

Clasificación de informantes bilingües precoces según su lengua materna y lengua de reproducción *

Classification of early bilingual informants
according to their mother language
and reproduction language

Irati De Pablo Delgado

Universidad del País Vasco,
Facultad de Educación de Bilbao,
Departamento de Didáctica
de la Lengua y la Literatura.

Miembro del equipo
de investigación EUDIA
irati.depablo@ehu.eus

Naia Eguskiza Sánchez

Universidad del País Vasco,
Facultad de Educación de Bilbao,
Departamento de Didáctica
de la Lengua y la Literatura.

Miembro del equipo
de investigación EUDIA

Iñaki Gaminde Terraza

Miembro del equipo
de investigación EUDIA

Recepción del artículo: 8 de febrero de 2019
Recepción de la versión revisada: 28 de agosto de 2019
Aceptación: 17 de septiembre de 2019
doi: 10.22201/enallr.01852647p.2019.70.874

Resumen

La finalidad de este trabajo es clasificar a jóvenes bilingües precoces en función del tiempo de reacción empleado desde la visualización de un estímulo mostrado hasta la emisión léxica correspondiente para observar las diferencias según su lengua materna (LM) y la lengua de reproducción (LR) empleada. Para ello, se han obtenido los datos de las emisiones léxicas de 40 alumnas de la Facultad de Educación de Bilbao a partir de un cuestionario de 34 estímulos habilitado en una secuencia de diapositivas, al que las informantes debían responder en el menor tiempo posible. Se ha realizado la clasificación de las informantes según los datos obtenidos mediante los siguientes métodos cuantitativos de conglomeración: método de conglomeración jerárquico, K-medias y el algoritmo FCM basado en la lógica difusa. En los resultados obtenidos se observa la clasificación de las informantes en grupos heterogéneos y diferencias significativas en las emisiones léxicas según la LM y la LR.

Palabras clave: léxico, método de conglomeración jerárquico, K-medias, algoritmo FCM

Abstract

The purpose of this work is to classify early bilingual young people according to the reaction time used from the visualization of a stimulus shown to the corresponding lexical emission and, with this, it is intended to observe the differences according to their mother tongue (LM) and the language of reproduction (LR) used. To this end, the data on the lexical emissions of 40 students of the Faculty of Education of Bilbao were obtained using a questionnaire of 34 stimuli through a sequence of slides, to which participants had to respond as quickly as possible. Participants were classified according to the data obtained through the following quantitative methods of conglomeration: hierarchical conglomeration method, K-means and the FCM algorithm based on fuzzy logic. In the results obtained, the participants were classified into heterogeneous groups and significant differences in lexical emissions according to LM and LR were observed.

Keywords: lexicon, hierarchical conglomeration method, K-means, FCM algorithm

1. Introducción

En la comunidad lingüística vasca han ocurrido cambios significativos en los últimos años. Dichos factores de cambio han sido, por un lado, la estandarización del euskera y, por otro, la adopción de los diversos modelos educativos bilingües dentro del marco formal y su posterior crecimiento. Estos hechos han derivado en el aumento de personas bilingües en la comunidad.

Las características de las personas bilingües son diversas, por lo que es posible hacer distinciones respecto al momento de adquisición de la lengua: los *bilingües precoces* (los informantes de la presente investigación) y los *bilingües tardíos*. Los bilingües precoces son hablantes que adquieren su lengua materna (LM) en el hogar y la segunda lengua, que llamaremos *lengua de reproducción* (LR), dentro del contexto formal educativo, alrededor de los 3 años de edad. Actualmente, en el contexto lingüístico vasco se diferencian principalmente dos grupos de bilingüismo precoz en función del orden de adquisición de cada una de las dos lenguas oficiales. Así, el primero de ellos es el compuesto por alumnado que ha adquirido una de las variedades tradicionales del euskera en el entorno familiar y que, a su vez, también ha adquirido el español (o francés) a edad muy temprana, bien en el ámbito educativo, bien en el social. El segundo grupo está formado por los hablantes que han adquirido el español por transmisión familiar y el euskera en el sistema educativo a edad muy temprana.

Los trabajos que analizan las producciones y emisiones de los jóvenes bilingües precoces en el contexto que se describe han sido abordados desde distintos enfoques y perspectivas (Eguskiza, De Pablo & Gaminde, 2017; Etxebarria, Gaminde & Olalde, 2016; Gaminde, 2010, 2011; Gaminde & Eguskiza, 2017). Sin embargo,

* Esta investigación se ha realizado dentro de los proyectos IT1028/16 (Gobierno Vasco) y GIU 16/22 (Universidad del País Vasco) y gracias al apoyo económico de la UPV/EHU para la contratación de personal investigador en formación predoctoral (PIF).

estas investigaciones no han profundizado en el campo del léxico (Gaminde Etxebarria, Eguskiza, Romero & Unamuno, 2016; Olalde, Gaminde & Etxebarria, 2016), el cual ha recibido atención únicamente de manera tangencial.

Con respecto a los métodos cuantitativos multivariantes de clasificación que se retoman en este artículo, son conocidos en diversos campos científicos y se han empleado dentro de la dialectología, concretamente en el área de la dialectometría (Clua, 2010; Goebel, 1978a, 1978b, 1981, 1991, 1992, 2010, 2018; Heeringa & Braun, 2003; Léonard, Heinsalu, Patriarca & Darlu, 2015; Nerbonne & Heeringa, 2001, 2010; Séguy, 1973; Valls, Nerbonne, Prokic, Wieling, Clua & Lloret, 2012). Generalmente las variedades diatópicas son objeto de análisis por medio de estos métodos, pero también se cuenta con estudios de clasificación de informantes (Gaminde, Olalde, Etxebarria, Eguskiza & Gaminde, 2017; Eguskiza, 2019).

