

Influencia del euskera estándar entre dos generaciones, alfabetizada y no alfabetizada, del valle de Arratia

Euskara estandarraren eragina Arratia bailarako bi belanalditan, alfabetizatu eta ez alfabetizatuan

Influence of literacy on two generations, literate and non-literate generation, of Arratia valley

Aitor Iglesias Chaves

UPV/EHU

aitor.iglesias@ehu.eus

<https://orcid.org/0000-0001-5381-5957>

Juan Abasolo Isasa

UPV/EHU

juan.abasolo@ehu.eus

<https://orcid.org/0000-0002-1911-4118>

Aintzane Etxebarria Lejarreta

UPV/EHU

aintzane.etxebarria@ehu.eus

<https://orcid.org/0000-0001-8418-6276>

Ariane Ensunza Aldamizetxebarria

ariane.ensunza@ehu.eus

<https://orcid.org/0000-0003-0908-7467>

EUDIA ikerketa-taldea

Recibido / Noiz jaso den: 01/4/2020

Aceptado / Noiz onartu den: 05/05/2020

Resumen

El objetivo del trabajo es comprobar la influencia de la alfabetización en el habla habitual en la zona vascohablante del valle de Arratia. Para ello, hemos analizado la distancia lingüística entre dos generaciones de dicho valle respecto al euskera estándar: una generación no alfabetizada en euskera (45-50 años); y la otra, que sí está alfabetizada (20-25 años).

La metodología para realizar el análisis de los datos se divide en tres pasos: la estandarización de los datos, el cálculo de la distancia lingüística; y, por último, un análisis de correspondencias múltiples (ACM). Con los datos estandarizados hemos realizado dos análisis: por una parte se ha calculado la distancia lingüística RIV mediante el programa Diatch y, en segundo lugar, se ha realizado un análisis multivariante para datos categóricos mediante FactoMineR programado en lenguaje R.

Los resultados nos muestran que sí existe un leve movimiento en las hablas de las generaciones alfabetizadas hacia la utilización de formas propias del estándar que son ajenas a las hablas de gente no escolarizada ni alfabetizada en idioma vasco.

Palabras clave

Educación, alfabetización, estándar.

Sumario

1. INTRODUCCIÓN. 2. MARCO TEÓRICO. 3. CORPUS. 4. METODOLOGÍA. 5. RESULTADOS. 6. CONCLUSIONES. REFERENCIAS.

Laburpena. Lan honen helburua da Arratiako bailarako euskararen aldaeran alfabetatzeak izan duen eragina egiaztatzea. Horretarako, euskara estandarrekiko distantzia linguistikoa aztertu da bailarako bi belaunalditan: bata euskaraz alfabetatu gabekoa (45-50 urte); bestea euskaraz alfabetatua (20-25 urte).

Datuen analisirako erabilitako metodologia hiru urratsetan dago banatua: datuen estandarizazioa, distantzia linguistikoaren kalkulua; eta korrespondentzia multipleen analisia. Behin datuak estandarizatu direnean, bi analisi egin dira: batetik, RIV distantzia kalkulatu da Diattech programarekin; eta, bestetik, datu kategorikoentzako aldagai anitzeko analisi bat egin da R lengoian programatutako FactoMineR bidez.

Datuek erakusten dute badagoela joera belaunaldi alfabetizatuen artean estandarrekoak diren formak erabiltzearena, herriko hizkeran ezezagunak direnak herritar alfabetatugabeengan.

Gako hitzak. Hezkuntza, alfabetizatzea, estandarra.

Abstract. The objective of this work is to check the influence of literacy on the vernacular of the Basque-speaking area of the Arratia valley. To this end, we have analyzed the linguistic distance between two generations of this valley with respect to standard Basque: a non-Basque-literate generation (45-50); and the other, which is literate in Basque (20-25 years).

The methodology for analyzing the data is divided into three steps: the standardization of the data, the calculation of the linguistic distance; and, finally, a multiple correspondence analysis (MCA). With the standardized data we have performed two analyzes: on the one hand the RIV linguistic distance has been calculated using the Diattech program and, secondly, a multivariate analysis has been performed for categorical data using FactoMineR programmed in the R language.

The results show us that there is, in fact, a slight movement in the vernacular of the literate generations towards the use of forms typical of the standard that are unknown in non-basque-literate generation's vernacular.

Keywords. Education, literacy, standard.

1. Introducción

El objetivo del trabajo es analizar la influencia de la alfabetización en el habla habitual en la zona vascohablante del valle de Arratia cercana a la Facultad de Educación de Bilbao puesto que ésta es una de las zonas de donde provienen gran parte de nuestro alumnado que sí utiliza el euskera como lengua de comunicación habitual. Para ello, hemos tenido que calcular la distancia lingüística entre dos generaciones de dicho valle respecto al euskera estándar. Una generación no alfabetizada en euskera, 45-50 años, y la otra, que sí está alfabetizada, 20-25 años, ya que en la escuela ha estudiado en euskera.

Trabajamos con la hipótesis de que habrá más distancia lingüística respecto al euskera estándar con la generación de mayor edad, puesto que creemos que la alfabetización influirá en la generación más joven haciendo que su habla se acerque al euskera estándar. Pero, por otro lado, queremos saber qué formas son las que se han introducido en el habla dialectal.

Para realizar este estudio hemos utilizado el corpus EAS (Euskararen atlas soziolinguistiko-dialektologikoa) (Aurrekotxea y Ormaetxea, 2006) recogido por

el grupo de investigación EUDIA de la UPV/EHU. De éste hemos utilizado las cuatro localidades de Arratia (Lemoa, Zeanuri, Dima y Zeberio). El corpus consta de un cuestionario de 202 preguntas, lo que ha permitido la elaboración de una base de datos con 1818 ítems. Los datos recogidos en el atlas han debido de ser tratados previamente para poder compararlos con el euskera estándar, ya que estaban transcritos de forma fonética.

Con los datos estandarizados se han realizado dos análisis. Por una parte, se ha calculado la distancia lingüística RIV y, en segundo lugar, se ha realizado un análisis multivariante para datos categóricos, Análisis de Concordancias Múltiples. En el primer caso, para hallar la distancia RIV, se ha calculado la matriz de distancias entre las hablas analizadas con el programa Diatech (Aurrekoetxea, Iglesias, Santander y Usobiaga, 2015b). El resultado del análisis de distancias se presenta en matriz y en agrupamientos clusters mediante dendrograma. En el segundo caso hemos utilizado FactoMineR (Husson, Josse, Le y Mazet, 2018), programado en lenguaje R, para hacer un análisis multidimensional categórico, siguiendo los planteamientos del análisis de componentes principales (Aurrekoetxea, Fernández-Aguirre, Modroño y Rubio, 2016) y replanteando el escalamiento multidimensional desarrollado por Embleton (1993). Se presentan los datos desde perspectivas monodimensionales mediante diagramas de puntos, *dotchart*, bidimensionales mediante biplot y una proyección tridimensional.

