

Artículo original

Marcos y metodologías para la clasificación dialectal (con aplicación al Caso Vasco)¹

Frameworks and methodologies for dialectal classification (with application to the Basque case)

Ariane Ensunza

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0003-0908-7467>

e-mail: ariane.ensunza@ehu.eus

Lorea Unamuno

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0001-6249-717X>

e-mail: lorea.unamuno@ehu.eus

Gotzon Aurrekoetxea

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0002-9562-4806>

e-mail: gotzon.aurrekoetxea@ehu.eus

Recibido: 26/5/2023
Aprobado: 19/7/2023

RESUMEN

En este artículo se plantea una revisión -que se pretende completa y clarificadora- de las diferentes clasificaciones y metodologías dialectales con aplicación al caso vasco. Para ello analizamos clasificaciones basadas en diferentes marcos teóricos en los que se han desarrollado los estudios sobre las clasificaciones de los dialectos fundamentados principalmente en dos direcciones: criterios extralingüísticos y lingüísticos. Entre los primeros se enclavan aquellas llevadas a cabo con criterios extralingüísticos (étnicos -tribus...-, geográficos, etc.) y los basados en criterios del hablante no especialista (dialectología perceptual), en la que la clasificación se basa en la percepción que los hablantes tienen de las diferencias y similitudes lingüísticas de su lengua (Preston 1999; 2010b). Entre las segundas clasificaciones se citan las llevadas con criterios estrictamente lingüísticos (tanto las basadas en la gramática histórica (clas. Filogenética, etc.), dialectología estructural, generativa, etc. en el que la selección de los rasgos lingüísticos depende del marco teórico en el que se desarrolla la investigación; y las basadas en criterios cuantitativos o dialectométricos, que opera con grandes bases de datos y que han sido creados con información recogida generalmente en los atlas lingüísticos (Goebel 1992 entre otros muchos). Además, determinamos las diferencias que hay entre estos tipos de clasificación y, por último, examinamos las distintas metodologías utilizadas para la clasificación dialectal.

Palabras clave: dialectos; clasificación; dialectología isoglótica; dialectología cuantitativa; vasco.

¹ Este trabajo se ha llevado a cabo con la financiación del proyecto GIU21/016 financiado por la UPV/EHU.

ABSTRACT

In this article we propose a review -which is intended to be complete and clarifying- of the different classifications and dialectal methodologies with application to the Basque case. For this purpose, we analyze classifications based on different theoretical frameworks in which the studies on dialect classifications have been developed, mainly based on two directions: extralinguistic and linguistic criteria. Among the former are those based on extralinguistic criteria (ethnic -tribes...-, geographical, etc.) and those based on non-specialist speaker criteria (perceptual dialectology), in which the classification is based on the perception that speakers have of the linguistic differences and similarities of their language (Preston 1999; 2010b). Among the second classifications are those based on strictly linguistic criteria (both those based on historical grammar (phylogenetic class, etc.), structural dialectology, generative, etc., in which the selection of linguistic features depends on the theoretical framework in which the research is developed; and those based on quantitative criteria or dialectometry, which operates with large databases that have been created with information generally collected in linguistic atlases (Goebel 1992 in many others). In addition, we determine the differences between these types of classification and, finally, we examine the different methodologies used for dialect classification.

Keywords: dialects; classification; isoglottic dialectology; isoglottic dialectology; quantitative dialectology; Basque.

INTRODUCCIÓN

Uno de los aspectos más interesantes de los estudios de la variación diatópica, geolingüística o geografía lingüística es la determinación y la clasificación de las áreas dialectales.

La gran duda de antaño sobre si se pueden clasificar áreas lingüísticas (Schuchardt 1870), la discusión sobre fronteras y *continuum* dialectal entre tipófilos y tipófobos (Ascoli vs. Paris-Meyer), etc. ha de ser tenida como totalmente superada. Aquella fue época de búsqueda de fronteras con datos reales. De hecho, no habrá dialectólogo que no conozca el mapa creado por Rosenqvist con diversas isoglosas a lo largo de Francia o La línea Specia-Rimini en Italia que separa la Romania occidental de la oriental, por poner dos ejemplos del mundo románico. Toda la controversia finalizó con la aseveración de Louis Gauchat (1903) que sí había fronteras dialectales. A diferencia de los comienzos de la dialectología como ciencia, hoy en día hay abundantísima y muy diversa bibliografía sobre las clasificaciones dialectales, demostración del avance en esta materia, enlazando diversas escuelas y teorías lingüísticas hasta hacer de la variación un eje central en la teoría de la lingüística (Hinskens 2018).

De hecho, la controversia de los principios de la dialectología fue un falso debate, porque todo se puede clasificar, tanto flora como fauna, cualquier campo del saber tiene sus mecanismos y su metodología para clasificar. No puede quedarse el lenguaje o la diferenciación lingüística fuera de la clasificación. Si bien es verdad que la actividad lingüística hablada es efímera, toda la producción en un determinado momento y recogido por medio de diversos instrumentos de grabación y/o transcripción, tomado como un muestreo, es susceptible de ser clasificable (ver un ejemplo en el monográfico X (2022) de la revista *Dialectologia* con una primera entrega del proyecto “DIACLEU. Dialect Classifications of Languages in Europe”: <http://www.edicions.ub.edu/revistes/dialectologiasp2022/> las clasificaciones dialectales en diversas lenguas europeas, entre las que se encuentra el vasco (Ensunza & Unamuno, 2023).

Aunque todas las lenguas desarrolladas y con literatura lingüística científica tienen estudios sobre las clasificaciones dialectales, se da una evolución en la manera de enfocar la clasificación de los dialectos. Esta evolución se plasma en los distintos marcos teóricos que se han creado a lo largo de la historia de la dialectología, basados en distintas escuelas, teorías o encuadres teóricos de la lingüística, como puede ser la gramática histórica, el estructuralismo, el generativismo, etc.