En el campo específico de la dialectología de la lengua vasca existen numerosos estudios que utilizan dichas metodologías (Aurrekoetxea, 1992, 1995; Aurrekoetxea & Videgain, 2017; Aurrekoetxea, Gaminde, Ormaetxea & Videgain, 2019). Incluso se ha desarrollado un *software* para realizar clasificaciones similares (Aurrekoetxea, Iglesias, Santander & Usobiaga, 2016; Iglesias, Clua, Salicrú & Usobiaga, 2017). Tales metodologías se han aplicado en diferentes áreas lingüísticas: fonética y fonología (Heeringa, 2004), prosodia (Aurrekoetxea, Gaminde, Gandarias & Iglesias, 2012; Aurrekoetxea, Gandarias, Gaminde & Iglesias, 2014; Iglesias & Aurrekoetxea, 2017), morfología (Aurrekoetxea, 2016; Aurrekoetxea & Videgain, 2016), sintaxis (Aurrekoetxea, 2019) y léxico (Aurrekoetxea, 2009; Aurrekoetxea, Ormaetxea & Videgain, 2018).

Los trabajos sobre el léxico en euskera se han realizado en dos vertientes principales: confección de diccionarios generales en la lengua estándar o sobre ciertas materias en específico (biología, física, música, deportes, etc.) y monografías sobre el léxico dialectal. En general, este tipo de trabajos no atienden la cuestión de la accesibilidad léxica, aspecto que entronca directamente con la

capacidad léxica de los hablantes y que debería tener impacto en la enseñanza de la lengua. Así, una de las aportaciones relevantes al respecto es la aplicación de este tipo de metodología para conocer la homogeneidad o dispersión de los hablantes en función del tiempo de reacción empleado en la emisión verbal de cierto estímulo visual para que, posteriormente, estos datos puedan utilizarse en nuevas aplicaciones didácticas del ámbito lingüístico.

Tomando en cuenta esta posibilidad, en este trabajo se presentan los resultados de la clasificación de hablantes bilingües precoces, euskera-español, por medio del análisis de la accesibilidad a algunos ítems de un campo semántico específico en su lengua materna y su lengua de reproducción. La hipótesis inicial es que pueden existir diferencias en la accesibilidad al léxico de los informantes según el empleo de la LM o la LR utilizada en la realización de una tarea. En general, puede pensarse que los hablantes mostrarán una mayor accesibilidad léxica cuando usan su lengua materna en contraste con el empleo de su lengua de reproducción. De modo que, si presentan un tiempo similar en sus emisiones, podría concluirse que tienen un nivel de accesibilidad parecido respecto a determinado ítem en ambas lenguas. Sin embargo, si difieren en el tiempo de emisión, se deduce un distinto nivel de accesibilidad del léxico y esto, a su vez, debería considerarse a nivel educativo, como prueba de la diversidad lingüística del alumnado.

El trabajo se divide en tres apartados a partir de este. En §2 se describe el corpus, la metodología y el diseño de la tarea utilizada para la obtención de datos; a continuación, en §3, se presentan los resultados y, por último, en §4, se explican las conclusiones.

2. Corpus y metodología

Esta sección se ha dividido en tres subapartados: en § 2.1, se expondrán las características de los informantes; después, en § 2.2, se describirá la metodología empleada para la recogida de los datos y, por último, en § 2.3, se explicará cómo se ha llevado a cabo el análisis de los mismos.

2.1. *Informantes*

Para simplificar las características de la muestra en este estudio se ha trabajado con personas que se denominan bilingües precoces que, como se mencionó previamente, han adquirido la L2 dentro del sistema educativo formal en torno de los 3 años de edad. Siguiendo esta definición, cabe señalar que prácticamente toda la comunidad de hablantes de la lengua vasca es bilingüe, ya sea euskera-español en la parte peninsular del País Vasco o euskera-francés en la parte continental. Dichos grupos bilingües se diferencian en lo que respecta al conocimiento de la lengua y al uso que se hace de la misma en las diferentes áreas geográficas (Eusko Jaurlaritza, Nafarroako Gobernua & Euskararaen Erakunde Publikoa, 2016). En el contexto donde se ha llevado a cabo la investigación, la adquisición de la segunda lengua ocurre por medio de la transmisión familiar, del contexto social y/o del formal educativo.

Se ha contado con la participación y colaboración de 40 alumnas de la Facultad de Educación de Bilbao. Se ha elegido a este grupo de informantes por su posible injerencia en la formación lingüística de futuras generaciones de hablantes bilingües. Las participantes nacieron entre los años 1992 y 1998, cursaron todos sus estudios en el modelo D y son bilingües precoces, es decir, han adquirido una de las dos lenguas, euskera o español, en edades muy tempranas.

2.2. *Confección del corpus*

Para la obtención de los datos, se ha empleado la misma metodología en ambas lenguas. Respecto al procedimiento, se ha explicado a las participantes que verían una serie de imágenes a través de diferentes diapositivas de Power Point y, al momento de visualizar el estímulo, debían emitir verbalmente la palabra correspondiente a la imagen con la mayor celeridad posible. Dichos estímulos fueron precedidos por un pitido inicial, que posteriormente fungió como

referencia para calcular el tiempo desde la audición del pitido hasta la emisión léxica correspondiente.

Los estímulos visuales se corresponden al campo semántico de las partes del cuerpo humano. La clasificación de los hablantes puede realizarse utilizando cualquier elemento lingüístico que, sujeto a un proceso de análisis y después de un proceso de modelización, pueda convertirse en un número. En este caso se ha elegido el campo semántico del cuerpo humano debido a que lo integran palabras de uso generalizado y a disposición de todos los hablantes, además de haber sido ampliamente estudiado en los trabajos de disponibilidad léxica, por ejemplo.

Para obtener los datos es posible utilizar diferentes métodos: respuesta a estímulos sin límite de tiempo, textos libres, traducción de palabras, etc. En este trabajo, al ser los términos léxicos sencillos y en principio de uso generalizado, se ha optado por un método que dé cuenta del grado de accesibilidad de los hablantes al léxico y, por tanto, se ha medido el tiempo de reacción utilizado desde que aparece el estímulo visual (simultáneamente con el sonido) hasta que se produce la respuesta. Por tanto, se considera que esta es la aportación novedosa del trabajo: no se mide la respuesta misma de cada informante sino el tiempo que emplea en responder. En las condiciones en que se realiza la tarea, el tiempo de reacción refiere a la accesibilidad al léxico, en tanto que los métodos cuantitativos de análisis deben revelar si existen diferencias según el uso de la lengua materna y/o la lengua de reproducción de los informantes.