Los resultados nos muestran que la hipótesis se cumple parcialmente. Sí existe un leve movimiento en las hablas de las generaciones alfabetizadas hacia la utilización de formas propias del estándar que son ajenas a las hablas de gente no escolarizada ni alfabetizada en idioma vasco.

Este trabajo además de la introducción consta de cinco partes. En primer lugar, se centra el marco teórico en el que se va a trabajar; después, se presenta el corpus lingüístico que se ha utilizado; en tercer lugar, la metodología con la que se ha llevado a cabo este examen; luego los resultados obtenidos; y para finalizar las conclusiones del estudio además de las referencias bibliográficas.

2. Marco teórico

De todos es sabido (Trudgill, 2003) que las lenguas están en continuo cambio y que si una lengua deja de cambiar, desaparece. La lengua es cambiante, cualquier hablante sabe que su lengua no es uniforme y que no siempre se usa igual (Martínez, 2010). «La lengua es variable y se manifiesta de modo variable» (Moreno, 1999). Estos cambios pueden suceder a lo largo de la historia, en diacronía, pero también pueden ser sincrónicos. Por otro lado, los cambios lingüísticos suceden por las condiciones de dos agentes. Un cambio puede venir de fuera, ser

un préstamo; o puede surgir dentro de la propia lengua, una analogía (Zarraña et al., 2010; Gaminde, Romero y Legarra, 2012).

En este trabajo, como mencionaremos más adelante, se recogen cambios que se producen en la morfología nominal, la morfología verbal, la sintaxis y el léxico. Estos cambios suceden en cinco niveles (Bulot y Blanchet, 2013):

Diacrónico: En este nivel la variación lingüística se produce a lo largo del tiempo y se conocen como *cronolecto* a las diferentes variantes.

- Diatópico: La variación diatópica se centra en las diferencias que se producen geográficamente entre hablantes de una misma lengua, y sus diferentes variantes se denominan *dialectos*.
- Diastrático: En el nivel diastrático o social las diferencias que surgen entre hablantes de una lengua por su pertenencia a un grupo social son conocidas como *sociolectos*.
- Diafásico: en este nivel se producen las variaciones debidas a los diferentes contextos en los que se utiliza el habla, son los diferentes registros que usa un hablante.
- Diagénico: En algunos contextos sociales los hombres y las mujeres no usan las mismas formas lingüísticas, son los *sexolectos*.

La mayoría de los trabajos que se han hecho en euskera han sido a nivel diatópico. Desde los primeros trabajos de Bonaparte (1863 y 1869) han sido y son innumerables los estudios que han examinado la variabilidad geográfica. Es en estos últimos años donde se observa un aumento en el estudio de la variabilidad sociolingüística con numerosos trabajos (Aurrekoetxea, Gandarias, Gaminde e Iglesias, 2014; Ensunza, 2016; Epelde y Oyharçabal, 2010; Gaminde, Etxebarria, Garay y Romero, 2015; Gaminde, Olalde, Etxebarria, Eguskiza y Gaminde, 2017; Ibarra, 2011; Iglesias y Unamuno, 2017; Lantto, 2012, 2015; Unamuno, Ensunza y Ormaetxea, 2012; Unamuno, Eguskiza e Iglesias, 2019).

En este trabajo vamos a examinar dos variables: una a nivel diatópico y otra a nivel diastrático. A nivel diatópico vamos a medir la distancia lingüística que hay entre varias poblaciones del valle de Arratia (Lemoa, Zeanuri, Dima y Zeberio). A nivel diastrático, mayormente, han sido objeto de estudio cuatro variables de origen social: edad, género, clase social y estudios. En nuestro caso vamos a tener en cuenta dos, la edad y los estudios, que, en realidad, y por la peculiaridad de la situación social de los sujetos de este estudio, están íntimamente relacionadas. En este trabajo se recogen datos de dos generaciones: una generación de 20-25 y otra de 45-50 años. La generación de 20-25 años ha estudiado en euskera y está alfabetizada en este idioma; sin embargo, la generación de 45-50 años no lo pudo hacer debido a la política lingüística que se llevó a cabo en el país, donde estaba prohibido estudiar en euskera. Es por esta razón por la cual edad y estudios, alfabetización en este caso, están estrechamente relacionadas. Para saber cuáles son las variables sociales que influyen en la lengua hay que conocer la organización

de la sociedad (Zarraña et al., 2010), ya que esta es un producto de las relaciones sociales de sus hablantes y estas características pueden intervenir como factores que afectan al uso de la lengua y, eventualmente, al cambio lingüístico (Hualde, Olarrea, Escobar y Travis, 2009). Como apunta Gaminde (2010), para entender los cambios lingüísticos y la situación de cada generación en el País Vasco, sobre todo, hay que tener en cuenta dos condicionantes cronológicos. El primero es el haber nacido antes o después de la Guerra Civil (en este trabajo las dos generaciones han nacido después de la guerra). Y el segundo es la creación del euskera batua o euskera unificado.

Desde los primeros textos recogidos en euskera hasta que comenzó a desarrollar su registro estándar el euskera ha sufrido una fuerte dialectalización. La falta de un modelo o fuerza unificadora ha hecho que a lo largo de los siglos cada modelo dialectal se haya ido alejando aún más de los otros. Esta situación ha provocado que haya una gran variabilidad geolingüística en esta lengua. Para la supervivencia de una lengua es imprescindible que haya una unificación y un proceso de normalización. Con la creación del euskera batua o unificado en la reunión de La Real Academia de la Lengua Vasca Euskaltzaindia el año 1968 en Arantzazu (Euskaltzaindia, 1968) se logra una lengua de prestigio que pone en marcha unas fuerzas unificadoras de los dialectos que hasta entonces apenas se había logrado. Aunque también surgieran controversias al respecto (Zuazo 2000 y 2005), son indiscutibles las ventajas que ha traído el euskera unificado. Aunque al principio este euskera se utilizaba principalmente a nivel escrito, este se ha ido extendiendo también a nivel oral y hoy en día se utiliza en casi todos los ámbitos públicos y también poco a poco se está introduciendo en el ámbito privado. En pocos años su influencia ha ido en aumento en el habla coloquial dialectal (Aurrekoetxea, 2004, 2006; Aurrekoetxea, Gaminde, Gandarias e Iglesias, 2015a; González, 2001; Maia y Larrea, 2010; Ensunza, 2018) y esta, sobre todo, ha sido marcada por su implantación en la educación formal. Es esta influencia la que queremos medir en este trabajo examinando dos generaciones: una alfabetizada en euskera y la otra no.

