DIFERENCIAS LÉXICAS ENTRE LOS BILINGÜES PRECOCES EN RELACIÓN CON EL TIEMPO DE REACCIÓN[1]

Naia Eguskiza
Universidad del País Vasco (UPV/EHU)
Irati de Pablo
Universidad del País Vasco (UPV/EHU)
Iñaki Gaminde
Universidad del País Vasco (UPV/EHU)
Aintzane Etxebarria
Universidad del País Vasco (UPV/EHU)

RESUMEN

La finalidad de este trabajo es constatar el empleo del léxico de las informantes, según la lengua de reproducción, contabilizando el tiempo de reacción desde la visualización del estímulo y la emisión léxica correspondiente. Para ello, se han obtenido los datos de la emisión léxica de 40 alumnas de la Facultad de Educación de Bilbao llevando a cabo la misma metodología en ambas lenguas. Se ha empleado un cuestionario de 34 entradas por medio de una secuencia de diapositivas. El análisis de los datos se ha realizado mediante tres métodos cuantitativos objetivables de conglomeración: (1) método jerárquico, (2) método de K medias y (3) algoritmo C medias basado en lógica difusa (fuzzy C-means, FCM). Entre los resultados encontramos ciertas convergencias en lo que respecta a la cantidad de grupos creados en ambos idiomas en el método jerárquico y el K medias. No obstante, no coinciden la totalidad de las palabras en los dos grupos. Y en lo que respecta al FCM, podemos observar diferencias en las palabras de transición.

Palabras clave: sociolingüística, léxico, bilingüismo precoz, tiempo de reacción

[1]Esta investigación se ha realizado dentro de los proyectos IT 1028/16 y GIU 16/22.

ABSTRACT

The purpose of this work is to verify the use of the lexicon of the informants, according to the language of reproduction, counting the reaction time from the visualization of the stimulus and the corresponding lexical emission. For this, the data of the lexical emission of 40 students of the Faculty of Education of Bilbao have been obtained carrying out the same methodology in both languages. A questionnaire of 34 entries was used through a sequence of slides. The analysis of the data was carried out using three objective quantitative methods of conglomeration: (1) hierarchical method, (2) medium K method and (3) average C algorithm based on fuzzy C-means, FCM). Among the results, we find certain convergences regarding the number of groups created in both languages in the hierarchical method and the average K. However, the totality of the words in the two groups do not coincide. And with regards to the FCM, we can observe differences in the transition words.

Keywords: sociolinguistics, lexicon, early bilingualism, reaction time

1. INTRODUCCIÓN

Son numerosos los trabajos realizados sobre la variación lingüística en euskera y abarcan diferentes aspectos de la lengua, ahora bien los que se refieren al léxico se circunscriben a aquellas investigaciones que tienen como objetivo la clasificación dialectal de las diferentes variedades de la lengua (dentro de éstos, el más importante realizado en los últimos años es el atlas dialectal de la lengua que viene realizando Euskaltzaindia desde el año 2008), así como una serie de monografías de carácter etnográfico (Barrutia, 2000; Gaminde, 2005; Zuazo, 2017). En cuanto a aquellos trabajos que utilizan métodos cuantitativos para la comparación de las variedades diatópicas de la lengua se ha producido un avance en los últimos años gracias a la introducción de métodos dialectométricos (Aurrekoetxea, 2016; Aurrekoetxea et al., 2014; Clua et al., 2017; Eguskiza et al., 2017).

El cambio producido en la comunidad lingüística vasca en las últimas décadas ha sido gracias a dos hechos de innegable importancia: la estandarización de la lengua y la implantación y posterior crecimiento exponencial de los modelos educativos en euskera. Eso ha producido un creciente interés en investigar las producciones de los llamados bilingües precoces (Etxebarria et al., 2016; Gaminde, 2011; Gaminde y Eguskiza, 2017; Gaminde et al., 2017); es decir, de las personas que teniendo como lengua materna el euskera o el español, han aprendido la segunda lengua dentro del sistema educativo en edades tempranas.

A pesar del creciente interés en este campo, el léxico sólo ha recibido atención de manera tangencial y son pocos los trabajos realizados en este ámbito (Gaminde et al., 2016a; 2016b).