CLASIFICACIONES BASADAS EN DIFERENTES MARCOS TEÓRICOS

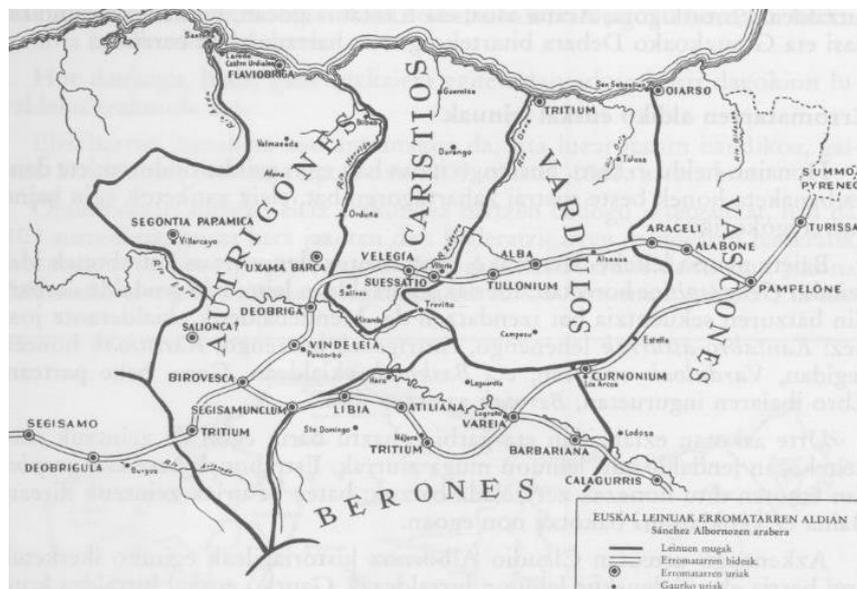
Siendo conscientes de que se trata de un tema muy controvertido, en el que no hay anclajes asumidos por una extensa gama de dialectólogos; a rasgos generales se podrían establecer diversos marcos teóricos en los que se basan las clasificaciones dialectales llevadas a cabo en la historia de la dialectología hasta el día de hoy. Estos marcos se basan en diferentes criterios que principalmente se pueden agrupar en dos grupos: criterios extralingüísticos y criterios lingüísticos. Tanto en un caso como en el otro se encuentra un enfoque global que intentamos abarcarlo en unos cuantos criterios.

En el marco del proyecto en curso DIACLEU (<https://www.diacl.eu/>), se ha decidido por un criterio amplio para modelar distintos marcos teóricos, haciendo abstracción e incluyendo todos ellos en cuatro grupos, a saber: clasificación etnológica-extralingüística, clasificación isoglótica, clasificación perceptual y clasificación cuantitativa o dialectométrica.

CLASIFICACIONES EXTRALINGÜÍSTICAS

Se incluyen en ella clasificaciones basadas mayoritariamente en aspectos no lingüísticos (o extralingüísticos), que pueden ser tanto culturales, como etnológicos, étnicos, geográficos o hechas por personas no especialistas o lingüistas. Este criterio ha sido utilizado, por ejemplo, en la lengua vasca para distinguir dialectos de acuerdo con tribus étnicas prehistóricas, que argumentado por Caro Baroja (1945) sintetizó en un mapa Mañarikua (1964) (figura 1). Según este trabajo, en las regiones en las que se asentaron en épocas prerromanas los carístios se sitúa el dialecto vizcaíno (dialecto más occidental del euskera); en el territorio de los várdulos se asienta actualmente el dialecto guipuzcoano (dialecto central) que ocupa territorios del sur y oeste de San Sebastián; en el territorio que fue ocupado por los vascones se sitúa el dialecto alto navarro, que ocupa regiones al norte y oeste de Pamplona.

Figura 1. Dialectos vascos atendiendo a criterios extralingüísticos



Mañarikua 1964 - <http://euskerarenjatorria.eus/?p=40110>

Como se puede comprobar, no se utiliza ninguna característica lingüística para la determinación y ubicación de los dialectos; es más, no hay evidencia ninguna de que dichas etnias utilizaran dicho dialecto, ni tan siquiera esa lengua. No hay constancia alguna de la lengua que hablaron dichos grupos étnicos. La coincidencia de las demarcaciones entre ciertos límites de grupos étnicos y los dialectos puede que no sea casual, pero no tiene por qué serlo.

Este tipo de clasificación se ha utilizado en entornos en donde la investigación antropológica se ha desarrollado antes que la investigación lingüística.

Cuando se utilizan criterios geográficos las motivaciones y los límites pueden ser diversos: uno de ellos es la nación en la que se encuentra; algo de ello podría decirse, según entendemos, saliendo del ámbito de nuestros conocimientos, probablemente dentro de las familias de las lenguas guaraníes podría identificarse esta característica.

CLASIFICACIONES ISOGLÓTICAS

Son clasificaciones llevadas a cabo, en su mayoría, dentro del marco de la dialectología tradicional espacial o geolingüística, basadas mayoritariamente en un punto de vista diacrónico, histórico, genealógico y teniendo en cuenta unas cuantas características lingüísticas. De hecho, en la dialectología tradicional hay variedad de submarcos de acuerdo con las distintas escuelas lingüísticas en las que se ha desarrollado el dialectólogo. Desde una clasificación basada en la lingüística histórica y comparativa (en la que los rasgos más determinantes de la evolución lingüística tienen mayor peso e importancia) hasta la lingüística geográfica en la que los rasgos se clasifican de acuerdo a su capacidad de configurar áreas dialectales, pasando por la lingüística estructural, diversas etapas de la generativa, etc. hay clasificaciones que tienen en cuenta el aspecto diacrónico y otros que se basan en aspectos estructurales, sistémicos y por lo tanto sincrónicos.

