

ÑEMITYRÃ

Revista Multilingüe de Lengua, Sociedad y Educación



ISSN: 2707-1642 (en línea)

ISSN: 2707-1634 (impreso)

ÑE'Ë
LENGUA
LANGUAGE
LÍNGUA
LANGUE
SPRACHE

Revista Multilingüe
de Lengua, Sociedad
y Educación

Volumen 5 - Número 2 - Agosto 2023

ÑE'Ë
LENGUA
LANGUAGE
LÍNGUA
LANGUE
SPRACHE

ÑEMITÿRÃ

Revista Multilingüe de
Lengua, Sociedad y Educación

ISSN (en línea): 2707-1642

ISSN (impreso): 2707-1634



ÑEMITÝRÃ

Revista Multilingüe de Lengua, Sociedad y Educación

Volumen 5 – Número 2 – Agosto 2023

Asunción, Paraguay

Edición: Instituto Superior de Lenguas, UNA, Py

AUTORIDADES:

Dra. Zully Concepción Vera

Rectora

Universidad Nacional de Asunción

Dr. Ricardo Pavetti

Decano

Facultad de Filosofía, UNA

EDITORA:

Valentina Canese, PhD

Instituto Superior de Lenguas, UNA

Teléfono: +595 21 207599

Avenida España 1009, Asunción, Paraguay

SOBRE LA REVISTA:

ÑEMITÝRÃ, Revista Multilingüe de Lengua, Sociedad y Educación se publica semestralmente en formato impreso y digital desde el Instituto Superior de Lenguas de la Facultad de Filosofía de la Universidad Nacional de Asunción, en Paraguay.

ÑEMITÝRÃ es una revista arbitrada que publica trabajos de investigación inéditos, notas y reseñas sobre temas de investigación en educación, cultura o sociedad en su interrelación con la lengua, ya sea inglés, alemán, portugués, francés, guaraní o español; con el objetivo de difundir y debatir avances e innovaciones. El contenido está dirigido a especialistas, investigadores, docentes y estudiantes del nivel superior

ÑEMITÝRÃ está editada por el Instituto Superior de Lenguas en conjunto con la Dirección de Investigación de la Facultad de Filosofía de la Universidad Nacional de Asunción.

Esta publicación tiene una licencia Creative Commons Atribución
No Comercial - Sin Derivadas 4.0 Internacional (CC BY-NC-ND 4.0)



EDITORA EN JEFE

Valentina Canese Caballero, PhD
Instituto Superior de Lenguas, UNA
Paraguay

EDITOR ASISTENTE

Carlos Anibal Peris, PhD
Instituto Superior de Lenguas, UNA
Paraguay

CONSEJO EDITORIAL

Resolución del Consejo Directivo N° 024-40-2018

Lic. Feliciano Acosta,
Instituto Superior de Lenguas, UNA
Paraguay

Dra. Ma. Eva Mansfeld,
Instituto Superior de Lenguas, UNA
Paraguay

Dr. Tadeo Zarratea,
Academia de la Lengua Guaraní
Paraguay

Dr. Carlos Paulino Ferreira,
Academia de la Lengua Guaraní
Paraguay

Mag. Rocio Mazzoleni,
Instituto Superior de Lenguas, UNA
Paraguay

Dra. Magdalena Rivarola,
Instituto Superior de Lenguas, UNA
Paraguay

Dra. Clarissa Menezes Jordão,
Universidade Federal de Paraná
Brasil

Dra. Katherine Mortimer,
University of Texas, San Antonio
USA

Dra. Susan Spezzini,
University of Alabama, Birmingham
USA

Mg. Valeria Vázquez Cubilla,
Instituto Superior de Lenguas, UNA
Paraguay

Dr. Luis Eduardo Wexell Machado,
Instituto Superior de Lenguas, UNA
Paraguay

Dr. Erwin Warkentin,
Memorial University
Canada

Dr. Friedhelm Guttandin,
Universidad Católica "Nuestra Señora de la
Asunción"
Paraguay

Dr. Paulo Astor Soethe,
Universidade de Paraná Brasil
Brasil

Dra. Gloria Medina,
Instituto Superior de Lenguas, UNA
Paraguay

Mg. Gabriela Rojas Caballero,
Instituto Superior de Lenguas, UNA
Paraguay

Dr. Guillaume Candela,
Universidad de Paris Sorbonne
Francia

Mg. Pablo Debuchy,
Instituto de Empresas de Madrid
España

Dra. Estela Mary Peralta,
Instituto Superior de Lenguas, UNA
Paraguay

Dr. José María Rodrigues,
Instituto Superior de Lenguas, UNA
Paraguay

Dr. José Antonio Alonso Navarro,
Universidad del Norte
Paraguay

Dra. Debora Racy Soares
Universidade Federal de Lavras
Brasil

Dr. Rubén González Vallejo
University College Dublin
Irlanda

CONTENIDO**DOSIER: PROSODIA Y VARIACIÓN LINGÜÍSTICA**

Prosodia y variación lingüística (Editorial) <i>Juan Abasolo</i> <i>Naia Eguskiza</i>	7
Motivación y estilo de aprendizaje en estudiantes de Educación Superior en la realización del Videocast <i>Aitor Iglesias</i> <i>Ariane Ensunza</i>	9
Análisis del ritmo y percepción intercultural de textos orales <i>Aintzane Etxebarria</i> <i>Naia Eguskiza</i>	21
El análisis lexicométrico, el corpus y el diccionario previo: caso de la lengua vasca <i>Juan Abasolo</i>	36
Contraste de métodos y variación lingüística <i>Iñaki Gaminde</i> <i>Naia Eguskiza</i>	52
Marcos y metodologías para la clasificación dialectal (con aplicación al Caso Vasco) <i>Ariane Ensunza</i> <i>Lorea Unamuno Gotzon</i> <i>Aurrekoetxea</i>	79
Notas sobre el Atlas Lingüístico Vasco <i>Xarles Videgain</i>	92

CONTENIDO

ARTICULOS CIENTÍFICOS VARIOS

Aspecto socio-cultural de la situación lingüística del Paraguay <i>Ígor Protsenko</i>	102
A memória discursiva em quadrinhos da Marvel sobre Thor <i>Bruno Aguinaldo Feitosa</i> <i>Rosemere de Almeida Agüero</i>	109
Educación bilingüe hacia la dignidad: Implicaciones para países polilingües como EE. UU. y Paraguay <i>Christian Faltis</i>	119
Writing Skill Training of tertiary students by using Letter Writing on All Topics App <i>Oraily Madruga Rios</i>	127
Architecture and Research Corpus Indexation of Texts by Yuri Shevelyov <i>Illia Danyliuk</i> <i>Anatolii Zahnitko</i> <i>Hanna Sytar</i>	137
Reseña de la novela de Miguel Zapata Ros: "CARADOC" (I y II parte) <i>José Antonio Alonso Navarro</i>	146

Editorial

Prosodia y variación lingüística

Juan Abasolo

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0002-1911-4118>

e-mail: juan.abasolo@ehu.eus

Naia Eguskiza

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0002-8928-7006>

e-mail: naia.egusquiza@ehu.eus

La diversidad lingüística es una riqueza que se manifiesta en cada rincón del mundo, y en este sentido, tanto Paraguay como el País Vasco se destacan como ejemplos emblemáticos. En Paraguay, la convivencia del guaraní junto al castellano no solo da cuenta de un bilingüismo constitucional desde la carta magna de 1992, sino también de una identidad y cultura profundamente arraigada en las sonoridades y saberes de sus lenguas. Por su parte, el País Vasco, con su milenaria lengua, presenta una experiencia única de resiliencia y hoy en día de revitalización lingüística que, ante los avatares de la historia, ha sabido mantener viva su esencia y hoy promover la coexistencia con el resto de las lenguas de su entorno.

Esta simbiosis lingüística, donde las lenguas comparten y modelan un territorio y a su gente, es el testimonio de la capacidad humana de adaptación y colaboración. En esta edición, buscamos tender un puente entre estos dos escenarios plurilingües, destacando los aprendizajes y desafíos que enfrentan y cómo estos pueden ser de relevancia mutua. Porque, al final del día, la lengua es más que un medio de comunicación: es identidad, historia y punto de encuentro.

El grupo de investigación EUDIA, que hoy tiene su sede en la Facultad de Educación de Bilbao de la Universidad del País Vasco (UPV/EHU), originalmente se creó para investigar la variación lingüística, específicamente del euskera, con la intención de especificar factores sociales o de otro tipo que la producen y el desarrollo de herramientas para esta investigación. Con una ampliación de objetivos se abrieron nuevas vías de investigación vinculadas a la didáctica de la lengua y el desarrollo de la competencia oral siempre en contextos multilingües como es el entorno en el que desarrollan su tarea los investigadores de diferentes universidades vascas. Este grupo refleja su compromiso con la conservación del patrimonio lingüístico y la colaboración internacional en el campo del multilingüismo.

El propósito central de esta colaboración entre el grupo EUDIA y el Instituto Superior de Lenguas de la UNA radica en tender puentes entre dos rincones del mundo que, aunque distantes geográficamente, comparten la riqueza y desafíos del multilingüismo. Al aportar investigaciones, EUDIA no solo pretende compartir descubrimientos y avances académicos, sino también proporcionar un reflejo de cómo se realiza eso en la lengua vasca. Creemos que este intercambio podría enriquecer a ambas partes: Paraguay puede obtener una visión fresca y diferente de la investigación en contextos formalmente bilingües, mientras que el País Vasco se beneficia al llevar su lengua y cultura a nuevos horizontes y recibir feedback de una nación

con su propia rica experiencia en el multilingüismo. Se busca generar un diálogo fructífero entre estas dos realidades y fortalecer la comunidad científica global en torno al estudio de las lenguas.

La serie de artículos que componen este número de ÑEMITÿRÃ abarcan una variedad temática que refleja la diversidad y profundidad con la que se aborda el estudio del euskera y el contexto plurilingüe en el País Vasco. Desde investigaciones sobre la percepción intercultural de textos orales y el ritmo de los textos orales, pasando por análisis lexicométricos y la creación de diccionarios de euskera para la utilización de herramientas de investigación, hasta estudios que profundizan en la variación lingüística y sus distintos métodos de análisis. Asimismo, se brinda una mirada crítica a las metodologías empleadas en la clasificación dialectal, poniendo de manifiesto las implicaciones de estas, desde el prisma de la realidad de los dialectos del euskera. Por último, se destaca la relevancia del Atlas Lingüístico Vasco como herramienta esencial para el estudio de la variación dialectal del euskera, mostrando la vastedad de datos que este compendio aporta a la investigación lingüística. En conjunto, estos trabajos ofrecen una panorámica holística sobre los desafíos, avances y particularidades del estudio del euskera, queriendo aportar también en la comunidad científica latinoamericana en general.