De esta manera, el corpus se constituye por 34 estímulos, que se muestran en el Cuadro 1.

Se ha conseguido un corpus total de 1360 datos, conformado por 34 estímulos y 40 informantes. Diez de las informantes de lengua materna euskera han realizado sus producciones en LM y las otras diez en LR. Del mismo modo, la mitad de las informantes de español como lengua materna han realizado las emisiones en dicha lengua y la mitad restante en euskera. En la Figura 1 se muestra esta distribución.

CUADRO 1. Orden de los estímulos visuales realizados

1 oreja	10 ojo	19 moratón	28 pierna
2 mano	11 espalda	20 campanilla	29 pantorrilla
3 dedo	12 riñón	21 rodilla	30 codo
4 boca	13 muela	22 pestañas	31 axila
5 nariz	14 uña	23 cabeza	32 venas
6 cuello	15 muslo	24 tobillo	33 orzuelo
7 brazo	16 colmillo	25 dedo del pie	34 ombligo
8 barbilla	17 frente	26 talón	
9 corazón	18 lengua	27 muñeca	

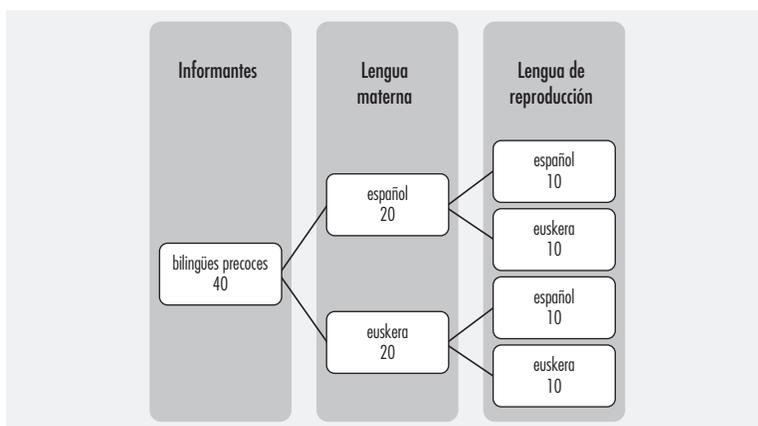


FIGURA 1. Distribución de las informantes según LM y LR

Todas las grabaciones han sido realizadas en contextos silenciosos empleando un ordenador y un micrófono externo omnidireccional de condensador (BOYA BY-M1). Las grabaciones se han realizado en el formato .wav para posterior análisis mediante el programa Praat (Boersma & Weenink, 2017). Cada emisión acústica ha sido anotada y etiquetada desde el pitido hasta el comienzo de la producción léxica de las informantes. Estos datos han sido vertidos a una matriz previamente creada mediante un *script* (Lennes, 2002), que ha facilitado la extracción automática de los datos en milisegundos (ms). Se han recogido todas las emisiones con excepción

de aquellas cuyas respuestas eran desconocidas por las informantes, las cuales se etiquetaron como nulas.

2.3. *Análisis de los datos*

El análisis de los datos se ha efectuado en dos planos. En el primero, se toma en cuenta la lengua materna y la lengua de reproducción de las informantes —ambas consideradas variables sobre las que se efectúan los diversos análisis del tiempo de reacción. En el segundo, se clasifican las informantes según dichos tiempos de reacción mediante tres métodos objetivables de conglomeración: el método jerárquico, el método de K-medias y el método de C-means basado en el algoritmo de la lógica difusa (FCM, por sus siglas en inglés). En todos los casos se ha utilizado la distancia euclídea y en el método de conglomeración jerárquico se ha empleado el método de Ward, por ser el más adecuado para explorar las similitudes y disimilitudes existentes entre los datos (Aurrekoetxea, 1995, 2005; Goebel, 1992, 2010; Pardo & Ruiz, 2002; Pérez, 2005).

Estos métodos parten de los elementos de un conjunto con atributos que los hacen pertenecer a algún tipo de subconjunto (en nuestro caso la lengua materna y la lengua de reproducción).¹ Con los elementos del conjunto y sus atributos, por una parte, se construye una matriz de datos de tantas filas como elementos del conjunto y tantas columnas como características investigadas. En esta matriz cada elemento se representa como un vector. La comparación de estos vectores por medio de los diferentes métodos de conglomeración que se describen a continuación es la base de la clasificación utilizada.

El primer método clasifica a las informantes en un único grupo, es decir, una determinada informante es parte de uno de los dos grupos, no dando información de la posible parte pertenecien-

¹ Otros atributos de orden social son el género, la edad, la clase social, etcétera.

te al otro grupo. La clasificación de las informantes es realizada en función de la distancia que hay entre el resto de elementos y sin establecer anteriormente la cantidad de grupos que se desea obtener. Así, a través de la realización del mismo se obtiene un dendrograma.

A diferencia del anterior, en el método K-medias, se debe elegir previamente la cantidad de conglomerados a obtener. Debido a que este destaca las igualdades y las diferencias entre los grupos, el investigador es quien decide la cantidad de los mismos.

El tercer método (FCM) permite al investigador elegir 1) la cantidad de grupos que desea crear y 2) el índice de pertenencia; esto es, a partir de cuándo un elemento es altamente perteneciente a un grupo. La pertenencia al grupo —cuya función se denota de la siguiente manera: $\mu_A: X \rightarrow [0,1]$ — abarca desde 0 a 1, siendo 0 la medida que demuestra que no pertenece a ese grupo en absoluto y siendo 1 cuando pertenece completamente al mismo. De manera que el informante que tenga más de 0.5 en alguno de ellos, por ejemplo 0.7, pertenece mayoritariamente a ese grupo en un 70%, pero también es, en menor medida, parte del otro en un 30%. En este estudio, se ha establecido la frontera de pertenencia en 0.85 y se ha denominado informantes de transición a quienes estén entre 0.5 y 0.85 debido a que solo pertenecen en cierta medida al grupo.