Son muchos factores los que pueden influir en la convergencia lingüística en los sitios que conviven diferentes variedades de la misma lengua: el prestigio, la enseñanza en la escuela, la utilización en los organismos oficiales (ayuntamiento, mancomunidad, gobierno, sanidad...), los medios de comunicación, etc. Normalmente la variedad estándar de una lengua se suele usar en todos los casos mencionados y suele ser de alto prestigio (Hernández-Campoy, 2011). En nuestro caso concreto, es la variedad estándar la que está presente en la enseñanza y en todas las comunicaciones oficiales que siempre son bilingües en castellano y euskera estándar. A excepción de algunos medios de comunicación locales, todos los demás usan también la lengua estándar. Referente al prestigio, Amorrortu (2001), teniendo en cuenta la opinión del alumnado universitario, no ve en peli-

gro la desaparición del dialecto vizcaíno ya que la valoración por parte de éste es positiva, tiene prestigio.

Para poder comparar y medir la influencia del estándar es imprescindible conocer detalladamente las variedades no estándares. En el caso de Arratia, afortunadamente, tenemos dos trabajos que recogen detalladamente las hablas de las localidades de Zeberio (Etxebarria, 1991) e Igorre (Iglesias, 2014), además de otros muchos trabajos que recogen datos lingüísticos más concretos (Etxebarria, 1995, 2016; Gaminde, Hernáez, Etxebarria y Gandarias, 1995; Hualde, 1992; Iglesias, 2008). Desgraciadamente, no todas las variedades lingüísticas han sido recogidas tan exhaustivamente, por lo que su convergencia es más difícil de estudiar por la falta de datos.

Todas las lenguas están en contacto con otras lenguas, y, por tanto, se influyen entre ellas. Los diferentes dialectos del euskera tienen como lenguas vecinas el español, francés y occitano, y tampoco puede escapar a la influencia de idiomas como el inglés.

La influencia ocurre tanto entre diferentes lenguas como dentro de la misma lengua, ya que las variedades del mismo idioma, así como el estándar (Auer e Hinskens, 1996), instan a cambios intralingüísticos. En este punto hay que mencionar dos términos: diglosia y diaglosia. Se dice que una lengua está en situación de *diglosia* cuando convive con otra lengua en un mismo territorio, y ese segundo idiomas tiene un dominio o preferencia sobre la primera (Ferguson, 1959). En el caso del euskera tenemos que distinguir dos territorios distintos, con el español y el francés como lenguas dominantes. La *diaglosia* (Bellman, 1998) es el desequilibrio entre variedades lingüísticas habladas en una comunidad bilingüe, una se considera variedad alta y la otra baja. La variedad alta habitualmente se utiliza en el lenguaje formal y la baja, por el contrario, en el lenguaje informal. Según Coyos (2010), en el caso del euskera, la diaglosia nos muestra cómo algunas características del estándar están penetrando en los diferentes dialectos, y la diglosia nos enseña la utilización estratificada de diferentes variantes de una misma lengua.

Entre la lengua estándar y los diferentes dialectos tenemos dos tipos de unificaciones: verticales y horizontales (Auer, 1998; Eguskiza, 2019). Las unificaciones *verticales* son las que suceden entre la lengua estándar y los dialectos. Esta convergencia surge debido a la fuerza centrípeta que ejerce la lengua de mayor prestigio, la de variedad alta. De este modo, los dialectos poco a poco hacen suyas características de la lengua estándar debido a su prestigio, y van perdiendo las suyas propias. A esta unificación también se la conoce como *convergencia dialectal monocéntrica* (Aurrekoetxea, 2006). A menudo, las unificaciones verticales acarrearán convergencias horizontales, ya que, al reducirse la distancia lingüística entre el estándar y los dialectos, también se reduce la distancia entre los propios dialectos al disminuir las formas que diferencian unos dialectos de otros (Auer y Hinskens 1996, Auer, Hinskens y Kerwill, 2005). Las convergencias *horizon-*

tales corresponden a las unificaciones entre diferentes dialectos. Este proceso de unificación, aunque como hemos mencionado pueda ser consecuencia de las relaciones entre la lengua estándar y los dialectos, se considera como un proceso entre dialectos que se encuentran en el mismo nivel. Según Aurrekotxea (2006), para que se produzcan procesos de convergencia y unificación además de haber una lengua estándar débil, tiene que haber una lengua policéntrica donde los dialectos deben tener sus zonas dinámicas propias y según el sentido que cojan estas zonas dinámicas surgirán convergencias o divergencias entre los dialectos. Estas convergencias pueden ser fuertes o débiles. Cuando la convergencia es fuerte puede llegar a hacer desaparecer el dialecto; y, cuando es débil, el dialecto solo recoge algunas características del estándar en detrimento de las suyas propias.

3. Corpus

Para realizar este estudio hemos utilizado el corpus EAS (*Euskararen Atlas Soziolingüistikoa* o *Atlas socio-geolingüístico del euskera*) (Aurrekotxea y Ormaetxea, 2006) recogido por el grupo de investigación EUDIA de la UPV/EHU entre los años 2006 y 2008. La recogida de este corpus tiene un doble objetivo: por una parte, realizar un examen sociolingüístico; y por otra, un examen geolingüístico. El objetivo de este corpus, desde esta doble visión sociolingüística y geolingüística, es estudiar y medir la influencia del euskera estándar en los últimos 40 años (Aurrekoetxea y Ormaetxea, 2006). Con este corpus podemos medir la variación lingüística mediante dos variables: la edad y el factor geográfico, ya que se ha recogido en 145 pueblos a hablantes de dos generaciones. Las dos generaciones examinadas en este trabajo comprenden las edades de 45-50 años y de 20-25 años. La generación de mayor edad, aunque no ha sido alfabetizada en euskera, ha visto cómo el euskera estándar se ha ido implantando a su alrededor, en diferentes aspectos de la vida cotidiana, pero sobre todo en la enseñanza de sus hijos e hijas. La generación más joven, sin embargo, ha recibido dos *input*. Por una parte, en el ámbito familiar han recibido como primera lengua su correspondiente variante dialectal; y, por otra parte, también han aprendido en la escuela el euskera estándar. Es, por tanto, que esta generación ha tenido que combinar estas dos variantes para hacer frente a las diferentes situaciones de comunicación a las que se han tenido que enfrentar. Las 145 localidades examinadas en el EAS son también localidades en las que se hicieron la encuesta del EHHA, *Euskararen Herrien Hizkuntza Atlasa* (Euskaltzaindia, 2008). Para nuestro estudio hemos utilizado las cuatro localidades de Arratia (Lemoa, Zeanuri, Dima y Zeberio), ya que este valle es uno de los más cercanos a la Facultad de Educación de Bilbao, donde se habla y se vive habitualmente en euskera, y es de allí de donde proviene gran parte del alumnado vascohablante activo.

La encuesta del proyecto EAS está basada en el cuestionario del EHHA (Aurrekoetxea y Videgáin, 1993). En total se recogen 202 características lingüísticas divididas del siguiente modo (Aurrekoetxea y Ormaetxea, 2006): 23 preguntas sobre la morfología nominal, 39 sobre la morfología verbal, 20 sobre la sintaxis y 120 sobre el léxico. Las cuatro localidades de Arratia recogen un total de 1818 ítems.

