



Sobre el valor de la dialectometría en la delimitación de las distancias lingüísticas*

On the Value of Dialectometry in the Delimitation of Linguistic Distance

GOTZON AURREKOETXEA OLABARRI

Universidad del País Vasco /

Euskal Herriko Unibertsitatea

gotzon.aurrekoetxea@ehu.es

ORCID iD: 0000-0002-9562-4806

Recibíu / Received: 1-VIII-2018

Aceptáu / Accepted: 20-XII-2018

RESUMEN. Este trabajo constituye una reflexión sobre el aporte de la dialectometría a la delimitación de las distancias lingüísticas, centrándose exclusivamente en el análisis cuantitativo de la variación geolingüística. Después de un somero repaso sobre las contribuciones de la dialectología tradicional y la dialectología aglomerativa, se analizan distintos aspectos que la dialectología cuantitativa ha proporcionado a esta disciplina: la isoglosa cuantitativa, la clasificación jerárquica de las hablas, la correlación entre la diferenciación lingüística y la distancia geográfica, etc.

Palabras clave: dialectología, dialectometría, clasificación jerárquica, isoglosa cuantitativa.

ABSTRACT. This paper is a reflection on the contribution of dialectometry to the delimitation of linguistic distance, with an exclusive focus on the quantitative analysis of geolinguistic variation. After a brief review of the contribution of traditional dialectology and agglomerative dialectology, different aspects that quantitative dialectology has contributed to this discipline are analyzed: the quantitative isogloss, the hierarchical classification of the speech, the correlation between linguistic differentiation and geographical distance, etc.

Keywords: dialectology, dialectometry, hierarchical classification, quantitative isogloss.

CÓMO CITAR / FORM OF CITATION: Aurrekoetxea, Gotzon (2019): "Sobre el valor de la dialectometría en la delimitación de las distancias lingüísticas", *Glosema*, 1, pp. 19-39. <https://doi.org/10.17811/glosema.1.2019.19-39>.

I. INTRODUCCIÓN

UNO de los aspectos más interesantes e importantes en la historia de la dialectología es la delimitación de las áreas dialectales y de las fronteras que las separan. Precisamente, desde el mismo nacimiento de esta disciplina se detectan controversias

* Mi agradecimiento tanto al editor como a los revisores por las sugerencias recibidas para mejorar y enriquecer la contribución.

sobre la existencia o no de fronteras dialectales. Desde el “no existen las fronteras dialectales y, por tanto, en sentido estricto, tampoco los dialectos” de Jordan (1973: 29), que resume perfectamente la idea que circuló durante mucho tiempo, hasta el día de hoy —en el que nadie discute la existencia o no de dialectos y de fronteras dialectales, sino la diferenciación que se debe hacer entre ellos—, hay un largo camino recorrido por la dialectología.

La existencia de fronteras dialectales, ya sean de un tipo o de otro, ha sido confirmada utilizando una herramienta que ha resultado fructífera para su determinación: las isoglosas. Esta herramienta ha posibilitado la demarcación de fronteras dialectales, basadas en unas pocas características lingüísticas escogidas al arbitrio del dialectólogo. A decir verdad, hay que reconocer que el uso de las isoglosas ha supuesto un paso importante en la historia de la dialectología, y los avances producidos gracias a este método han impulsado un significativo progreso en el conocimiento de la realidad dialectal de cada idioma.

Sin embargo, el recurso a las isoglosas, inapelable en tiempos en los que escaseaban datos y conocimiento de la realidad dialectal, ha quedado un tanto obsoleto, desde el punto de vista científico, con la llegada de los atlas lingüísticos y la consiguiente posibilidad de manejar grandes cantidades de datos, lo que hasta ese momento era impensable. Han sido muchas las voces que han criticado el método isoglótico: unos le atribuyen falta de objetividad en la elección de los rasgos lingüísticos (Goossens 1977); otros critican que muchas isoglosas no coinciden en el mismo espacio geográfico (Kessler 1995); algunos hacen hincapié en la dificultad de trazar una frontera en zonas de transición, o en la debilidad científica que supone la elección de solo unos pocos rasgos (Inoue 1996), etc. (ver también Nerbonne *et al.* 1996: 186).

Ante este estado de cosas, y después de algunos intentos de análisis cuantitativos en la delimitación de áreas dialectales, J. Séguy (1973b) acuñó, como es bien sabido, la denominación de *dialectometría* (DM en adelante) para referirse al estudio cuantitativo de las diferencias lingüísticas entre localidades; por su parte, Hans Goebel (1992, 2005) llevó la nueva metodología a un estadio inimaginable unos años antes, con la incorporación de la clasificación jerárquica de los dialectos y el estudio de la correlación entre la diversidad lingüística y la distancia geográfica. Posteriormente la dialectometría se ha diversificado con diversas técnicas y programas informáticos en local (VDM) y online (Diatech, Gabmap), incluyendo más posibilidades de análisis y utilizando diversas unidades de distancia, con diferentes algoritmos de clasificación, etc. Se ha pasado de una dialectología basada en unos pocos rasgos a una dialectología que maneja gran cantidad de datos.

2. LA APORTACIÓN DE LA DIALECTOLOGÍA TRADICIONAL

Antes de nada, hay que subrayar que la aportación de la dialectología tradicional a la delimitación de las fronteras dialectales ha sido ingente. Ha logrado configurar áreas dialectales en todos los idiomas en los que se ha desarrollado, definiendo conceptos como ‘dialecto’ y ‘subdialecto’ referidos a determinadas áreas que presentaban singularidades idiomáticas, basándose en la mayoría de los casos en muy pocos datos lingüísticos y siguiendo los métodos atomistas de los primeros dialectólogos.

El dialectólogo tradicional, con apoyo en los principios de la gramática histórica, considera el lenguaje como un conjunto de rasgos que tienen una historia y un desarrollo propios. No concibe el lenguaje como un sistema, tal como se entiende desde la lingüística estructural. Por tanto, no delimita áreas correspondientes a sistemas lingüísticos diferentes, sino áreas dialectales que contienen ciertos rasgos lingüísticos que se estiman propios de ellas.

En la delimitación de fronteras, el procedimiento consiste en trazar isoglosas de los rasgos elegidos por los dialectólogos. La isoglosa se ha convertido en una herramienta fundamental para el establecimiento de las fronteras. Cuantas más isoglosas pasen por un determinado espacio geográfico, más importante será la frontera que crean. De acuerdo con ello, puede haber fronteras bruscas (abruptas) o suaves. Y en aquellos espacios en los que no se pueden fijar fronteras se habla de zonas de transición. El solapamiento y la acumulación de muchas isoglosas indica la presencia de una frontera dialectal, y su dispersión señala zonas de transición o ausencia de una frontera clara entre dos áreas dialectales.

El dialectólogo tradicional resalta el valor cualitativo de los rasgos lingüísticos. Para él unos rasgos son más importantes que otros atendiendo a criterios propios de la gramática histórica y dependiendo tanto de su papel en la evolución de la lengua como del espacio geográfico en el que actúan. Cuanto más antigua sea su documentación y cuantos más fenómenos lingüísticos haya originado (desarrollando posteriormente más formas o variaciones lingüísticas), tanto más importante será dicho rasgo. Por otra parte, cuanto mayor espacio geográfico abarque, mayor será su importancia; un fenómeno reducido a un espacio pequeño no tiene suficiente capacidad para producir efectos a gran escala, y por lo tanto su importancia será menor.