En América Latina, y especialmente en Paraguay, la convivencia de lenguas en un mismo territorio ha sido una constante histórica, cultural y social. Esta realidad lingüística, que implica retos y oportunidades, se refleja también en el País Vasco, donde el euskera hoy coexiste y se desarrolla en un entorno plurilingüe. Para investigadores paraguayos y latinoamericanos, los estudios presentados en esta edición podrían ofrecer una mirada detallada sobre algunas aristas del proceso de investigación, herramientas y reflexiones que podrían ser de interés en sus propias investigaciones.

Más allá de la geografía y las especificidades culturales, las lenguas son puentes que conectan realidades, experiencias y conocimientos. Para la comunidad científica en general, estos trabajos representan una oportunidad de establecer diálogos interdisciplinarios, crear conexiones entre distintas áreas de investigación y reflexionar sobre la importancia del multilingüismo en la construcción de sociedades más inclusivas y diversas. Desde esta perspectiva, invitamos a todos los lectores a adentrarse en estos estudios, a identificar puntos de encuentro y a establecer nuevas rutas de investigación que enriquezcan aún más el vasto panorama lingüístico y cultural de nuestra región.

La colaboración entre el grupo EUDIA y el ISL no es solo un acto académico, sino también un símbolo del poder unificador de las lenguas y de la investigación. Las voces del País Vasco resuenan ahora en Paraguay y en toda Latinoamérica, ofreciendo una perspectiva sobre la riqueza del estudio de la lengua vasca y también su contexto plurilingüe. Estas interacciones académicas transnacionales tienen el potencial de abrir caminos inexplorados y fomentar nuevas inquisiciones en territorios lingüísticos comunes y divergentes.

Nuestro deseo es que estos estudios no solo informen, sino que también inspiren. Que despierten la curiosidad de investigadores de distintos rincones del mundo para adentrarse en el estudio del euskera, del guaraní, o de cualquier otra lengua que coexista en un contexto bilingüe o plurilingüe. Al final del día, son estas conexiones y descubrimientos los que fortalecen los lazos académicos y culturales, construyendo puentes de entendimiento y cooperación entre diferentes regiones del mundo. Esperamos que esta edición sea solo el comienzo de muchas más colaboraciones fructíferas y de un diálogo constante en el vasto mundo del multilingüismo.

Artículo original

Motivación y estilo de aprendizaje en estudiantes de Educación Superior en la realización del Videocast¹

Motivation and learning style in higher education students making a videocast

Aitor Iglesias

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0001-5381-5957>

e-mail: aitor.iglesias@ehu.eus

Ariane Ensunza

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

e-mail: ariane.ensunza@ehu.eus

Recibido: 11/4/2023
Aprobado: 25/6/2023

RESUMEN

En el marco de un Proyecto de Innovación Educativa en la UPV/EHU, alumnos de tres facultades dentro de diferentes asignaturas dirigidas entre otras al desarrollo de competencias comunicativas, realizaron un videocast. En este trabajo se presentan las diferencias encontradas en los resultados de los cuestionarios que analizan la motivación (EMSI) y el desarrollo de los estilos de aprendizaje (CHAEA), según la facultad de pertenencia del alumnado. Se presentan diferencias estadísticamente significativas tanto en los estilos de aprendizaje como en algunas dimensiones de la motivación.

Palabras clave: innovación educacional; videocast; motivación; estilo de aprendizaje; enseñanza superior.

ABSTRACT

Within the framework of an Educational Innovation Project in the UPV/EHU, students from three faculties within different subjects directed among others to the development of communicative skills, made a videocast. This article presents the differences found in the results of the questionnaires that analyses motivation (EMSI) and the development of learning styles (CHAEA), according to the faculty to which student body belongs. There are statistically significant differences both in learning styles and in some dimensions of motivation.

Keywords: educational innovations; videocast; motivation; teaching; learning styles; higher education.

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

INTRODUCCIÓN

El Espacio Europeo de Educación Superior (EEES) fomenta facilitar a los universitarios y las universitarias el acceso a gran variedad de programas, a oportunidades para llevar a cabo estudios multidisciplinares, al perfeccionamiento de idiomas y a la habilidad para utilizar las nuevas tecnologías informativas. En el año 2009 el Espacio Europeo de Educación Superior (EEES) ve la necesidad de una reforma curricular continuada orientada hacia el desarrollo de resultados del aprendizaje potenciando tanto la innovación como la formación investigadora el alumnado.

El Ministerio de Educación, Cultura y Deporte comparte estos objetivos y asume la responsabilidad de modificar la estructura universitaria para la integración del sistema español en el Espacio Europeo de Enseñanza Superior, por lo tanto, por medio de la Ley Orgánica 6/2001, de 21 de diciembre, de Universidades se adoptan medidas para que las universidades impulsen la realización de programas dirigidos a la renovación metodológica de la enseñanza universitaria para el cumplimiento de los objetivos de calidad del Espacio Europeo de Educación Superior. El Sistema Universitario Vasco en el último Plan Universitario 2015-2018 plantea cuatro objetivos estratégicos por medio de distintas líneas de actuación, en el ámbito docente el objetivo es responder a las necesidades de formación superior de la sociedad vasca con modelos y metodologías docentes innovadores, y se especifican las siguientes líneas de actuación:

- L1. Formación, capacitación e innovación en capacidades y modelos educativos
- L2. Impulso del emprendimiento y de la inserción laboral
- L3. Consolidación de las estructuras y herramientas que favorecen la movilidad de estudiantes

Este modelo educativo impulsa la enseñanza-aprendizaje cooperativa y dinámica por medio de metodologías activas y con el apoyo de las tecnologías de la información y comunicación en un contexto de enseñanza plurilingüe, y además interviene de manera cooperativa toda la comunidad educativa: alumnado, profesorado, personal de administración y servicios[†].

Las metodologías didácticas han pasado de utilizar materiales más tradicionales como libros, transparencias y vídeos a otros que se pueden emplear en un espacio y tiempo indeterminado como pueden ser los podcasts de vídeo o videocast, se puede acceder a ellos en cualquier dispositivo y permite la reproducción de archivos audiovisuales en cualquier aparato como teléfonos móviles, smartphones y netbooks (Piñeiro Otero, 2012:1-2). Así mismo, es importante conocer y emplear el lenguaje audiovisual, puesto que las tecnologías de la información y los medios de comunicación exigen una nueva alfabetización (UNESCO, 2005; Pérez Tornero, 2012). Tal y como señala Pérez-Tornero (2005) la enseñanza con este tipo de herramientas se fundamenta en una metodología de investigación y construcción del conocimiento, con ejercicios de lectura crítica de los medios y actividades de producción y expresión.

Aguaded y Guerra (2012) defienden la idea de que por medio de la educación en medios se afrontan habilidades y competencias necesarias para la sociedad en la que vivimos, y subrayan la importancia de formar a las personas desde el ámbito educativo tanto formal como informal para desarrollar el pensamiento crítico y las habilidades comunicativas de producción y de comprensión.

[†] Ver: <https://www.ehu.eus/es/web/sae-helaz/ikd-curriculum-garapena-oinarriak>

En este contexto se ha llevado a cabo un proyecto innovador en el que participan alumnos de distintas áreas de estudio y que está basado en el aprendizaje cooperativo y dinámico para conseguir el reto de crear podcast de vídeo. El aprendizaje basado en retos ayuda a trabajar competencias como el trabajo colaborativo y multidisciplinario, la toma de decisiones, la comunicación, la ética y el liderazgo (Malmqvist, Radberg & Lundqvist, 2015), y además aporta algunos beneficios que Martín Álvarez (2014) haciendo referencia a Nie, Casmore y Cane (2008) menciona:

La motivación (la perspectiva afectiva a través de las nuevas tecnologías, el trabajo en equipo y de la interacción), la cognición (a través de la investigación, el debate, la relación del podcast con contenidos ya aprendidos y por aprender, el hallazgo de diferentes puntos de vista para trabajar un mismo tema, las preguntas de sus compañeros, la compartición de materiales...) y la metacognición (a través de la reflexión). (Martín Álvarez, 2014: 5)

De todas formas, todos aprendemos de manera diferente, con ejercicios de distinto tipo, e incluso a todos no nos gustan los mismos métodos ni lugares para aprender. Con todo esto podemos hablar de diferentes estilos de aprendizaje (Alonso et. al., 2007) que en la literatura científica se han definido y descrito de muchas formas (Hunt, 1979; Schmeck, 1982; Gregorc, 1979; Claxton y Ralston, 1978; Riechmann, 1979; Butler, 1982; Smith, 1988; Kolb, 1984; Leichter, 1973). Alonso et. al. (2007) mencionan que una de las definiciones más claras y ajustadas es la que propone Keefe (1988):

Los Estilos de Aprendizaje son los rasgos cognitivos, afectivos y fisiológicos, que sirven como indicadores relativamente estables, de cómo los discentes perciben, interaccionan y responden a sus ambientes de aprendizaje. (Alonso et. al., 2007: 48)

Según Alonso et. al. (2007), los estilos de aprendizaje pueden ser de tres tipos teniendo en cuenta las características del estudiante; (1) el estilo activo, que tiene como características principales ser animador, improvisador, descubridor, arriesgado y espontáneo; (2) el estilo reflexivo, que se caracteriza principalmente por ser ponderado, concienzudo, receptivo, analítico y exhaustivo; (3) el estilo teórico, cuyas características fundamentales son metódico, lógico, objetivo, crítico y estructurado. Podemos averiguar el estilo de aprendizaje que más predomina por medio del cuestionario CHAEA (Cuestionario Honey-Alonso de Estilos de Aprendizaje) en un único estudiante o en un grupo de alumnos y alumnas, y consta de tres partes que son: cuestiones acerca de datos personales y socio-académicos de los alumnos; instrucciones para la realización y 80 ítems con respuestas dicotómicas (+ o -); por último, el perfil de Aprendizaje numérico y gráfico (Alonso et. al., 2007). El baremo que se emplea para la interpretación de los resultados es el siguiente (Tabla 1):

Tabla 1. Baremo para la interpretación de los datos del cuestionario CHAEA

	10%	20%	40%	20%	10%
	Preferencia				
	Muy Baja	Baja	Moderada	Alta	Muy alta
Activo	0-6	7-8	9-12	13-14	15-20
Reflexivo	0-10	11-13	14-17	18-19	20
Teórico	0-6	7-9	10-13	14-15	16-20
Pragmático	0-8	9-10	11-13	14-15	16-20

Fuente: Alonso et. al., 2007: 114.