4. Metodología

La metodología de este trabajo se divide en tres pasos: la normalización o estandarización de los datos, el cálculo de la distancia lingüística; y, por último, un análisis de correspondencias múltiples (ACM).

El primer paso es la normalización o estandarización de los datos. Para poder obtener la distancia lingüística que hay entre los datos obtenidos en el EAS de varios hablantes de diferentes generaciones y localidades y el euskera estándar, tenemos que tener datos comparables, transcritos de la misma manera. Los datos del euskera estándar evidentemente están transcritos con las reglas de ortografía que rigen el idioma y que las marca la Real Academia de la Lengua Euskaltzaindia (Mitxelena, 1968); pero, sin embargo, los datos del EAS están transcritos fonéticamente. Para estandarizar estos datos se ha actuado de la siguiente manera: en las respuestas referidas al léxico y a la sintaxis se han obviado todas las variantes fonéticas. Así, respuestas como /okela/ y /okola/ se han estandarizado en una misma forma «okela» ya que lexicalmente son iguales. En las respuestas referidas a la morfología nominal y verbal nos hemos fijado exclusivamente en los morfemas que se buscan en cada pregunta. De este modo, respuestas como /itʃaʒórantza/ e /itʃoʒórantza/ se han estandarizado como «itsasorantza», ya que lo pertinente de esta pregunta es el morfema *-rantza* para poder compararlo con el morfema *-rantz* de la palabra estándar «itsasorantz».

El segundo paso es calcular la distancia lingüística que hay entre el euskera estándar y cada localidad y generación. Para ello es necesario crear una matriz de distancia y aplicar los algoritmos adecuados a nuestros datos. Este método se empezó a aplicar en la segunda mitad del siglo XX y se ha ido desarrollando hasta hoy en día con ayuda de la informática. Sin duda, primeramente, tenemos que mencionar a Jean Séguy (1973), autor del *Atlas linguistique et ethnographique de la Gascogne* y del término *dialectometría*, que como bien explica Goebel (2010), que buscó una posibilidad que con métodos cuantitativos le permitiera comprender de forma global esa variabilidad; es decir, desistiendo de muchos detalles cualitativos, pero desgraciadamente no pudo desarrollar este empeño debido a su temprano fallecimiento. Fue Goebel (1984) quien sí llevó a cabo este proyecto publicando sus principios teóricos y metodológicos; y, posteriormente,

desarrollando el programa informático Visual Dialectometry (VDM)¹. Este programa que trabaja con matrices de datos ofrece varios algoritmos de evaluación. Las medidas de similaridad RIW, GIW y DEM calculan en la matriz de datos la similaridad entre las distintas localidades. También da la posibilidad de trabajar con los algoritmos de segmento MINMWMAX, MWMAX y MED, y otros varios cálculos (mínimo, máximo, desviación estándar, asimetría...). Hoy en día hay dos escuelas que siguen el trabajo de Goebel, la de Groningen y la de la UPV/EHU, dirigidas por John Nerbonne y Gotzon Aurrekoetxea, respectivamente. El grupo de Groningen ha desarrollado la aplicación web Gabmap² que utiliza la distancia *Levenshtein* ya que esta aplicación está preparada para transcripciones fonéticas. La distancia *Levenshtein* es el número mínimo de operaciones requeridas para transformar una cadena de caracteres en otra. Por su parte, el grupo de investigación EUDIA de la UPV/EHU ha desarrollado el software Diatech³, que además de la distancia *Levenshtein*, también da la oportunidad de usar las distancias RIV (Relative Index Value) y WIV (Weighted Identity Value). La distancia RIV es una distancia de igualdad o desigualdad absoluta, es decir, si los datos que se comparan son exactamente iguales o diferentes. Esta distancia se utiliza normalmente cuando tenemos los datos estandarizados. Para medir las distancias lingüísticas es necesario tener en cuenta cómo están transcritos los datos con los que se va a trabajar. Como hemos mencionado anteriormente, nuestros datos han sido estandarizados y por esa razón vamos a usar la medida de distancia RIV. Por último, tenemos que tener en cuenta qué medida tomar para el proceso de agrupación si queremos crear un clúster jerárquico. Tenemos que elegir el algoritmo de clasificación de distancia más adecuado. Por ejemplo, el programa Diatech da tres opciones de algoritmos: *ward*, *average* y *complete*. Para crear nuestro clúster vamos a usar el algoritmo *ward* ya que es el que muestra una varianza interna mínima (Ward, 1963).

El tercer y último paso es un análisis de correspondencias múltiples (ACM) en inglés *Multiple Correspondences Analysis* (MCA). El análisis de correspondencias es un método de análisis de datos que representa gráficamente tablas de datos (Greenacre, 2008). Técnicamente, el ACM es la aplicación de análisis de correspondencias simples (ACS) al estudio de tablas lógicas donde se considera un n.º cualquiera de variables cualitativas. En un análisis de correspondencias tenemos una representación gráfica mediante un diagrama de dispersión, que no es sino un mapa de datos. En nuestro trabajo con los datos ya estandarizados crearemos una matriz para poder realizar un ACM sobre una tabla disyuntiva

¹ <http://ald.sbg.ac.at/dm/Engl/VDM/default.htm>

² https://gabmap.nl/?page_id=378

³ <http://eudia.ehu.es/diatech/index>

completa (Lebart, Morineau y Warwick, 1984). Mediante este análisis obtendremos las formas lexicales, morfológicas y sintácticas que se han importado del euskera estándar y las que no, las que se mantienen en su variante dialectal pese a la influencia del estándar.

5. Resultados

Una vez normalizados, estandarizados, los datos, en este punto ofreceremos primeramente los resultados obtenidos con el cálculo de la distancia lingüística utilizando la medida RIV y el clúster jerárquico cuyo agrupamiento ha sido obtenido con el método de varianza mínima de Ward; y, en segundo lugar, facilitaremos los resultados obtenidos con el análisis de correspondencias múltiples.

En la Tabla 1 podemos ver la matriz de distancias obtenidas con la medida RIV. En ella aparecen el euskera estándar y las dos generaciones examinadas en las cuatro localidades de Arratia, un total de nueve elementos entre los cuales se ha obtenido la distancia lingüística.