En cuanto a la denominación, puesto que el término de “dialectología tradicional” no convence a algunos y otros se posicionan en contra, o la dialectología basada en (unas pocas) características (“(single) feature-based dialectology”), preferimos utilizar el término de ‘dialectología isoglótica’. Es cierto que se usa mucho el término “(single) feature-based dialectology” (sobre todo en la literatura anglosajona). Se puede decir que casi todos ellos tienen una orientación diacrónica (bien por dar mayor importancia a fenómenos lingüísticos que han perdurado más que otros a lo largo de la evolución de la lengua, bien porque se han tenido como más determinantes a los fenómenos más innovadores en detrimento de los más conservadores), si bien es cierto que las clasificaciones llevadas a cabo en parámetros estructurales o generativos son eminentemente sincrónicos.

Es precisamente esta terminología la que se ha utilizado en el proyecto DIACLEU arriba mencionado para designar las clasificaciones llevadas a cabo en este marco teórico y la que hemos adoptado en esta contribución. Esta denominación tiene en cuenta la única herramienta que ha sido usada para la diferenciación lingüística en este marco. La isoglosa de cada rasgo lingüístico ha ido configurando una geografía lingüística con múltiples espacios lingüísticos diferenciados (ver por ejemplo Brun-Trigaud 2023:138, el mapa que Rosenqvist creó en 1919 con isoglosas en territorio galo-romance), y la suma de estas isoglosas particulares ha configurado áreas dialectales. La necesidad de encontrar áreas lingüísticas ha hecho que el dialectólogo reúna las características que configuren áreas homogéneas en la que las isoglosas cubran territorios semejantes, que posteriormente han recibido la denominación de dialectos y subdialectos, de acuerdo con la opinión o creencia del dialectólogo.

Este tipo es el más común en las lenguas en las que ha habido al menos una clasificación dialectal, por ejemplo, en el galo-romance (Brun-Trigaud 2023), el italiano (Cugno 2023), o en el noruego (Nesse & Høyland 2023).

En vasco tenemos dos clasificaciones isoglóticas: la primera de ellas es debida a L. L. Bonaparte (1868) que distingue 8 dialectos, 25 subdialectos y 51 variedades. Y la segunda de K. Zuazo (1998 y años sucesivos) que distingue 5 dialectos. Con respecto a la diferente cantidad de dialectos, el territorio navarro (en rojo) en el que Bonaparte distinguía dos dialectos Zuazo distingue un único, y en el territorio labortano-bajo navarro (en lila) en la que Bonaparte distinguía 3, Zuazo los agrupa en uno (figura 2).

Figura 2. Clasificación de los dialectos vascos



Zuazo 2019.

Es conocido que en esta metodología generalmente se usan unas pocas características lingüísticas, pero difieren mucho de unas clasificaciones a otras: por ejemplo, en las dos clasificaciones llevadas a cabo mediante esta metodología se han utilizado muchas características. Aunque en el caso de Bonaparte no se sabe a ciencia cierta cuántos utilizó (se cree que alrededor de 80), Zuazo ha utilizado 108 (mapa 2).

CLASIFICACIONES PERCEPTUALES

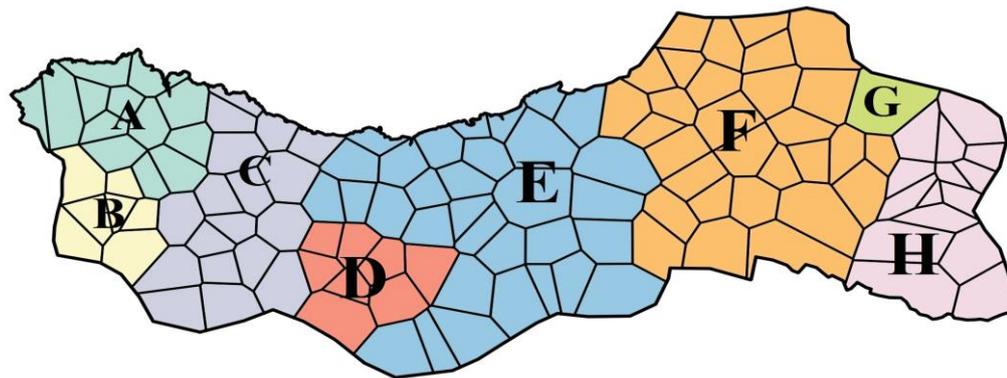
Las clasificaciones perceptuales son aquellas que se basan en las creencias, percepciones u opiniones de hablantes no-lingüistas; dicho de otro modo, clasificaciones llevadas a cabo mediante técnicas usadas en la dialectología perceptual. El máximo impulsor y referente de la dialectología perceptual es sin duda Dennis Preston (1989, 1999, 2002, 2010a). Actualmente, aparte de Preston, se pueden citar a Chris Montgomery (2007, 2011) y Montgomery & Beal (2011). Las clasificaciones perceptuales que se han llevado a cabo respecto a la lengua castellana no han sido prolíficas y no destacan en este campo, aunque últimamente están aflorando diversas investigaciones de las que se verá qué relevancia alcanzan (Calaza Díaz et al. 2016; Díaz-Campos y Navarro-Galisteo 2009; Moreno Fernández & Moreno Fernández 1999 etc.).

Para la detección y clasificación de los dialectos o áreas lingüísticas es sumamente práctico el método *little arrow method*. Para conseguir estos límites subjetivos de las áreas dialectales es suficiente una pregunta: “¿entre los pueblos de tu área en cuál se habla como en tu pueblo o muy parecido?”. Las respuestas logradas se colocan en un mapa, en el que las localidades se unen mediante flechas que parten de la localidad de origen (localidad encuestada) y el o los pueblos donde el lugareño ha señalado que se habla igual o muy parecido. Si el informante menciona muchas localidades y/o la variedad de la localidad de referencia no difiere mucho de las de su alrededor, será origen de muchas flechas. De todas formas, ciertas zonas del mapa no estarán surcadas por flechas y habrá espacios blancos (sin flechas). Estas zonas, de una u otra manera, indican fronteras lingüísticas entre áreas lingüísticas diferentes. Si se trazan líneas en estas zonas blancas se crearía el mapa mental de los hablantes sobre las fronteras de los dialectos de esa lengua. Mediante este método se logra una demarcación subjetiva del mapa de los dialectos, que puede o no coincidir con la demarcación hecha por los especialistas en el tema.