En palabras de Sabogal et. al. (2011), la manera de enseñar y aprender puede variar la motivación del alumnado siendo esta variable un elemento esencial para entender el rendimiento académico en el ámbito universitario. Podemos definir la motivación de la siguiente manera:

The self-determination theory (Deci & Ryan, 1985) constitutes a theoretical framework that helps to understand motivation within the educational setting (Deci, Vallerand, Pelletier, & Ryan, 1991). Self-determination implies the sense of freedom of choice and of feeling free to do what one has decided to do. (Martín-Albo y col., 2009: 800)

Se puede saber qué tipo de motivación tiene el alumnado por medio del cuestionario EMSI (Martín-Albo, et. al., 2009), y las respuestas de los 16 ítems que la componen responden a la siguiente pregunta: “¿Por qué participas actualmente en esta actividad?”. Los ítems que responden a esta pregunta hacen referencia a cuatro tipos de motivación propuestas en la teoría de la autodeterminación de Deci y Ryan (1985): motivación, regulación externa, motivación identificada y motivación extrínseca.

Por lo tanto, por medio de la aplicación los cuestionarios CHAEA (Alonso, et. al., 2007) y EMSI (Martín-Albo, et.al., 2009) se intenta conseguir los siguientes objetivos: definir el estilo de aprendizaje de los alumnos de distintas áreas y especificar el tipo de motivación de dichos alumnos en la consecución del proyecto que tiene como finalidad la creación de los videocast.

Por último, en las siguientes líneas se especifica el proceso de aplicación del proyecto, los análisis y los resultados de los datos.

IMPLEMENTACIÓN DEL PROYECTO INNOVADOR Y RESULTADOS

En las siguientes líneas se especifican el proceso de aplicación del proyecto, el análisis y los resultados en cuanto al área de estudio, motivación y estilo de aprendizaje.

Proceso de aplicación y características del proyecto

Los y las estudiantes siguieron los siguientes pasos para conseguir el reto que se les planteó:

- Presentación del reto por parte del profesorado.
- Colaboración entre el alumnado de distintas disciplinas y generación de ideas.
- Investigación y revisión en cuanto a los temas que se tratan, al lenguaje, al tipo de texto y a la parte técnica de estos recursos.
- Grabación de los vídeos dirigidos a niños de Educación Primaria.

En la tabla 2 se especifican las competencias básicas y transversales que se trabajaron por medio de este proyecto en el tiempo de 7 semanas, el número de alumnos que tomaron parte según el área de estudio, el grado en el que estaban matriculados y las asignaturas en las que se implementó el proyecto:

Tabla 2. Competencias trabajadas en cada grado y asignatura

<p>Facultad: Facultad de Ciencias Sociales y de la Comunicación. Grado: Grado en Comunicación Audiovisual. Asignatura: Técnicas de Creación Audiovisual. Competencias básicas: Conocer y aplicar la tecnología y los recursos para la realización y diseño de proyectos audiovisuales. Competencias transversales: Trabajo en equipo; Innovación. Tarea: Creación del vídeo según los textos recibidos. Tiempo: 7 semanas Número de alumnos: 57</p>
<p>Facultad: Facultad de Ciencia y Tecnología. Grado: Grado en matemáticas; Grado en física. Asignatura: Comunicación en lengua vasca: ciencia y tecnología. Competencias básicas: Buscar, comprender, sintetizar y examinar de manera crítica la información sobre ciencia; proyectos de investigación e informes técnicos, explicar por medio del texto oral y escrito los resultados de laboratorio; explicar temas del ámbito de las ciencias teniendo en cuenta el contexto comunicativo. Competencias transversales: Trabajo en equipo; Innovación. Tarea: Creación de los textos de las áreas de matemáticas y ciencias para el vídeo. Tiempo: 7 semanas Número de alumnos: 27</p>
<p>Facultad: Facultad de Educación de Bilbao. Grado: Grado en educación primaria. Asignatura: Desarrollo de la competencia Comunicativa –Lengua Castellana y Lengua Vasca- en Educación Primaria. Competencias básicas: Examinar, procesar y crear textos del contexto académico. Competencias Transversales: Trabajo en equipo; Innovación. Tarea: Creación de los textos literarios para el vídeo. Tiempo: 7 semanas Número de alumnos: 80</p>

Fuente: elaboración propia con base en lo investigado.

ANÁLISIS DE LOS DATOS Y RESULTADOS

Una vez realizado el proyecto, los alumnos y las alumnas que tomaron parte cumplieron los cuestionarios EMSI (Martín-Albo, et. al., 2009) y CHAEA (Alonso, et. al., 2007), los datos recogidos se analizaron por medio del lenguaje estadístico R (R Core Team 2018), mediante la interface RStudio (RStudio Team 2015), utilizando los paquetes asociados base, nortest (Gross and Ligges 2015), vcd (Meyer, Zeileis, and Hornik 2006), ggplot2 (Wickham 2016) y likert (Bryer and Speersschneider 2016).

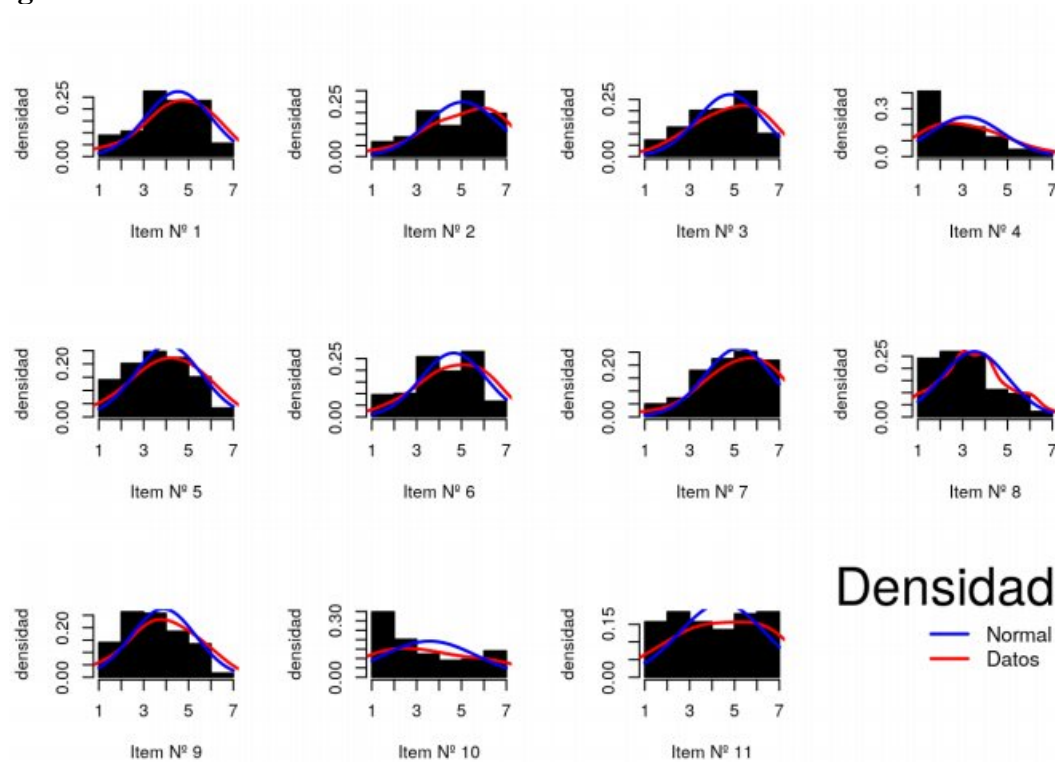
Dada la naturaleza de los datos recogidos, cantidad y escala, se podría optar por un tipo de análisis aplicando técnicas paramétricas, de mayor potencia y menos robustez a los incumplimientos de los presupuestos de normalidad y homocedasticidad, pero también podrían utilizarse técnicas no paramétricas, que son menos robustas a la hora de detectar diferencias significativas, pero también son más robustas ante distribuciones no normales o la falta de homogeneidad en la varianza de los datos.

Para esta primera toma de decisión se utilizaron técnicas de exploración visual de la distribución de los datos, y se estudió si los datos se distribuyen de acuerdo a la curva normal o si, por el contrario, existen evidencias que demuestren que los datos no se ajustan a la distribución normal, con un nivel de confianza del 95%.

Para el contraste de la normalidad de los datos se realizó el test de Kolmogorov-Smirnov con la modificación de Lillefors para media y varianza desconocidas.

Se puede aseverar que las distribuciones de las respuestas (Figura 1) del cuestionario EMSI no se ajustan a la distribución normal. En la inspección gráfica esto se ve inequívocamente en la distribución de los ítems 4, 10 y 11. Asimismo, el contraste estadístico nos reporta que tampoco en el resto de los ítems se puede asumir una distribución normal, con un nivel α del 0.05.

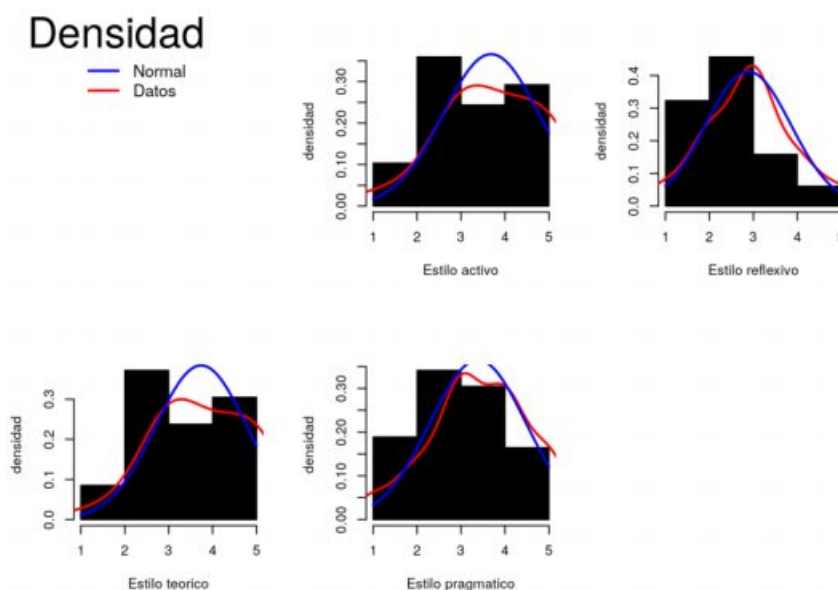
Figura 1. Distribución de los datos EMSI.