Tabla 1. Matriz de distancias lingüísticas obtenidas con la medida RIV

	Dima joven	Dima mayor	Euskera estándar	Lemoa joven	Lemoa mayor	Zeanuri joven	Zeanuri mayor	Zeberio joven	Zeberio mayor
Dima joven	0	-	-	-	-	-	-	-	-
Dima mayor	29.82	0	-	-	-	-	-	-	-
Euskera estándar	58.07	65.03	0	-	-	-	-	-	-
Lemoa joven	21.17	33.42	56.65	0	-	-	-	-	-
Lemoa mayor	28.62	26.83	60.70	26.16	0	-	-	-	-
Zeanuri joven	24.28	30.51	55.03	22.00	27.46	0	-	-	-
Zeanuri mayor	31.80	26.58	60.82	29.30	27.93	27.97	0	-	-
Zeberio joven	32.33	31.54	59.65	30.82	31.88	26.11	30.73	0	-
Zeberio mayor	30.80	28.31	62.39	29.65	31.54	28.63	34.55	29.57	0

Como nos demuestra la matriz de distancias (Tabla 1), las distancias lingüísticas más grandes son con el euskera estándar, marcadas en color azul, siendo la más alta la que hay entre la generación mayor de Dima y el estándar (65,03), seguida por la de la generación mayor de Zeberio (62,39) y después la generación mayor de Zeanuri (60,82) y por último la generación mayor de Lemoa (60,70). Las generaciones jóvenes presentan más proximidad al euskera estándar que las

de sus mayores, aunque estas distancias también sean muy altas. Sería la generación joven de Zeanuri la que más próxima estaría al euskera estándar (55,03). Dejando de lado las comparaciones con el euskera estándar, los dos sujetos que menor distancia lingüística presentan son los jóvenes de Dima y Lemoa (21,17) seguidos muy de cerca por los jóvenes de Zeanuri y Lemoa (22,00).

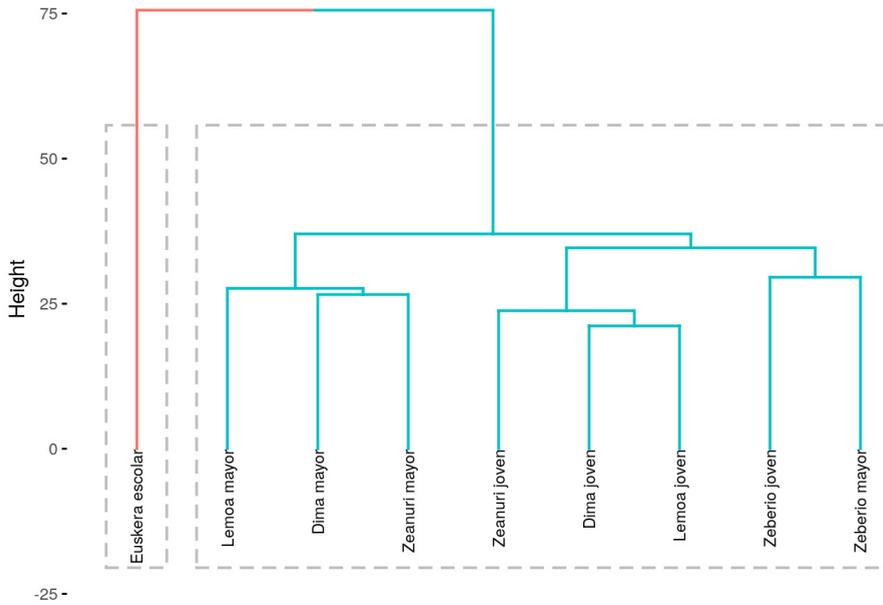


Figura 1: Dendrograma obtenido por el método de Ward.

En la Figura 1 vemos el dendrograma que representa el clúster jerárquico obtenido por el método de Ward, buscando construir agrupaciones que tengan la menor variabilidad interna. En él se ven claramente dos grupos: por una parte, el euskera estándar; y, por otra parte, las hablas del valle de Arratia. Si nos fijamos exclusivamente en las hablas de Arratia, podemos observar también dos grupos. En el dendrograma tenemos un grupo a la izquierda con las generaciones mayores de Lemoa, Dima y Zeanuri; y a la derecha están las generaciones jóvenes de las cuatro localidades más la generación mayor de Zeberio. Por lo tanto, casi hay una completa división entre generaciones exceptuando la mayor de Zeberio, que si se permite la expresión, parecería ser un mayor «moderno».

En el gráfico de puntos (Fig. 2) obtenemos la representación de la distancia lingüística donde se muestra la acumulación o tendencia, la variabilidad o disper-

sión y la forma de distribución. La principal ventaja de una gráfica de puntos es que muestra la distribución de los datos, conservando el valor individual de cada uno de ellos.

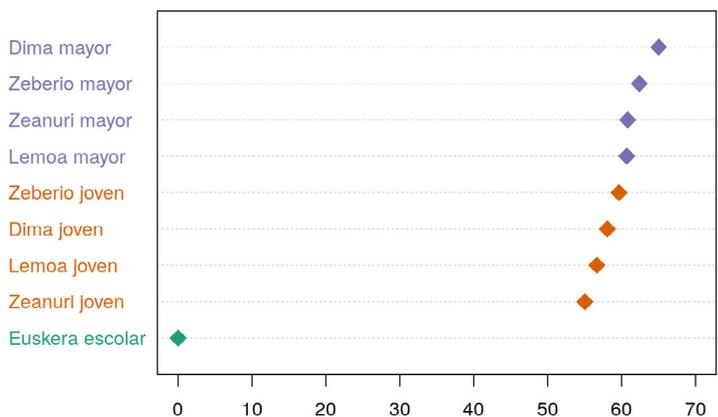


Figura 2: Representación de la distancia lingüística mediante un gráfico de puntos.

Podemos definir el ACM como el AC de la matriz binaria. Este análisis proporciona coordenadas para las 202 filas correspondientes a las preguntas realizadas y las 9 columnas correspondientes a los informantes. Este análisis, por tanto, tendrá ocho dimensiones ($n - 1$, es decir, $9 - 1$). Como sabemos, la suma de todas las dimensiones nos ofrece el 100% de la información. La primera dimensión, que es la que más información contiene (Greenace, 2008), nos ofrece un 32,98% (Fig. 3).

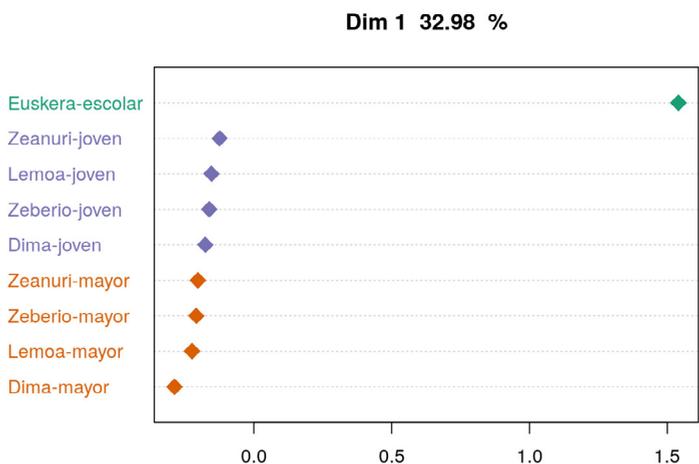


Figura 3: Representación de la 1ª dimensión.

En la Figura 4 tenemos dos mapas del ACD mostrando la primera y segunda dimensión. Como sabemos, la suma de todas las dimensiones nos ofrece el 100% de la información. La primera dimensión, explicando un 32,98% más la segunda con un 12,03%; en total, estos dos mapas del ACM nos explican 45,01% de la información total.