En este nivel no existe la diferenciación entre “lengua” y “habla” que se dio con el estructuralismo, ni la que propone el generativismo entre “competencia” y “actuación”, a pesar de que muchos dialectólogos hayan desarrollado su actividad tras el establecimiento de dichas dicotomías. La dialectología tradicional se ha mantenido mayoritariamente fiel a los postulados de la gramática histórica.

Revela también otras características, que a los efectos de esta aportación son tangenciales y no revisten especial relevancia, tales como la búsqueda del informante “puro” (los NORMs de los que hablan Chambers y Trudgill [1980: 35]), la prioridad de los rasgos más conservadores frente a los más innovadores (de ahí la importancia otorgada a las zonas aisladas), la predominancia de las zonas rurales sobre las más pobladas, etc.

Si bien en algunas latitudes no se ha utilizado ningún sistema cuantitativo para la delimitación de las fronteras lingüísticas, en otras ha habido propuestas interesantes: en la dialectología alemana, por ejemplo, fue Karl Haag quien propuso un sistema cuantitativo sugiriendo una poligonación de los mapas y distinguiendo diversos niveles de frontera de acuerdo a la cantidad y frecuencia de las isoglosas que convergían entre dos localidades encuestadas (Schrambke 2010). En la segunda mitad del siglo XX, Reed y Spicer (1952) proponen medir la variación en zonas de transición con el sistema que denominaron “the statistical method of correlation”, y a su vez definieron la correlación como “the study of simultaneous variation of two or more variates” (*ib.*: 350). Atwood (1955) analizó la variación fonológica en Bélgica sobre la base del *Atlas Linguistique de la Wallonie (ALW)*, trazando en él más de quinientas isoglosas fonológicas distribuidas en seis grupos y cartografiando las fronteras con líneas de diferente grosor. Shaw (1974) da cuenta de algunas aplicaciones estadísticas en el mundo anglosajón, pero quizás sea Charles L. Houck (1967, 1969 —su tesis doctoral—, 1985, 1986), junto con Davis y Horvath (1999), uno de los que más ha trabajado este aspecto. Houck ha utilizado diversas técnicas estadísticas para el análisis de la variación lingüística, tales como el coeficiente phi o coeficiente de correlación de Mathews, la correlación tetracórica, el análisis factorial, el análisis de clúster, el *Multidimensional Scaling*, etc.

3. LA DIALECTOLOGÍA AGLOMERATIVA

Aunque en español no se ha utilizado habitualmente esta denominación, en inglés es sumamente usual a partir de la primera década de este siglo su equivalente *Aggregate Dialectology*, que se contrapone a (*Single*) *Feature-Based Variation Studies* (Nerbonne 2009). La acuñación *dialectología aglomerativa* indica claramente el propósito de superar la concepción científica anterior para presentar una dialectología cimentada en el valor del uso de grandes masas de datos, en vez de determinadas características escogidas subjetivamente por el dialectólogo.

Este cambio radical viene inducido por la búsqueda de la objetividad en el estudio de la variación geolingüística. Ello implica dos decisiones sumamente importantes en el devenir de esta ciencia: el uso de todos los datos disponibles, en vez de una

selección de los mismos, y el empleo de técnicas cuantitativas a la hora de medir las diferencias lingüísticas entre localidades y de acometer la tarea clasificadora. Este planteamiento rompe claramente con el modo tradicional de investigación para entrar en una práctica investigadora nueva y desconocida, gracias al empleo de métodos estadísticos. La dialectología aglomerativa se fundamenta, pues, en dos pilares: el uso de grandes masas de datos y la cuantificación automática para la medición de las diferencias lingüísticas entre localidades y la clasificación jerárquica de las hablas.

El primer pilar implica el uso, en el sentido más exhaustivo, de los atlas lingüísticos. A pesar de que los primeros proyectos de atlas lingüísticos no suponían la elaboración de bases de datos, tan pronto como la informática se puso al alcance de los dialectólogos, los proyectos atlantográficos comenzaron a usarla para crear dichas bases y proyectarlas cartográficamente. No hay que olvidar que los primeros dialectómetros tuvieron que generar sus bases de datos partiendo de información extraída de atlas lingüísticos ya publicados (por ejemplo, J. Séguy con su *ALG [Atlas linguistique de la Gascogne]*, y H. Goebel con su proyecto de dialectometrización del *AIS [Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale]* o del *ALF [Atlas linguistique de la France]*). Todo el tiempo y todo el esfuerzo humano y económico invertidos en estas tareas han merecido la pena, sin ninguna duda. Sin atlas lingüísticos o proyectos semejantes de recogida y ordenación de los datos, es impensable hablar de DM: los atlas lingüísticos son un requisito básico. En realidad, la DM ha dado vida a los atlas lingüísticos y ha proporcionado una metodología para su explotación integral. De esta manera, quedan atrás los lamentos de J. Séguy (1973b: 3) sobre su nula utilización, cuando afirmaba que “les riches collections que constituent les atlas linguistiques restent sous-exploitées”; o los de M. Alvar (1984: 112-113) en el mismo sentido:

Los atlas abruma por la inmensa cantidad de datos que encierran y si no facilitamos su consulta, quedarán como gigantescos archivos no utilizados. Desgraciadamente, tal es la situación en el mundo hispánico.

Por tanto, la DM ha dado sentido a los atlas lingüísticos. La dialectología tradicional no disponía de las herramientas adecuadas para el análisis integral de la inmensa cantidad de datos que albergan. La DM proporciona esa herramienta y posibilita la proyección cartográfica sintética de los mismos.

El segundo pilar implica el uso de la cuantificación informática. Fue el mismo creador del término *dialectometría* —J. Séguy— el que, a pesar de haber llevado a cabo todo su trabajo dialectométrico manualmente, creía indispensable el uso de los ordenadores. Según sus palabras, él tuvo que utilizar “el ordenador de los pobres”,

es decir, su cerebro, pero reivindicaba la informática para las tareas de la cuantificación. De hecho, estas tareas implican dos pasos. El primero es la búsqueda de una unidad de medida de las diferencias lingüísticas entre localidades. Séguy (1973a: 21) propuso el uso de la *distancia de Hamming* y, al hacerlo, dio un paso de gigante, pues por primera vez en la historia de la dialectología se postulaba una unidad para medir dicha magnitud. Este hecho tiene una relevancia extraordinaria en el desarrollo de la cuantificación de la variación lingüística, no tanto por haber o no acertado con la unidad de medida determinada, sino por el hecho mismo de haber planteado una unidad de medida. El segundo fue el hecho de aplicar esta unidad de distancia característica a característica, localidad por localidad.