3. Resultados

A continuación, se procederá a explicar los resultados. Primeramente, en §3.1, se expondrá el análisis del tiempo de reacción de las informantes para después clasificarlas según los diferentes métodos expuestos en el apartado metodológico, atendiendo a los resultados obtenidos de las respuestas en LM y LR.

Los resultados se expondrán haciendo referencia en todo momento a los siguientes cuatro grupos:

- 1) Informantes con lengua materna español y lengua de reproducción español.

- 2) Informantes con lengua materna español y lengua de reproducción euskera.
- 3) Informantes con lengua materna euskera y lengua de reproducción español.
- 4) Informantes con lengua materna euskera y lengua de reproducción euskera.

3.1. Análisis del tiempo de reacción

En este apartado, se muestran los datos generales (respuestas válidas y nulas) en función de los cuatro grupos identificados previamente. Enseguida, se plantea el análisis de las respuestas nulas de cada una de las informantes.

En el Cuadro 2 se exponen la cantidad y el porcentaje de las respuestas válidas y nulas en cada grupo.

CUADRO 2. Respuestas nulas y válidas por lengua materna y lengua de reproducción

GRUPOS	VÁLIDAS		NULAS		TOTAL	
	N	%	N	%	N	%
1	331	97.35	9	2.65	340	100
2	284	83.53	56	16.47	340	100
3	314	92.35	26	7.65	340	100
4	318	93.53	22	6.47	340	100

En cada uno de los grupos ha habido un total de 340 respuestas posibles: en el grupo uno se han obtenido 331 respuestas válidas y 9 nulas, en el segundo grupo 284 válidas y 56 nulas, en el tercer grupo 314 válidas y 26 nulas, y en el cuarto grupo 318 válidas y 22 nulas.

Si se analizan los datos generales en función de la LM y LR, puede observarse que la media es de 324.5, es decir, de un 95.44%, en el caso de las informantes que han respondido en su lengua materna, y de 299 respuestas válidas, esto es, un 87.94%, cuando lo

han hecho en su segunda lengua. Por lo tanto, se concluye que las respuestas válidas aumentan cuando las informantes responden en su LM.

Respecto a la LM, es decir, si es español o euskera, se observa cómo las informantes de LM español obtienen un 90.44% de respuestas válidas frente a 92.44% de las informantes con LM euskera, logrando estas últimas un mayor porcentaje de respuestas válidas.

En cuanto a las respuestas nulas, se ha realizado la prueba de Mann-Whitney con el fin de identificar si existen diferencias estadísticamente significativas, y se han obtenido los datos presentados en el Cuadro 3.

CUADRO 3. Significación (Mann-Whitney) según las respuestas nulas, comparación entre grupos

GRUPOS	1	2	3	4
1		*0.02	*0.012	0.684
2	*0.02		0.144	*0.020
3	*0.012	0.144		0.338
4	0.684	*0.020	0.338	

Como se observa en el Cuadro 3, los resultados estadísticamente significativos (*) son los obtenidos en la comparativa entre el grupo 1 con el 2, el 1 con el 3 y el 2 con el 4.

En lo que respecta al tiempo de reacción, en la Figura 2 y el Cuadro 4 se expone la desviación estándar, el mínimo, el máximo y el rango de las respuestas en función del grupo.

Para el análisis del tiempo de reacción se ha aplicado la prueba HSD de Tukey con el fin de observar si existen diferencias estadísticamente significativas entre los grupos. Los datos obtenidos se muestran en el Cuadro 5.

Los grupos que presentan significación estadística entre sí (*) son el grupo 1 con el 2, el 2 con el 3 y el 2 con el 4.

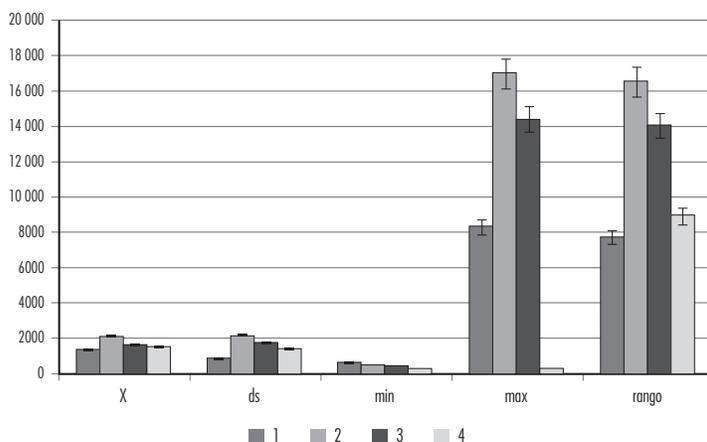


FIGURA 2. Media, desviación estándar, mínimo, máximo y rango de las respuestas según los cuatro grupos

CUADRO 4. Media, desviación estándar, mínimo, máximo y rango de las respuestas según los cuatro grupos

GRUPOS	\bar{x}	DS	MIN.	MAX.	RANGO
1	1304.41	830.26	600.00	8290.00	7690.00
2	2059.71	2082.66	440.00	16970.00	16530.00
3	1589.58	1697.38	380.00	14400.00	14020.00
4	1463.83	1349.09	260.00	260.00	8900.00

CUADRO 5. Significación (HSD de Tukey) según el tiempo de reacción

GRUPOS	1	2	3	4
1		*0.000	0.086	0.549
2	*0.000		*0.001	*0.000
3	0.086	*0.001		0.733
4	0.549	*0.000	0.733	

3.2. Clasificación de las informantes según la lengua materna y la lengua de reproducción

Para realizar la clasificación de las informantes según la LM y la LR, se han utilizado tres métodos cuantitativos de conglomeración: el clúster jerárquico, K-medias y Fuzzy C-means (FCM).