Solo hay una clasificación de este tipo en vasco, se trata de la que publicaron Aurrekoetxea et al. (2017) basándose en el método del *little arrow method* se plasmaron las respuestas en un mapa subjetivo de dialectos en el que se distinguieron 8 variedades diferentes (figura 3). De acuerdo con la percepción de los hablantes vascos esta lengua estaría dividida

en 8 dialectos: tres de ellos (los grupos A, B y C) se hallarían encuadrados en lo que tradicionalmente ha sido llamado dialecto vizcaíno, el dialecto guipuzcoano y el navarro conformarían un único dialecto (dialecto E), exceptuando la parte sur-occidental (dialecto D). En la parte continental el dialecto F agruparía los dialectos tradicionales labortano y bajo-navarro, exceptuando el mixano (dialecto G) y el dialecto tradicional suletino que conformaría el dialecto H.

Figura 3. Clasificación subjetiva de los dialectos vascos



Aurrekoetxea et al. (2017).

CLASIFICACIONES AGLOMERATIVAS

Clasificaciones dialectales basadas en grandes cantidades de datos y herramientas estadísticas. En este tipo de clasificaciones dialectales hay un cambio de paradigma: ya no se basan en unos pocos rasgos lingüísticos, como en las clasificaciones isoglóticas, ni en isoglosas. El punto de partida son grandes masas de datos (recogidos en la mayoría de los casos en los atlas lingüísticos). Las bases de datos diseñadas para un tratamiento cuantitativo pueden ser creadas tanto a nivel superficial de los datos (nivel fonético) como a nivel profundo (nivel fonológico), según los intereses y el fin de la investigación. Si el objetivo es una investigación del nivel superficial o fonético de los datos se crearán bases fonéticas u ortográficas, con una transcripción de lo más fidedigna de las respuestas de los hablantes. Si, al contrario, el objetivo es una investigación a nivel profundo del idioma la base de datos ha de ser elaborada de acuerdo con un marco teórico determinado, puesto que las divergencias a nivel superficial no son las mismas que a nivel profundo.

En cuanto a la denominación, preferimos la clasificación aglomerativa (“aggregate dialectology”, Nerbonne 2008), a pesar de que esta denominación se refiere más a la metodología que al marco teórico en sí. Entre las características más sobresalientes se trata de una clasificación eminentemente sincrónica; con datos sincrónicos y análisis sincrónico de los mismos.

Otra de las características de este tipo de clasificaciones es que se trabaja con grandes bases de datos, por lo que también ha sido denominado “dialectología de corpus” (corpus-based: Szmrecsanyi 2008, 2011). A día de hoy, uno de los requisitos imprescindibles es el uso de grandes masas de datos si el objetivo es lograr una clasificación dialectal fiable, si bien el mero hecho de contar con una gran base de datos no implica que la muestra sea representativa. En lingüística, y por supuesto en dialectología, no se puede contar con el total de los datos, por lo que siempre serán una muestra; por ello, es imprescindible que la muestra sea representativa del conjunto del sistema que se trata de analizar.

A dichas bases de datos se les aplica una unidad de distancia lingüística. En dialectología

se usan varios tipos de unidades, tanto las “distancias de edición”, como las nominales o categóricas (ver, por ejemplo, Aurrekoetxea 2022:47), con ello se consiguen por una parte distancias fonéticas y por otra distancias lingüísticas. Dependiendo del tipo de base de datos se elegirá una u otra. Aplicando una de estas medidas de diferenciación lingüística a la base de datos se logra la matriz de distancias, en las que se expresan las distancias (o diferencias) lingüísticas entre las localidades analizadas. A la matriz de distancias se les aplican los procedimientos matemáticos y estadísticos (análisis multivariantes), y cartografía automatizada que desembocan en diversos mapas cuantitativos y diagramas. La dialectometría ha desarrollado diversos tipos de mapas: mapas en los que se expresan distancias lingüísticas partiendo de una localidad, mapas en la que se expresan las correlaciones entre la distancia geográfica y la lingüística, mapas de áreas, etc. Diversas escuelas dialectométricas han elaborado aplicaciones específicas que guían y ayudan al especialista en su análisis cuantitativo: por ejemplo, la escuela de Salzburgo la herramienta VDM, la escuela de Groningen (la Gabmap), la UPV/EHU Diotech, entre otros. La primera herramienta funciona en local y las otras dos *online*. Cada uno de ellos obedece a principios y objetivos diferentes[†]. Para llevar a cabo una clasificación de los dialectos el procedimiento habitual de los trabajos dialectométricos suele ser el análisis de conglomerados (o *cluster analysis*); un procedimiento determinista que actualmente se suele complementar con otros tipos como el escalamiento multidimensional (MDS). El uso de herramientas que utilicen procedimientos estadísticos no implica que el lingüista deba ser experto en estadística para hacer dialectometría. Pero sí debe saber interpretar lingüísticamente los resultados del análisis.

Este tipo de clasificaciones supera la fase de la dialectología basada en las isoglosas, puesto que no analiza cada rasgo lingüístico, sino todos ellos en su conjunto. Se pasa de un sistema en el que el investigador usa determinados rasgos lingüísticos, elegidos de acuerdo a criterios muchas veces no especificados, a otro en el que se toman todos los datos disponibles. En el primero puede darse el caso de que los rasgos hayan sido recogidos por diferentes dialectólogos con diferentes objetivos, mientras que en el segundo suelen ser recogidos mediante atlas lingüísticos, atendiendo a criterios establecidos anteriormente. En el primer caso no hay una red de localidades homogénea, mientras que en el segundo se suele establecer antes del comienzo de la recogida de datos. Por todo ello, este tipo de clasificaciones cambia por completo el paradigma de la investigación.