Séguy superó muchas trabas y concepciones erróneas en su búsqueda de un análisis sintético de los datos suministrados por los atlas: una de ellas fue la idea de que cualquier forma que no fuera la transcripción lo más exacta y fiel posible de las respuestas recogidas en el mapa se consideraba una “interpretación”:

La définition de ces deux secteurs va immédiatement évoquer, dans l'esprit du lecteur, le verbe interpréter et ses dérivés. Nous refusons cette étiquette. En matière de cartographie linguistique, on en fait un étrange abus: est qualifiée d'interprétative toute carte qui n'est pas un simple report de données point par point, à la manière de Gilliéron. [...] Compter, traduire en chiffres et en graphismes n'est pas interpréter. Le volume VI de l'ALG n'enfreint donc pas les normes du genre: nous ne faisons qu'élaborer une matière informe et insaisissable et que mener les données quantitatives jusqu'à leur formalisation terminale et mathématiquement nécessaire. (Séguy 1973b: 2)

Esta idea, que ha predominado en la romanística hasta nuestros días, ha impedido en cierta manera, o al menos ha retrasado, la aparición de otras formas de representar los datos en los atlas lingüísticos. Una vez resuelto el problema de la medición de las diferencias dialectales y una vez cuantificadas las diferencias entre localidades, la dialectometría ha ido generando diversos tipos de análisis: desde la dialectometría interpuntual e isoglótica, a la dendrográfica (con diversos algoritmos) y la multidimensional, pasando por la correlativa. Las técnicas han ido variando y enriqueciéndose paulatinamente hasta nuestros días.

Si bien se considera a Jean Séguy el creador de la DM, ha sido Hans Goebel el que ha desarrollado esta disciplina, partiendo de una dialectometría interpuntual y llegando a la clasificación jerárquica (*cluster analysis*) y a la DM correlacional. Paralelamente, se han ido desarrollando otras técnicas estadísticas, entre las que destacan, por ser las más usadas, la clasificación *fuzzy* y la MDS (*Multidimension Scaling*), que visibilizan mejor las zonas de transición.

Una de las mayores virtudes del desarrollo de la DM ha sido la creación de herramientas específicas para el análisis sintético. Así, se pueden citar la herramienta *VDM*, de la escuela de Salzburgo; la *Gabmap*, de la escuela de Groninga; y *Diatech*,

de la Escuela de la Universidad del País Vasco. La primera trabaja en local y las dos últimas en web.

4. LA APORTACIÓN DE LA DM A LA DELIMITACIÓN DE FRONTERAS DIALECTALES

Antes de nada, es necesario recordar que la DM es, simplemente, una metodología que transforma datos lingüísticos nominales en datos numéricos, a fin de medir cuantitativamente las diferencias lingüísticas entre localidades y realizar clasificaciones entre las hablas. Se trata de una metodología que puede ser aplicada correctamente dependiendo de las habilidades del dialectólogo. Como en investigaciones de otra índole, no es achacable a la dialectometría el uso incorrecto de un método, que será, en todo caso, responsabilidad del dialectólogo. No se le puede pedir a un método cuantitativo, como alguna vez se ha hecho, que no tenga en cuenta el análisis cualitativo. Será potestad del investigador, de acuerdo con los fines de su investigación, el uso de uno y otro método o la combinación de ambos.

Aclarado esto, el método dialectométrico, desde el punto de vista científico, supone un avance en toda regla cuando se requiere conocer cuantitativamente la diferenciación lingüística entre dos localidades o establecer una clasificación de las variedades dialectales. Nunca hasta su aparición se ha avanzado tanto en la cuantificación de la variación geolingüística.

Ciertamente, existe alguna desconfianza hacia la DM entre los dialectólogos tradicionales; no se puede ocultar esta realidad. Se trata de una comprensible desconfianza a ciertas técnicas y métodos desconocidos y nunca aplicados hasta hace poco. Algunos dialectólogos formados en la escuela filológica tradicional, cuyo interés lingüístico se centra en la diacronía y despliegan una metodología orientada a la resolución de problemas en el ámbito de la evolución de la lengua, recelan de metodologías que recurren a la cuantificación. No ven la necesidad de grandes masas de datos, ni de una cuantificación de los mismos, puesto que en su praxis científica no tienen necesidad de ello. Forma parte de su acervo científico el tomar en cuenta rasgos técnicos tenidos por más importantes en el desarrollo evolutivo de la lengua. La preeminencia de ciertos rasgos sobre otros viene dada por diferentes factores: unas veces se basa en su presencia en el uso diario del idioma (en la *performance*), mientras que en otras puede tratarse de una característica o ley lingüística (*competence*) que se presenta en un área determinada.

Pero cuando se trata de “medir” la diferencia lingüística entre localidades o entre áreas dialectales es ineludible el uso de la cuantificación. Y para ello hay que superar el estadio de *Single Feature-Based Dialectology* y llegar a la dialectología aglomerativa,

de suerte que resulte posible plantearse otros objetivos, como apuntan Nerbonne *et al.* (2010: 41):

More exact techniques, and especially computational techniques, serve to broaden the empirical base that dialectology can effectively build on, improve the replicability of data analysis techniques, and enable more abstract questions to be addressed with empirical rigor.

La primera gran aportación en cuanto a los resultados es la noción de *isoglosa cuantitativa*. Mediante una acumulación de todas las isoglosas posibles, se llega a conformar una isoglosa que no marca el límite de un único rasgo o fenómeno lingüístico, sino un número determinado de rasgos que pasan entre dos localidades analizadas. Séguy (1973a), en un suplemento al volumen VI del *Atlas lingüistique de la Gascogne (ALG)*, publicó los primeros resultados de su investigación dialectométrica. En él traza líneas rectas entre localidades vecinas, y sobre ellas añade el valor numérico de los rasgos lingüísticos que las separan. Este trabajo constituye el primer gran paso hacia una dialectología cuantitativa, al plasmar en un mapa la diferenciación lingüística en un lenguaje numérico. Pero fue un trabajo posterior (Goebel 1981) el que visualizó por primera vez en un mapa el concepto de isoglosa cuantitativa aplicando diferentes grosores a los polígonos o diagramas de Voronoi.

Aparte de la representación en mapas de las diferencias lingüísticas entre localidades mediante distintas técnicas de visualización, el investigador tiene la posibilidad de consultar tablas de distancias lingüísticas, como si se tratara de las tablas de distancias kilométricas entre localidades. Nunca hasta este momento había podido el dialectólogo precisar tanto la variación geolingüística o comparar las distancias lingüísticas entre localidades. Las diferencias se convierten en distancias lingüísticas concretas, tangibles y demostrables, pues la metodología utilizada posibilita el retorno a los datos e incluso la confección de un listado de los diversos rasgos diferenciales entre dos localidades.

Si el investigador necesita saber dónde se encuentran los límites entre áreas dialectales, la DM posibilita la clasificación jerárquica. Se trata de la segunda gran aportación de la DM a la dialectología: el análisis de clúster o clasificación jerárquica aglutina todas las hablas en grupos jerárquicamente organizados, comenzando por hablas más afines entre sí (con mayor número de características comunes), y finalizando el análisis con la agrupación de todas las hablas en un único grupo.

La DM resuelve también la relación entre la distancia geográfica y la diferencia lingüística. Para ello se usa la *dialectometría correlativa* (Goebel 2005), mediante la cual se pone en relación la distancia geográfica o euclídea (o cualquier otro tipo de medida, como el tiempo necesario para ir de una localidad a otra, etc.) y la diferencia lingüística medida con la unidad elegida por el investigador. Así, la DM resuelve la

observación por todos aceptada y nunca confirmada de que la diferencia lingüística crece a medida que lo hace la distancia geográfica.

