para determinar la existencia de un área dialectal autónoma.

Examinemos a continuación la composición de cada una de las agrupaciones:

A – comprende no sólo la zona del Rosellón en territorio francés sino lo que se conoce actualmente como catalán septentrional de transición, ya en tierras catalanas. Observemos como en el diagrama aparece también, por la intersección entre puntos de A y de B, delimitada una clara zona de transición.

B – corresponde al denominado catalán oriental (y actualmente catalán central). En este caso no existe intersección, y por tanto una clara zona de transición, ni con el área del catalán occidental (D) ni con el valenciano (C), mediante la superposición de los puntos de cada zona. Sin embargo, es curioso observar como este área se extiende hacia el sur, incluyendo dos localidades (92: Falset y 95: Tortosa), consideradas pertenecientes lingüísticamente al catalán occidental, y, en el caso de Tortosa, como una zona de transición con el valenciano. Quizá podremos obtener una respuesta desde el punto de vista lingüístico.

C – comprende la zona valenciana, sin hacer ninguna diferenciación entre los tres subdialectos que componen esta área (septentrional, apitxat y meridional). Es necesario hacer constar la ascensión hacia el norte de la frontera hasta una localidad propiamente del catalán occidental como es Fraga, donde puede preverse un área de transición.

D – integra las localidades correspondientes al catalán occidental, excepto 93: Gandesa, 91: Riba-roja y 89: Fraga, que sitúan en la zona valenciana.

E – delimita el área de las Islas Baleares, sin mostrar subdivisiones entre las islas de Mallorca, Menorca e Ibiza.

G – configura la localidad del Alguer en la isla de Cerdeña.

F – comprende *grosso modo* el centro y el norte del área de transición de la franja de Aragón.

Es interesante establecer una comparación con los resultados del análisis de cluster propuesto mediante el programa VDM de Hans Goebel. Observemos como, mediante el cálculo RIW (relative identity value), obviando las respuestas múltiples, una clasificación en ocho grupos ofrece una distribución parcialmente distinta de la anteriormente propuesta:

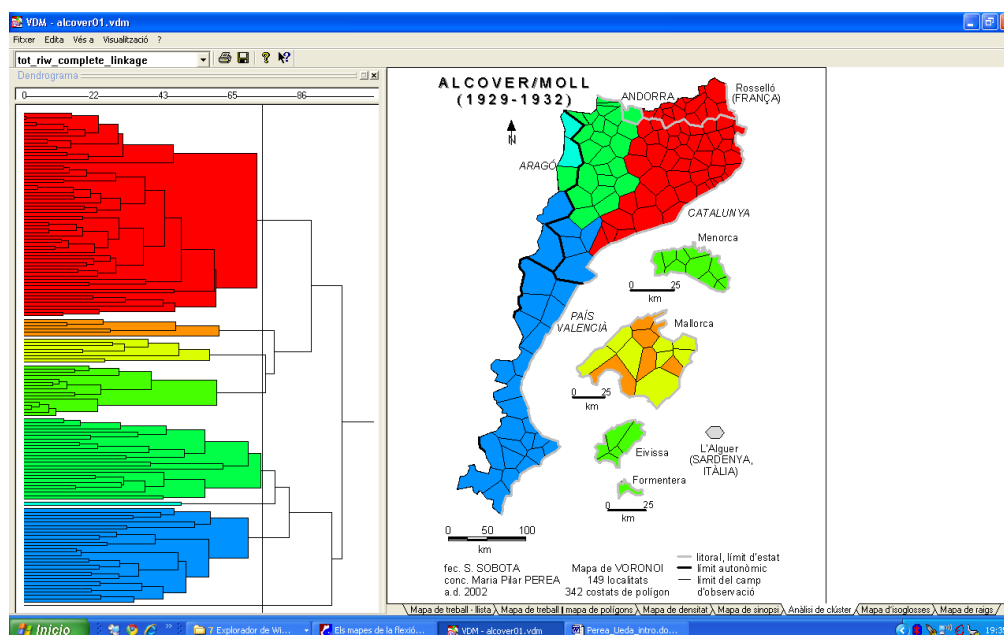


Fig. 3.1. Análisis de cluster según la propuesta de Goebel

El color rojo aún inesperadamente las áreas del pirenaico oriental y del catalán oriental. El color azul muestra el área valenciana. El color verde, con subdivisiones

dentro de la isla de Mallorca, corresponde a las variedades insulares. El dialecto alguerés, como es previsible, se muestra aislado de los demás. En último término también se aprecia en color azul claro un área que puede asociarse con la transición entre catalán y aragonés; y en este punto existe coincidencia con nuestros resultados plasmados en F.