3.2.1. Clúster jerárquico

Con el *software* de análisis estadístico SPSS se obtiene el clúster jerárquico. Inicialmente, se exponen los datos obtenidos mediante el análisis de conglomerados jerárquicos, por medio del cual se ha obtenido el dendrograma de la Figura 3.

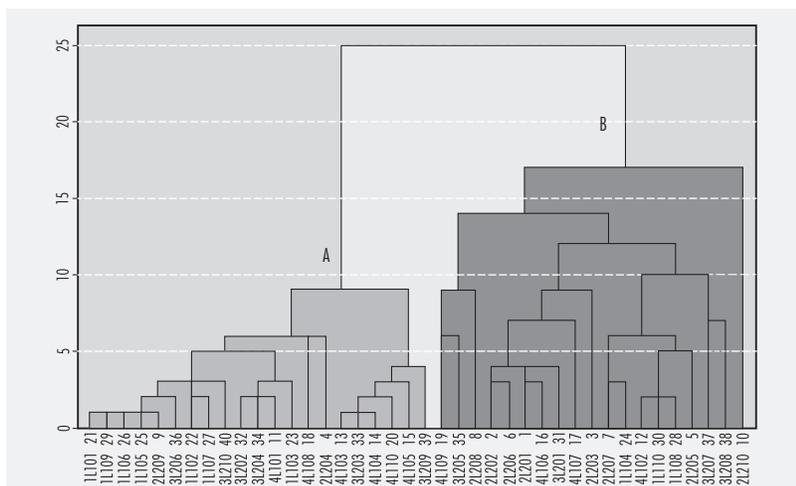


FIGURA 3. Dendrograma del clúster jerárquico

En el dendrograma se pueden observar los dos principales grupos creados por medio del análisis jerárquico, a los que se han denominado A y B. Con estos resultados se ha planteado el Cuadro 6, donde se indica la pertenencia del número de informantes por cada

CUADRO 6. Cuantificación de los datos obtenidos en el dendrograma del clúster jerárquico

	LM		TOTAL	LR		TOTAL
	1	4		2	3	
A	7	6	13	2	6	8
B	3	4	7	8	4	12

conglomerado. Los datos obtenidos no son significativos estadísticamente (χ^2 (a.m. 3) 5.915; $p = 0.116$).

Como se indica en el Cuadro 6, las informantes con LM español y LR español, se han clasificado 7 en el grupo A y 3 en el grupo B. Asimismo, las informantes con LM español y LR euskera, se han clasificado 6 en el grupo A y 4 en el grupo B. En el caso de las informantes con LM euskera y LR euskera, se han encontrado 2 en el grupo A y 8 en el grupo B. Por último, aquellas informantes con LM euskera y LR euskera, se han situado 6 en el grupo A y 4 en el grupo B.

3.2.2. K-medias

Para la aplicación de este método de conglomeración, se ha optado por crear dos grupos según los obtenidos en el análisis jerárquico. En el Cuadro 7 se muestra el número de informantes por conglomerado.

CUADRO 7. Cantidad de informantes en cada grupo tras la realización del método K-medias

	LM		TOTAL	LR		TOTAL
	1	4		2	3	
A	3	4	7	8	4	12
B	7	6	13	2	6	8

Como se ve en el Cuadro 7, las informantes con LM español y LR español, se clasifican 3 en el grupo A y 7 en el grupo B. Aquellas

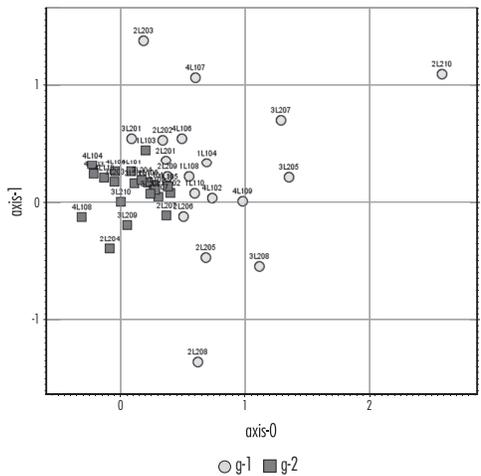


FIGURA 4. Dispersión de los resultados obtenidos con el método K-medias

informantes con LM español y LR euskera, se han clasificado 4 en el grupo A y 6 en el grupo B. En el caso de las informantes con LM euskera y LR euskera, se han encontrado 8 en el grupo A y 2 en el grupo B. Por último, aquellas informantes con LM euskera y LR euskera, se han catalogado 4 en el grupo A y 6 en el grupo B. Los resultados del análisis K-medias coinciden plenamente con los resultados del análisis jerárquico, expuesto en §3.1. A partir de este método se realizó el gráfico de dispersión de escalamiento multidimensional que se presenta a continuación (véase Figura 4).

En esta figura se muestran los dos grupos con diferente iconicidad y tono: A está representado por un cuadrado y B por un círculo. Además, se observa que el último grupo de círculos es más disperso que el representado por cuadrados.

3.2.3. Fuzzy C-means (FCM)

Según el método de conglomeración basado en la lógica difusa (FCM), al utilizar dos grupos de conglomeración se obtienen los datos correspondientes al Cuadro 8, donde aquellos elementos con

CUADRO 8. Clasificación de las informantes según LM y L2 en dos grupos según el método FCM

		LM		TOTAL	L2		TOTAL
		1	4		2	3	
A	≥ 0.85	4	6	10	1	6	7
	< 0.85	3	0	3	1	1	2
B	≥ 0.85	3	3	6	5	3	8
	< 0.85	0	1	1	3	0	3

función de pertenencia menor a 0.85 se consideran como elementos del conglomerado pero de transición, ya que una parte de su función de pertenencia está incluida en el otro conglomerado.