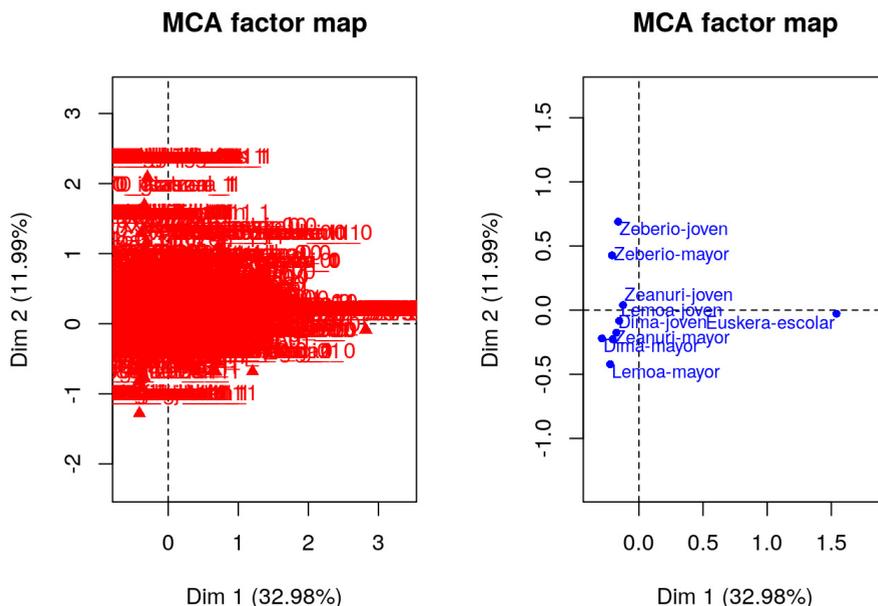


Figura 4: Mapa del ACM.

En el mapa de la izquierda aparecen representadas todas las formas que se han utilizado en este estudio distribuidas en función de las dos primeras dimensiones. En el de la derecha aparecen representadas las nueve hablas también distribuidas en función de las dos primeras dimensiones.

Con los dos mapas de la Figura 4 se ha obtenido un gráfico exploratorio biplot (Fig. 5) de las dos primeras dimensiones. Este biplot nos permite mostrar gráficamente la información de las filas (las respuestas obtenidas) y las columnas (informantes) de nuestra matriz de datos multivariantes. Los datos representados en este gráfico cuanto más alejados del 0 aparecen más explican las diferencias.

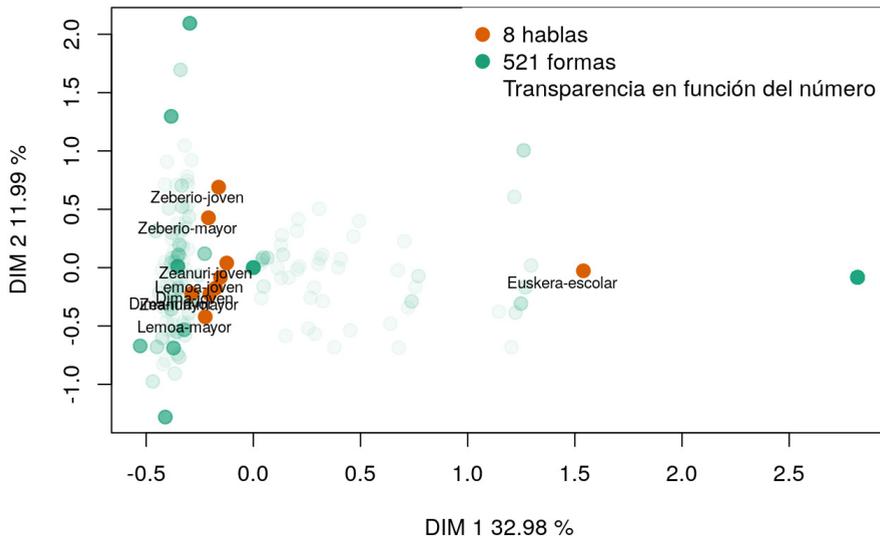


Figura 5: Gráfico exploratorio biplot de las dos primeras dimensiones.

Además de un biplot, también tenemos la opción de ver la variabilidad de los datos explicada por las tres primeras dimensiones (Fig. 6).

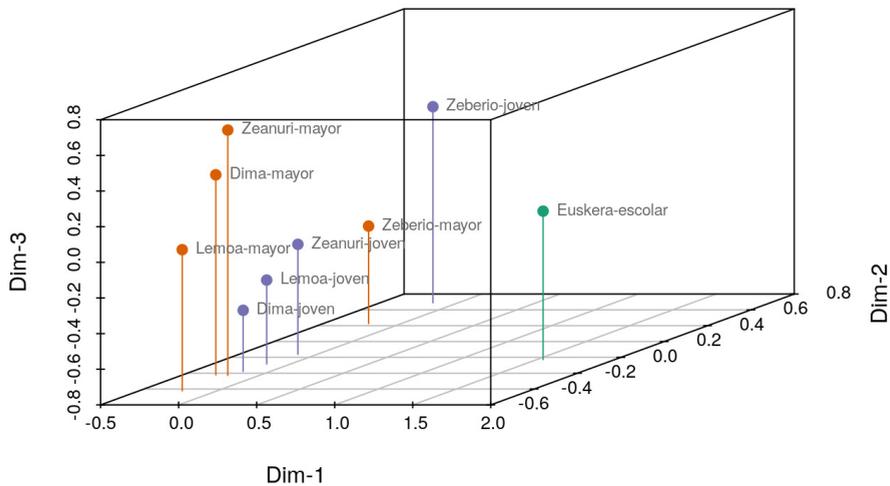


Figura 6: Las tres primeras dimensiones del ACM de Arratia y euskera estándar.

Las tres dimensiones del ACM nos explican el porcentaje de la variabilidad de los datos en un 56,55% y nos ofrecen un gráfico tridimensional.



Figura 7: Aportación de las formas variables a la 1ª dimensión.

El ACM nos permite discernir qué formas del estándar no se utilizan en Arratia, qué formas del estándar se están difundiendo en Arratia, cuáles son las formas que solo se utilizan en Arratia y sus diferentes niveles de extensión entre las diferentes localidades y generaciones; y por último, las formas que comparten el estándar y las hablas de Arratia (Fig. 7). Atendiendo al objetivo de este trabajo, la influencia del estándar, podemos obtener una lista (además de las otras menciones anteriormente) de las formas del estándar preguntadas en la encuesta del EAS que ya se han introducido en el habla de Arratia sustituyendo a las propias: *abuxtu, ahal + nor, alfer, berdin + NorNori, ditu, errauts, errieta egin, gazta, gurin, haragi, hasperen, inbidia, janarazi, jaurti, korrika, logela, nekatu, nindoan, odoloste, ostadar, pospolo, uholde, urratu, zaldar, zaunka egin, zenuen, zerua, zeuk, xirimola, zituen, zuhaitz*. De las 202 características preguntadas en el EAS, éstas serían las formas estándares que han sustituido a las propias dialectales. Como se puede observar tenemos formas pertenecientes al léxico, otras pertenecientes a la sintaxis, a la morfología verbal y también a la morfología nominal.