En nuestro caso, para conocer el nivel de disponibilidad, hemos optado por medir el tiempo de reacción que necesitan los informantes para responder por medio del léxico a los estímulos visuales recibidos.

El objetivo de este trabajo es constatar si existen diferencias significativas en el tiempo de reacción entre la visualización del estímulo y la emisión léxica correspondiente, producida por hablantes precoces de euskera y español en ambas lenguas.

2. CORPUS Y METODOLOGÍA

El cuestionario utilizado se compone de 34 preguntas (tabla 1) presentado a los informantes por medio de una secuencia de diapositivas. En cada una de ellas se mostraba un dibujo que correspondía a una de las preguntas del cuestionario y a la que los informantes debían responder lo más rápido posible. En el momento de mostrarse el estímulo visual se emitía un pitido instantáneo que era grabado junto con la respuesta dada (formato *.wav*).

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

Tabla 1. Orden de los estímulos visuales

Para la obtención de los datos se contó con la colaboración de 40 informantes, todas ellas alumnas de la escuela universitaria de magisterio de Bilbao nacidas entre los años 1992 y 1998. 20 de ellas han sido de lengua materna euskera y 20 de español, todas ellas bilingües precoces que han realizado sus estudios en el modelo D.

Las grabaciones se etiquetaron utilizando el programa Praat (Boersma y Weenink, 2016). Posteriormente se utilizó un script para obtener las medidas de duración automáticamente (milisegundos). En los casos en que los informantes no conocían la respuesta, ésta se consideró nula.

En el análisis de los datos que presentamos en este trabajo se han utilizado diversos métodos cuantitativos. Se han clasificado las palabras según las similitudes en el tiempo de reacción para producirlas y en el número de respuestas nulas; para esto, se han utilizado tres métodos de conglomeración: el método jerárquico, el método de K medias y el método FCM. En todos los casos la distancia utilizada ha sido la euclídea y el método de Ward de conglomeración en el caso de los conglomerados jerárquicos.

3. RESULTADOS

En este apartado presentamos los resultados de la investigación atendiendo a las especificidades halladas según la lengua empleada por los informantes.

3.1 Análisis del léxico

En esta sección analizamos las diferencias léxicas encontradas para cada lengua teniendo en cuenta el número de respuestas vacías o nulas y las válidas empleadas por los informantes. Los resultados obtenidos se muestran en la tabla 2. Esta diferencia es estadísticamente significativa de acuerdo con la prueba de X^2 (chi = 19,235 (gl. 1) p = 0,000) según la lengua de reproducción.

	Nulas	%	Válidas	%	Total
Español	35	5,15%	645	94,85%	680
Euskera	80	11,76%	600	88,4%	680
Total	115	8,46%	1245	91,54%	1360

Tabla 2. Número y porcentajes de respuestas nulas y válidas por lengua

Asimismo, se presentan en el gráfico 1 los datos generales de las respuestas en ambas lenguas.

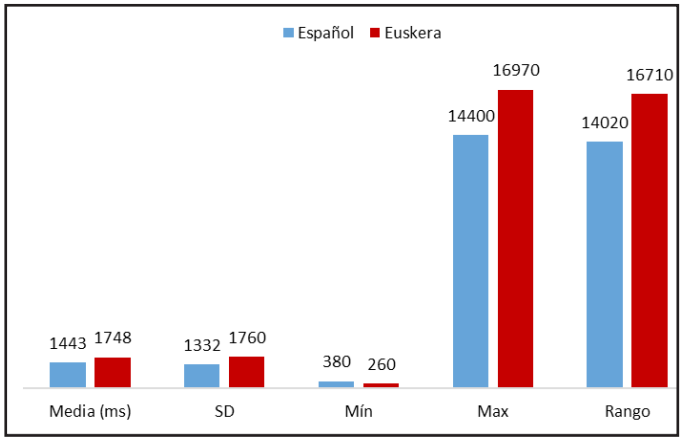

Gráfico 1. Porcentajes de respuestas nulas por lenguas

3.2 Análisis del tiempo de reacción en español

El número de respuestas válidas recogidas en español ha sido de 645, la media del tiempo de reacción es de 1443ms, la desviación es de 1332ms, el mínimo 380ms, el máximo 14400ms y el rango 14020ms. Si hacemos una clasificación de las palabras de acuerdo con las medias que presentan, obtenemos los resultados de la tabla 3.