b) De tipo lingüístico

La divisoria – que responde tanto a la intuición como a la realidad lingüística del momento en que Alcover desarrolló las encuestas – entre el pirenaico oriental (incluyendo el área actual correspondiente al catalán de transición) y el catalán oriental, que aparece en nuestro análisis, muestra la coincidencia que existe en algunas formas verbales del presente de indicativo de todas las conjugaciones. Esta coincidencia se manifiesta tanto en la primera persona – cf. el análisis de cluster patronizado en la figura 2.3b –: *podí* ‘I can’, *beuí* ‘I drink’, *riuí* ‘I laugh’, *deuí* ‘I must’, que es compartida por las localidades próximas a los Pirineos, en el territorio francés (en el norte) y en el catalán (en el sur) – como en la segunda– *vales* ‘you worth’, *riués* ‘you laugh’, *beues* ‘you drink’, *caues* ‘you fall’, *podes* ‘you can’, *neixí* ‘I was born’, *perdí* ‘I lose’, *senti* ‘I feel’. Estas formas contrastan con las propias del catalán oriental: primera persona: *puc* ‘I can’, *bec* ‘I drink’, *ric* ‘I laugh’, *dec* ‘I must’; segunda persona: *vals* ‘you worth’, *rius* ‘you laugh’, *beus* ‘you drink’, *caus* ‘you fall’, *pots* ‘you can’, *neixo* ‘I was born’, *perdo* ‘I lose’, *sento* ‘I feel’. El contraste entre estos resultados, caracterizados por las desinencias *-i* (en pirenaico oriental) y *-u* (en catalán oriental), en la primera persona, y la presencia vs. la ausencia del morfo *-e-* [ə] de modo/tiempo en la segunda persona (*vales*, *riués* vs. *valØs*, *riuØs*) de verbos de la segunda y la tercera conjugación, y otros de características similares, provocan la división entre los dos dialectos.

El análisis de cluster patronizado también pone de manifiesto la transición entre los dos dialectos citados, a través de la existencia de formas con desinencia *-i* y con el morfo *-e-* [ə], que son utilizadas en algunas localidades del norte del catalán oriental (22: Ribes de Freser, 24: Ripoll): *dormí* ‘I sleep’, *cusi* ‘I sew’; (22: Ribes de Freser, 24: Ripoll, 26: la Pobla de Lillet): *dormes* ‘you sleep’, *sentés* ‘you feel’, las cuales pertenecen al actual catalán septentrional de transición.

El área correspondiente al catalán oriental forma un conjunto bastante homogéneo con relación a las formas verbales, y no merece un comentario especial.

En cuanto al valenciano, se observan soluciones homogéneas en las primeras personas del presente de indicativo: *bullc* ‘I boil’, *cusc* ‘I sew’, *mereixc* ‘I deserve’, *fuigc* ‘I escape’, *naixc* ‘I was born’, *perc* ‘I lose’, compartidas por todas las localidades. Tienen también una extensión casi general las formas *correc* ‘I run’, *obric* ‘I open’, *dormc* ‘I sleep’ o *partixc* ‘I divide’. Las formas de segunda persona también tienen un uso mayoritario: *naixes* ‘You were born’, *partixes* ‘you divide’ o *obris* ‘you open’. Es igualmente apreciable como las formas *creem* ‘we believe’, *creeu* ‘you believe’, *naixes* ‘you were born’, *naix* ‘’, *naixen* ‘’, que se corresponden con las localidades 93: Gandesa, 91: Riba-roja y 89: Fraga, anteriormente citadas, parecen configurar una área de transición hacia el valenciano. Nuevos análisis aplicados a otros tiempos verbales pueden corroborar esta hipótesis.