6. Conclusiones

El objetivo del presente trabajo era comprobar la influencia del euskera estándar en dos generaciones del valle de Arratia que es uno de los lugares más cercanos a la Facultad de Educación de Bilbao, donde se utiliza como lengua habitual el euskera. Para ver esta influencia hemos utilizado dos técnicas. Primeramente, hemos calculado la distancia lingüística entre todas las hablas; y, posteriormente, hemos realizado un ACM para poder observar cuáles son las formas que el estándar ha prestado al habla dialectal.

El cálculo de las distancias lingüísticas nos ha corroborado la hipótesis que habíamos formulado; es decir, que el habla de las generaciones jóvenes se encuentra más cercana al estándar que el habla de las generaciones mayores. Como hemos mencionado, estas generaciones de jóvenes entre 20 y 25 años han sido alfabetizadas en euskera puesto que han realizado todos sus estudios en este idioma. Esta alfabetización es, sin duda, la clave de esta mayor influencia del estándar sobre el habla dialectal.

El ACM nos ha proporcionado, entre otras, la lista de las formas preguntadas en la encuesta del EAS que ha prestado el estándar al habla dialectal. En ella hemos comprobado que ha habido influencia en todas las áreas: en el léxico, en la sintaxis, la morfología verbal y también en la morfología nominal. Hay que tener en cuenta que en la encuesta del EAS no se preguntan el mismo número de ítems para el léxico, la morfología o la sintaxis, por esta razón, conviene obtener los porcentajes de influencia de estas formas que conforman la lista obtenida por el ACM. Así, hemos podido observar que respecto al léxico el estándar influye en un 18,48%; en la morfología nominal la influencia es del 8,69%; en la morfología verbal del 10,52% y en la sintaxis del 14,28%. Aunque se pueda apreciar una diferencia en estos porcentajes, aún es pronto para poder sacar un patrón y aseverar, por ejemplo, que la influencia del estándar es mayor en el léxico que en otros apartados de la gramática porque este trabajo se ha focalizado en una zona concreta y con una determinada encuesta, el EAS. En un futuro, para poder observar si en el resto de localidades de Euskal Herria se sigue un patrón en los préstamos del estándar, habría que aplicar este mismo método de trabajo en toda la encuesta; o, por lo menos, a las localidades que conforman un dialecto. Sin embargo, los datos obtenidos seguirían siendo los referidos a una encuesta en concreto, donde solo se realizan una serie de preguntas. Sin duda alguna, a falta de otra gran macroencuesta, estos deben ser los primeros pasos a seguir para poder analizar en profundidad la influencia del estándar en las hablas dialectales.

Referencias

- Amorrortu, E. (2001). Unibertsitate-ikasleen euskalki eta batuarekiko jarrerak. En *Euskalkia eta Hizkuntza* (pp. 61-80). Bilbao: Mendebalde Kultura Alkartea.
- Auer, P. (1998). Dialect Levelling and the Standard Varieties in Europe. *FL*, 32(1-2), 1-9.
- Auer, P. e Hinskens, F. (1996). The Convergence and Divergence of Dialects in Europe. New and not so New Developments in an Old Area. *Sociolingüística*, 10, 1-30.
- Auer, O., Hinskens, F. y Kerswill, P. (2005). *Dialect change: convergence and divergence in European languages*. Cambridge: Cambridge University Press.
- Aurrekoetxea, G. (2004). Estandar eta dialektoen arteko bateratze joerak (ikuspuntu teorikotik begirada bat). *Uztaro*, 50, 45-57.
- Aurrekoetxea, G. (2006). Hizkuntza estandarren eta dialektoen arteko bateratze joerak. *ASJU*, 40, 133-160.
- Aurrekoetxea, G. y Videgain, X. (1993). Euskal Herriko Hizkuntz Atlasa. Galdesorta. *Euskera*, 38, 529-647.
- Aurrekoetxea, G. y Ormaetxea, J. L. (2006). 'Euskararen atlas sozio-geolingüistikoa' ikerketa-proiektua. *Euskalingua*, 9, 157-163.
- Aurrekoetxea, G., Gandarias, L., Gaminde, I. e Iglesias, A. (2014). Variación prosódica en vasco: áreas acentuales. En Y. Congosto, M. L. Montero y A. Salvador (Eds.), *Fonética experimental, educación superior e investigación. III. Prosodia* (pp. 11-28). Madrid: Arco.
- Aurrekoetxea, G., Gaminde, I., Gandarias, L. e Iglesias, A. (2015a). Euskararen azen-tuera: hizkera tradizioaletatik euskara estandarera. En M. J. Ezeizabarrena y R. Gómez (Eds.), *Eridenen du zersaz kontenta: sailkideen omenaldia Henrike Knörr irakasleari (1947-2008)* (pp. 71-96). Bilbao: UPV-EHU.
- Aurrekoetxea, G., Iglesias, A., Santander, G. y Usobiaga, I. (2015b). Diatech muga dialektalak zehazteko tresna. En G. Aurrekoetxea, A. Romero y A. Etxebarria (Eds.), *Linguistic Variation in the Basque and Education I.* (pp. 132-144). Bilbao: UPV/EHU.
- Aurrekoetxea, G., Fernandez-Aguirre, K., Modroño, J. I. y Rubio, J. (2016). Principal Components and clustering on lexical distances from standard Basque to local varieties. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 3(1), 5-22.
- Bellmann, G. (1998). Between Base Dialect and Standard Language. *FL*, 32(1-2), 23-34.
- Bonaparte, L. L. (1863). *Carte des sept Provinces Basques*. London: Stanford's Geographical Stablishment.
- Bonaparte, L. L. (1869). *Le verbe basque en tableaux*. London: Strangeways & Walden. (Euskaltzaindia (reeds.) (1991). *Opera Omnia Vasconice*. Bilbao: Euskaltzaindia.
- Bulot, T. y Blanchet, P. (2013). *Une introduction à la sociolinguistique: pour l'étude des dynamiques de la langue française dans le monde*. Paris: Archives contemporaines.
- Coyos, J. B. (2010). Hizkuntza ukipena. En *Soziolinguistika eskuliburua* (pp. 135-184). Gasteiz: Eusko Jaurlaritza.
- Eguskiza, N. (2019). *Hizkuntza aldakortasun geo-soziolinguistikoa Arratian* (tesis doctoral). Leioa, UPV/EHU.