5. DE VUELTA A LA DIALECTOLOGÍA

A día de hoy es insoslayable la utilización de grandes masas de datos y de la cuantificación en la labor de medir las diferencias o distancias lingüísticas entre variedades o hablas. Por otra parte, si el objetivo del investigador es contar las diferencias lingüísticas entre variedades o hablas locales, inexorablemente tiene que elegir una unidad de medida. En tercer lugar, el dialectólogo tiene que decidir el nivel lingüístico en el que se va a llevar a cabo la medición. Una vez resueltas esas cuestiones previas, comienza la tarea propiamente dicha: *a)* preparar la base de datos en la que se fundamentará su análisis, *b)* determinar la técnica estadística adecuada, y *c)* decidir el procedimiento para visualizar los resultados.

Un requisito básico de la actividad científica es la objetividad de los datos, la cual depende de la cantidad y la calidad de los mismos. Es necesario un volumen suficientemente grande de ellos para que los análisis estadísticos sean estables y consistentes. A día de hoy es inasumible el uso de ciertos rasgos elegidos al arbitrio del investigador para llevar a cabo cualquier análisis. La subjetividad en la selección de datos es una característica superada hace tiempo en la investigación. Por lo que se refiere a la calidad de los datos, resulta capital la metodología aplicada en su recogida. Es la piedra angular de toda investigación: una buena recogida de datos es la condición básica de una buena investigación, y sin ella el trabajo posterior no sirve de nada o de muy poco.

El segundo requisito es el marco teórico de la investigación. No hay investigación alguna (ni lingüística —tampoco dialectológica— ni de ninguna ciencia) que no disponga de un marco teórico. El dialectólogo debe elegir uno, ya sea la lingüística histórica, el estructuralismo, las distintas ramas de la gramática generativa (minimalismo, teoría de la optimidad...), etc. Inmediatamente después ha de decidir el objeto a medir: si la medición se va a llevar a cabo en el nivel de los datos (nivel fonético) o en el nivel de los rasgos lingüísticos (nivel fonológico). En la dialectología hay ejemplos de los dos tipos: algunos miden las distancias lingüísticas en el nivel de los datos, mientras que otros proponen distinguir entre los datos y los rasgos lingüísticos. Los primeros provienen directamente de las respuestas recogidas en los trabajos de campo; los segundos se logran mediante un análisis lingüístico de los mismos. En otras palabras, la medición se puede hacer en el nivel fonético o en el nivel fonológico. A este respecto, se ha de determinar cómo se articula el sistema lingüístico de cada localidad o área geográfica. La diferencia entre estos dos niveles es grande, como

se ha comprobado en los trabajos hasta ahora publicados (Clua 2010, Aurrekoetxea 2016).

El tercer requisito para acometer una actividad mensurativa es la elección de la unidad de medida; es decir, ¿cómo se van a contar las diferencias lingüísticas? En la DM se utilizan varias unidades de medida: Séguy (1973a) empleó la *distancia de Hamming*; su discípulo D. Philips (1983), la *distancia CHI2*; Goebel (1987, 1992: 436-438) utiliza dos unidades: el *índice relativo de identidad-IRI* (RIV en textos en inglés) y el *índice ponderado de identidad-IPI* (WIV en textos en inglés). Por su parte, los hermanos Hoppenbrouwers (1988, 1994) propusieron el *Frecuency method*, que mide las frecuencias de características fonéticas. El grupo de Groninga usa el *algoritmo de Levenshtein* (Kruskal 1983, Nerbonne y Heeringa 2010). La *distancia Manhattan* también ha sido utilizada alguna vez (Jain y Dubes 1988); el *método COD* ha sido empleado en catalán (Valls y Clua 2010), etc. No todas las unidades de distancia sirven para todo, ya que cada una de ellas actúa de manera diferente y atiende a necesidades distintas: unas son unidades categóricas (el RIV o IRI, por ejemplo), otras numéricas (la distancia de Hamming o el algoritmo de Levenshtein); unas son más apropiadas para medir datos fonéticos, otras para datos fonológicos. Es responsabilidad del investigador decidir la unidad que debe utilizar en su investigación de acuerdo a los fines establecidos (ver Heeringa y Prokić 2018 para un resumen de las medidas usadas en DM).

La cuarta tarea que debe abordar es la construcción de la base de datos que será el fundamento de su análisis cuantitativo, eligiendo la herramienta informática más adecuada. Cada programa informático exige un modelo de base de datos. El dialectólogo modelará dicha base con la información que sea necesaria para el correcto análisis cuantitativo. Creada la base de datos, el investigador debe implementarla en el programa dialectométrico elegido, determinar la técnica estadística adecuada y decidir el procedimiento de visualización de los resultados.

6. REFLEXIONES SOBRE LA APORTACIÓN DE LA DM A LA DIALECTOLOGÍA

Cuando han pasado ya más de cuarenta y cinco años desde la publicación del primer artículo sobre la DM (período suficiente para su desarrollo), estimo necesario reflexionar sobre su aportación a la dialectología y sobre su futuro.

Indudablemente, la DM ha proporcionado a la dialectología los mecanismos y herramientas necesarios para precisar las diferencias lingüísticas entre variedades dialectales. Fue el primer paso de la DM, en lo que se conoce como *dialectometría no dendrográfica*: el hecho de posibilitar la cuantificación de las diferencias lingüísticas. Es, a mi modo de ver, el mayor avance dado por la dialectología desde sus inicios.

El acontecimiento fundamental es el “invento” de una unidad de medida lingüística, sin la cual no habría habido posibilidad de progresar en la delimitación de las diferencias entre variedades. Una vez creada una unidad, los sucesivos pasos se fueron dando de forma previsible: distancias cuantitativas entre localidades, isoglosas cuantitativas, etc. La cuantificación de las diferencias lingüísticas es la mayor aportación de la DM a la dialectología. Es una revolución en toda regla, que deja atrás un siglo de carencias en el estudio preciso de las diferencias dialectales.

Una vez que se logró expresar la distancia lingüística en términos cuantitativos, se abrió un abanico inmenso de posibilidades. Y así, se pudo confirmar, con datos concretos, la suposición de que la diferencia lingüística iba pareja con la distancia geográfica; es decir, que cuanto mayor era la distancia geográfica entre dos localidades, mayor era la diferencia lingüística. Séguy (1971: 348-349), después de un análisis con datos de diversos atlas lingüísticos, logró crear una curva mediante una ecuación matemática en la que quedaba demostrado que la diferencia lingüística crecía con la distancia geográfica, aunque con muchos matices. Más tarde, tanto H. Goebel (2005) como otros (Nerbonne *et al.* 2010: 52) han analizado dicha curva con más datos y métodos disponibles (para un análisis más extenso sobre el tema, ver Aurrekoetxea 2016).

Entre otras posibilidades importantes que abría la cuantificación, estaba la capacidad de crear una clasificación jerárquica de las variedades dialectales, algo que era inconcebible hasta ese momento. Pero, una vez que se contó con la unidad de medida de la diferencia, lo cual permitía medir cuantitativamente las diferencias lingüísticas entre localidades, solo quedaba un paso: disponer de la herramienta estadística. Fue H. Goebel (1991) el primer investigador que consiguió una clasificación de este tipo. Se trata del avance con mayor proyección que se ha dado en la dialectología del siglo XX, pues abre una vía para poder graduar las fronteras dialectales y discernir científicamente las diferentes áreas dialectales en dialectos, subdialectos, etc.