Grupo	Palabras
(1)	nariz (714), ojo (732), dedo (743), boca (785), mano (785), lengua (883), corazón (1043), oreja (1060), muela (1072), brazo (1111), axila (1186), espalda (1192), moratón (1193), ombligo (1204), cabeza (1210), uña (1234), pierna (1356) y pestaña (1413)
(2)	vena (1507), barbilla (1517), frente (1543), muslo (1559), rodilla (1577), dedo pie (1579), muñeca (1667), orzuelo (1696), colmillo (1725), cuello (1733), campanilla (2075), codo (2246), riñón (2477), tobillo (2487), pantorrilla (2655) y talón (3215)

Tabla 3. Medias del tiempo de reacción de cada estímulo

En el gráfico 2 se puede ver el dendograma resultante del análisis jerárquico y en la tabla 4 las palabras pertenecientes a cada grupo. Si comparamos los resultados con los obtenidos en la tabla 2, constatamos que las palabras que aparecen en el grupo 2 están situadas todas por encima de la media y presentan las medias más altas.

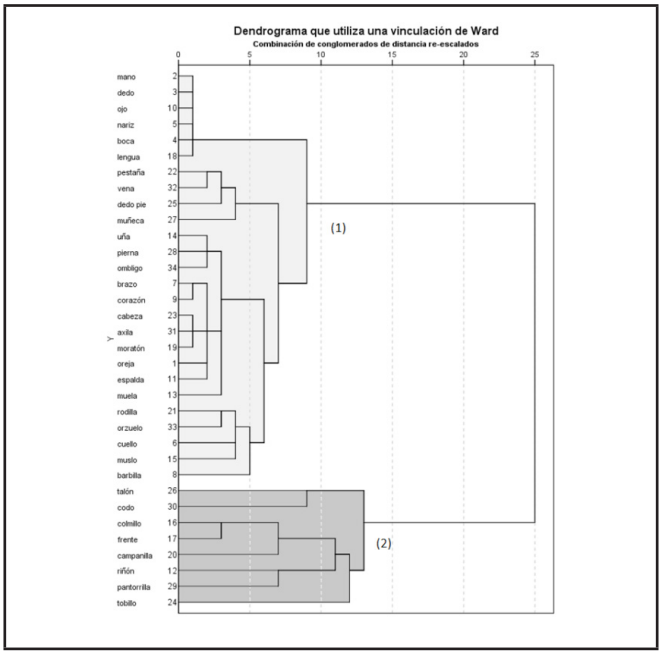

Gráfico 2. Dendograma de los principales grupos de palabras en español

Grupo	Palabras
(1)	oreja, mano, dedo, boca, nariz, cuello, brazo, barbilla, corazón, ojo, espalda, muela, uña, muslo, lengua, moratón, rodilla, pestaña, cabeza, dedo pie, muñeca, pierna, axila, vena, orzuelo y ombligo
(2)	riñón, colmillo, frente, campanilla, tobillo, talón, pantorrilla y codo

Tabla 4. Clasificación de las palabras según el método jerárquico

Si se realiza el análisis de conglomerados por el método de K medias asignando 2 conglomerados al inicio, se obtienen los resultados de la tabla 5 que, como se puede comprobar, no difieren de los obtenidos por medio del método jerárquico de conglomeración (tabla 3).

Grupo	Palabras
(1)	oreja, mano, dedo, boca, nariz, cuello, brazo, barbilla, corazón, ojo, espalda, muela, uña, muslo, lengua, moratón, rodilla, pestaña, cabeza, dedo pie, muñeca, pierna, axila, vena, orzuelo y ombligo
(2)	riñón, colmillo, frente, campanilla, tobillo, talón, pantorrilla y codo

Tabla 5. Clasificación de las palabras según el método K medias

Finalmente, cuando utilizamos el método de conglomeración FCM encontramos ciertas diferencias; así en la tabla 6 podemos observar que, asignando un índice pertenencia al grupo superior a 0,85 en el grupo "1", se obtiene un subgrupo de palabras que forma un área de transición (en la tabla se muestra el índice de pertenencia de cada palabra).