Fuente: elaboración propia con base en lo investigado.

En las respuestas del cuestionario CHAEA (Figura 2) se puede intuir una distribución diversa a la normal en las escalas de los estilos de aprendizaje teórico y activo. Aplicando a los resultados el test de hipótesis de una distribución diferente a la normal de Kolmogorof y Smirnov con la modificación de Lillefors se puede comprobar que la totalidad de las escalas tienen distribuciones que no se ajustan a la distribución normal. Estos primeros resultados nos llevan a presentar un análisis estadístico no paramétrico.

Figura 2. Distribución de los datos CHAEA.



Fuente: elaboración propia con base en lo investigado.

Partiendo de la premisa de que el alumnado de las diferentes áreas de estudio puntuará diversamente en las distintas escalas de motivación y también de desarrollo de los estilos de aprendizaje, se opta por aplicar pruebas estadísticas no paramétricas.

RESULTADOS SEGÚN LA MOTIVACIÓN

En primer término, presentamos los resultados del test de Kruskal-Wallis para detectar diferencias entre las respuestas a los ítems y la facultad de filiación del alumnado.

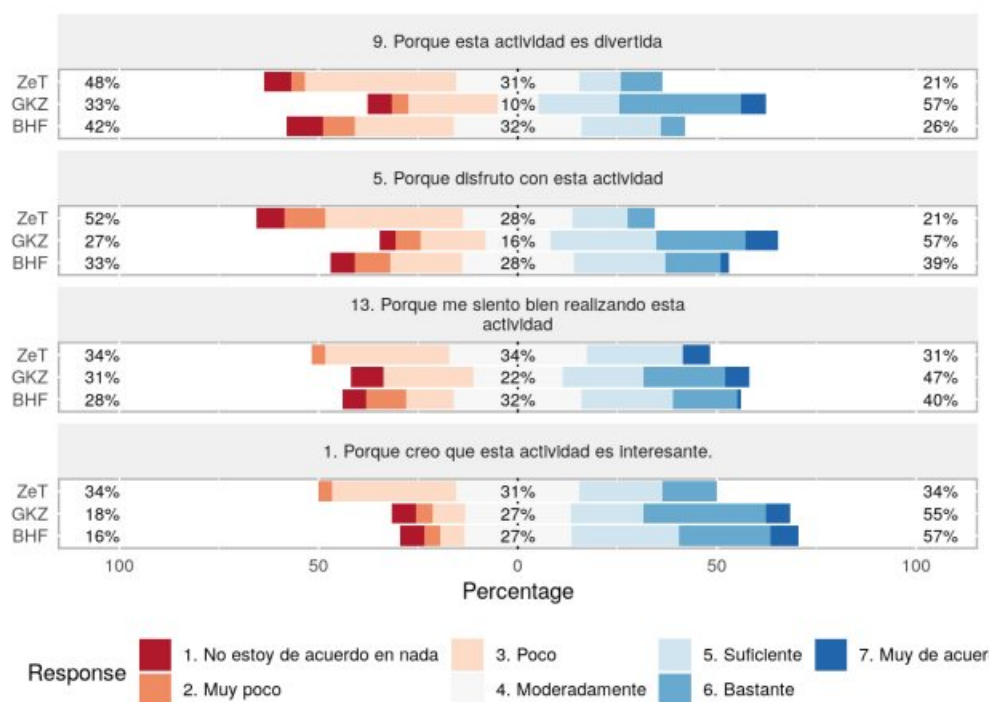
Tabla 3. Resultados según la motivación del alumnado.

Ítem	Chi cuadrado	p-valor	Sig.
1. Porque creo que esta actividad es interesante.	5.601	0.061	
2. Porque se supone que debo hacerlo.	1.276	0.528	
3. Por mi propio bien	3.869	0.144	
4. Puede que haya buenas razones para realizar esta actividad, pero yo no veo ninguna	3.670	0.160	
5. Porque disfruto con esta actividad	9.943	0.007	**
6. Porque creo que esta actividad es buena para mí	4.682	0.096	
7. Porque es algo que tengo que hacer	0.407	0.816	
8. Realizo esta actividad, pero no estoy seguro de si vale la pena	0.313	0.855	
9. Porque esta actividad es divertida	11.738	0.003	**
10. Por decisión personal	3.080	0.214	
11. Porque no tengo otra alternativa	2.320	0.313	
12. No lo sé. No veo qué me aporta esta actividad	3.023	0.221	
13. Porque me siento bien realizando esta actividad	1.426	0.490	
14. Porque creo que esta actividad es importante para mí	10.385	0.006	**
15. Porque creo que tengo que hacerlo	5.139	0.077	
16. Hago esta actividad, pero no estoy seguro de que sea conveniente continuar con ella	0.157	0.925	

Del total de los ítems solo se detectaron diferencias significativas en tres. Dos que miden la motivación intrínseca (los ítems 5 y 9) y otro que mide la regulación identificada (14).

Las diferencias significativas que se encontraron entre los alumnos de ZeT (Ciencia y Tecnología) y GKZ (Facultad de Ciencias Sociales y de la Comunicación) pertenecen al campo de la motivación intrínseca y se encontraron en un ítem: el quinto. También se encontraron diferencias significativas entre la GKZ y las otras dos facultades, en este caso en el noveno ítem (Figura 3).

Figura 3. Distribución de las respuestas acerca de motivación intrínseca según facultad de pertenencia.



Fuente: elaboración propia con base en lo investigado.

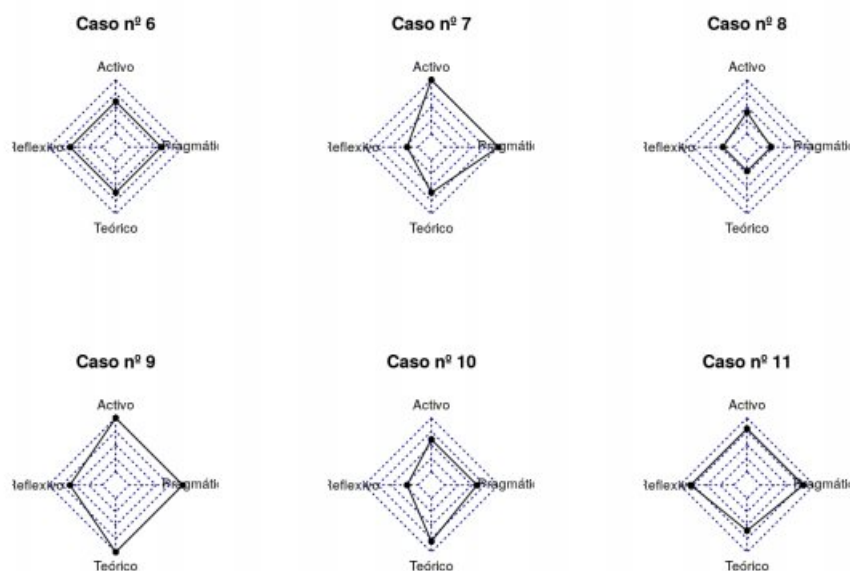
En el quinto ítem se plantea la cuestión de si disfrutan con la actividad, a la que cerca de dos tercios de los alumnos/las alumnas de la GKZ respondieron positivamente y poco más de la mitad respondieron negativamente en la ZeT. En el noveno ítem se plantea si es considerada divertida la actividad; al responder sobre la cuestión, el alumnado de la GKZ contesta mayormente que sí, con más de la mitad de las respuestas con valoraciones positivas y solo un cuarto en la BHF (Facultad de Educación de Bilbao) y un quinto en la ZeT. En esas facultades las respuestas negativas prácticamente doblan a las positivas y son más que las respuestas de carácter neutro.

Entre el alumnado de las carreras de educación y comunicación las valoraciones positivas sobre la importancia de la tarea en su propio desarrollo es mayormente positivo, con valoraciones positivas que pasan o rozan a la mitad del total. Por su lado, las valoraciones positivas de los alumnos de la ZeT son solo un sexto de las valoraciones totales y casi la mitad valora negativamente la actividad en función de su desarrollo. Del total de las diferencias que se encontraron, las únicas que son estadísticamente significativas, con un nivel de confianza del 95%, son las que se encontraron entre los encuestados de la Facultad de Educación de Bilbao (BHF) y los de la Facultad de Ciencia y Tecnología (ZeT).

RESULTADOS SEGÚN LOS ESTILOS DE APRENDIZAJE

El cuestionario CHAEA mide los estilos de aprendizaje desde una perspectiva cuatridimensional, en la que al unisono reporta en una escala del 1 al 5 cuál es el nivel de desarrollo de cada uno de los estilos de aprendizaje. En la figura 4 se muestran los resultados de 6 casos elegidos al azar, se puede apreciar que, si bien algunos son coincidentes en sus medidas, otros no lo son y que los diferentes desarrollos no parecen estar relacionados.

Figura 4. Resultados de estilo de aprendizaje de seis casos.



Fuente: elaboración propia con base en lo investigado.

Así, para poder contrastar de una manera equivalente las respuestas recogidas mediante este cuestionario, se optó por calcular cuáles son los estilos con mayor desarrollo en cada uno de los encuestados y, en base a ese dato, identificar si hay estilos de aprendizaje más prominentes entre el alumnado de alguna de las facultades, y en caso de que sí lo hubiera, identificar de qué estilo se trata.

El resumen de este análisis lo vemos en la tabla 4, en la que se identifican los niveles más desarrollados de sus hábitos de estudios según su pertenencia a una u otra facultad.

