

ALGUNAS CONSIDERACIONES SOBRE LA CLASIFICACIÓN DIALECTAL DE LAS VARIETADES DEL CATALÁN*

Esteve Clua (Universitat Pompeu Fabra, Barcelona), esteve.clua@upf.edu

Maria-Rosa Lloret (Universitat de Barcelona), mrosa.lloret@ub.edu

0. INTRODUCCIÓN

En el ámbito de la geolingüística románica —y en este sentido la lengua catalana no fue ninguna excepción—, los estudios dialectales tradicionales, siguiendo criterios básicamente cualitativos, realizaban las clasificaciones de las variedades geográficas a partir de la noción de isoglosa. Pero desde principios de los años setenta del siglo pasado, se han ido desarrollando diversos métodos de descripción y clasificación dialectal basados en criterios cuantitativos, que comparten la característica de tratar de forma sintética y global el conjunto de datos de un corpus o atlas dialectal. También los caracteriza el hecho de fundamentarse en el concepto de distancia lingüística.

Este tipo de análisis de la variación, generalmente de carácter cuantitativo, recibe el nombre de dialectometría y supuso un cambio radical en relación a los estudios geolingüísticos anteriores, basados en el concepto de isoglosa. Todos los estudios sobre las agrupaciones dialectales del catalán realizados por el grupo de investigación que trabaja con el Corpus Oral dialectal del catalán contemporáneo (COD) (*vid.* <http://www.ub.edu/lincat/> y también http://www.iula.upf.edu/rec/vdm_val/) se han inscrito de lleno en la disciplina dialectométrica.

El objetivo del presente trabajo es describir el análisis dialectométrico aplicado al COD y presentar los resultados más relevantes en relación a las clasificaciones dialectales del catalán realizadas desde la perspectiva de los estudios geolingüísticos clásicos. Además relacionaremos estos resultados con el nuevo

* Este trabajo forma parte de los proyectos de investigación coordinados FFI2010-22181-C03-02 y FFI2010-22181-C03-03, financiados por el MICINN y el FEDER, y del grupo de investigación consolidado 2009SGR521, financiado por la Generalitat de Catalunya. El análisis dialectométrico al que nos referimos en este trabajo ha sido descrito también en Clua *et alii* (2009) y en Clua (2010). Desde aquí queremos agradecer a los profesores Miquel Salicru y Sergi Civit del Departamento de Estadística de la Universitat de Barcelona toda la ayuda que nos han prestado en el diseño y ejecución del tratamiento estadístico de los datos del COD.

proyecto de investigación (DiLET), que a partir de la actualización del corpus pretende analizar la variación dialectal del catalán tanto desde el eje espacial como desde el eje temporal.

Para ello empezaremos por presentar sucintamente las características de nuestro corpus oral (§1); a continuación (§2), describiremos el análisis dialectométrico llevado a cabo con los datos del COD, que, a diferencia de lo que ocurre con otros tratamientos cuantitativos de esta índole, se caracteriza por la realización de un análisis lingüístico previo. Seguidamente (§3), expondremos las consideraciones más relevantes en cuanto a la estructura de interdistancias lingüísticas de las variedades del catalán; y, finalmente (§4) presentaremos las propuestas de futuro del nuevo proyecto DiLET.

1. EL CORPUS ORAL DIALECTAL DEL CATALÁN CONTEMPORÁNEO (COD)

Desde 1991, el Departamento de Filología Catalana de la Universitat de Barcelona compiló y sistematizó en bases de datos el Corpus Oral Dialectal del catalán contemporáneo (COD), que contiene información de las seis principales variedades dialectales del catalán: alguerés (la variedad de la ciudad sarda de l'Alguer), balear, central, norte-occidental, rosellonés y valenciano.¹ Este corpus se constituyó a partir de una serie de entrevistas que se llevaron a cabo en las 82 capitales de comarca (o ciudades equivalentes) del ámbito lingüístico catalán.

Consta de materiales fonéticos y morfológicos obtenidos a través de un cuestionario de unos 600 ítems y de un conjunto de textos espontáneos.² Los datos están recogidas en ocho bases de datos diferentes: morfología verbal regular, aspectos fonéticos, clíticos pronominales, artículos, posesivos, pronombres personales tónicos, demostrativos y locativos.³ Y el número de registros que contiene cada una de estas bases de datos es el de (1).

En cuanto al cuestionario, a pesar de que constaba de un núcleo común de preguntas para poder establecer la adecuada comparación entre las diferentes

¹ Sobre las clasificaciones de las variedades del catalán, *vid.* Colomina i Castanyer (1999) y Veny (1982).

² En cuanto a estos textos, *vid.* Viaplana & Perea (2003) y también el apartado DADES de <http://www.ub.edu/lincat/>.

³ Para una descripción más amplia del COD *vid.* Lloret & Perea (2002). Existe una edición en CD-Rom de los datos del COD en Viaplana *et alii* (2007); *vid.* también <http://www.ub.edu/lincat/>.

variedades del catalán, también tenía partes específicas dirigidas a recoger los aspectos idiosincrásicos de cada variedad.