Grupo		Palabras
(1)	<0,85	oreja, mano, dedo, boca, nariz, cuello, brazo, barbilla, corazón, ojo, espalda, muela, uña, muslo, lengua, moratón, rodilla, pestaña, cabeza, dedo pie, pierna, pantorrilla, axila, vena, orzuelo y ombligo
	>0,85	riñón (0,76), campanilla (0,8) y tobillo (0,68)
(2)	<0,85	colmillo, frente, talón, codo
	>0,85	muñeca (0,62)

Tabla 6. Clasificación de las palabras según el método FCM

En resumen, para el español obtenemos una clasificación en dos grupos de palabras; en el grupo principal se aglomeran 26 palabras y 8 en el secundario; estas palabras se corresponden con las que tienen un mayor tiempo de reacción. Estos grupos coinciden utilizando el método jerárquico de conglomeración K medias. El método FCM muestra pequeñas divergencias pero también una mayor precisión al asignar los elementos a los conglomerados.

3.3 Análisis del tiempo de reacción en euskera

El número de respuestas válidas recogidas en euskera ha sido de 600, la media del tiempo de reacción es de 1748ms, la desviación 1760ms, el mínimo 260ms, el máximo 16970ms y el rango 16710ms. Si hacemos una clasificación de las palabras de acuerdo con las medias que presentan, obtenemos los resultados de la tabla 7.

Grupo	Palabras
(1)	ojo (693), nariz (734), boca (738), mano (757), oreja (864), dedo (959), brazo (1077), lengua (1085), muela (1141), muñeca (1277), cabeza (1300), espalda (1381), dedo pie (1431), codo (1494), rodilla (1521), cuello (1581), barbilla (1649) y moratón (1695)
(2)	pierna (1753), uña (1788), corazón (1875), ombligo (1891), muslo (2033), pestaña (2051), frente (2266), talón (2386), vena (2525), riñón (2756), colmillo (2793), tobillo (3000), axila (3115), campanilla (3169), pantorrilla (3718) y orzuelo (3953)

Tabla 7. Medias del tiempo de reacción de cada estímulo

En el gráfico 3 se muestra el dendograma resultante del análisis jerárquico y en la tabla 8 las palabras pertenecientes a cada grupo.

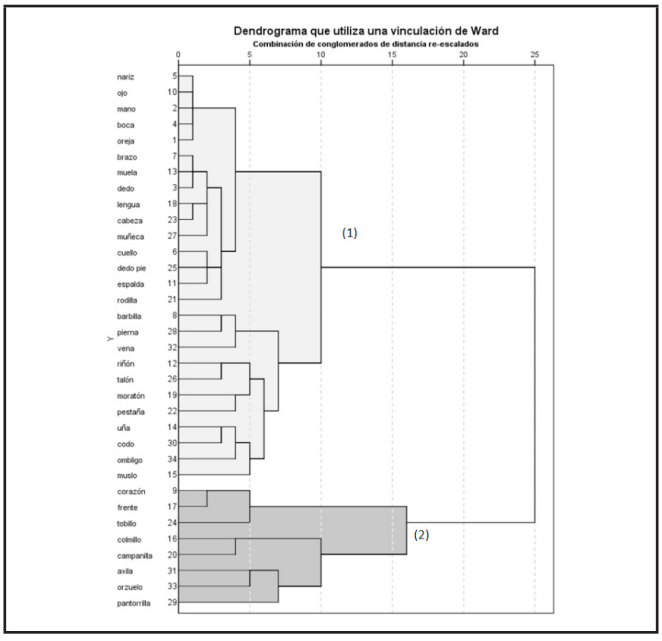

Gráfico 3. Dendograma de los principales grupos de palabras en euskera

Grupo	Palabras
(1)	oreja, mano, dedo, boca, nariz, cuello, brazo, barbilla, ojo, espalda, riñón, muela, uña, muslo, lengua, moratón, rodilla, pestaña, cabeza, dedo pie, talón, muñeca, pierna, codo, vena y ombligo
(2)	corazón, colmillo, frente, campanilla, tobillo, pantorrilla, axila y orzuelo

Tabla 8. Clasificación de las palabras en dos grupos según el método jerárquico de conglomeración

Al realizar el análisis de conglomerados por el método de K medias, se obtienen los resultados de la tabla 9 que, al igual que en español, coinciden con los obtenidos por medio del método jerárquico de conglomeración (tabla 3).