En el Cuadro 8 podemos ver la división de dos grupos: grupo A y grupo B. Cada uno de los grupos tiene dos columnas, la primera perteneciente a las informantes que han respondido en su LM y la segunda perteneciente a las informantes que han respondido en su L2. Además, se presentan dos subcategorías dentro del grupo: una de ellas (≥ 0.85) muestra que las informantes poseen una función de pertenencia alta al grupo, esto es, mayor que 0.85, y la segunda (< 0.85) indica que las informantes son de elementos de transición del conglomerado, ya que la función de pertenencia es menor a 0.85.

Así pues, en el grupo A las informantes que han respondido en su LM y de alta función de pertenencia al grupo son 10 (4 con LM español y 6 con LM euskera). Las informantes de transición son 3, todas ellas de LM español.

Dentro del grupo A, las informantes que han respondido en su LR, se encuentra una informante con LM español que ha respondido en euskera y 6 vascoparlantes que han respondido en español. Con una pertenencia inferior a 0.85, es decir, entre las informantes de transición hay una informante con lengua materna español que ha respondido en euskera y 1 vascoparlante que ha respondido en español.

En lo que respecta al grupo B, las informantes con alta pertenencia al grupo (≥ 0.85) son 6 (3 de LM español y 3 de LM euskera);

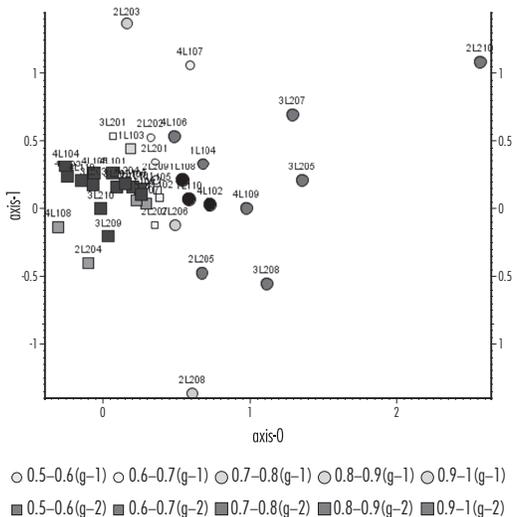


FIGURA 5. Dispersión de los datos tras la aplicación del método FCM

y respecto a las informantes de transición en su LM se encuentra una con LM euskera.

Dentro del grupo B, las informantes que han respondido a los estímulos en su LR, con una función de pertenencia alta, son 8 (5 con LM español y respuestas en euskera y 3 LM euskera y LR español). Las informantes de transición son tres, todas ellas con LM español y LR euskera. Con base en estos datos, se ha creado el gráfico de dispersión que se muestra en la Figura 5, donde podemos ver representados cada uno de los grupos en dos formas geométricas diferentes. El grupo A se muestra con círculos, mientras que el grupo B por cuadrados. Además de ello, la intensidad de los grises y el tamaño de las figuras varía en función del grado de pertenencia al grupo. Así, aquellas con mayor tamaño y tono más intenso son aquellos con mayor pertenencia al grupo, y cuanto más claros y de menor tamaño, tienen menor pertenencia al grupo.

Si comparamos los resultados obtenidos en el análisis de conglomerados jerárquicos y K-medias con el análisis FCM se puede observar una diferencia importante en los componentes de cada

conjunto, ya que, en el análisis FCM, hay una informante con LM euskera que se clasifica en el otro conglomerado a diferencia de los análisis anteriores.

4. Conclusiones y trabajos futuros

Este estudio se ha llevado a cabo en un territorio en el que conviven dos lenguas: el euskera y el español. Sin embargo, el tiempo de reacción ante los estímulos léxicos se ve influenciado en función de la exposición de las informantes a cada una de las lenguas en contacto. Este aspecto se ha corroborado en el presente trabajo, donde se observa que mediante los tres métodos de conglomeración empleados se crean distintas agrupaciones de los hablantes en función de las variables objeto de estudio: la lengua materna y la lengua de reproducción.

En los resultados obtenidos se confirma que los grupos formados por las informantes son parcialmente heterogéneos. Dentro de dicha heterogeneidad se constata que el grupo A, conformado por informantes de LM euskera, muestra mayor homogeneidad en la reproducción léxica en ambas lenguas en comparación con el grupo B, formado por hablantes de LM español. Es decir, las informantes del grupo B muestran mayor tiempo de reacción en sus producciones léxicas en euskera que en su LM, mientras que en el caso de las informantes con LM euskera la diferencia entre ambos idiomas no es considerable.

Por lo tanto, pese a que el contexto es bilingüe, las adquisiciones léxicas no lo son, ya que varía el tiempo de reacción según la LM y la LR, debido a la situación diglósica en la que el español es la lengua hegemónica. Por lo tanto, resulta necesario que los resultados de este estudio y de los posibles trabajos futuros sean considerados en la práctica docente para dar respuesta a la diversidad lingüística del alumnado.

En futuras investigaciones, resulta pertinente el aumento del corpus y el análisis comparativo entre los resultados obtenidos en función del género de los informantes. Además, los métodos de

conglomeración presentados pueden ser aplicables a otros contextos bilingües e incluso expandidos a otras áreas de la lengua.