En el momento en el que se llevaron a cabo las encuestas los informantes tenían entre 30 y 45 años, que en la mayoría de territorios era la edad límite para contar con personas que no habían sido escolarizadas en catalán; además, se seleccionaron informantes con un nivel cultural medio. Los criterios utilizados para la selección de localidades e informantes se debían a la voluntad de compilar un corpus dialectal más urbano que los tradicionales, que mostrara de una manera clara cuáles eran las características del catalán contemporáneo.

(1)

Base de datos	Número de registros
Verbos	68262
Fonética	55270
Clíticos pronominales	24766
Artículos	5049
Posesivos	5223
Pronombres personales	1608
Demostrativos	1512
Locativos	803

2. EL ANÁLISIS DIALECTOMÉTRICO APLICADO A LOS DATOS DEL COD

Con la finalidad de caracterizar el análisis llevado a cabo, nos centraremos, en principio, en uno de los aspectos que singulariza nuestro trabajo dialectométrico: el análisis lingüístico previo al tratamiento estadístico.

Si alguna cosa caracteriza globalmente los diferentes métodos cuantitativos de descripción y clasificación dialectal que han proliferado durante las últimas cuatro décadas en Europa, es el hecho de compartir la adopción del concepto de distancia lingüística como pieza básica para la descripción de la variación lingüística. El concepto de distancia, adoptado del ámbito científico del análisis de datos, se asocia generalmente en nuestro campo a la cuantificación de las similitudes o diferencias que existen entre variedades lingüísticas en relación a un conjunto de datos.

Pero ¿cómo determinamos las diferencias o similitudes entre variedades lingüísticas a partir de las cuales establecer el tratamiento cuantitativo? ¿Reflejan las representaciones fonéticas de los ítems de un determinado atlas o corpus todas las diferencias existentes entre las variedades lingüísticas? Vamos a intentar responder a estas preguntas a partir de los ejemplos siguientes en los que contrastamos las formas del sustantivo *nas* ‘nariz’ en dos variedades lingüísticas.

(2)	<u>Variedad 1</u>	<u>Variedad 2</u>
a.	nas [ˈnas] ‘nariz’	nas [ˈnas]
b.	nas petit [ˈnaspeˈtit] ‘nariz pequeña’	nas petit [ˈnaspeˈtit]
c.	nas gran [ˈnazˈɣran] ‘nariz grande’	nas gran [ˈnazˈɣran]
d.	nas ample [ˈnaˈzample] ‘nariz ancha’	nas ample [ˈnaˈsample]

Si contrastamos las realizaciones fonéticas del sustantivo aislado (2a) en las dos variedades, vemos que no presentan ninguna diferencia. Tampoco hay diferencias cuando aparece delante de otra palabra empezada por consonante sorda (2b). Cuando precede una palabra empezada por consonante sonora (2c) tampoco encontramos diferencias entre ambas variedades, pero en este caso podemos observar que el sonido sibilante del final del sustantivo presenta una realización sonora [z], que contrasta con las realizaciones sordas de este segmento en los contextos anteriores. Finalmente en (2d) sí que podemos observar una realización diferente del sustantivo en las dos variedades contrastadas: mientras que la variedad 1 presenta una realización sonora de la sibilante final, en la variedad 2 encontramos una realización sorda.

A partir de las primeras observaciones podríamos concluir, en principio, que estas dos variedades no presentan ninguna diferencia en la representación fonética del sustantivo *nas* [ˈnas], porque los tres segmentos que integran esta palabra coinciden totalmente; pero la coincidencia no puede ser total si tenemos en cuenta el comportamiento del último segmento, el sibilante [s], que se sonoriza siempre

delante de un segmento sonoro [\pm vocálico] en la variedad 1, mientras que solo lo hace delante de un segmento no vocálico en la variedad 2. Es decir, las dos variedades coinciden totalmente en los segmentos fonéticos que constituyen el sustantivo *nas*, pero los procesos fonológicos que afectan a dichos segmentos son diferentes. Si entendemos que estos procesos fonológicos son básicos para comprender la estructura sonora de las variedades lingüísticas, tendremos que concluir asimismo que no podemos pasarlos por alto al intentar captar las diferencias existentes entre ellas.