Grupo	Palabras
(1)	oreja, mano, dedo, boca, nariz, cuello, brazo, barbilla, ojo, espalda, riñón, muela, uña, muslo, lengua, moratón, rodilla, pestaña, cabeza, dedo pie, talón, muñeca, pierna, codo, vena y ombligo
(2)	corazón, colmillo, frente, campanilla, tobillo, pantorrilla, axila y orzuelo

Tabla 9. Clasificación de las palabras según el método K medias

Por último, al utilizar el método de conglomeración FCM encontramos un mayor grado de precisión al asignar cada elemento a su conglomerado correspondiente. En la tabla 10 se muestra el índice de pertenencia de cada palabra cuando es menor que 0,85.

Grupo		Palabras
(1)	<0,85	oreja, mano, dedo, boca, nariz, cuello, brazo, barbilla, ojo, espalda, riñón, muela, uña, muslo, lengua, moratón, rodilla, pestaña, cabeza, dedo pie, talón, muñeca, pierna, codo, vena y ombligo
	>0,85	pantorrilla (0,73)
(2)	<0,85	corazón, colmillo, axila, frente y tobillo
	>0,85	campanilla (0,82) y orzuelo (0,78)

Tabla 10. Clasificación de las palabras según el método FCM

Para el euskera obtenemos una clasificación en dos grupos de palabras; en el grupo principal se aglomeran 26 palabras y 8 en el secundario; las palabras se corresponden con las que tienen un mayor tiempo de reacción y, por tanto,

una disponibilidad menor. Estos grupos coinciden utilizando el método jerárquico de conglomeración así como el de K medias.

3.4 Comparación de los datos de las dos lenguas

En esta sección exponemos las convergencias y divergencias principales que se han observado. En el análisis basado en la comparación de las medias del tiempo de reacción de cada palabra, hemos podido comprobar que en ambas lenguas se da una coincidencia en la pertenencia al grupo en 22 palabras (64,71%). Las mayores diferencias las encontramos en aquellas palabras que pertenecen al grupo de mayor tiempo de reacción y, por tanto, de menor disponibilidad léxica.

Al aplicar los diversos métodos de conglomeración, las palabras se clasifican en dos grupos en todos los casos y en ambas lenguas. Otra coincidencia destacable es que en el grupo "1" aparecen 26 palabras y en el grupo "2" 8 en las dos lenguas. También resulta reseñable la coincidencia que se produce tanto con el método de conglomeración jerárquico como con el de K medias y el parecido de las divergencias obtenidas por el método de FCM.

El número de palabras que coinciden en la pertenencia a cada grupo en ambas lenguas es de 28 (82,35%), mientras que las no coincidentes (6, 17,65%) son 6.

4. CONCLUSIONES Y TRABAJOS FUTUROS

Además de las aportaciones de datos, este trabajo ha pretendido ser un estudio piloto de cara a futuras investigaciones. El empleo de múltiples métodos cuantitativos de objetivación hace que la mirada del investigador/a pase a un segundo plano, elevando así el carácter objetivo de los datos (Eguskiza, Gaminde y De Pablo, 2017).

Los tres métodos de conglomeración aquí empleados pueden ser complementarios entre ellos ya que, basándonos en el método jerárquico podemos posteriormente establecer los grupos a crear por los dos métodos restantes (K medias, FCM). Además, el método basado en la lógica difusa nos permite conocer las palabras de transición.

Somos conscientes de las limitaciones que tiene nuestra investigación, ya que todos nuestros informantes han sido personas de género femenino y la temática ha sido enfocada únicamente a las partes del cuerpo. Con lo cual, dejamos para futuras investigaciones la ampliación de la muestra, el campo léxico y el corpus y la introducción del género como variable.