Actualmente el vasco cuenta con una clasificación dialectal basada en métodos cuantitativos (Aurrekoetxea et al. 2019), llevado a cabo con la herramienta Diotech, en el que se distinguen 3 dialectos y 7 subdialectos. Para la clasificación de los dialectos vascos los datos han sido tomados del *Atlas Lingüístico del Vasco (EHHA)*. En total 649 conceptos o nociones recogidas en 145 localidades, lo que hace más de 100.000 datos en la base de datos que se ha utilizado. Una vez construida la base de datos se le aplica la unidad de distancia lingüística para conseguir la matriz de distancias lingüísticas entre localidades, a las que se les aplica el análisis de conglomerados multivariante. Es sabido que en esta técnica pueden ser usados diversos algoritmos (*Ward* o varianza mínima, vecino más alejado, o más próximo, distancias mínimas, distancias máximas, distancias entre centroides, distancias ponderadas...). En este caso se aplicaron varios de ellos y se eligió el que más algoritmos proporcionaban (figura 4).

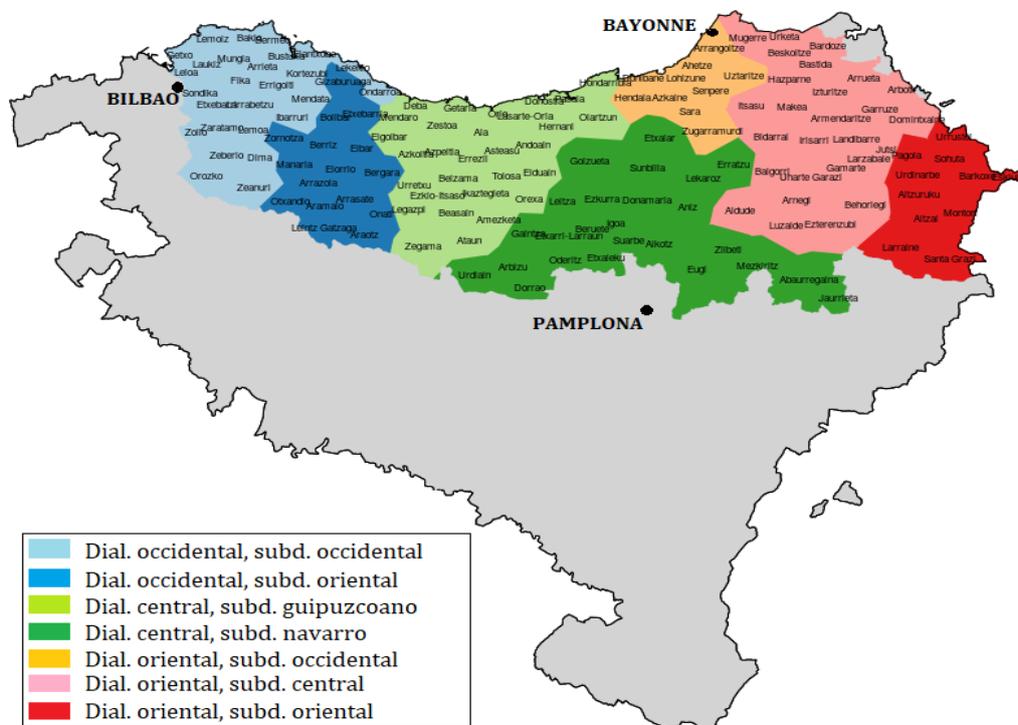
¿QUÉ DIFERENCIA HAY ENTRE ESTOS TIPOS DE CLASIFICACIÓN?

A pesar de que hayan sido utilizados para distintas clasificaciones dialectales, las diferencias entre estos cuatro tipos de clasificación varían mucho tanto en cuanto al marco

[†] Los mapas tanto de la clasificación perceptual como de la dialectométrica se han elaborado con el programa Diotech (<http://eudia.ehu.es/diotech/index>)

teórico en el que se basan como en la metodología que usan. Un esquema-resumen puede ser útil para diferenciar los marcos teóricos utilizados (tabla 1):

Figura 4. Mapa de los dialectos y subdialectos



Aurrekoetxea et al. 2019.

Tabla 1. Tipos de marcos teóricos usados en la clasificación dialectal

Clasificación extralingüística	<ul style="list-style-type: none"> - Criterios etnológicos - Criterios geográficos - Criterios étnicos - Criterios culturales
Clasificación isoglótica	<ul style="list-style-type: none"> - Dialectología de características seleccionadas - Clasificación filogenética - Lingüística histórica - Lingüística comparativa - Lingüística geográfica
Clasificación perceptual	<ul style="list-style-type: none"> - Hablantes no-lingüistas - Clasificación impresionista - Clasificación subjetiva
Clasificación dialectométrica	<ul style="list-style-type: none"> - Grandes masas de datos - Unidad de distancia lingüística - Matrices de distancias - Análisis multivariante - Clasificación de conglomerados - Escalamiento multidimensional

Elaboración con base al trabajo de investigación.

Como se puede ver, estos tipos se diferencian en el marco teórico en el que se basan. Creemos que hay que priorizar este punto de vista, aunque también se pueden analizar atendiendo a la metodología desplegada en su análisis.

De acuerdo con el marco teórico, estos cuatro tipos de clasificación se pueden reducir a dos, siguiendo criterios estrictamente lingüísticos:

- Clasificaciones basadas en criterios estrictamente no lingüísticos o extralingüísticos: entre aquellos citaríamos a la clasificación extralingüística y perceptual.
- Clasificaciones basadas en criterios únicamente lingüísticos: la clasificación isoglótica y cuantitativa.