Para poder tener en cuenta todas estas diferencias, en nuestros estudios aplicamos a los datos fonéticos del corpus un análisis lingüístico basado en el modelo de la fonología generativa clásica, porque además de que nos permite captar este tipo de diferencias, que serían imperceptibles a partir de un simple análisis fonético, también nos permite discriminar las diferencias superficiales o predecibles (fonéticas), que expresan las regularidades de las lenguas (y de las variedades que las componen), de las diferencias subyacentes o impredecibles (fonológicas), que afectan a la estructura léxica o gramatical de las palabras.⁴ Desde nuestro punto de vista, la distinción entre estos dos niveles de análisis es fundamental para determinar la distancia lingüística entre variedades.⁵

Ejemplificaremos esta distinción entre diferencias fonológicas y fonéticas con ejemplos de la morfología verbal extraídos de Clua (2010). El catalán presenta distintas terminaciones para la 1ª persona del singular del presente de subjuntivo: unas variedades presentan $-[e]$ (en la mayor parte del área valenciana), otras $-[i]$ (variedades del catalán oriental) y otras $-[a]$ (en algunas variedades del catalán norte-occidental). Ninguna de estas diferencias, sin embargo, puede ser atribuida a un fenómeno sistemático de la fonología del catalán; es decir, en estas variedades no existe ningún proceso regular por el cual $-e/$ se convierta en $-[i]$ o en $-[a]$, o viceversa. Estas diferencias, por tanto, han de ser atribuidas directamente a la estructura léxica de las palabras. Podemos verlo en los ejemplos de (3), correspondientes al verbo *cantar*.

⁴ Vid. Lloret & Viaplana (1998).

⁵ Vid. Clua (1999) y Viaplana (1999).

(3) Diferencias Fonológicas

1ª persona del singular del Presente de Subjuntivo de *cantar*

- a. Variedad 1: [ˈkante] /e/
- b. Variedad 2: [ˈkanti] /i/
- c. Variedad 3: [ˈkanta] /a/

Otro ejemplo de variación que afecta a las terminaciones verbales nos lo ofrecen las formas del gerundio. Muchas variedades, como es el caso del catalán de Barcelona, presentan una –[n] final en esta forma verbal; en cambio la mayoría de las variedades valencianas y baleares presentan –[nt] al final de las formas de gerundio. Si realizamos una cuantificación de las diferencias a partir de los datos fonéticos, estas diferencias son similares a las que veíamos en (3). Desde la óptica generativa, sin embargo, la alternancia [n]~[nt] que se observa en la variedad 1 (4a) puede ser atribuida a una única forma fonológica /nt/, que se realiza como [n] en posición final de palabra y como [nt] cuando le sigue un clítico, ya que en estas variedades opera un proceso fonológico de simplificación del grupo consonántico [nt] en posición final de palabra, que no se produce cuando la forma verbal va seguida de un clítico pronominal; en cambio en las variedades de (4b) no se produce ningún tipo de alternancia, ya que estas variedades no conocen el proceso de simplificación del grupo [nt].

(4) Diferencias Fonéticas

- a. Variedad 1:
 - cantant ‘cantando’ [kanˈtan] /nt/
 - cantant-ho ‘cantándolo’ [kanˈtanto]
- b. Variedad 2
 - cantant [kanˈtant] /nt/
 - cantant-ho [kanˈtanto]

Desde nuestra perspectiva las diferencias observadas en (3) son fonológicas, solo predecibles a partir de la estructura morfológica del verbo catalán que puede marcar el Presente de Subjuntivo con [e], [i] o [a] dependiendo de los dialectos;

mientras que en (4) las diferencias son meramente fonéticas y se pueden predecir por un proceso fonológico. Creemos que se trata de una distinción pertinente que debe tenerse en cuenta en la cuantificación de la distancia lingüística entre variedades, ya que de lo contrario el resultado del análisis cuantitativo puede apartarse considerablemente de la realidad. Así pues, el análisis lingüístico nos permite, por un lado, captar similitudes o diferencias entre variedades que a simple vista fonética serían imperceptibles, y por el otro, nos permite distinguir entre diferencias estructurales (las que aquí denominamos fonológicas) y diferencias predecibles (fonéticas) a partir de los procesos fonológicos sistemáticos que caracterizan las variedades lingüísticas.

En trabajos anteriores hemos argumentado y justificado la pertinencia del análisis lingüístico previo para realizar un tratamiento cuantitativo adecuado de la distancia lingüística. En Clua & Lloret (2006) lo hicimos a partir del análisis de la distancia lingüística en los clíticos pronominales, una de las áreas del catalán que presenta más variación fonética. En este caso, se trataba de lo que podríamos catalogar de pequeño *ensayo de laboratorio* a partir de la variación observada en estos elementos en tres variedades lingüísticas del catalán, entre las cuales medíamos y representábamos la distancia lingüística, primero, a partir de los datos fonéticos y, a continuación, a partir de los datos analizados fonológicamente. Las diferencias obtenidas en ambos tratamientos hacían aconsejable partir de un análisis lingüístico de los datos fonéticos del corpus.