5. Referencias

- AURREKOETXEA, GOTZON (1992). Nafarroako euskara: azterketa dialektometrikoa. *Uztaro*, 5, 59–109.
- AURREKOETXEA, GOTZON (1995). *Bizkaieraren egituraketa geolinguistikoa*. Lejona: Universidad del País Vasco.
- AURREKOETXEA, GOTZON (2005). Nafarroako euskararen sailkapenaz. En Pilartxo Etxebarria & Henrike Knörr (Eds.), *Nerekin yaio nun. Txillardegiri omenaldia* (pp. 109–124). Bilbao: Euskaltzaindia.
- AURREKOETXEA, GOTZON (2009). Iparraldeko hizkeren sailkapena (1): lexikoa. *ASJU*, 38(1), 287–330.
- AURREKOETXEA, GOTZON (2016). Analysis of the morphological variation of Basque. *Dialectologia et Geolinguistica*, 24(1), 21–41.
- AURREKOETXEA, GOTZON (FEBRERO, 2019). Classification of syntax in Basque. Ponencia presentada en el taller Language Documentation, 7: Documenting Variation, Bolzano, Italia.
- AURREKOETXEA, GOTZON; GAMINDE, IÑAKI; GANDARIAS, LEIRE, & IGLESIAS, AITOR (2012). Prosodic variation in the Basque language: Stress areas. En Xosé Afonso Álvarez Pérez, Ernestina Carrilho & Catarina Magro (Eds.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)* (pp. 35–51). Lisboa: Centro de Lingüística da Universidade de Lisboa.
- AURREKOETXEA, GOTZON; GAMINDE, IÑAKI; ORMAETXEA, JOSÉ LUIS, & VIDEGAIN, XARLES (2020). Hizkeren sailkapen berria. En Juan Abasolo, Irati De Pablo & Ariane Ensunza (Eds.), *Hezkuntzari buruzko ekarpenak/Contributions on education* (pp. 5–18). Bilbao: Universidad del País Vasco. Recuperado de: https://web-argitalpena.adm.ehu.es/pasa_pdfFin.asp
- AURREKOETXEA, GOTZON; GANDARIAS, LEIRE; GAMINDE, IÑAKI, & IGLESIAS, AITOR (2014). Variación prosódica en vasco: áreas acentuales. En Yolanda Congosto Martín, Ma. Luisa Montero Curiel & Antonio Salvador Plans (Eds.), *Fonética experimental, educación superior e investigación, III: prosodia* (pp. 11–28). Madrid: Arco Libros.

- AURREKOETXEA, GOTZON; IGLESIAS, AITOR; SANTANDER, GOTZON, & USOBIAGA, IKER (2016). Diatech tresna informatikoaren bertsio berria. En Aitor Iglesias, Asier Romero & Ariane Ensunza (Eds.), *Linguistic variation in the basque language and education-II/Euskararen bariazioa eta bariazioaren irakaskuntza-II* (pp. 16–24). Bilbao: Universidad del País Vasco.
- AURREKOETXEA, GOTZON; ORMAETXEA, JOSÉ LUIS, & VIDEGAIN, XARLES (2018). Euskalkien sailkapen zientifikoa: lexikoa. En Lorea Unamuno, Asier Romero, Aintzane Etxebarria & Aitor Iglesias (Eds.), *Linguistic variation in the basque and education-III/Euskararen bariazioa eta bariazioaren irakaskuntza-III* (pp. 126–140). Universidad del País Vasco.
- AURREKOETXEA, GOTZON, & VIDEGAIN, XARLES (2016). Euskararen bariazio geo-morfologiaren azterketa. En Aitor Iglesias, Asier Romero & Ariane Ensunza (Eds.), *Linguistic variation in the basque language and education-II/Euskararen bariazioa eta bariazioaren irakaskuntza-II* (pp. 25–36). Bilbao: Universidad del País Vasco.
- AURREKOETXEA, GOTZON, & VIDEGAIN, XARLES (2017). Le traitement dialectométrique du corpus Bourciez de 1895. *Géolinguistique*, 17, 49–98.
- BOERSMA, PAUL, & WEENINK, DAVID (2017). *Praat: Doing phonetics by computer* [Computer program]. Ámsterdam: Universidad de Ámsterdam. Versión 6.0.36. Recuperado de <http://www.praat.org>
- CLUA, ESTEVE (2010). Relevancia del análisis lingüístico en el tratamiento cuantitativo de la variación dialectal. En Gotzon Aurrekoetxea & Jose Luis Ormaetxea (Eds.), *Tools for linguistic variation* (pp. 151–166). Bilbao: Universidad del País Vasco.
- IGLESIAS, AITOR; CLUA, ESTEVE; SALICRÚ, MIQUEL, & USOBIAGA, IKER (2017). Nola identifikatu ezazgarri esanguratsuenak bariazio dialektalean: diatech nabarmentzeko aukera. En Iglesias, Aitor & Ensunza, Ariane (Eds.), *Gotzon Aurrekoetxea lagunarterik hara*. Bilbao: Universidad del País Vasco.
- EGUSKIZA SÁNCHEZ, NAIA (2019). *Hizkuntza aldakortasun geo-soziolinguistikoa Arratian*. (Tesis doctoral inédita). Universidad del País Vasco, Leioa.
- EGUSKIZA, NAIA; DE PABLO, IRATI, & GAMINDE, IÑAKI (2017). Hizkuntz hizkeren sailkapena metodo kuantitatiboen bidez. En Cristina Arriaga & Asier Romero (Eds.), *XXIV Jornadas de Investigación en Psicodidáctica* (pp. 403–416). Bilbao: Universidad del País Vasco.