METODOLOGÍA EN LA CLASIFICACIÓN DIALECTAL

Las clasificaciones dialectales también se pueden diferenciar en la metodología utilizada en cada una de ellas. Para analizar las distintas metodologías utilizadas para la clasificación dialectal en este trabajo se concentrará en las dos clasificaciones dialectales llevadas a cabo en los marcos estrictamente lingüísticos; es decir, las clasificaciones isoglóticas y las clasificaciones aglomerativas.

METODOLOGÍA DE LAS CLASIFICACIONES ISOGLÓTICAS

Se trata de una metodología basada en rasgos seleccionados por el investigador lo que en terminología inglesa se ha denominado (*single*)-*feature based dialectology*. El primer paso de esta metodología consiste en seleccionar los rasgos lingüísticos en los que se basará la diferenciación lingüística y la clasificación dialectal. La selección de los rasgos puede obedecer a más de un factor: la importancia del rasgo en la evolución de la lengua, el peso en la estructuración del espacio debido al rasgo en cuestión y/o la facilidad de la obtención de los rasgos, entre otros factores. En la mayoría de los casos el investigador elige entre los rasgos que tiene a mano las que considera que son más fundamentales o necesarios para la diferenciación dialectal.

Las características lingüísticas usadas pueden ser tanto fonológicas, morfológicas, lexicales como sintácticas, aunque en la mayoría de los casos suele ser una mezcla de todas ellas. Hay muy pocas clasificaciones dialectales llevadas a cabo basándose únicamente en una de estas categorías: por ejemplo, solo en características fonológicas.

Esta metodología está orientada en la mayoría de los casos a la diacronía, y utiliza argumentos y características más importantes del devenir de la evolución lingüística de la lengua; sin embargo, también hay clasificaciones creadas desde el punto de vista sincrónico, aunque son más esporádicas. Cuando la investigación tiene una orientación diacrónica, la lengua no es considerada como un sistema con una estructura lingüística sino como una suma de características. Cuando se trabaja desde la sincronía generalmente la lengua suele ser analizada como un sistema (ver por ejemplo Newton 1972).

Es de reseñar que en este tipo de clasificaciones no suele haber información relativa a las técnicas de clasificación de los dialectos y subdialectos. En las publicaciones de esta índole no aparece en detalle cuál ha sido el procedimiento llevado a cabo para designar un área dialectal como “dialecto” y/o “subdialecto”. El lector debe suponer que las isoglosas marcadas por ciertas características han sido consideradas como las fronteras entre dialectos, pero no hay constancia fehaciente de ello.

Entre los rasgos de la “dialectología isoglótica” se pueden citar:

- Selección de características lingüísticas; el dialectólogo elige las características lingüísticas a tener en cuenta de acuerdo a sus intereses y disponibilidad de los rasgos.
- Clasificación basada en pocas características lingüísticas (Inoue 1996a: 74).

- Predominancia de unas características ante otras, en general basándose en su importancia en la evolución de la lengua. Sin embargo como Chambers y Trudgill (2004:97) afirman “Yet in the entire history of dialectology, no one has succeeded in devising a satisfactory procedure or a set of principles to determine which isoglosses or which bundles should outrank some others”.
- Arbitrariedad en la elección de los rasgos lingüísticos.
- Subjetividad (riesgo de) en la elección de las características a tener en cuenta. Ya lo decía nuestro gran maestro Mitxelena “... se emplean en realidad, aquellos criterios que en la práctica se han demostrado más manejables, que no son necesariamente los mejores en teoría” (1976:319).
- Las isoglosas no siempre coinciden (Kessler 1995: 60), con lo que hay que elegir o rechazar algunas de ellas.
- Como consecuencia de ello, muchas isoglosas no dividen de forma nítida áreas lingüísticas (Kessler 1995: Ídem).
- Ello conlleva a la dificultad para delimitar fronteras (Kessler 1995: Íbidem);
- Subjetividad en la determinación de “dialectos” y “subdialectos”; y a parte de ello, en la mayoría de los casos no hay argumentación científica para diferenciarlos;
- Imposibilidad de delimitar el grado de divergencia entre dialectos: “No puedo entrar aquí a discutir hasta qué punto son divergentes los distintos dialectos vascos. Me limitaré, pues, a decir que una consideración científica de este y otros problemas semejantes exige, como requisito indispensable, algún procedimiento operativo de cuantificar las distancias lingüísticas, y remitiré a la propuesta de nuestro malogrado amigo Jean Séguy, (...). A falta de alguna suerte de medida, habrá que referirse a estimaciones del orden de magnitud, hechas a ojo de buen o mal cubero” (Mitxelena 1981:295).
- Imposibilidad de jerarquizar las fronteras.

Mitxelena da en el clavo cuando se refiere que son “hechas a ojo de buen o mal cubero”, porque entre otras cosas, la dialectología isoglótica adolece de una medida de diferenciación lingüística que haga posible concretar dicha diferenciación.

METODOLOGÍA DE LAS CLASIFICACIONES AGLOMERATIVAS

Con el objetivo de superar las debilidades de la dialectología tradicional o isoglótica se ha desarrollado lo que en inglés ha sido designado como “aggregate dialectology” (Nerbonne 2008), o dialectología aglomerativa. También ha sido llamada “dialectología de corpus” (Szmrecsanyi 2011) y se podría utilizar también el término de dialectología sintética, porque se trata de hacer una síntesis de todos los datos que se han recogido en un corpus.

Se trata de un cambio radical en la metodología para la elaboración y determinación de áreas dialectales, en el análisis de la influencia de la distancia geográfica en la diferenciación lingüística. Estas diferencias se pueden concretar en: a) la elección de los rasgos lingüísticos, que ha sido tildada de “subjetivo” y que lastra todo el trabajo posterior; b) el uso de pocos rasgos lingüísticos; c) el uso de la subjetividad en la determinación de las áreas dialectales; d) el no uso de la cuantificación en la diferenciación lingüística. Todos ellos se pueden resumir en dos puntos: rasgos lingüísticos utilizados y cuantificación de la diversidad lingüística.