Dado que algunas veces se había esgrimido que en un tratamiento cuantitativo global (con grandes cantidades de datos lingüísticos como las que pueden ofrecer un atlas o un corpus como el COD) este tipo de diferencias (entre los resultados obtenidos a partir de los datos fonéticos o fonológicos) podía ser totalmente inapreciable e intrascendente, en Clua (2010) realizamos un tratamiento cuantitativo contrastado de los datos de nuestro corpus. En estos trabajos, contrapusimos los resultados de la representación de la distancia lingüística de los datos de este corpus previamente analizadas fonológicamente con los obtenidos a partir de los mismos datos, pero sin el análisis fonológico previo. El resultado de la comparación corrobora las hipótesis a las cuales habíamos llegado con los *ensayos de laboratorio* anteriores; en el sentido de que los resultados del análisis dialectométrico son considerablemente diferentes si partimos de datos fonéticos o si

lo hacemos después de analizar fonológicamente estos mismos datos. Véase en este sentido las figuras 1 y 2; en la figura 1 representamos la distancia lingüística de las variedades del catalán a partir de los datos analizados fonológicamente, mientras que la figura 2 representa la distancia lingüística a partir de los datos fonéticos.

A partir del análisis lingüístico de los datos obtenemos la variables que van a ser objeto de comparación para determinar la distancia lingüística entre las variedades del catalán. Estas variables son las formas subyacentes de los diferentes morfemas y los procesos fonológicos que los afectan. Es decir que no comparamos formas nominales o verbales, sino los segmentos morfológicos que las constituyen. Así, en el caso de la morfología verbal los segmentos que se han tenido en cuenta han sido los de (5).⁶ En el caso de la flexión nominal hemos contrastado los segmentos: raíz, género ([±femenino]) y número.

(5) Segmentos morfológicos de las formas verbales

	Raíz	Extensión	Vocal Temática	TAM	Número Persona
cantaran 'cantaran'	kant		a	rá	n
serveixis 'sirvas'	serb	éʃ		i	s

Como ya hemos apuntado, además de la comparación de los segmentos morfológicos también se tuvieron en cuenta los diferentes procesos fonológicos que los afectan y que permiten explicar su realización fonética. Entre muchos otros, algunos de los procesos que hemos utilizado como base de comparación en la flexión verbal son: desacentuación de la raíz (*canto* [kánto] pero *cantava* [kantáβa]), ensordecimiento de obstruyentes en situación final de palabra (*begui* 'beba' [béʎi] pero *bec* 'bebo' [bék]), elisión de *-r* en situación final de palabra (*cantar-la* 'cantarla' [kantárla] pero *cantar* [kantá]), etc. En la flexión nominal hemos tenido en cuenta, por ejemplo, fenómenos de elisión de vocales en contacto y de inserción de epéntesis para reparar estructuras silábicas anómalas.

⁶ *Extensión* es un sufijo post-radical que en catalán determina la subclase verbal en algunas conjugaciones; *TAM* representa las categorías de tiempo, aspecto y modo

Para establecer la comparación entre estos elementos y poder definir la distancia lingüística hemos utilizado el siguiente índice de distancia.⁷

(6)

$$dist(i,j) = \frac{\sum_{k=1}^{long} dif_k(i,j)}{long} \times 100$$

Es decir, la distancia lingüística entre dos variedades (i, j) es igual al sumatorio (Σ) de las diferencias en cuanto a una variable k entre las variedades (i, j), dividido por $long$, que es la longitud (número de sonidos) de cada segmento morfológico comparado.

La elección del índice de similitud / distancia tiene una gran relevancia para el resultado final del proceso, puesto que según el tipo de índice establecido los resultados pueden variar considerablemente; sobre todo cuando las matrices de comparación presentan casillas nulas, debido a respuestas erróneas en la entrevista o a la carencia de respuestas, o cuando los datos implican respuestas múltiples. En estos casos, la medida de similitud que se suele utilizar es el porcentaje de las coincidencias, en relación con el total de elementos comparados entre dos variedades.

Una vez definida la distancia lingüística, es necesario obtener su representación gráfica. Las matrices de similitudes / diferencias, resultantes de la comparación de las características lingüísticas de las diferentes variedades a clasificar, determinan las interdistancias lingüísticas entre estas variedades, pero, cuando se trabaja con cantidades importantes de datos, no permiten una buena interpretación ni una buena visualización de los resultados. Como se puede comprobar en (7), que es un pequeño fragmento de una de las matrices de similitudes con las que hemos trabajado, a partir de un conjunto de datos de este tipo podemos discernir cuáles son las variedades que presentan más divergencias (en este caso serían las variedades de l'Alguer y Felanitx, que presentan un 33,67 % de divergencias en el conjunto de los datos del COD analizados), pero resulta

⁷ La definición de la distancia lingüística se ha realizado con el programa Microsoft® Excel. Concretamente, hemos utilizado unas macros de Excel diseñadas por el profesor Sergi Civit del Departamento de Estadística de la Universitat de Barcelona.

imposible establecer una clasificación global y captar adecuadamente las relaciones de interdistancias entre variedades.