Aparte de los avances descritos, la DM ha aportado la objetividad de la que carecía la dialectología tradicional en su tarea de definición de las fronteras. Se ha pasado de usar unos cuantos rasgos elegidos por el dialectólogo al manejo de ingentes cantidades de datos, recogidos en la mayoría de los casos en los atlas lingüísticos. Dado que la actividad científica se caracteriza por la utilización de grandes masas de datos, esta última dificultad o hace imposible la distorsión que puede provocar la utilización de datos subjetivos o aleatorios.

Otra de las ventajas que ha aportado la DM es el rigor científico que proporciona la estadística descriptiva o deductiva, rigor que debe caracterizar a cualquier actividad científica. Si defendemos la dialectología como la ciencia de la variación lingüística, es indispensable actuar con rigor científico. Con ello no queremos decir que toda la

dialectología o todo análisis de fronteras deban ser cuantitativos; lo que afirmamos es que en nuestro objetivo de delimitar fronteras o de clasificar las variedades y dialectos, necesariamente hemos de usar métodos cuantitativos. Cuando se trabaja con fronteras se opera con grados de diferenciación entre áreas dialectales, y en ese terreno no hay otro camino que la cuantificación. El análisis de clúster o clasificación jerárquica aporta el rigor mínimo necesario, independientemente de los algoritmos o del tipo de clúster que se utilice. El tipo de algoritmo adecuado en cada caso otorgará un plus de adecuación a los fines de la investigación, sin duda, pero lo fundamental aquí es poder proporcionar una clasificación jerárquica y un análisis de clúster.

Además del rigor, la DM contribuye con precisión y fiabilidad a la delimitación de las fronteras, empezando por su mera existencia. Tanto el análisis de clúster como el MDS (*Multidimensional Scaling*) reflejan no solo el tipo de frontera, sino la existencia misma de frontera. Las fronteras pueden ser de diversas clases, como bien distingue la dialectología tradicional: abruptas, nítidas, difusas, graduales, escalonadas, etc. El análisis de clúster reflejará cada tipo de frontera, dependiendo de la longitud de los brazos o líneas horizontales. Si bien existen técnicas automatizadas para la validación del clúster, aquí se utilizará la técnica manual con el fin de no agobiar al lector con tecnicismos innecesarios para la correcta comprensión del problema: cuanto mayor sea la longitud de las líneas horizontales del dendrograma, mayor será la diferencia lingüística entre las hablas o áreas dialectales representadas.

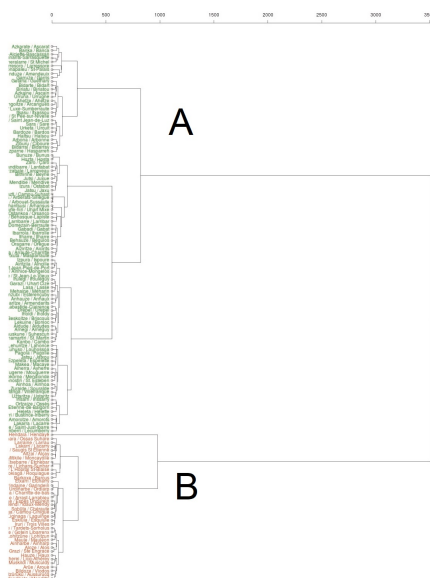


FIGURA 1. Dendrograma producido por el análisis de clúster determinista

En el dendrograma de la fig. 1 hay dos líneas horizontales paralelas que se unen en la parte derecha del gráfico. Esas dos líneas horizontales representan dos áreas dialectales con diferencias lingüísticas sustanciales, ya que las líneas son largas. Si hubiera que delimitar áreas dialectales teniendo en cuenta todas las hablas representadas en el dendrograma (cada línea horizontal que nace en la parte izquierda del dendrograma representa una localidad), de acuerdo con la diferenciación lingüística entre ellas, lo más idóneo sería la delimitación en dos áreas: el grupo A abarcaría 100 hablas, y el grupo B, 38. El grupo A es más compacto que el B, porque sus localidades se agrupan en un nivel inferior a las del B. Los diversos tipos de fronteras de la dialectología tradicional se transforman con el análisis de clúster determinista en verdaderas o falsas fronteras (más abajo se definirá mejor este concepto). Aquí lo que se determina es que, para que haya frontera, tiene que haber una clara diferenciación lingüística entre áreas dialectales. Este tipo de frontera generalmente se dibuja como abrupta en clústeres deterministas.

Si en vez de un clúster determinista se lleva a cabo un análisis MDS (fig. 2), las localidades se presentan en un espacio continuo, sin ningún tipo de marcación de frontera. En este tipo de configuraciones la distancia o la diferenciación lingüística entre localidades se representa con la distancia espacial entre las localidades, no mediante ningún tipo de marcación.

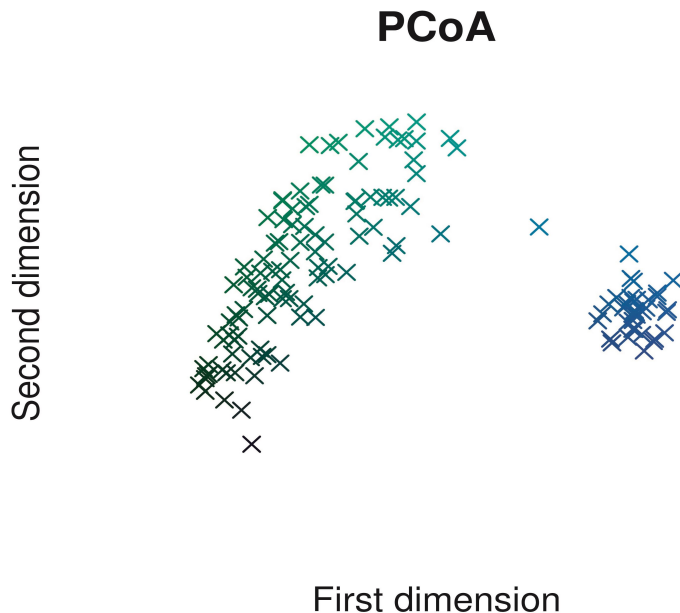


FIGURA 2. MDS de los datos de Bourciez

En efecto, las dos figuras se fundamentan en la misma base de datos —la de Bourciez (*cf.* Aurrekoetxea y Videgain 2009)—, pero difieren en la técnica de presentación de los datos. En la primera figura se determina perfectamente la configuración de los grupos, pero no se visualiza espacialmente la distancia o la diferencia lingüística. En la MDS, sin embargo, dicha visualización es clara. Como se puede apreciar en la fig. 2, las localidades se agrupan en dos áreas: una a la derecha de la figura y la segunda en el centro-izquierda. Entre dichos grupos se encuentra una localidad equidistante, lo cual indica que dicha localidad se halla lingüísticamente entre las dos áreas dialectales. Las localidades de este tipo representan un dilema para aquellos investigadores que en su praxis necesitan encuadrar todas las localidades en algún grupo dialectal.

El lector avezado puede dudar de las distintas representaciones de los datos e incluso puede desdeñar sus resultados o el papel de la estadística en la dialectología. Es cierto que estas dos representaciones, y muchas otras parecidas o análogas que se podrían proponer, difieren por presentar distintos resultados con los mismos datos. Ahondando en el tema, se podría afirmar que no resuelven el problema con toda la exactitud, puesto que diversos algoritmos y diversas técnicas ofrecen resultados parecidos (en la mayoría de los casos, no coincidentes) y en escasas ocasiones resultados iguales. Siendo todo esto cierto, conviene volver la vista al pasado y contemplar el recorrido hecho por la dialectología: de aquellos mapas dialectales, basados en unos pocos rasgos considerados, certeramente o no, como inamovibles y fiel reflejo de la realidad lingüística, hasta estos mapas, basados en grandes bases de datos y creados por procedimientos estadísticos, hay un largo camino.