REFERENCIAS BIBLIOGRÁFICAS

Aurrekoetxea, G. 2016. "Distantzia geografikoaren eta hizkuntza distantziaren arteko korrelazioa". En G. Aurrekoetxea, J.M. Makazaga y P. Salaberri (arg.). *Hire Bordatxoan: Txipi Ormaetxea Omenduz*. Bilbao: UPV/EHU.

Aurrekoetxea, G., Gandarias, L., Gaminde, I., Iglesias, A. 2014. "Variación prosódica en vasco: áreas acentuales". En Y. Congosto Martín, M.L. Montero Curiel y A. Salvador Plans (eds.), *Fonética Experimental, Educación Superior e Investigación*. Madrid: Arco Libros.

Barrutia, E. 2000. *Bizkaiko arrantzaleen hiztegia. Léxico marinero vizcaino*. Labayru Ikastegia eta BBK: Bilbo.

Boersma, P. y Weenink, D. 2016. *Praat: Doing Phonetics by Computer, Version 5.1*. URL: http://www.praat.org.

Clua, E., Iglesias, A., Usobiaga, I. y Salicrú, M. 2017. "Nola identifikatu ezaugarri esanguratsuenak bariazio dialektalean: Dia Tech nabarmentzeko aukera". En A. Iglesias y A. Ensunza (ed.), *Gotzon Aurrekoetxea lagunarterik hara*. Bilbao: UPV/EHU.

Eguskiza, N., de Pablo, I., Gaminde, I. 2017. "Hizkuntz hizkeren sailkapena metodo kuantitatiboen bidez". En *XXIV Jornadas de Psicodidáctica.* Bilbao: UPV/EHU, 403-416.

Etxebarria, A., Gaminde, I. Olalde, A. y Gaminde, U. 2016. "Hizkuntza aldakortasuna Larrabetzuko aditz morfologian". En A. Iglesias, A. Romero, A. Ensunza (eds.), *Linguistic variation in the Basque language and education – II / Euskararen bariazioa eta bariazioaren irakaskuntza – II*. Bilbao: UPV/EHU.

Euskaltzaindia. 2008. *Euskararen Herri Hizkeren Atlasa*. Euskaltzaindia Bilbo.

Gaminde, I. 2005. *Lezama Berbarik Berba*. Zorrizketan eta Lezamako Udala.

Gaminde, I. y Eguskiza, N. 2017. "Arratiako gazteen aditz laguntzaileen bariazioaz". En A. Iglesias y A. Ensunza (eds.), *Gotzon Aurrekoetxea lagunarterik hara*. Bilbao: UPV/EHU.

Gaminde, I., Olalde, A., Gaminde, U. y Etxebarria, A. 2016a. "Populazioaren sailkapena lexikoaren arabera. IKASTORRATZA". *e-Revista de Didáctica*, 16: 68-84. URL: http://www.ehu.es/ikastorratza/16_alea/4.pdf.

Gaminde, I., Etxebarria, A., Eguskiza, N., Romero, A. y Unamuno, L. 2016b. "Lexikoaren bariazioa eta multzokatze-azterketa". *Euskalingua* 28: 19-23.

Gaminde, I., Etxebarria, A., Romero, A. y Eguskiza, N. 2017. "Características de la competencia prosódica de jóvenes bilingües vascos en la lectura en voz alta: Las cumbres tonales". *RLA. Revista de Lingüística Teórica y Aplicada*, Concepcion (Txile), 55 (1), I Sem., 35-52.

Gaminde, I., Romero, A., Garay, U. y Etxebarria, A. 2011. "Los tonos de frontera de las oraciones interrogativas absolutas producidas por hablantes bilingües vasco-español" *UNAM* 54, 61-79.

Hernández Muñoz, N. 2006. *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos*. Salamanca: Universidad.

López Morales, H. 1995. "Los estudios de disponibilidad léxica: pasado y presente". *Boletín de Filología de la Universidad de Chile (Homenaje a Rodolfo Oroz)* 35: 245-259.

Zuazo, K. 2017. *Mendebaleko Euskara*. Donostia: Elkar.