(7) Matriz de distancias

	L'Alguer	Ciudadella	Eivissa	Felanitx	Formentera	Manacor	Maó	Palma
L'Alguer	0,00							
Ciudadella	33,01	0,00						
Eivissa	30,53	6,43	0,00					
Felanitx	33,67	9,62	7,96	0,00				
Formentera	29,59	8,00	3,43	8,75	0,00			
Manacor	32,91	8,29	6,20	4,88	7,11	0,00		
Maó	33,27	4,93	9,90	12,59	11,15	10,43	0,00	
Palma	32,21	6,44	4,77	4,38	5,84	3,17	10,03	0,00

Por eso es necesario buscar una representación de las distancias lingüísticas de la matriz que permita visualizar con claridad las relaciones de proximidad o lejanía entre las variedades lingüísticas analizadas. Se trata, pues, de conseguir una representación en un espacio de dimensión reducida (cómo por ejemplo, el plano) que, con un mínimo de distorsión de la estructura de interdistancias original, permita una interpretación adecuada de los datos.

Para ello nos hemos servido del *Cluster Análisis* y hemos usado un algoritmo de clasificación basado en el método UPGMA (*Unweighted Pair-Group Method Using Arithmetic Averages*),⁸ que ha sido contrastado ampliamente en aplicaciones de la taxonomía numérica en múltiples disciplinas. Para la evaluación de la distorsión entre las representaciones y la distancia original, aplicamos el coeficiente de correlación cofenética, con unos resultados que corroboran la fidelidad de las representaciones jerárquicas en relación con la distancia lingüística de partida.⁹

⁸ Vid. Sneath & Sokal (1973).

⁹ La definición de la distancia lingüística se ha realizado con el programa Microsoft® Excel. El análisis de conglomerados y la representación gráfica de la distancia lingüística se ha llevado a cabo con el sistema de análisis multivariante GINGKO (Departament de Biologia Vegetal, Universitat de Barcelona <http://biodiver.bio.ub.es/vegana/index.html>).

3. CLASIFICACIÓN DE LAS VARIEDADES DEL CATALÁN A PARTIR DE LA DISTANCIA LINGÜÍSTICA

El dendrograma o representación arbórea obtenida en nuestro análisis dialectométrico de los datos del COD es el que presentamos en la Figura 1. En él se puede ver claramente que en el ámbito de la lengua catalana hay cuatro áreas dialectales bastante definidas y una variedad que mantiene una distancia considerable de todo el resto, la variedad de la ciudad sarda de l'Alguer. En cuanto a los cuatro grupos, hay un primer bloque formado por las variedades baleares (mallorquín, menorquín e ibicenco), un segundo bloque que comprende el catalán central y el septentrional, un tercer grupo en el que se integran tanto el catalán norte-occidental como la mayoría de las variedades de la provincia de Castelló, y, finalmente, un cuarto y último conjunto que comprende el resto de variedades valencianas. Otra variedad que aparece aislada, sin agruparse con ninguna otra, es la de Benabarri, que se sitúa entre las variedades norte-occidentales y las variedades valencianas; el aislamiento de esta población pone de manifiesto que sigue vigente su carácter de habla de transición hacia el aragonés.

Las características de estos cuatro grupos principales son las siguientes:

a) Las variedades de las islas Baleares y Pitiusas forman un clúster muy definido y con una clara estructura interna: las variedades de la isla de Mallorca se agrupan entre ellas, y ocurre lo mismo en los casos de Eivissa y Formentera y de Ciutadella y Maó en Menorca. Emergen, por lo tanto, las tres principales variedades balearicas: el mallorquín, el menorquín y el ibicenco.

b) Las variedades del catalán central constituyen un clúster especialmente homogéneo. De hecho, tan sólo el habla de Solsona se sitúa a una cierta distancia del resto, posiblemente debido a su carácter de variedad de transición entre los subdialectos central y norte-occidental. En cuanto a las variedades tradicionalmente adscritas al llamado catalán septentrional de transición se diluyen, en el presente dendrograma, en el clúster homogéneo del catalán central. Parece dudoso, cuando menos, defender la vigencia de este subdialecto a partir de los datos del COD.

Las variedades adscritas al catalán septentrional (Perpinyà, Prada, Sallagosa y Ceret), en territorio del estado francés, se agrupan en primera instancia con el catalán central, pero a una distancia considerable, y entre ellas tampoco se aprecia la homogeneidad que caracterizaba el clúster del catalán central.

c) En el ámbito correspondiente a las llamadas variedades occidentales, en primer lugar, emerge un clúster que engloba tanto las variedades norte-occidentales como aquellas hablas de transición al valenciano. En este conglomerado además de las variedades estrictamente norte-occidentales encontramos el grupo formado por el tortosino estricto (Amposta, Tortosa y Vinaròs) y, un último grupo que incluye las variedades norteñas de Castelló. Las variedades de este último grupo, que a menudo han sido llamadas variedades de transición entre el catalán y el valenciano, se unen claramente al clúster de variedades norte-occidentales, de forma que creemos poder aclarar considerablemente la filiación de estas hablas, que a menudo ha sido objeto de discusión porque, dependiendo de la isoglosa utilizada (1ª persona del presente de indicativo *canto / cante*; 3ª persona del mismo tiempo *cante / canta*; 2ª persona del imperfecto de subjuntivo *cantares / cantesses* ...), se habían vinculado más a las variedades norte-occidentales o a las variedades valencianas.

d) Las hablas de las actuales provincias de València y Alacant (y también la variedad de Borriana) conforman, finalmente, un clúster que presenta, a su vez, una subdivisión en dos grupos: uno mayoritario, que engloba las variedades del valenciano central y meridional, y otro formado por las cuatro localidades, representativas del llamado valenciano *apitxat*: València, Llíria, Sagunt y Alzira.