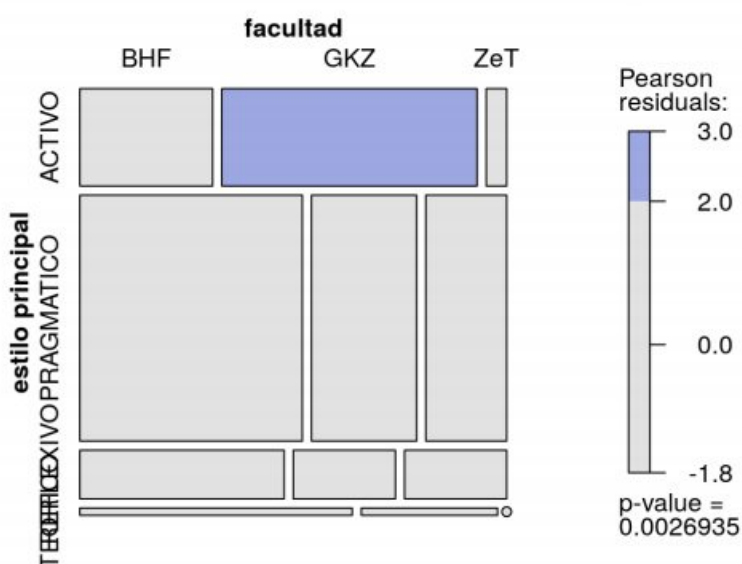
Tabla 4. Distribución en porcentajes de los estilos de aprendizaje según facultades.

	Activo	Pragmático	Reflexivo	Teórico
BH F	7.93	33.54	6.10	1.22
GK Z	15.24	15.85	3.05	0.61
ZeT	1.22	12.20	3.05	0.00

Fuente: elaboración propia con base en lo investigado.

Se aplicó a estos datos un estadístico de χ^2 el que, debido a la falta de casos de estilo prominentemente pragmático en la facultad ZeT, fue replicado 100.000 veces siguiendo la técnica de Montecarlo, para obtener un valor de probabilidad. Así, con una $\chi^2=20.068$ y un p-valor=0.00275 basado en las citadas réplicas, se encontró que sí hay diferencias significativas con respecto al desarrollo de hábitos de estudios según la facultad de pertenencia (Figura 5).

Figura 5. Diferencias significativas según el estilo de aprendizaje.



Fuente: elaboración propia con base en lo investigado.

Para identificar entre qué características de la distribución se encuentran estas diferencias, se procedió a utilizar el análisis visual de los residuos de la distribución χ^2 (Gráfico 4). El color azul del cuadrante GKZ/Activo de la imagen, da la información de que la proporción que encontramos ahí es mayor a la esperada. La tonalidad y la barra de la derecha de la imagen nos especifican que el tamaño del efecto de esta característica es moderado. Por lo que, en base estadística podemos asegurar que entre los estudiantes de GKZ hay un desarrollo moderadamente mayor que en BHF y ZeT de los hábitos de estudio activos, con un nivel de confianza del 95%.

CONCLUSIONES

Actualmente, el sistema universitario vasco que sigue las directrices del Espacio Europeo de Educación Superior (EEES), actúa de manera activa en la formación, capacitación e innovación en capacidades y modelos educativos. Por lo tanto, impulsa la enseñanza/aprendizaje cooperativa y dinámica por medio de metodologías activas. En el proyecto que se presenta siguiendo este marco pedagógico se utilizan como recurso las tecnologías de la información y de la comunicación.

Por lo tanto, el alumnado de tres áreas de estudio participa en un proyecto basado en el reto de crear un videocast, con la finalidad de trabajar, entre otras, las siguientes competencias: (A) La competencia comunicativa tanto escrita como oral y (B) La competencia de conocer y aplicar la tecnología y los recursos para la realización y diseño de proyectos audiovisuales. Una

vez realizado el proyecto se ha querido definir el estilo de aprendizaje de los participantes y el tipo de motivación, puesto que todos aprendemos de manera diferente, de la misma manera que no nos motiva lo mismo.

Los resultados de los análisis estadísticos nos muestran que entre el alumnado que participó en el Proyecto de Innovación Educativa los pertenecientes a la Facultad de Comunicación y Ciencias Sociales son mayormente de carácter más activo en sus estilos de aprendizaje y a la vez valoran la actividad como agradable. Entre los alumnos de la Facultad de Educación de Bilbao, a su vez, lo que se valora distintivamente es el aporte de la actividad en su formación profesional; en este alumnado se aprecia una tendencia a los estilos pragmáticos de aprendizaje, si bien no quedó corroborado esto por los análisis estadísticos. El alumnado de la Facultad de Ciencia y Tecnología identificó mayormente la actividad como tediosa, identificando a agentes externos como determinantes de que la tuvieran que realizar; con respecto a este alumnado tampoco se puede asegurar más que una tendencia a estilos reflexivos de aprendizaje, que no fueron corroborados por el análisis estadístico.

REFERENCIAS

- Aguaded, I., & Guerra, S. (2012). Razones para una educación mediática en la sociedad multipantallas. *Sphera Pública*, 12, 21-39.
- Alonso, C., Gallego, D. J., & Honey, P. (2007). Los estilos de aprendizaje. *Procedimientos de diagnóstico y mejora*.: Ediciones Mensajero.
- Bryer, J., & Speerschneider, K (2016). Likert: Analysis and Visualization Likert Items. [R package version 1.3.5]. <https://CRAN.R-project.org/package=likert>
- Butler, A. (1982). Learning Style Across Content Areas. NASSP.
- Claxton, C. S., & Ralston, Y. (1978). Learning Styles: Their Impact on Teaching and Administration. American Association for Higher Education.
- Deci, E. L., & Ryan, R. M. (2008). Self-Determination Theory: A Macrotheory of Human Motivation, Development, and Health. *Canadian Psychology*, 49(3), 182-185. doi:10.1037/a0012801.
- Departamento de Educación. (2018). Sistema Universitario Vasco. Planes Universitarios. Plan Universitario 2015-2018. <http://www.euskadi.eus/sistema-universitario-vasco-plan-universitarios/web01-a2hunib/es/>
- Gregorc, A. F. (1979). Learning/Teaching Styles: Potent Forces Behind Them. *Educational Leadership*, 5, 234-236.
- Gross, J., & Ligges, U. (2015). Nortest: Tests for Normality. [R package version 1.0-4]. <https://CRAN.R-project.org/package=nortest>
- Hunt, D. E. (1979). Learning Styles and student needs: An introduction to conceptual level. NASSP.
- Kolb, D.A. (1984). *Experimental Learning: Experience as the source of Learning and Development*. Englewood Cliffs. Prentice-Hill.
- Leichter, H. J. (1973). The concept of Educative Style. *Teachers College Record*, 75(2), 239-250.
- Ley Orgánica 6/2001, de 21 de diciembre, de Universidades. Boletín Oficial del Estado. Madrid, 24 de diciembre de 2001, (307) 1-58. Obtenido el 12 de mayo de 2019, desde HYPERLINK "<https://www.boe.es/buscar/pdf/2001/BOE-A-2001-24515-consolidado.pdf>" <https://www.boe.es/buscar/pdf/2001/BOE-A-2001-24515-consolidado.pdf>
- Malmqvist, J., Rådborg, K. K., & Lundqvist, U. (2015). Comparative Analysis of Challenge-Based Learning Experiences. Proceedings of the 11th International CDIO Conference, Sichuan (China): Chengdu University of Information Technology. Obtenido el 20 de marzo de 2019, desde: http://rick.sellens.ca/CDIO2015/final/14/14_Paper.pdf
- Martín Álvarez, J. (2014). Diseño de un proyecto de podcasting para la mejora de la expresión oral en la enseñanza de las lenguas. *Revista Científica de opinión y divulgación*. 28, 1-15.
- Martín-Albo, J., Núñez, J.L., & Navarro, J.G. (2009). Validation of the Spanish version of the Situational Motivation Scale (EMSI) in the educational context. *The Spanish journal psychology* 12(2), 799-807.
- Meyer, D., Zeileis, A., & Hornik, K. (2006). The Strucplot Framework: Visualizing Multi-Way Contingency Tables with Vcd. *Journal of Statistical Software*, 17 (3).

- Ministros Europeos de Educación Superior (2001). Declaración de Praga, 2001 Hacia el Área de la Educación Superior Europea. Obtenido el 20 de marzo 2019, desde <http://www.eees.es/es/documentacion>
- Pérez Tornero, J. M. (2005). Hacia un nuevo concepto de educación en medios. *Comunicar: Revista científica iberoamericana de la comunicación y educación*, 24, 21-24.
- Pérez Tornero, J. M. (2012). Un nuevo horizonte para la educación en medios. *Sphera Pública*, 12, 5-17.
- Piñeiro, T., & Caldevilla, D. (2011). Podcasting didáctico: una aproximación a su uso en el ámbito didáctico de la universidad española. *Sapiens: Revista Universitaria de Investigación*, 12(2), 14-30.
- R Core Team. (2018). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Riechmann, S. W. (1979). Learning Styles: Theri Role in Teaching Evaluation and Course Design. ERIC.
- RStudio Team. (2015). RStudio: Integrated Development Environment for R. RStudio Inc.
- Sabogal, L. F., Barraza, E., Hernández, A., & Zapata, L. (2011). Validación del cuestionario de motivación y estrategias de aprendizaje. *Psicogente*, 14(25), 36-50.
- Schmeck, R. R. (1982). Inventory of Learning Processes. *Educational Psychology* 11(34), 343-362.
- Smith, R. M. (1988). Learning how to learn. U.K. Open University Press.
- UNESCO. (2005). Informe mundial de la UNESCO. Hacia las sociedades del conocimiento. Ediciones de la UNESCO.
- Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.

Artículo original

Análisis del ritmo y percepción intercultural de textos orales¹

Rhythm analysis and intercultural perception of oral texts

Aintzane Etxebarria

Universidad del País Vasco (UPV/EHU), España

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

e-mail: aintzane.etxebarria@ehu.eus

Naia Eguskiza

Universidad del País Vasco (UPV/EHU), España

<https://orcid.org/0000-0002-8928-7006>

e-mail: naia.egusquiza@ehu.eus

Recibido: 21/4/2023

Aprobado: 15/6/2023

RESUMEN

Esta investigación muestra los resultados de la percepción intercultural de textos orales ateniendo a la descripción de un rasgo suprasegmental: el ritmo. El corpus lo comprenden nueve grabaciones de canciones breves, juegos de palabras y poesías recitadas por hablantes nativos de árabe, bereber, noruego, polaco, alemán, portugués, lituano y rumano. Así mismo, se han recogido 100 respuestas a un test de percepción diseñado por los investigadores y basado en las muestras obtenidas. El test lo han respondido 100 futuros maestros y maestras bilingües de la Comunidad Autónoma del País Vasco. Los resultados muestran que el alumnado identifica el tipo de texto en la mayoría de las lenguas aun no conociéndola, excepto en el caso del lituano. Esto nos lleva a deducir que las características suprasegmentales de los tipos de texto nos pueden ayudar a identificar lo que escuchamos aun no conociendo la lengua.