En cuanto al aspecto lingüístico, no se trata ya de una selección de rasgos llevada a cabo por el investigador; es más, se rehúye de ello. A pesar de tener en cuenta que en cuestiones de lengua no es posible analizar toda la lengua (el universo), y hay que utilizar una muestra, se intenta ser lo más 'objetivo' posible, sin que quepa un atisbo de subjetividad en la elección de la muestra. Con este objetivo se parte de la toma en consideración de todos los datos de que dispone el investigador en base a criterios objetivos y establecidos de antemano. Estos datos

deben cumplir ciertos requisitos, como:

- Que sean sincrónicos
- Que conformen una red extensa de localidades
- Que sean coherentes desde el punto de vista sociolingüístico
- Que hayan sido recogidos utilizando iguales o parecidas técnicas

Los atlas lingüísticos suelen cumplir estos requisitos y por lo tanto suelen ser el punto de partida de los análisis y clasificaciones aglomerativas.

Es un cambio de paradigma del análisis de las características individualmente se pasa al análisis de grandes masas de datos. Nace con el objetivo de superar el (*single*)-*feature based dialectology*, impulsando una mayor abstracción en el análisis de los datos. Este cambio de lo particular a lo colectivo simboliza claramente Nerbonne (2008:2) cuando dice que:

Please note that we have not accused dialectology of not identifying important signals of provenance. We rather formulate the charge that single feature based dialectology fails to isolate these signals analytically and that aggregation is the key methodological step needed to enable analytical progress.

Analizando grandes cantidades de datos se puede llegar más fácilmente a proponer y formular leyes generales sobre la variación teniendo en cuenta toda la diversidad lingüística que se encuentra en una lengua y no solamente unas cuantas características. Hay que agregar muchos rasgos en cada variedad, para detectar de forma fiable las relaciones que tienen las distintas variedades entre sí. Precisamente porque como dijo ya Bloomfield (1933:328) “every word has its own history”. Efectivamente, hay que superar lo individual y elevar el nivel de abstracción, y usar técnicas de análisis más potentes que tengan en cuenta todas las características lingüísticas (posibles) de una variedad, para analizarlas en su conjunto en vez de individualmente, ver la configuración geográfica que crean, y las distintas relaciones que puede haber entre la variación lingüística y otros aspectos sociales, sociológicos, históricos, geográficos, etc.

Como asevera Francis (1983:235)

The advantage of maps of this sort is that they show objectively and quantitatively the changing degrees of linguistic variation across an area while avoiding the misleading implications which too often arise from isoglosses, whether single or in bundles.

Otro de los aspectos a destacar en la dialectología agregada es la cuantificación. Por ello también ha sido llamada dialectología cuantitativa. De hecho, los dos aspectos son inseparables, puesto que es prácticamente imposible la gestión de grandes masas de datos sin instrumentos automatizados o informáticos.

Hay muchas técnicas cuantitativas de agregación (ver entre otros Goebel 1992, Heeringa 2004; Nerbonne & Kretschmar 2006), tanto para detectar áreas dialectales, como para jerarquizar fronteras (Aurrekoetxea et al 2019; Nerbonne, Kleiweg & Manni 2007) o para analizar los *continuum* lingüísticos (Heeringa & Nerbonne 2001).

En resumen, la dialectología cuantitativa necesita de:

- Base de datos amplia; un corpus estructurado de datos, producto de la agregación de los datos
- Unidad de distancia lingüística para saber cómo contar las diferencias lingüísticas
- Matriz de distancias entre variedades
- Sistemas cuantitativos para el análisis de las diferencias lingüísticas (análisis de conglomerados, escala multidimensional, etc.)
- Cartografía computarizada

CONCLUSIONES

Las diferentes clasificaciones dialectales llevadas a cabo en distintas lenguas han sido encasilladas en cuatro grupos, atendiendo al marco teórico en el que han sido confeccionadas. De los cuatro grupos dos de ellos se proveen de características propiamente lingüísticas y han sido creadas basándose en criterios de especialistas en la materia: las clasificaciones isoglóticas y las clasificaciones dialectométricas. Se ha demostrado que las diferencias entre éstas últimas, tanto en el marco teórico como en el metodológico, son muy grandes. El avance y la sofisticación en la clasificación dialectal permite a esta ciencia lograr resultados cualitativamente mejores y más fiables, porque entre otras cuestiones, el uso de sistemas cuantitativos para analizar la diferenciación lingüística tiene mayor grado de fiabilidad y aporta la necesaria reproductibilidad.