Esta clasificación difiere considerablemente de la mayoría de divisiones dialectales del catalán que se habían postulado tradicionalmente desde la dialectología clásica, a partir del concepto de isoglosa. Unas divisiones dialectales del dominio catalán que solían establecer dos grandes grupos: las variedades orientales y las occidentales¹⁰. Esta clasificación estaba basada en un número limitado de isoglosas y principalmente, se ceñía a los diferentes procesos de reducción vocálica, por cuya acción los sistemas vocálicos tónico y átono de estas variedades difieren sustancialmente.¹¹ Si tenemos en cuenta que cuando se postuló esta clasificación tradicional, los estudios dialectales se basaban únicamente en los datos fonéticos, no es de extrañar que esta clasificación tradicional en dos grandes grupos de variedades dialectales sea también la que aparece reflejada en el dendrograma de la Figura 2, que corresponde a la distancia lingüística a partir de los datos del COD sin el análisis lingüístico.

¹⁰ Vid. Colomina i Castanyer (1999)

¹¹ Vid. Veny (1982: 17-18).

4. EL PROYECTO DiLET: UN NUEVO ANÁLISIS DE LA DISTANCIA LINGÜÍSTICA

Desde el presente año con el proyecto DiLET (*Distancia lingüística en los ejes espacial y temporal: aspectos fonológicos y morfológicos del catalán*) pretendemos contribuir a ampliar el conocimiento sobre la variación lingüística en general y sobre la distancia entre variedades lingüísticas, en particular, desde una doble perspectiva: la espacial y la temporal. Con esta finalidad nos proponemos constituir un nuevo corpus oral de las variedades geográficas del catalán.

Este nuevo corpus de voz nos habrá de permitir, en primer lugar, definir la distancia lingüística existente en la actualidad entre dichas variedades geográficas, mediante un análisis dialectométrico adecuado; y, posteriormente, a partir del contraste con el Corpus Oral Dialectal del catalán contemporáneo (COD), determinar la distancia lingüística existente entre ambos momentos de la evolución temporal de estas variedades y, por consiguiente, el cambio lingüístico que han experimentado.

Aunque solo han transcurrido dos décadas desde que se inició la constitución del COD, consideramos que durante este periodo la escolarización en catalán y la generalización de determinadas variedades de prestigio en los medios de comunicación audiovisual de los diferentes territorios donde se habla catalán han podido propiciar procesos de nivelación y convergencia dialectal, en definitiva, procesos de cambio lingüístico, dignos de ser estudiados.

En la actualidad, y como primer paso del proyecto, hemos elaborado el cuestionario y hemos empezado a implementarlo en diferentes poblaciones para testarlo. Esta encuesta dialectal nos habrá de permitir constituir un corpus oral representativo de las variedades del catalán y a la vez comparable con el COD. Por esta razón el cuestionario elaborado, aunque más extenso que el del COD, comprende todos los ámbitos lingüísticos que aparecían en aquel para así poder comparar los mismos datos lingüísticos en dos periodos de tiempo diferentes.

La encuesta se realizará en las 82 capitales de comarca (o ciudades equivalentes) de todo el ámbito de la lengua catalana y en cada población se entrevistará como mínimo a tres informantes de 30 a 45 años y de nivel sociocultural medio.

En segundo lugar, nos centraremos en la creación de una base de datos para el corpus que facilite la accesibilidad y la explotación por parte de la comunidad

científica. A continuació, treballarem en la transcripció i en la adequació de los materials orals para el corpus. Y, finalmente, iniciarem la dialectometrització de los datos para determinar la distancia lingüística.