Palabras clave: prosodia; ritmo; competencia comunicativa intercultural; Euskera.

ABSTRACT

This research shows the results of cross-cultural perception of spoken texts by focusing on the description of a suprasegmental feature: rhythm. The corpus consists of nine recordings of short songs, puns and poems recited by native speakers of Arabic, Berber, Norwegian, Polish, German, Portuguese, Lithuanian and Romanian. In addition, 100 responses were collected to a perception test designed by the researchers and based on the samples obtained. The test was taken by 100 future bilingual teachers in the Autonomous Community of the Basque Country. The results show that pupils identify the type of text in most languages even if they do not know the language, except in the case of Lithuanian. This leads us to deduce that the suprasegmental characteristics of text types can help us to identify what we hear even if we do not know the language.

Keywords: prosody; rhythm; intercultural Communicative competence; Basque.

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

INTRODUCCIÓN

La competencia comunicativa intercultural, más allá de la competencia comunicativa, tiene como fin la comprensión y relación con personas de otras culturas (Byram, 1997). El concepto de competencia comunicativa intercultural surgía a mediados del siglo XX en los Estados Unidos. A posteriori se extendió el interés a otros lugares (Alsina, 1996). Actualmente, sin embargo, el debate se ha centrado en el tipo de actitudes que deben fomentarse en la sociedad receptora de personas migrantes, y es por ello que cobra interés la faceta afectiva de la competencia comunicativa intercultural (Sanhueza, & Cardona, 2009).

Se habla de comunicación intercultural cuando la comunicación interpersonal está influenciada por factores cuyo origen se encuentra en la percepción diferencial de la cultura y sus valores (Ruiz-Bernardo, 2012). Favorecer la comunicación intercultural implica conocer las creencias, visión de la vida, valores, costumbres, hábitos y estilos de vida de otras personas y tener la voluntad de superar las barreras que esa diversidad cultural pueda suponer. Para eso, son necesarios diversos conocimientos, actitudes y habilidades que componen la competencia comunicativa intercultural (Peñalva & Leiva, 2019; Pinilla, 2021; Vilà, 2006). Vilà (2007) define la comunicación intercultural de este modo:

La comunicación interpersonal en la que intervienen personas con unos referentes culturales suficientemente diferentes como para que se autoperciban, teniendo que superar algunas barreras personales y/o contextuales para llegar a comunicarse de forma efectiva (p.16).

Sin embargo, hay autores que defienden la redundancia de competencia comunicativa intercultural, ya que consideran que la competencia comunicativa *per se* lleva consigo la notación de interculturalidad sin necesidad de añadirla (Alsina, 2012; Jiménez & Sercu, 2007).

Los modelos de competencia comunicativa intercultural más destacables son el de Bennett (1998) y el de Chen y Starosta (1996). El primero se centra en la dimensión cognitiva mientras que el segundo se centra en tres dimensiones: la cognitiva, conductual y afectiva. La competencia afectiva permitirá a las personas, por un lado, ser suficientemente sensibles para el reconocimiento y el respeto de las diferencias culturales y, por otro lado, disfrutar de la comunicación interpersonal (Alsina, 2012). Vilà (2006), por su parte, también defiende tres componentes de distinta índole: aspectos cognitivos, de carácter afectivo y comportamentales.

Según Vilà (2007), para que la comunicación sea auténticamente intercultural deben darse dos elementos fundamentales: (1) Las personas que emprenden la aventura de comunicarse pertenecen a referentes culturales distintos, y además perciben tales diferencias culturales. (2) Las personas que entran en contacto valoran un grado aceptable o suficiente de comprensión mutua y de satisfacción en sus relaciones interpersonales, superando algunos de los obstáculos presentes en el intercambio intercultural. En este intercambio cultural es importante que se dé la mediación intercultural que se desarrolla como estrategia o herramienta para dar respuesta a poblaciones culturalmente diversas, principalmente personas de origen inmigrante. Es por eso que la mediación intercultural surge ante una necesidad de abordaje en un contexto de diversidad cultural. Esta mediación comenzada en organizaciones y ONGs ha dado paso al abordaje desde el ámbito administrativo. Giménez (1997) plantea la mediación intercultural como una modalidad de intervención de terceros en situaciones sociales de multiculturalidad significativa, orientada hacia la consecución del reconocimiento del otro y el acercamiento de las partes, la comunicación y comprensión mutua, el aprendizaje y desarrollo de la convivencia, la regulación de los conflictos y la adecuación institucional, entre actores sociales o institucionales culturalmente diferenciados (García & Barragán, 2004).

En el Marco Común Europeo de Referencia para las Lenguas la mediación se introdujo en la enseñanza y aprendizaje de lenguas, distanciándose del modelo de las cuatro destrezas, como uno de los modos de comunicación: comprensión, interacción, producción y mediación.

El uso de la lengua implica la mediación y combina la comprensión, la expresión y la interacción. Una lengua, además de para comunicar un mensaje, también se utiliza para desarrollar una idea o para facilitar la comprensión y la comunicación. El enfoque adoptado para la mediación no se reduce únicamente a la mediación interlingüística, sino que incluye también la mediación relacionada con la comunicación y el aprendizaje, así como la mediación social y cultural. Este enfoque es debido a la mayor diversificación que encontramos hoy en día en las aulas. La mediación forma parte del aprendizaje y, especialmente, del aprendizaje de lenguas (Ministerio de Educación y Formación Profesional e Instituto Cervantes, 2021).

La mediación intercultural consta de mediadores culturales que tienen como función recibir y transmitir información de forma comprensible y eficaz, para asumirla, codificarla y utilizarla (Rodríguez & Souto, 2009). Es posible que muchos docentes tengan la función de mediadores sociales en los centros educativos y que seguramente no conozcan todas las lenguas que se hablan en el ámbito académico. En estos casos la percepción intercultural a través de las características suprasegmentales que constituyen la competencia prosódica como parte de la competencia comunicativa (Hymes 1971, 1972) puede ayudar a interpretar la función pragmática de los textos que se escuchan.

PERCEPCIÓN PROSÓDICA: DEFINICIÓN Y CARACTERÍSTICAS

La percepción del habla se describe como un proceso de escucha, transformación y comprensión de la información lingüística:

La transformación de una señal acústica, producida por una persona, en un mensaje comunicativo, recibido por otra. Este modo particular de escucha implica una “sintonización” del aparato auditivo humano con las características acústicas particulares de la señal del habla, que permite la extracción y comprensión de información a partir del procesamiento de la señal. (Barón et al., 2014:74)

Si la percepción es adecuada, se descifra el sentido del mensaje que el emisor quiere hacer entender (función pragmática), puesto que “la intencionalidad del hablante, las características del contexto y los condicionantes interpretativos del oyente intervienen necesariamente en su conformación final” (Gutiérrez Ordóñez, 2010, p.322). La prosodia es uno de los pilares básicos en la percepción y en general en el desarrollo de la competencia comunicativa oral para poder alcanzar una comunicación con éxito, ya que comprende la melodía, tono, velocidad, pausas e inflexiones conferidas al discurso que ayudan a comprender y dar a entender, entre otras características, las emociones y los estados afectivos del hablante (Saá, 2001). Así mismo, es una de las características más importantes del mensaje oral, puesto que tiene una evidente función pragmática (Wharton, 2012). Wichmann y Blakemore (2006) llegan a la conclusión de que los elementos suprasegmentales contribuyen a deducir el sentido de una oración a través de la percepción:

...prosody/intonation is a kind of paralinguistic overlay on an utterance whose referential meaning has already been identified by other means. However, (...) prosody can also contribute to the identification of utterance meaning. While some papers are concerned with continuously variable parameters, such as relative pitch range, others focus on the role that phonological choices can play in conveying meaning. The choice of a rise or fall, or the placement of a pitch accent, may be as important a cue to speaker meaning as its phonetic realisation. (Wichmann & Blakemore, 2006:1537)

El estudio de las lenguas tanto de su sistema como de su proceso de instrucción se realiza desde el punto de vista de la producción y de la percepción; se han analizado las propiedades fonéticas y de los procesos fonológicos, y además se han estudiado la percepción del oyente y

se han realizado investigaciones que aportan información sobre la interpretación de los enunciados en las interacciones comunicativas (Gil, 2015).

La mayoría de estos estudios se fueron llevados a cabo por medio de diseños experimentales (Cortés, 2005). Algunos de ellos estudian la percepción, y las dificultades de la lengua en los niveles segmentales y suprasegmentales de los alumnos de segundas lenguas (Cortés, 2001; Kimura et al., 2010; Martínez Celadrán, 1993; Martínez y Rojas, 2011; Muñoz, 2015; Romero, 1988, Ruiz, 2018). Otros investigan los niveles mencionados en los hablantes nativos como, por ejemplo, en aquellos que tienen como lengua materna el español, e intentan responder a la cuestión de si son capaces de distinguir los acentos tonales por medio de los test de percepción (Díaz Campos y Ronquest, 2007). En lo que se refiere a las investigaciones de la lengua vasca, se han obtenido resultados a través de las pruebas de percepción en la identificación de las emociones (Gaminde et al., 2015), donde los resultados del test indicaron que la emoción mejor percibida fue la tristeza y la peor el enfado; y en la valoración de los distintos tipos de oraciones se demostró la influencia de la lengua materna (Gaminde et al., 2011); por último, hay quienes desarrollan o crean herramientas para auxiliar a investigadores de distintas áreas para que puedan realizar test de percepción (Díaz Campos y Ronquest, 2007).

Las pruebas de percepción se realizan a través de test específicos, muchos de ellos manipulados en los laboratorios de fonética, y tienen como objetivo “evaluar el proceso perceptivo donde el examinador recurre a tareas con estímulos contrastivos que eliciten respuestas observables” (Cervera e Ygual, 2001, p.57). Con todo esto, la cuestión que se intenta responder en este trabajo es si los futuros maestros y maestras que estudian los Grados de Educación Infantil y Primaria perciben el tipo de texto oral que escuchan en una lengua totalmente desconocida[†], para saber si la melodía de la lengua que producen los elementos suprasegmentales, contribuye a la percepción del tipo de texto.