- Embleton, S. (1993). Multidimensional Scaling as a Dialectometrical Technique: Outline of a Research Project. En R. Köhler y B. B. Rieger (Eds.), *Contributions to Quantitative Linguistics* (pp. 267-276). Dordrecht: Kluwer.
- Ensunza, A. (2016). The sociolinguistic variation of palatalization: the case of Gernika-Lumo. *Dialectología*, 16, 71-91.
- Ensunza, A. (2018). The influence of Standard Basque in Busturialdea. Part 1: demonstrative pronouns. *Dialectología*, 22, 17-29.
- Epelde, I. y Oyharçabal, B. (2010). Perpaus barneko hizkuntza lerratzeak (HL) ekialdeko euskal elebidunetan. *ASJUren gehigarriak*, 52, 55-64.
- Etxebarria, J. M. (1991). *Zeberio haraneko euskararen azterketa etno-linguistikoa*. Euba: Ibaizabal.
- Etxebarria, J. M. (1995). *Gorbeia inguruko etno-ipuñ eta esaundak*. Bilbao: Labayru-BBK.
- Etxebarria, J. M. (2016). *Gorbeia inguruko etno-ipuñ eta esaundak II*. Bilbao: Iker.
- Euskaltzaindia (1968). Arantzazu-ko biltzarra. *Euskera*, 13, 137-265.
- Euskaltzaindia (2008). *Euskararen herri hizkeren atlasa I*. Bilbao: Euskaltzaindia.
- Ferguson, Ch. (1959). Diglossia. *Word*, 15, 325-340.
- Gaminde, I. (2010). Bizkaiko gazteen prosodiaz. Bilbao: Mendebalde Kultura Alkartea / Bizkaiko Foru Aldundia.
- Gaminde, I., Etxebarria, A., Garay, U. y Romero, A. (2015). La influencia de la variación diatópica y la lengua materna en la percepción de emociones en la lengua vasca. *Círculo de Lingüística Aplicada a la Comunicación*, 63, 152-173.
- Gaminde, I., Hernández, I., Etxeberria, P. y Gandarias, R. (1995). Zeanuriko intonazioaz. *Ele*, 16, 97-121.
- Gaminde, I., Olalde, A., Etxebarria, A., Eguskiza, N. y Gaminde, U. (2017). *Hizkuntza aldakortasuna Larrabetzun*. Bilbao: Larrabetzuko Udala.
- Gaminde, I., Romero, A. y Legarra, H. (2012). *Gramatika eta hizkuntz bariazioa Bermeon*. Bilbao: Erroteta.
- Goebel, H. (1984). *Dialektometrische Studien. Anhand italomannischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*. Tübinga: Niemeyer.
- Goebel, H. (2010). Introducción a los problemas y métodos según los principios de la escuela dialectométrica de Salzburgo. En G. Aurrekoetxea y J. L. Ormaetxea (Eds.), *Tools for linguistic variation* (3-39 pp.). Bilbao: ASJU LIII.
- González, S. (2001). Barietatea eta estandarra. *Enseiukarreen*, 15, 89-144.
- Greenacre, M. (2008). *La práctica del análisis de correspondencias*. Bilbao: Fundación BBVA.
- Hernández-Campoy, J. M. (2011). Variation and Identity in Spain. En M. Díaz-Campos (Ed.), *The Handbook of Hispanic Sociolinguistics* (pp. 704-727). Malden, MA: Wiley-Blackwell.
- Hualde, J. I. (1992). Notas sobre el sistema acentual de Zeberio. *ASJU*, 27, 767-776.
- Hualde, J. I., Olarrea, A., Escobar, A. M. y Travis, C. E. (2009). *Introducción a la lingüística hispánica*. Cambridge: Cambridge University Press.
- Husson, F., Josse, J., Le, S. y Mazet, J. (2018). FactoMineR: Multivariate Exploratory Data Analysis and Data Mining (Version 1.41). Recuperado de <https://CRAN.R-project.org/package=FactoMineR>

- Ibarra, O. (2011). *Nafar euskaldun gasteen ahozko diskurtsoa*. Iruñea: Universidad Pública de Navarra.
- Iglesias, A. (2008). Igorre accent. *Euskalingua*, 13, 16-26.
- Iglesias, A. (2014). *Igorreko hizkeraren asterketa dialektologikoa* (tesis doctoral). Leioa, UPV/EHU.
- Iglesias, A. y Unamuno, L. (2017). Hizkuntza bariazioa Dimako hizkeran. En A. Iglesias y A. Ensunza (Eds.), *Gotzon Aurrekoetxea lagunarterik hara* (pp. 223-234). Bilbao: UPV-EHU.
- Lantto, H. (2012). Code-switching, swearing and slang: The colloquial register of Basque in Greater Bilbao. *International Journal of Bilingualism*, 0-0: 1-16. doi 10.1177/1367006912457274
- Lantto, H. (2015). Conventionalized code-switching: Entrenched semantic-pragmatic patterns of a bilingual Basque-Spanish speech style. *International Journal of Bilingualism*, 19(6), 753-768.
- Lebart, L., Morineau, A. y Warwick, K. (1984). *Multivariate Descriptive Statistical Analysis*. New York: J. Wiley.
- Maia, J. y Larrea, K. (2010). Las relaciones entre la lengua estándar y los dialectos en el proceso de normalización de la lengua vasca. En M. R. Caballero y M. J. Pinar (Eds.), *Modos y formas de la comunicación humana* (pp. 627-634). Ciudad Real: Universidad de Castilla-La Mancha.
- Martínez, L. (2010). Hizkuntza ukipena. En Zarraga et al. (Eds.), *Soziolinguistika eskuliburua* (pp. 85-134). Vitoria-Gasteiz: Eusko Jaurlaritza.
- Mitxelena, L. (1968). Ortografía. *Euskera*, 13, 203-219.
- Moreno, F. (1999). *Lenguas para fines específicos (VI). Investigación y enseñanza*. Alcalá de Henares: Universidad de Alcalá.
- Séguy, J. (1973). La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de linguistique romane*, 37, 1-24.
- Trudgill, P. (2003). *A glossary of sociolinguistics*. Oxford: Oxford University Press.
- Unamuno, L., Ensunza, A. y Ormaetxea, J. L. (2012). EAS project: first overview of the syntax variation in the Basque language. En A. Pérez, X. Afonso, E. Carrilho y C. Magro (Eds.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*. Lisbon, 2011 (pp. 225-235). Lisboa: Centro de Linguística da Universidade de Lisboa.
- Unamuno, L., Eguskiza, N. e Iglesias, A. (2019). Hizkuntza-bariazioa Igorreko hizkeran. En A. Etxebarria, A. Iglesias, A. Leğarra y A. Romero (Eds.), *Traineru bete lagun: Iñaki Gaminde omenduz* (pp. 377-395). Bilbao: UPV/EHU.
- Ward, J. H. Jr. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58, 236-244.
- Zarraga, A., Coyos, J. B., Hernández, J. M., Joly, L., Larrea, I., Martínez, L., Uranga, B. y Bilbao, P. (2010). *Soziolinguistika eskuliburua*. Vitoria-Gasteiz: Eusko Jaurlaritza.
- Zuazo, K. (2000). *Euskararen sendabelarrak*. Irun: Alberdania.
- Zuazo, K. (2005). *Euskara batua: ezina ekinez egina*. Donostia: Elkar.