El catalán occidental se muestra, a través del análisis de cluster patronizado, en coincidencias absolutas en el área con las formas *vac* ‘I go’, *vai* ‘I go’, *deiem* ‘we say’, *deieu* ‘you say’, y en coincidencias mayoritarias del tipo *coiem* ‘we cook’, *coieu* ‘we cook’, *partixo* ‘I divide’ o *sinto* ‘I feel’. Existe un mayor grado de dispersión en formas como *aurixo* ‘I open’, *aubrixo* ‘I open’ o *bullixo* ‘I boil’, propias de determinadas localidades de la zona. Las localidades 92: Falset y 95: Tortosa se distancian de 89: Fraga y 91: Riba-roja d’Ebre. Estos dos últimos dos puntos, junto con 93: Gandesa, forman un subgrupo dentro del grupo C, apartándose de las localidades restantes del mismo grupo e indican una zona de transición. Las cinco localidades mencionadas (Falset, Tortosa, Fraga, Riba-roja d’Ebre y Gandesa) coinciden en algunas formas y se diversifican también en otras.

El balear, como agrupación dialectal, manifiesta un conjunto importante de realizaciones comunes, especialmente en las formas de primera persona de presente de indicativo sin desinencia: *cant* ‘I sing’, *obr* ‘I open’, *fuig* ‘I escape’; en la segunda y tercera personas del plural del presente de indicativos de los verbos *prendre* ‘to take’ (*prenim* ‘we take’, *preniu* ‘you take’), *dir* ‘to say’ (*deim* ‘we say’, *deis* ‘you say’), *fer* ‘to make’ (*feim* ‘we make’, *feis* ‘you make’) o *creure* ‘to believe’ (*creim* ‘we believe’, *creis* ‘you believe’).

El alguerés consta en el análisis de cluster patronizado con una representación única que recoge las formas exclusivas e idiosincrásicas de esta variedad dialectal. Este comportamiento único se muestra especialmente los presentes de indicativo: *beun* ‘they drink’, *caun* ‘they fall’, *rien* ‘they laugh’; y en la flexión de los verbos *romir* (Est. Cat.

dormir ‘to sleep’: *rom* ‘I sleep’), *paltir* (Est. Cat. *partir* ‘to divide: *palteix* ‘I divide’) o *pelda* (Est. Cat. *perdre* ‘to lose: *pelt* ‘I lose’).

Las localidades estudiadas correspondientes a la franja de Aragón son las siguientes: 69: Vilaller, 71: Bonansa, 78: Benavarri y 82: Tamarit de Llitera. Las tres primeras se enmarcan en la agrupación F y la última forma en solitario la agrupación H. Aunque presentan coincidencias formales en la realización de algunas personas, la localidad de Tamarit muestra diferencias notables en la flexión de algunos verbos (*corrs* ‘you run’, *corr* ‘he runs’, *aurisses* ‘you hear’, *aurís* ‘he hears’, *aurissen* ‘they hear’, así como en la presencia de una extensión velar en formas como *coneguem* ‘we know’ / *conegueu* ‘you know’, *entenguem* ‘we understand’ / *entengueu* ‘you understand’, *valguem* ‘we cost’ / *valgueu* ‘you cost’, *prenguem* ‘we take’ / *prengueu* ‘you take’, alternando con soluciones sin extensión), lo cual la aleja de la agrupación anterior (que presentan resultados exclusivos sin la aparición de la extensión velar: *coneixem* / *coneixeu*, *entenem* / *enteneu*, *valem* / *valeu*, *prenem* / *preneu*). Con la aplicación de la metodología estadística usada con las formas de presente de indicativo a todos los datos verbales que Alcover recopiló se podrá plantear la existencia de esta localidad o bien de manera aislada o bien incorporada a los puntos restantes que configuran la franja aragonesa.

4. Conclusiones

1. Esta primera aproximación al tratamiento de los datos verbales de Alcover ha demostrado la idoneidad del método aplicado a respuestas múltiples.

2. El análisis de cluster que se ha utilizado refleja la realidad de la variación lingüística que experimentaba el verbo catalán a principios del siglo XX. Otros cálculos basados en el uso de respuestas únicas ofrecen sólo resultados parciales que no son absolutamente representativos.