- ETXEBARRIA, AINTZANE; GAMINDE, IÑAKI, & OLALDE, ANDER (2016). Hizkuntza aldakortasuna Larrabetzuko aditz morfologian. En Aitor Iglesias, Asier Romero & Ariane Ensunza (Eds.), *Linguistic variation in the Basque language and education -II/Euskararen bariazioa eta bariazioaren irakaskuntza-II* (pp. 104–119). Bilbao: Universidad del País Vasco.
- EUSKO JAURLARITZA; NAFARROAKO GOBERNUA, & EUSKARARAEN ERAKUNDE PUBLIKOA (2016). VI. Inkesta Soziolinguistikoa. Euskararen eremo osoa. Recuperado de https://www.irekia.euskadi.eus/uploads/attachments/9954/VI_INK_SOZLG-EH_eus.pdf?1499236557
- GAMINDE, IÑAKI (2010). *Bizkaiko gazteen prosodiaz: euskaraz eta gaztelaniaz*. Bilbao: Mendebalde Kultura Alkartea.
- GAMINDE, IÑAKI (2011). Noraka euskal azentuak? En Gotzon Aurrekoetxea & Iñaki Gaminde (Eds.), *Prosodiaz eta Hezkuntzaz I. Jardunaldiak/I. Jornadas sobre Prosodia y Educación* (pp. 63–85). Lejona: Universidad del País Vasco.
- GAMINDE, IÑAKI, & EGUSKIZA, NAIA (2017). Arratiako gazteen aditz laguntzaileen bariazioaz. En Aitor Iglesias & Ariane Ensunza (Eds.), *Gotzon Aurrekoetxea lagunarterik hara* (pp. 141–154). Bilbao: Universidad del País Vasco.
- GAMINDE, IÑAKI; ETXEBARRIA, AINTZANE; EGUSKIZA, NAIA; ROMERO, ASIER, & UNAMUNO, LOREA (2016). Lexikoaren bariazioa eta multzokatze-azterketa. *Euskalingua*, 28, 19–32.
- GAMINDE, IÑAKI; OLALDE, ANDER; GAMINDE, URSUA, & ETXEBARRIA, AINTZANE (2016). Populazioaren sailkapena lexikoaren arabera. *Ikastorratza, e-Revista de Didáctica*, 16, 68–84.
- GAMINDE, IÑAKI; OLALDE, ANDER; ETXEBARRIA, AINTZANE; EGUSKIZA, NAIA, & GAMINDE, URSUA (2017). *Hizkuntza aldakortasuna larrabetzun*. Bilbao: Larrabetzuko Udala.
- GOEBL, HANS (1978a). La dialectométrie appliquée à l'«ALF» (Normandie). En Alberto Várvaro (Ed.), *Atti del XIV Congresso Internazionale di Linguistica e Filologia Romanza (Napoli 15–20 Aprile, 1974)* (Vol. II, pp. 165–196). Amsterdam: John Benjamins.
- GOEBL, HANS (1978b). Analyse dialectométrique de quelques points de l'ais (italien standard, valdotain, provençal, alpin, turinois, milanais). En Gianrenzo Clivio & Giuliano Gasca (Eds.), *Lingue e dialetti nell'arco alpino occidentale: atti del Convegno Internazionale di Torino*,

- 12-14 aprile 1976 (pp. 282–294 + mapas). Turín: Centro Studi Piemontesi.
- GOEBL, HANS (1981). Éléments d'analyse dialectométrique (avec application à l'AIS). *Revue de Linguistique Romane*, 45, 349–420.
- GOEBL, HANS (1991). Una classificazione gerarchica di dati geolinguistici tratti dall' AIS. Saggio di dialettometria dendrografica. *Linguistica*, 31(1), 341–352.
- GOEBL, HANS (1992). Problèmes et méthodes de la dialectométrie actuelle (avec application à l' AIS). En Real Academia de la Lengua Vasca (Ed.), *Nazioarteko dialektologia biltzarra. Agiriak/Actes du Congrès International de Dialectologie* (pp. 429–475). Bilbao: Real Academia de la Lengua Vasca.
- GOEBL, HANS (2010). Introducción a los problemas y métodos según los principios de la Escuela Dialectométrica de Salzburgo (con ejemplos sacados del 'Atlante Italo-Svizzero', AIS). En Gotzon Aurrekoetxea & Jose Luis Ormaetxea (Eds.), *Tools for linguistic variation* (pp. 3–40). Bilbao: Universidad del País Vasco.
- GOEBL, HANS (2018). Dialectometry. En Charles Boberg, John Nerbonne & Dominic Watt (Eds.), *The handbook of dialectology* (pp. 123–142). Hoboken: John Wiley.
- HEERINGA, WILBERT (2004). *Measuring dialect pronunciation differences using Levenshtein distance* (Tesis doctoral inédita). Rijksuniversiteit Groningen, Groningen.
- HEERINGA, WILBERT, & BRAUN, ANGELIKA (2003). The use of the Almeida-Braun system in the measurement of Dutch dialect distances. *Computers and the Humanities*, 37(3), 257–271.
- IGLESIAS, AITOR, & AURREKOETXEA, GOTZON (2017). Intonazioaren egituraketa geolinguistikoa. En Aintzane Etxebarria & Naia Eguskiza (Eds.), *Bariazioa esaldien intonazioan* (pp. 277–295). Bilbao: Universidad del País Vasco.
- LENNES, MIETTA (2002). *Segment data script*. GNU General Public License.
- LÉONARD, JEAN LÉO; HEINSALU, ELS; PATRIARCA, MARCO, & DARLU, PIERRE (2015). Modeling regional variation from EAS: Complexity and communal aggregates. En Gotzon Aurrekoetxea, Asier Romero & Aintzane Etxebarria (Eds.), *Linguistic variation in the basque an education-I/Euskararen bariazioa*

eta bariazioaren Irakaskuntza-I (pp. 145–172) Bilbao: Universidad del País Vasco.

- NERBONNE, JOHN, & HEERINGA, WILBERT (2001). Computational comparison and classification of dialects. *DiG*, 9, 69–83.
- NERBONNE, JOHN, & HEERINGA, WILBERT (2010). Measuring dialect differences. En Peter Auer & Jürgen Erich Schmidt (Eds.), *Language and space. An international handbook of linguistic variation. Volume I: Theories and methods* (pp. 550–567). Berlín: De Gruyter Mouton.
- PARDO, ANTONIO; & RUIZ, MIGUEL (2002). *SPSS 11 Guía para el análisis de datos*. Madrid: McGraw-Hill.
- PÉREZ LÓPEZ, CÉSAR (2005). *Métodos estadísticos avanzados con SPSS*. Madrid: Thomson.
- SÉGUY, JEAN (1973). La dialectométrie dans l'Atlas linguistique de la Gascogne. *Revue de Linguistique Romane*, 37, 1–24.
- VALLS, ESTEVE; NERBONNE, JOHN; PROKIC, JELENA; WIELING, MARTJIN; CLUA, ESTEVE, & LLORET, MARIA ROSA (2012). Applying the Levenshtein distance to Catalan dialects: A brief comparison of two dialectometric approaches. *Verba*, 39, 35–61.