MÉTODO

Para la recogida del corpus con el que se ha desarrollado este trabajo, han participado nueve personas hablantes de árabe, bereber, noruego, polaco, alemán, portugués, ruso, lituano y rumano, a las cuales se les han grabado varios textos orales en su lengua materna. Además, con las muestras recogidas se ha diseñado un test de percepción al que han contestado 100 alumnos bilingües futuros docentes de la Comunidad Autónoma Vasca (CAV), con el fin de saber si los futuros maestros y maestras identifican el tipo de texto, canciones, juegos de palabras o poesías, que escuchan en una lengua que ninguno de estos participantes entendía. Estos 100 alumnos y alumnas bilingües de la CAV son estudiantes de los Grados de Educación Infantil y Educación Primaria, con edades comprendidas entre 19 y 25 años. La mayoría de las que han respondido la encuesta son mujeres (78%), el 67% de los encuestados tiene como lengua materna el castellano, pero todos ellos son bilingües (vasco y castellano), y el 72% se ha escolarizado en el modelo D, es decir la lengua vehicular en su educación fue el euskera.

CARACTERÍSTICAS DEL CORPUS RECOGIDO

El corpus consta de nueve muestras grabadas a informantes de diferentes lenguas. Estas muestras se agrupan en canciones cortas (árabe, noruego y polaco), juegos de palabras (alemán, ruso y portugués) y breves poesías recitadas (bereber del Rif, rumano y lituano). Los hablantes de estas lenguas son hombres y mujeres jóvenes originarias de los países mencionados y los textos orales (canciones, juego de palabras y pequeñas poesías recitadas) pertenecen al patrimonio cultural inmaterial de estos pueblos.

En la tabla 1 se presentan las características de las muestras:

[†] Los encuestados no conocían ninguna de estas lenguas, aunque sí habían oído hablar sobre ellas.

Tabla 1. Características de la muestra

Tipo	Lengua	Código de la muestra	Número de pausas	Número de intervalos vocálicos	Número de intervalos consonánticos
Canción	Árabe	C-ara	11	55	52
Canción	Noruego	C-nor	16	32	27
Canción	Polaco	C-pol	7	29	31
Juego de palabras	Alemán	J-deu	8	13	21
Juego de palabras	Portugués	J-por	5	15	13
Juego de palabras	Ruso	J-rus	3	18	18
Pequeña poesía recitada	Bereber	R-ber	2	8	8
Pequeña poesía recitada	Lituano	R-lit	4	19	18
Pequeña poesía recitada	Rumano	R-ron	5	28	27

Elaboración propia con base en lo investigado.

Seguidamente, se presentan los textos de las grabaciones. Primero se muestran las canciones cortas en árabe, noruego y polaco. Posteriormente, los juegos de palabras en alemán, portugués y ruso. Por último, se muestran las rimas en bereber, rumano y lituano. Primero se ofrece la transcripción y seguido se ofrece la traducción al castellano, exceptuando en el caso del bereber, ya que no hemos obtenido la transcripción, debido al bajo nivel de alfabetización entre los hablantes de esta lengua minorizada en nuestro entorno.

CANCIONES CORTAS

Árabe:

شاطر شاطر شاطر شاطر
اللي يدرس في مدرستو، بنحبو وبنضل نولو
شاطر شاطر شاطر شاطر
اللي بيسعد ريفياتو
واللي بيما يحمسشي صحباتو
شاطر شاطر

Trasliteración:

shatir shatir shatir shatir allly yudras fi madrastu, binahbu wabinadl nawlw
shatir shatir shatir shatir allly biseud rifyatu wallly byma yahmasshi sahbatu
wabandalah wanaquluh
shatir shatir

Traducción al castellano:

Shater Shater Shater Shater, que enseña en su escuela, Benhabbu y Bandallo, Shater,
Shater, Shater, que es feliz, Rivato, y no tiene amigos.

Noruego:

Hode, skulder, kne og tå,
kne og tå
Hode, skulder, kne og tå,
kne og tå
Øyne, ører, kinn å klappe på
Hode, skulder, kne og tå, kne og tå.

Traducción al castellano:

Cabeza, hombro, rodilla y dedo del pie,

rodilla y dedo del pie
 Cabeza, hombro, rodilla y dedo del pie,
 rodilla y dedo del pie
 Ojos, orejas, barbilla para acariciar
 Cabeza, hombro, rodilla y dedo del pie, rodilla y dedo del pie.

Polaco:

Sto lat, sto lat - niech żyje, żyje nam
 Sto lat, sto lat - niech żyje, żyje nam!
 Jeszcze raz, jeszcze raz - niech żyje, żyje nam
 Niech żyje żyje nam

Traducción al castellano:

Feliz cumpleaños, feliz cumpleaños, feliz cumpleaños
 Cien años, cien años, ¡viva, viva nosotros!
 Una vez más, una vez más: viva, viva nosotros
 Viva nosotros

JUEGOS DE PALABRAS

Alemán:

Fischers Fritz físcht frische Fische.
 Frische Fische físcht Fischers Fritz.

Traducción al castellano:

Fischers Fritz pesca pescado fresco.
 Fritz, el pescador de pescado fresco.

Portugués:

O rato roeu a rolha da garrafa do rei da Rússia

Traducción al castellano:

El ratón roe la tapa del garrafón del rey de Rusia

Ruso:

Полетели, полетели, полетели - и на головку сели!

Transliteración:

Poleteli, poleteli, poleteli - i na golovku seli!

Traducción al castellano:

Volamos, volamos, volamos y nos sentamos en la cabeza.

RIMAS

Bereber[‡]:

Traducción al castellano:

Más rojo que la henna ...

Rumano:

Melc, melc codobelc,

Tu te duci la baltă

Azi bei apă caldă

Tu te duci la Dunăre

Azi bei apă tulbure

Traducción al castellano:

Melc, melc codobelc,

Vas a ir al charco

Hoy estás bebiendo agua caliente

Vas al Danubio

Hoy bebes agua turbia

Lituano:

Eni beni diki daki

urbi šurbi šmiki šmaki,

Aus baus bus medaus,

stora boba kiaulę pjaus.

Traducción al castellano:

Algunos de ellos fueron muy generosos

urbi shurbi shmiki shmiki shmaki,

Aus será miel,

El gordo boba cosechará el cerdo.

El test de percepción se llevó a cabo en la Facultad de Educación de Bilbao de la UPV/EHU. El alumnado escuchó los textos y eligió una de las opciones que se les daba (canciones cortas, juegos de palabras o breves poesías recitadas) mediante un cuestionario on-line del cual se extrajeron los resultados.

RESULTADOS

En este apartado se muestran, por un lado, los análisis de los resultados de las métricas rítmicas y, por otro lado, los resultados de percepción de los estímulos.

ANÁLISIS DE LOS RESULTADOS DE LAS MÉTRICAS RÍTMICAS

En esta sección se aborda el análisis de los resultados de la aplicación de las diferentes métricas que se utilizan para el análisis rítmico de la lengua, mediante el análisis de las etiquetas aplicadas a las muestras a través del software Correlatore (Mairano & Romano, 2009).

Primeramente, presentamos los resultados obtenidos en las métricas propuestas por Ramus (1999) en las que se presentan tanto los porcentajes de los intervalos vocálicos (%V) de las muestras como las desviaciones estándar de los componentes vocálicos (ΔV) y consonánticos (ΔC), en la tabla 2, figura 1 y figura 2.

[‡] Esta muestra corresponde al tarifit o rifeño, que es una variedad de la lengua bereber hablada en la región del Rif.

El %V da unos resultados muy próximos en la mayoría de los textos orales analizados, salvo en alemán y bereber en el caso de los componentes vocálicos. Sin embargo, en lo que se refiere a los elementos consonánticos, se aprecia una mayor diferencia, sobre todo en las canciones polacas y noruegas.

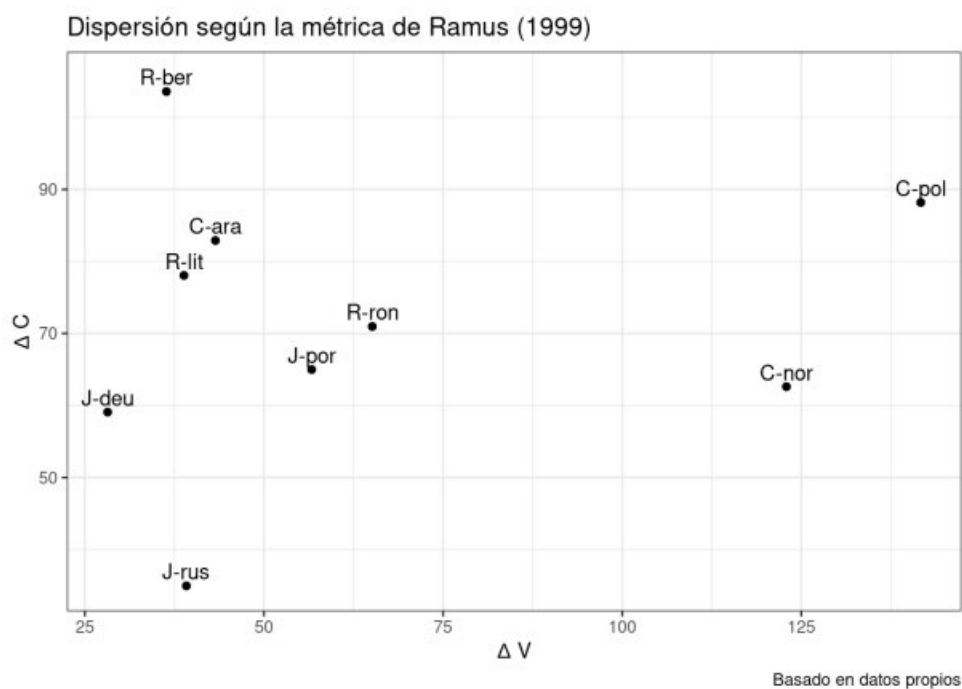
Tabla 2. Resultados de la métrica de Ramus (1999)

Muestra	%V	ΔV	ΔC
C-ara	41.98		43.22
C-nor	63.64		122.92
C-pol	64.69		141.67
J-deu	21.86		28.18
J-por	47.87		56.66
J-rus	55.29		39.17
R-ber	27.31		36.38
R-lit	42.89		38.83
R-ron	49.94		65.12

Elaboración propia con base en lo investigado.

En la tabla 2 y en la figura 1, se muestran los resultados obtenidos con las métricas de Ramus (1999). En la figura 1, se aprecia que por un lado están las pequeñas poesías recitadas en bereber, noruego y lituano, la canción en árabe y los juegos de palabras en alemán, ruso y portugués, y, por otro lado, las canciones en polaco y noruego.

Figura 1. Resultados de la métrica de Ramus (1999), dispersión según desviación estándar de los componentes



Elaboración propia con base en lo investigado.