REFERENCIAS

- Aurrekoetxea, G. (2022). Métodos de la dialectología cuantitativa. *EliES (Estudios de Lingüística del Español)* 45, 43-69 [https://infoling.org/elies/]
- Aurrekoetxea, G., Gaminde, I., Ormaetxea, J. L., Videgain, X. (2019), *Euskalkien sailkapen berria*. UPV/EHU.
- Aurrekoetxea, G., Arandia, A., Camino, M., & Etchebest, X. (2017), “Dialectología perceptual del vasco: aplicación del Little Arrow Method”. In Florentino Paredes García, Ana M. Cestero Mancera e Isabel Molina Martos (eds.), *Investigaciones actuales en Lingüística*. Vol. V: Sobre variación geolectal y sociolingüística, Alcalá de Henares: Universidad de Alcalá, Servicio de Publicaciones, 51-66.
- Bloomfield, L. (1933). *Language*. Dialect geography, 321-345.
- Bonaparte, L. L. (1868). *Carte des sept provinces basques montrant la délimitation actuelle de l’euscara et sa division en dialectes, sous dialectes et variétés*. Stanford’s Geographical Establishment.
- Brun-Trigaud, G. (2023). “Gallo-Roman dialects classifications”, in G. Aurrekoetxea, A. Ensuzza, J. Scofic & H. Van de Velde (eds.) *Dialects classifications in European Languages*. *Dialectología Special issue 10*, 127-154.
- Calaza Díaz, L., Suárez S., Crujeiras, R. M., Rodríguez, A., Sousa, X., & Ríos, J. R. (2016). A method for processing perceptual dialectology data. *Actas xii congreso galego de Estatística e investigación de operacións*, Lugo, 282-291.
- Caro Baroja, J. (1945). *Materiales para una historia de la lengua vasca en su relación con la latina*. Universidad de Salamanca.
- Chambers & Trudgill, P. (2004). *Dialectology*. CUP.
- Cugno, F. (2023). “Italian dialect classifications”, in G. Aurrekoetxea, A. Ensuzza, J. Scofic & H. Van de Velde (eds.) *Dialects classifications in European Languages*. *Dialectología Special issue 10*, 195-228.
- Díaz-Campos, M., & Navarro-Galisteo, I. (2009). “Perceptual Categorization of Dialect Variation in Spanish”. In Joseph Collentine et al. (eds.), *selected Proceeding of the 11th Hispanic linguistics symposium*, Somerville, MC: Cascadilla Proceeding Project, 179-195.
- Ensuzza, A., & Unamuno, L. (2023). “Basque dialects classifications”, in G. Aurrekoetxea, A. Ensuzza, J. Scofic & H. Van de Velde (eds.) *Dialects classifications in European Languages*. *Dialectología Special issue 10*, 69-91.
- Francis, W. N. (1983). *Dialectology: An Introduction*. University of Alabama Press.
- Gauchat, Louis (1903). *Gibt es Mundartengrenzen?*. *Archiv für das Stuëium der neueren Sprachens*, 1903.
- Hearinga, W. (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*. Groningen University.
- Hearinga W., & J. Nerbonne (2001). *Dialect areas and dialect continua*. *Language Variation and Change* 13(3), 375-400 .
- Hinskens, F. (2018). “Dialectology and Formal Linguistic Theory: The Blind Man and the Lame”, in Charles Boberg, John Nerbonne and Dominic Watt (eds.). *The Handbook of Dialectology*, First Edition. John Wiley & Sons, 88-105.
- Inoue, F. (1996). *Computational Dialectology*. *Area and Culture Studies* 52, 67-102.

- Kessler, B. (1995), Computational Dialectology in Irish Gaelic”, EACL '95: Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics. March 1995, 60–66. [https://doi.org/10.3115/976973.976983]
- Mañaricua, A. (1964). Obispos de Álava, Guipúzcoa y Vizcaya hasta fines del s. X. Vitoria-Gasteiz.
- Mitxelena, K. (1976). La fragmentación dialectal: conocimientos y conjeturas. REL VI, 309-324.
- Mitxelena, K. (1981). Lengua común y dialectos vascos. Anuario del Seminario Julio Urkijo-ASJU 15, 289-313.
- Montgomery, C. (2007). Northern English dialects: A Perceptual Approach. University of Sheffield.
- Montgomery, C. (2011). Starburst charts: Methods for investigating the geographical perception of and attitudes towards speech samples. Studies in Variation, contacts and change in English, 7.
- Montgomery, C., & Beal, J. (2011). Perceptual Dialectology. Cambridge University Press, 121-148.
- Moreno Fernández, J., & Moreno Fernández, F. (1999). “Madrid perceptions of regional varieties in Spain”, Daniel Long y Dennis Preston (eds.), Handbook of Perceptual dialectology, Amsterdam: John Benjamins, 295-320.
- Nerbonne, J. (2008). “Variation in the Aggregate: An Alternative Perspective for Variationist Linguistics”. In Kees Dekker, Alasdair MacDonald and Hermann Niebaum (eds.) Northern Voices: Essays on Old Germanic and Related Topics offered to Professor Tette Hofstra Leuven: Peeters, 365-382.
- Nerbonne, J., & Kretschmar, W. (2006). Progress in Dialectometry: Toward Explanation. Literary and Linguistic Computing 21, No. 4, 387-397.
- Nerbonne, J., Kleiweg, P., Heeringa, W., & Manni, F. (2007). Projecting Dialect Distances to Geography: Bootstrap Clustering vs. Noisy Clustering. Annual Conference of the Gesellschaft für Klassifikation, 647-654.
- Nesse, A., & Høyland, B. (2023). Norwegian dialect classifications. Dialectologia, Monográfico X, 255-298.
- Newton, B. (1972). The Generative Interpretation of Dialect: A Study of Modern Greek Phonology. Cambridge University Press.
- Preston, D. R. (1989). Perceptual dialectology: nonlinguists’ views of areal linguistics. Foris Publications.
- Preston, D. R. (ed.) (1999). Handbook of Perceptual dialectology, Vol 1, Amsterdam: John Benjamins.
- Preston, D. R. (2002). “Perceptual dialectology: Aims, methods, findings”, Jan Berns y Jaap van Marle (eds.), Present-day dialectology: Problems and findings, Berlin: Mouton de Gruyter, pp. 57-104.
- Preston, D. R. (2010a). Language, people, salience, space: Perceptual dialectology and language regard. Dialectologia, 5, 87-131.
- Preston, D. R. (2010b). “Perceptual dialectology in the 21st century”, Ada C. Anders, Markus Hundt y Alexander Lasch (eds.), Perceptual dialectology. neue Wege der dialektologie, Berlin: Mouton de Gruyter, 1-30.
- Schuchardt, H. (1870). Über die klassifikation der Eomanischen mundarten. Graz.
- Szmrecsanyi, B. (2008). Corpus-based dialectometry: aggregate morphosyntactic variability in British English dialects. International Journal of Humanistic Arts Computer, 2, 279-296.
- Szmrecsanyi, B. (2011). Corpus-based dialectometry – a methodological sketch. Corpora, 6, 45-76.
- Zuazo, K. (1998). Euskalkiak, gaur. Fontes linguae Vasconum 78, 191-233.
- Zuazo, K. (2019). Standard Basque and Its Dialects. Routledge.