3. La aplicación del método a las formas verbales completas de presente de indicativo de 29 verbos de 128 localidades ha mostrado una configuración geográfica en áreas dialectales que no se aleja de la divisoria tradicional en que Alcover dividió el dominio lingüístico catalán: pirenaico oriental, catalán oriental, catalán occidental, valenciano, balear y alguerés. Sin embargo, en el análisis de cluster aparecen otras dos áreas, situadas geográficamente en la franja de Aragón, que se ha integrado tradicionalmente al catalán occidental. Un examen más completo permitirá resolver dos cuestiones que aun quedan pendientes: 1) la existencia independiente de estas dos áreas

o la unificació de les mateixes en una sola que se alargue de nord a sud; y 2) en qualsevol dels dos casos, la constatació de la existència d'un nou dialecte corresponent a esta àrea, la qual fins ara se subdividia en tres subdialectes (pallares, ribagorçano y leridano), però que considerava la relevància de la frontera catalano-aragonesa solament com a àrea de transició.

4. En quant a les àrees de transició, l'anàlisi ha deixat entreveure agrupacions de localitats, integrades en determinades àrees, que poden considerar-se de caràcter transicional. El ús de tots els dades en els càlculs estadístics permetrà definir-les més clarament i remarcarà els elements lingüístics que permeten configurar-les.

References

- ALCOVER, A. M. & MOLL, F. de B. 1929-1933. *La flexió verbal en els dialectes catalans*, Barcelona: Anuari de l'Oficina Romànica (v. II (1929) [73] 1- [184] 112, v. III (1930) [73] 1- [168] 96, v. IV (1931) [9] 1- [104] 96, v. V (1932) [9] 2 - [72] 64).
- ELLEGARD, Alvar. 1959. "Statistical measurement of linguistic relationship." *Language*, 35, 131-156.
- GOEBL, Hans. 1997. "Some Dendographic Classifications of the Data of CLAE 1 and CLAE 2", in W. Viereck and H. Ramisch (eds.), *The Computer Developed Linguistic Atlas of England 2*, Tübingen: Max Niemeyer Verlag, 22-32.
- GOEBL, Hans, 2003. "Regards dialectométriques sur les données de l'Atlas Linguistique de la France (ALF): Relations quantitatives et structures de profondeur", *Estudis Romànics*, XXV, 61-117.
- HERDAN, Gustav. 1964. "Mathematics of genealogical relationship between languages", *Proceedings of the 9th International Congress of Linguistics*, The Hague: Mouton, 51-60.
- KROEBER, Alfred L. 1960. "Three quantitative classifications of Romance", *Romance Philology*, 14, 189-195.
- KROEBER, Alfred L. and CHRETIEN, C. D. 1937. "Quantitative classification of Indo-European languages", *Language*, 13, 83.
- KROEBER, Alfred L. and CHRETIEN, C. D. 1960. "Statistics, Indo-European and taxonomy." *Language*, 36, p. 1-21.
- PEREA, Maria-Pilar. 1999. *'La flexió verbal en els dialectes catalans' d'A. M. Alcover i F. de B. Moll. Compleció i ordenació*, Barcelona: Institut d'Estudis Catalans, CD-ROM.
- PEREA, Maria-Pilar. 2001. *'La flexió verbal en els dialectes catalans' d'A. M. Alcover i F. de B. Moll. Les dades i els mapes*, Palma de Mallorca: Conselleria d'Educació i Cultura. Govern de les Illes Balears, CD-ROM.

- PEREA, Maria-Pilar. 2008a. "Catalan verb morphology and dialectometric analysis", in G. Blaikner-Hohenwart, E. Bortolotti, R. Franceschini et alii (ed.), *Ladinometria, Festschrift für Hans Goebel zum 65. Geburtstag*, 2, Salzburg: Universität Salzburg, 61-78.
- PEREA, Maria-Pilar. 2008b. "Retrieving the sound: applying speech synthesis to dialectal data", comunicació presentada a *The Thirteen International Conference on Methods in Dialectology*, 4-8 agost 2008, Leeds (en prensa).
- PEREA, Maria-Pilar. 2004. "New Techniques and Old Corpora: La flexió verbal en els dialectes catalans (Alcover-Moll, 1929-1932). Systematisation and Mapping of a Morphological Corpus", *Dialectologia et Geolinguistica*, 12, 25-45.
- UEDA, Hiroto. 1995. "Zonificación del español. Palabras y cosas de la vida urbana", *Lingüística (ALFAL)*, 7, pp.43-86.
- UEDA, Hiroto. (en prensa) "Métodos de tratamiento informatizado de datos lingüísticos hispánicos: Observaciones inmediatas y abstracciones matemáticas", Curso de verano en la Universitat Jaume Primer, Castellón de la Plana, España, 2009/7/8.