El uso de la estadística automatizada proporciona también fiabilidad al objetivo de delimitar fronteras. A pesar de que la comunidad lingüística ha venido asumiendo acríticamente las distintas distribuciones y clasificaciones geolectales de la dialectología tradicional, estas últimas carecen, desde la atalaya actual, de suficiente fiabilidad. Entendemos que, si durante largo tiempo no han recibido apenas críticas, esto se ha debido más a la ausencia de métodos y técnicas adecuados que a la confianza generada por tales clasificaciones. Sin embargo, la fiabilidad es un requisito imperativo en toda empresa científica que se precie, y la DM la aporta. ¿Qué puede ofrecer más fiabilidad que una clasificación jerárquica con técnicas que son utilizadas en todas las ciencias, apoyadas en bases de datos extraordinariamente más grandes que las empleadas en la delimitación de fronteras en la dialectología tradicional? A nuestro modo de ver, la respuesta es clara: se puede afirmar que cualquier mapa o resultado obtenido apoyándose en grandes bases de datos y técnicas cuantitativas es más fiable y se acerca más al conocimiento de la realidad lingüística que cualquier mapa obtenido por métodos tradicionales.

El hecho de fundamentar la investigación sobre la variación lingüística en los datos ordenados en una base incide en un hecho importante, aunque no sea visible a simple vista: mientras que el dialectólogo tradicional tomaba un área geográfica para después determinar sus rasgos lingüísticos, el dialectómetro parte del análisis de los datos lingüísticos para llegar a configurar áreas dialectales (Aurrekoetxea 2010, Nerbonne 2011: 222). El factor geográfico se sitúa en un segundo plano, dando prioridad a los datos lingüísticos; lo primordial es el estudio de la configuración de los datos. Se debe analizar primeramente qué datos lingüísticos están correlacionados entre sí, cuál es la imbricación entre ellos, cuál es el grado de implicación entre distintos rasgos, y, una vez realizado este análisis, se proyectan los resultantes grupos de rasgos que tienen mucha o gran imbricación en un espacio geográfico para componer áreas dialectales. El área dialectal debe ser la plasmación geográfica del análisis diatópico. El espacio geográfico no debe ser la condición que reúnan los rasgos lingüísticos, sino su consecuencia. Al inicio está el análisis de los datos lingüísticos, y su consecuencia se refleja en un espacio geográfico determinado. En otras palabras: se parte de la lingüística para llegar a la geografía, y no al revés.

Todas estas aportaciones de la DM confieren a la dialectología un valor científico del que carecía la dialectología tradicional, a nuestro modo de ver. La dialectología moderna, en tanto que configurada como una ciencia de la variación lingüística, debe garantizar un valor científico incuestionable.

7. LOS RETOS DEL FUTURO

Es indudable que con las técnicas actuales las fronteras dialectales se determinan de manera mucho más fidedigna, pues están construidas con métodos rigurosos y científicos. La potencialidad de la dialectología cuantitativa es extraordinaria, pues ofrece rigurosidad, precisión y fiabilidad. Pero el dialectólogo actual carece de una heurística con la que pueda denominar las áreas dialectales creadas. En este aspecto no se ha avanzado ni un milímetro: no sabemos qué son los nuevos grupos dialectales que hemos creado, no sabemos cómo denominarlos, no sabemos dónde comienzan los dialectos o los subdialectos, ni dónde terminan. Desconocemos qué requisitos tiene que cumplir un área dialectal para ser denominada *dialecto*, *subdialecto* o *variedad*. ¿Estamos mejor o peor que antes? Realmente es una situación un tanto peculiar, porque a pesar de que a día de hoy se realicen infinidad de investigaciones dialectométricas, no hay prácticamente ningún trabajo acerca de la denominación de esas áreas dialectales. Se ha avanzado técnicamente muchísimo, pero muy poco en el aspecto teórico.

El dendrograma de la fig. 1 presenta dos áreas dialectales claras, con un nivel de diferenciación alto, que podría traducirse en una propuesta según la cual estas dos áreas configuran dos dialectos. Ambas áreas concentran una diferenciación lingüística superior al 28 %. El complemento de la fig. 1 es la fig. 2, en la que aparece la misma información pero con una visualización espacial de la diferenciación lingüística. Con las dos figuras sobre la mesa, la conclusión es clara: en el espacio geográfico estudiado, y de acuerdo con los datos utilizados, nos encontramos con dos áreas bien delimitadas que, sin riesgo de confusión, se podrían llamar *dialectos*.

Esta conclusión, sin embargo, no se ha explicitado en los trabajos dialectométricos, hasta el momento. Por ello, afirmamos que los investigadores que emplean la cuantificación en los estudios sobre la variación geolingüística tienen un déficit teórico o heurístico: no se ha desarrollado una metodología que permita trasplantar a la lingüística los datos proporcionados por las técnicas estadísticas. No se sabe cómo denominar las áreas o los grupos de variedades que aparecen en los resultados estadísticos. No hay un mecanismo válido para la interpretación lingüística de los datos que proporciona la estadística.

Pero, en vez de indagar en esta dirección, nos estamos enmarañando con complicaciones técnicas acerca de las clasificaciones (clústeres deterministas, *fuzzies*, *Spectral Graph Clustering*, *Principal Component Analysis*, *Bipartite Spectral Graph Partitioning*, *Bootstrap Clustering*, *Noisy Clustering*, etc.), que mejoran sin duda el análisis, pero habría que ver si el esfuerzo invertido compensa los resultados. No estamos sugiriendo ahorrar los esfuerzos necesarios para mejorar las técnicas de clasificación de variedades dialectales, no es esa la cuestión. Lo que proponemos es trabajar en la dirección de saber interpretar los resultados de las técnicas utilizadas profusamente en las últimas décadas, profundizar en los aspectos teóricos de la variación, indagar en los requisitos necesarios para que determinadas áreas dialectales puedan ser llamadas *dialectos* o *subdialectos*. Las herramientas que utilizamos, y sus resultados, de por sí no proporcionan la solución a nuestras inquietudes, pues son únicamente técnicas que nos aportan abundante y fehaciente información, pero que no pueden obviar la necesidad de interpretar los resultados obtenidos. A este respecto, a día de hoy son más las preguntas que las respuestas o soluciones. Este es uno de los retos de tiene la dialectología actual.

Por otra parte, los dialectólogos que trabajan en DM dudan de la robustez de las configuraciones dialectales logradas por métodos estadísticos de clústeres jerárquicos. Ante este hecho se propone el uso de clúster *fuzzy* con análisis de las localidades frontera (Clua *et al.* 2017) para diferenciar entre un *continuum* lingüístico en el que prácticamente no existe frontera lingüística y una frontera propiamente dicha en la que se pueda distinguir un tipo de ruptura lingüística entre dos áreas dialectales. La

delimitación entre frontera dialectal y zona de transición es también fundamental en el desarrollo de esta disciplina. Si bien se han dado los primeros pasos, la cuantificación debe llevar al dialectólogo a estadios de más seguridad en la configuración de las condiciones propias tanto de las fronteras como de las zonas de transición.