REFERENCIAS BIBLIOGRÁFICAS

- Chambers, J. K. y P. Trudgill (1980): *Dialectology*. Cambridge University Press, Cambridge.
- Clua, E. (1999): “Distància lingüística i classificació de varietats dialectals”, en *Caplletra. Revista Internacional de Filologia*, nro. 26. Publicacions de l’Abadia de Montserrat / Institut de Filologia Valenciana, Barcelona / València, pp.11-26.
- Clua, E. (2010): “Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal”, en Aurrekoetxea, G. y J:L: Ormaetxea (eds.), *Tools for Linguistic Variation*. Universidad del País Vasco, Bilbao, pp.151-166.
- Clua, E. y M-R. Lloret (2006): “New tendencies in geographical dialectology: The Catalan Corpus Oral Dialectal (COD)”, en Montreuil, J.P. (ed.), *New Perspectives on Romance Linguistics. Vol. 2: Phonetics, phonology and dialectology*. Editorial John Benjamins, Amsterdam/Philadelphia, pp. 31-47.
- Clua, E.; M-R. Lloret, y E. Valls (2009): “Análisis lingüístico y dialectométrico del corpus oral dialectal (COD)”, en Cantos Gómez, P. y A. Sánchez Pérez (eds.), *A survey on corpus-based research*. Asociación Española de Lingüística del Corpus, Murcia, pp. 1033-1045.
- Colomina i Castanyer, J. (1999): *Dialectologia catalana. Introducció i guia bibliogràfica*. Servei de Publicacions de la Universitat d’Alacant, Alacant.
- Lloret, M-R. y M.P. Perea (2002): “A Report on the *Corpus Oral Dialectal del Català Actual* (COD)”, en *Dialectologia et Geolinguística*, nro.10. Editorial Mouton de Gruyter, Berlin, pp. 1-18.
- Lloret, M.R. y J. Viaplana (1998): “*Variació morfofonològica. Variants morfològiques*”, en *Caplletra. Revista Internacional de Filologia*, nro. 25. Publicacions de l’Abadia de Montserrat / Institut de Filologia Valenciana, Barcelona / València, pp. 43-62.
- Sneath, P.H.A. y R.R. Sokal (1973): *Numerical Taxonomy. The Principles and Practice of Numerical Classification*. Editorial W. H. Freeman and Company, San Francisco.

- Veny, J. (1982): *Els parlars catalans (Síntesi de dialectologia)*. Editorial Moll, Palma.
- Viaplana, J. (1999): *Entre la dialectologia i la lingüística*. Publicacions de l'Abadia de Montserrat, Barcelona.
- Viaplana, J. y M.P. Perea (eds.) (2003): *Textos orals dialectals del català sincronitzats. Una selecció*. PPU, Barcelona.
- Viaplana, J.; M-R. Lloret; M. P. Perea y E. Clua (2007): *COD. Corpus Oral Dialectal*. PPU, Barcelona. [Publicación en CD-Rom].

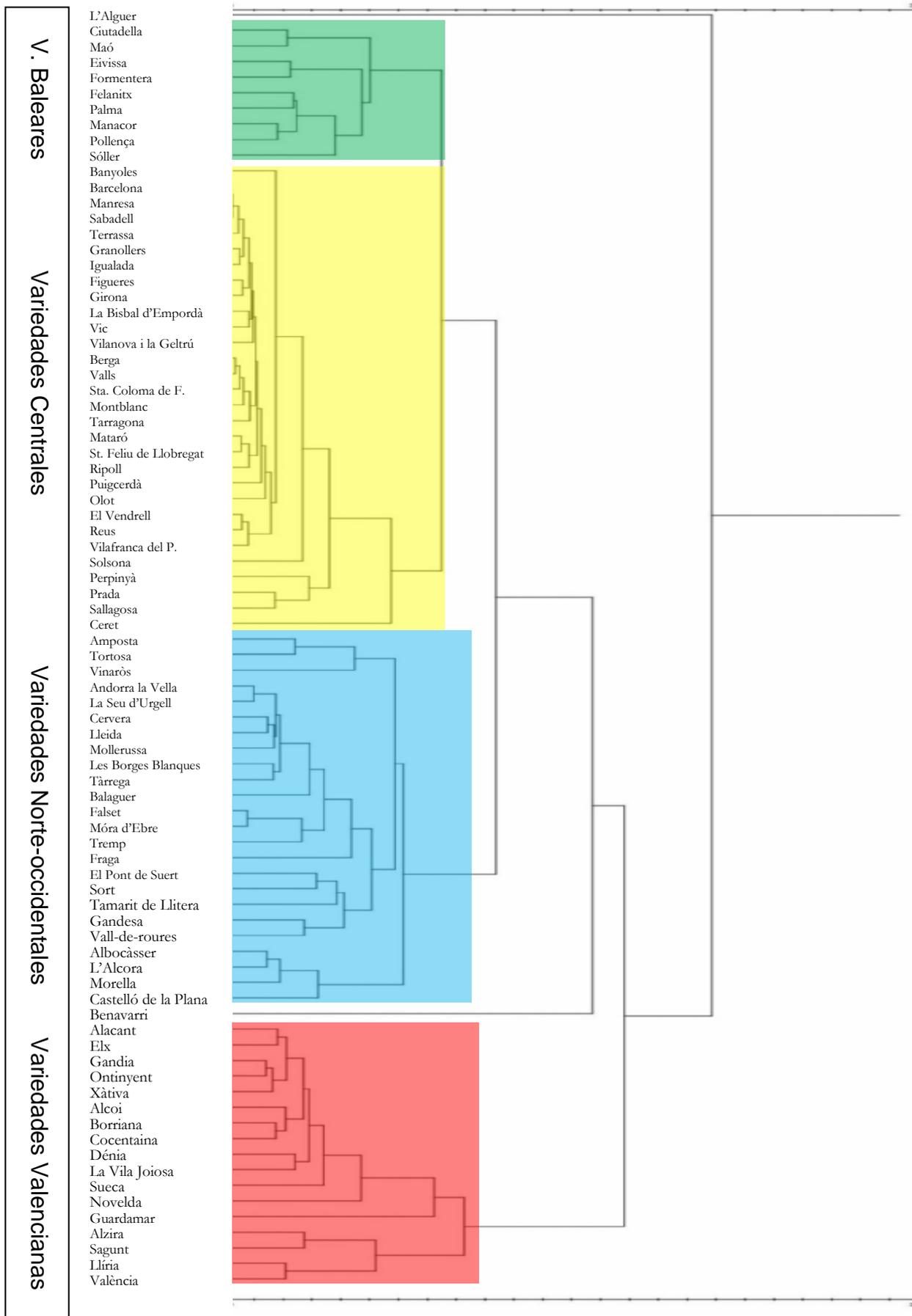


Figura 1. Distancia lingüística a partir de los datos del COD analizados fonológicamente

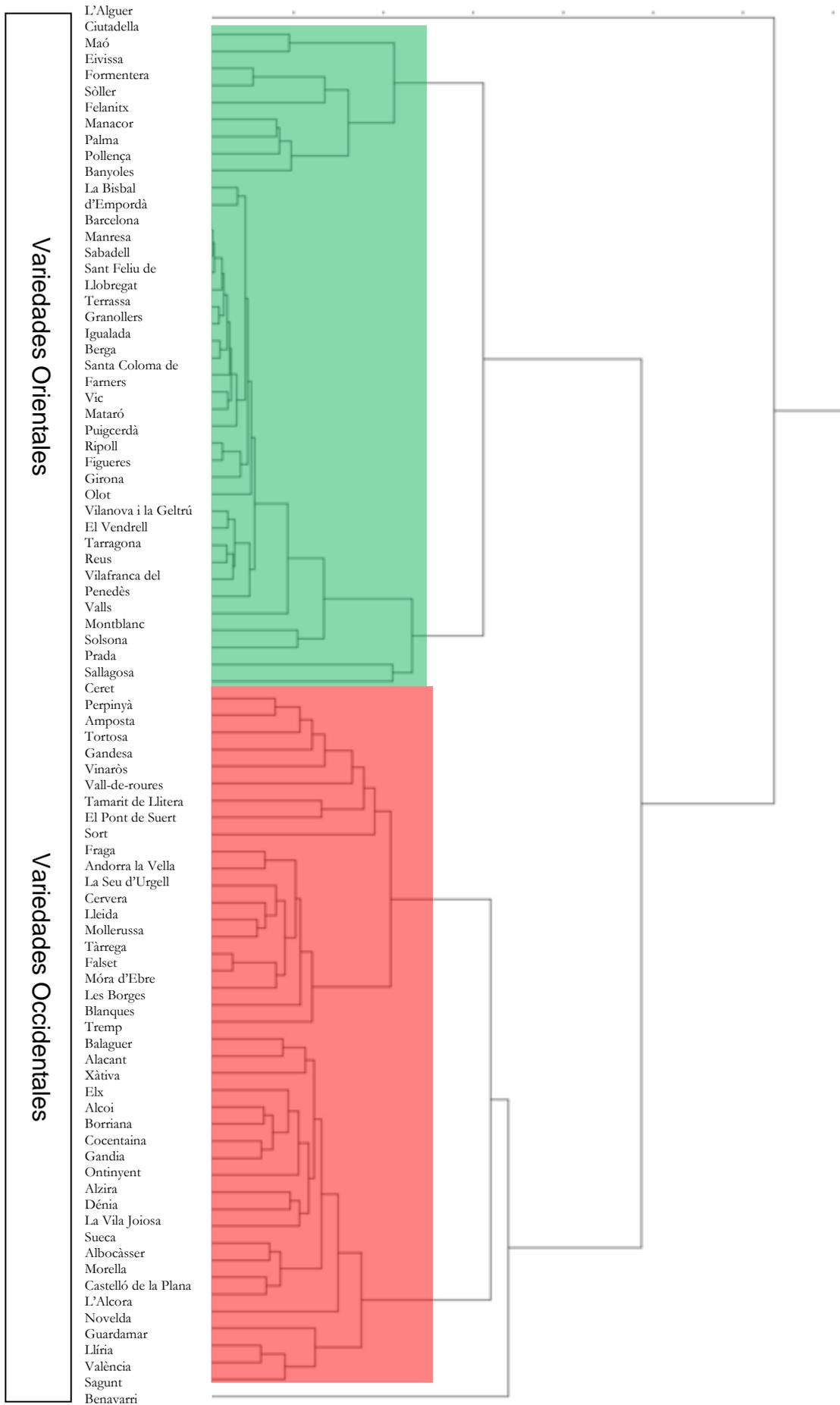


Figura 2. Distancia lingüística a partir de los datos fonéticos del COD