Como tercer reto, se podría citar el problema de las múltiples respuestas (MR). Entendemos que hay MR cuando en la misma localidad o a partir del mismo informante se recoge, para un mismo concepto, más de un vocablo o más de una realización fonética o morfológica de una palabra. Por poner un ejemplo, en el atlas lingüístico vasco, para la forma auxiliar verbal intransitiva del potencial hipotético, se han documentado dos respuestas en la localidad vizcaína de Zeberio: *leitike* y *lei* (EHHA, vol. VI: mapa n.º 1302); así, tanto *etorri leitike* como *etorri lei* son usadas para expresar ‘podría venir’. Cuando en la misma localidad se registra más de una respuesta a una pregunta del cuestionario, y no hay diferenciación ni sociolingüística ni estilística, nos encontramos ante dos formas que coexisten. Esta realidad lingüística no ha sido tenida en cuenta mayormente en la dialectología tradicional, y, cuando ha sido tratada, ha recibido la denominación de *polimorfismo* (Allières 1991, Aurrekoetxea 2019); y aunque en la DM se está incluyendo en los paquetes informáticos a día de hoy, hay diversidad de planteamientos (Aurrekoetxea *et al.* 2013, Nerbonne 2017) que en un futuro próximo deberían llegar a una interpretación y cuantificación aceptada por la comunidad científica.

Otro de los requisitos para que nuestra labor sea considerada científica es la portabilidad de los resultados. Este aspecto está muy unido al núcleo central de la definición de *dialecto*, y por tanto es tan espinoso como aquel. La dialectometría o dialectología cuantitativa debe proporcionar las técnicas necesarias para un correcto análisis de áreas dialectales y escalas de fronteras, de tal forma que una determinada frontera pueda ser asociada al concepto ‘dialecto’, otra a ‘subdialecto’, etc. Sin embargo, es tarea del dialectólogo, en una escala de fronteras, decidir el nivel de dicha escala para cada concepto (‘dialecto’, ‘subdialecto’, etc.). Un ejemplo a modo ilustrativo: si en una determinada lengua se estima que un dialecto tiene una escala de diferenciación lingüística, pongamos por caso, de un 30 % con respecto al dialecto contiguo, y en otra lengua se estima en un 60 %, es indudable que no estamos hablando de una entidad comparable, puesto que la diferencia entre estas dos realidades es doble. ¿Es posible en ciencia usar la misma denominación para dos realidades tan diferentes? A mi modo de ver, usar la misma denominación para dos realidades como las descritas aquí es causa de confusión y no ayuda a un uso racional del término.

Otro de los aspectos susceptibles de estudio es el de la frecuencia de las formas. En los análisis a nivel de pronunciación es muy interesante, para una diferenciación

de las variedades, la frecuencia de uso tanto del vocabulario como de formas gramaticales. La incorporación de la frecuencia en los análisis estadísticos reforzaría positivamente los resultados dialectométricos.

Por último, la DM del futuro debería tener la capacidad de revertir los datos y lograr automáticamente las características lingüísticas de cada área dialectal. De hecho, en el plano conceptual no tiene gran dificultad, y su coste podría ser asumible por diversos equipos dialectométricos.

REFERENCIAS BIBLIOGRÁFICAS

- ALLIÈRES, Jacques (1991): “Statut et limites du polymorphisme morphologique: le verbe dans la *Grammaire cantabrique basque* de Pierre d’Urte (1712), seconde partie (les auxiliaires)”, en Joseba A. Lakarra (ed.), *Memoriae L. Mitxelena magistri sacrum*, Donostia: Gipuzkoako Foro Aldundia, vol. 2, pp. 767-812.
- ALVAR, Manuel (1984): “Automatización de los índices en los atlas lingüísticos”, en Manuel Alvar, *Informática y lingüística*, Málaga: Ágora, pp. 107-119.
- ATWOOD, E. Bagby (1955): “The phonological divisions of Belgo-Romance”, *Orbis*, 4, pp. 367-389.
- AURREKOETXEA, Gotzon (2010): “Sobre la dialecticidad de los dialectos”, en Ana María Cano González (ed.), *Homenaxe a Xosé Lluís García Arias*, Uviéu: Academia de la Llingua Asturiana, t. 1, pp. 53-77.
- AURREKOETXEA, Gotzon (2016): “Distantzia geografikoaren eta hizkuntza distantzien arteko korrelazioa” [correlación entre distancias geográficas y distancias lingüísticas], en Gotzon Aurrekoetxea, Jesús Mari Makazaga y Patxi Salaberri (eds.), *Txipi Ormaetxea omenduz. Hire bordatxoan*, Bilbao: Universidad del País Vasco / Euskal Herriko Unibertsitatea, pp. 55-72.
- AURREKOETXEA, Gotzon (2019): “Hizkuntza bariazioa eta erantzun anitzak” [variación lingüística y respuestas múltiples], en Aintzane Etxebarria, Aitor Iglesias, Hiart Legarra y Asier Romero (eds.), *Traineru bete lagun. Inñaki Gaminde omenduz*, Bilbao: Universidad del País Vasco / Euskal Herriko Unibertsitatea, pp. 171-189.
- AURREKOETXEA, Gotzon, Karmele FERNÁNDEZ-AGUIRRE, Jesús RUBIO, Borja RUIZ y Jon SÁNCHEZ (2013): “DiaTech: A new tool for dialectology”, *Literary and Linguistic Computing*, 28(1), pp. 23-30. <https://doi.org/10.1093/lcl/fqs049>.
- AURREKOTXEA, Gotzon y Charles VIDEGAIN (2009): “Le projet Bourciez: traitement géolinguistique d’un corpus dialectal de 1895”, *Dialectologia*, 2, pp. 81-111.
- CHAMBERS, J. K. y Peter TRUDGILL (1980): *Dialectology*, Cambridge: Cambridge University Press.
- CLUA, Esteve (2010): “Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal”, en Gotzon Aurrekoetxea y Jose Luis Ormaetxea (eds.), *Tools for Linguistic Variation*, Bilbao: Universidad del País Vasco / Euskal Herriko Unibertsitatea, pp. 151-166.
- CLUA, Esteve, Aitor IGLESIAS, Iker USOBIAGA, y Miquel SALICRÚ (2017): “Nola identifikatu ezaugarri esanguratsuenak bariazio dialektalean: DiaTech nabarmentzeko aukera” [cómo identificar los rasgos más significativos en la variación dialectal: DiaTech como opción para destacarlos], en Aitor

- Iglesias y Ariane Ensunza (eds.), *Gotzon Aurrekoetxea lagunarterik hara*, Bilbao: Universidad del País Vasco / Euskal Herriko Unibertsitatea, pp. 25-34.
- DAVIS, Laurence M., Charles L. HOUCK y Veronika K. HORVATH (1999): "The East-West New England dialect boundary: Another look at the evidence", *Leeds Studies in English*, 30, pp. 123-135.
- EHHA = EUSKALTZAINDIA (2008): *Euskararen Herri Hizkeren Atlas* [atlas lingüístico vasco], Bilbao: Euskaltzaindia.
- GOEBL, Hans (1981): "Éléments d'analyse dialectométrique (avec application à l'AIS)", *Revue de Linguistique Romane*, 45, pp. 349-420.
- GOEBL, Hans (1987): "Points chauds de l'analyse dialectométrique: pondération et visualisation", *Revue de Linguistique Romane*, 51, pp. 63-118.
- GOEBL, Hans (1991): "Una classificazione gerarchica di dati geolinguistici tratti dall'AIS. Saggio di dialettometria dendrografica", *Linguistica*, 31, pp. 341-351. <https://doi.org/10.4312/linguistica.31.1.341-352>.
- GOEBL, Hans (1992): "Problèmes et méthodes de la dialectométrie actuelle", en Gotzon Aurrekoetxea y Charles Videgain (eds.), *Actas del Congreso Internacional de Dialectología*, Bilbao: Euskaltzaindia, pp. 429-476.
- GOEBL, Hans (2005): "La dialectométrie corrélative: un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme", *Revue de Linguistique Romane*, 69, pp. 321-367.
- GOOSSENS, Jan (1977): *Deutsche Dialektologie*, Berlin / New York: De Gruyter.
- HEERINGA, Wilbert y Jelena PROKIĆ (2018): "Computational dialectology", en Charles Boberg, John Nerbonne y Dominic Watt (eds.), *The Handbook of Dialectology*, Oxford: Wiley Blackwell, pp. 330-347. <https://doi.org/10.1002/9781118827628.ch19>.
- HOPPENBROUWERS, Cor y Geer HOPPENBROUWERS (1988): "De feature-frequentie methode en de classificatie van Nederlandse dialecten", *TABU, Bulletin voor Taalwetenschap*, 18(2), pp. 51-92.
- HOPPENBROUWERS, Cor y Geer HOPPENBROUWERS (1994): "De indeling van de zuidoostelijke steektalen", *TABU, Bulletin voor Taalwetenschap*, 24(2), pp. 37-63.
- HOUCK, Charles L. (1967): "A computerized statistical methodology for linguistic geography: A pilot study", *Folia Linguistica*, 1, pp. 80-95. <https://doi.org/10.1515/flin.1967.1.1-2.80>.
- HOUCK, Charles L. (1969): *A Statistical and Computerized Methodology for Analyzing Dialect Materials* [Ph. D. diss.]: Iowa: University of Iowa.
- HOUCK, Charles L. (1985): "The Iowa Northern-Midland boundary revisited: A computerized and statistical analysis". Midwest American Dialect Society Meeting, St. Louis, noviembre.
- HOUCK, Charles L. (1986): "Multidimensional scaling as a statistical analytical procedure in a traditional dialect study". Midwest American Dialect Society Meeting, Chicago, noviembre.
- INOUE, Fumio (1996): "Computational dialectology (1)", *Area and Culture Studies*, 52, pp. 67-102.
- IODAN, Iorgu (1973 [1967]): *Lingüística románica*, Madrid: Alcalá.
- JAIN, Anil K. y Richard C. DUBES (1988): *Algorithms for Clustering Data*, New Jersey: Prentice Hall.
- KESSLER, Brett (1995): "Computational dialectology in Irish Gaelic", en *Proceedings of the European Association for Computational Linguistics (EACL, Dublin)*, San Francisco: Morgan Kaufmann, pp. 60-67. <https://doi.org/10.3115/976973.976983>.

- KRUSKAL, Joseph B. (1983): "An overview of sequence comparison", en David Sankoff y Joseph B. Kruskal (eds.), *Time Warps, String Edits, and Macro Molecules. The Theory and Practice of Sequence Comparison*, Reading (MA): Massachusetts: Addison-Wesley, pp. 1-40.
- NERBONNE, John (2009): "Data-driven dialectology", *Language and Linguistics Compass*, 3(1), pp. 175-198. <https://doi.org/10.1111/j.1749-818X.2008.00114.x>.
- NERBONNE, John (2011): "How much does geography influence language variation?", en Peter Auer, Martin Hilpert, Anja Stukenbrock y Benedikt Szmrecsanyi (eds.), *Space in Language and Linguistics. Geographical, Interactional, and Cognitive Perspectives*, Berlin / New York: De Gruyter, pp. 220-236. <https://doi.org/10.1515/9783110312027.222>.
- NERBONNE, John (2017): "Respecting local variation", en Aitor Iglesias y Ariane Ensunza (eds.), *Gotzon Aurrekoetxea lagumarterik hara*, Bilbao: Universidad del País Vasco / Euskal Herriko Unibertsitatea, pp. 13-24.
- NERBONNE, John y Wilbert HEERINGA (2010): "Measuring dialect differences", en Peter Auer & Jürgen Erich Schmidt (eds.), *Language and Space. An International Handbook o Linguistic Variation. Volume I: Theories and Methods*, Berlin: De Gruyter, pp. 550-567. <https://doi.org/10.1515/9783110220278.550>.
- NERBONNE, John, Jelena PROKIĆ, Martijn WIELING y Charlotte GOOSKENS (2010): "Some further dialectometrical steps", en Gotzon Aurrekoetxea y Jose Luis Ormaetxea (eds.), *Tools for Linguistic Variation*, Bilbao: Universidad del País Vasco / Euskal Herriko Unibertsitatea, pp. 41-56.
- NERBONNE, John, Wilbert HEERINGA, Erik van den HOUT, Peter van der KOOI, Simone OTTEN y Willem van de VIS (1996): "Phonetic distance between Dutch dialects", en Gert Durieux, Walter Daelemans y Steven Gillis (eds.), *CLIN VI. Papers from the Sixth CLIN Meeting*, Antwerp: University of Antwerp, pp. 185-202.
- PHILIPS, Dennis (1983): "Automating linguistic atlases", en Sarah K. Burton y Douglas D. Short (eds.), *Sixth International Conference on Computers and the Humanities*, Rockville (MD): Computer Science Press, pp. 488-511.
- REED, David W. y John L. SPICER (1952): "Correlation methods of comparing idiolects in a transition area", *Language*, 28, pp. 348-359. <https://doi.org/10.2307/410105>.
- SCHRAMBKE, Renate (2010): "Language and space: Traditional dialect geography", en Peter Auer y Jürgen Erich Schmidt (eds.), *Language and Space. An International Handbook o Linguistic Variation. Volume I: Theories and Methods*, Berlin: De Gruyter, pp. 87-107. <https://doi.org/10.1515/9783110220278.87>.
- SÉGUY, Jean (1971): "La relation entre la distance spatiale et la distance lexicale", *Revue de Linguistique Romane*, 35, pp. 335-357.
- SÉGUY, Jean (1973a): *Atlas linguistique de la Gascogne. Complément du volume VI: Notice explicative*, Paris: Centre National de la Recherche Scientifique.
- SÉGUY, Jean (1973b): "La dialectométrie dans l'Atlas linguistique de la Gascogne", *Revue de Linguistique Romane*, 37, pp. 1-24.
- SHAW, David (1974): "Statistical analysis of dialectal boundaries", *Computers and the Humanities*, 8(3), pp. 173-177. <https://doi.org/10.1007/BF02402137>.
- VALLS, Esteve y Esteve CLUA (2010): "Distància de Levenshtein vs. 'mètode COD': dos sistemes de mesura de la distància fonètica aplicats al Corpus Oral Dialectal", en Emili Casanova Herrero y

Cesáreo Calvo Rigual (eds.), *Actas del XXVI Congreso Internacional de Lingüística y de Filología Románicas*, Berlin: De Gruyter, vol. 6, pp. 255-268. <https://doi.org/10.1515/9783110299991.255>.