Siguiendo la propuesta de métrica planteada por Dellwo (2006), se propone un replanteamiento de ΔC estudiando la covarianza en la distribución, mencionada como varcoC y varcoV el replanteamiento para el análisis de los intervalos vocálicos. Así, seguidamente presentamos en la tabla 3 y la figura 2 los resultados de esa métrica.

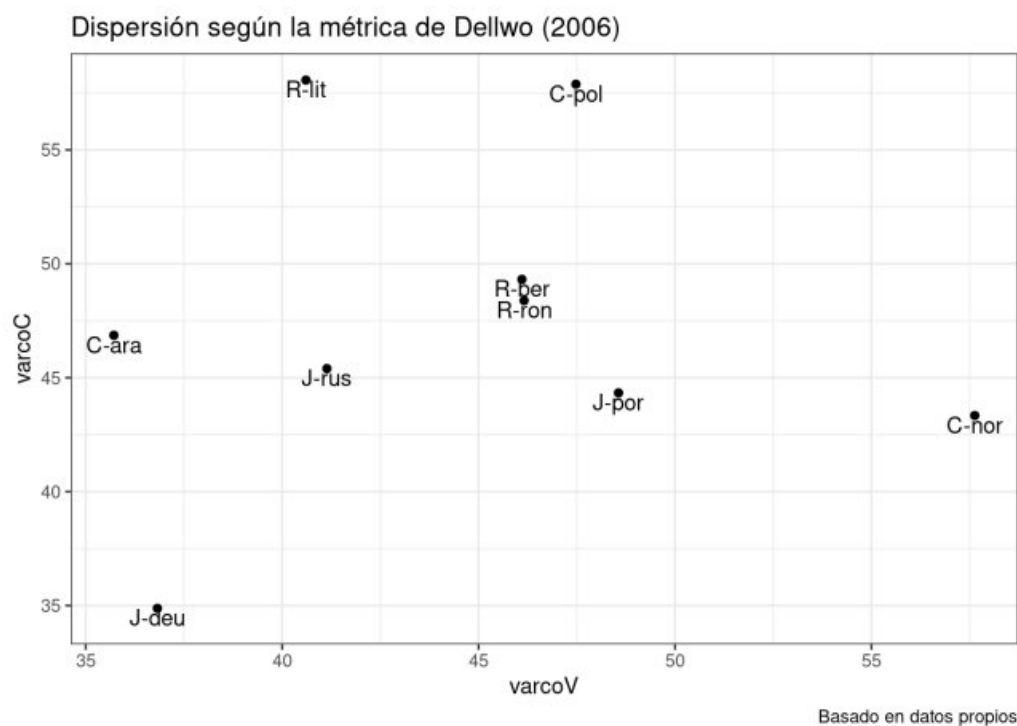
Tabla 3. Resultados de la métrica de Dellwo (2006)

Muestra	varcoV	varcoC
C-ara	35.71	46.86
C-nor	57.63	43.34
C-pol	47.48	57.88
J-deu	36.82	34.88
J-por	48.56	44.34
J-rus	41.14	45.40
R-ber	46.10	49.32
R-lit	40.60	58.06
R-ron	46.16	48.39

Elaboración propia con base en lo investigado.

En la tabla 3 y figura 2 se observa que la canción de noruega presenta las puntuaciones más altas en el varcoV y la pequeña poesía recitada en lituano. La dispersión según la métrica de Dellwo (2006) que se muestra en la figura 2 nos indica dos extremos; por una parte, están la pequeña poesía recitada en lituano y la canción en polaco, y en el otro extremo, está el juego de palabras en alemán.

Figura 2. Distribución de la muestra según valores varco vocálico y consonántico



Elaboración propia con base en lo investigado.

Desde un planteamiento de estandarización, Grabe y Low (2002) propusieron las métricas de índices de variabilidad por pares, abreviado PVI (*Pairwise Variability Indices* en inglés), en la que se oponen las variaciones de la longitud de los espacios vocálicos a los espacios intervocálicos, excluyendo las pausas. Así se disponen las métricas *rPVI* (medida no normalizada, del inglés *raw*) y *nPVI* (medida normalizada, del inglés *normalized*). Los resultados recogidos en la muestra se presentan en la tabla 4 y en la figura vinculada 3.

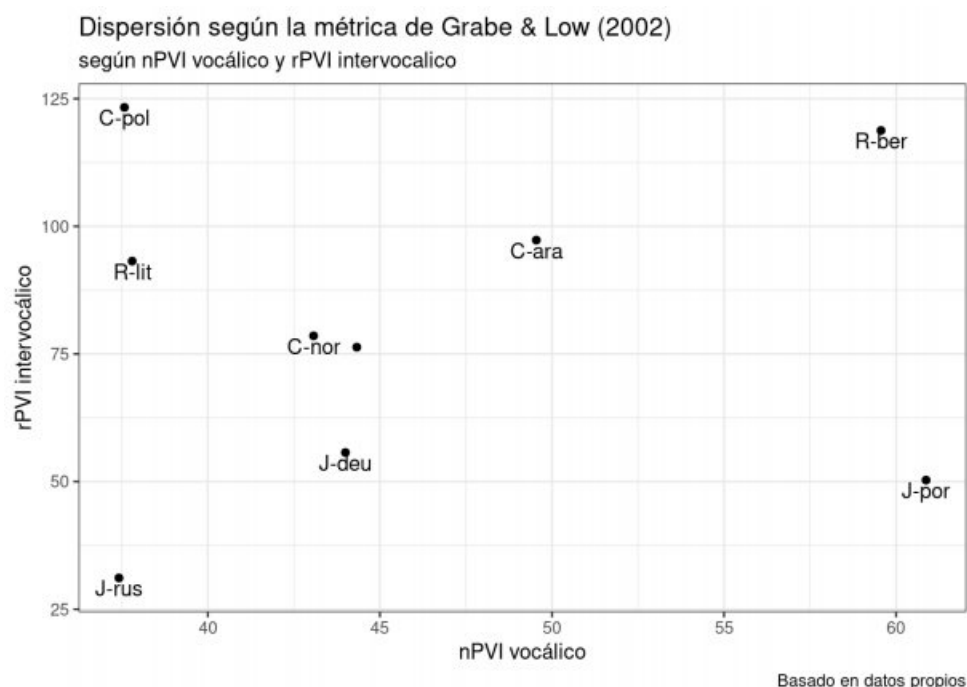
Tabla 4. Resultados de la métrica de Grabe & Low (2002)

Muestra	nPVI V	rPVI C
C-ara	49.55	97.29
C-nor	43.07	78.55
C-pol	37.57	123.30
J-deu	44.00	55.70
J-por	60.87	50.29
J-rus	37.42	31.13
R-ber	59.56	118.77
R-lit	37.80	93.18
R-ron	44.33	76.32

Elaboración propia con base en lo investigado.

Tal y como se aprecia en la tabla 4 y figura 3, los valores más altos en nPVI V se dan en el juego de palabras del portugués y en rPVI C en la pequeña poesía recitada del bereber. En la figura 3 se muestra de manera gráfica la dispersión según la métrica de Grabe y Low (2002).

Figura 3. Distribución de la muestra según los índices de variabilidad por pares.



Elaboración propia con base en lo investigado.

Como puede observarse en la figura 3, la poesía recitada del bereber y el juego de palabras del portugués se alejan del resto.

En este apartado se han presentado los análisis de las métricas propuestas por los siguientes autores: Ramus (1999), Dellwo (2006) y Grabe & Low (2002). Se observa que algunas producciones orales presentan una estructura silábica compleja, y como consecuencia se dan valores más altos en ΔC y varco C. Es el caso de las canciones noruegas y polacas (ΔC canción noruega: 122, 92; y en la canción polaca: 141, 67). En cuanto al valor más alto del varco C, se da en la canción polaca y en la pequeña poesía recitada en lituano. El valor varco C más bajo se muestra en el juego de palabras del alemán. Por último, el valor rPVI, que es indicador de estructuras silábicas más complejas, es más elevado en la canción polaca, en la pequeña poesía recitada en bereber y en lituano.

RESULTADOS DE PERCEPCIÓN DE LOS ESTÍMULOS

En el siguiente apartado se muestran los resultados obtenidos en el test de percepción por los futuros maestros y maestras.

Tabla 5. Resultados del test de percepción según lengua y tipo de texto

	Frecuencia	Porcentaje de aciertos
ARABE		
Juego de palabras	95	95,0
Recitado	5	5,0
BEREBER		
Canción	96	96,0
Juego de palabras	1	1,0
Recitado	3	3,0
ALEMÁN		
Juego de palabras	87	87,0
Recitado	13	13,0
RUMANO		
Canción	93	93,0
Juego de palabras	3	3,0
Recitado	4	4,0
NORUEGO		
Canción	100	100,0
PORTUGUÉS		
Juego de palabras	16	16,0
Recitado	84	84,0
RUSO		
Canción	2	2,0
Juego de palabras	80	80,0
Recitado	18	18,0
POLACO		
Canción	77	77,0
Juego de palabras	21	21,0
Recitado	2	2,0

Elaboración propia con base en lo investigado.

Como se puede observar en la tabla 5, los resultados del test de percepción nos dan a entender que los futuros maestros y maestras identifican el tipo de texto (canción, juego de palabras y pequeñas poesías recitadas) en la mayoría de las lenguas que desconocen. Sin embargo, en lo que respecta al lituano se observa que se han equivocado más, puesto que principalmente confundían la canción con juego de palabras, tal y como se muestra en la tabla 6:

Tabla 6. Resultados del lituano

	Frecuencia	Porcentaje	% válido	% acumulado válido
Canción	20	20,0	20,0	20,0
Juego de palabras	29	29,0	29,0	49,0
Recitado	51	51,0	51,0	100,0
Total	100	100,0	100,0	

Elaboración propia con base en lo investigado.

CONCLUSIONES

Cualquier subdisciplina de la lingüística, ya sea semántica, fonología o pragmática, está relacionada con procesos interculturales. Así, al igual que es posible impartir morfosintaxis o morfosintaxis intercultural, también se puede impartir una pragmática, una pragmática comparativa o una pragmática intercultural (Barquero, 2020). Según Escandell-Vidal (2015), el oyente percibe de inmediato:

La existencia de una estrecha conexión entre ciertas propiedades del sonido y la interpretación de las acciones, es decir, entre los aspectos prosódicos y los pragmáticos. El análisis científico de las relaciones entre prosodia y pragmática, sin embargo, no es en modo alguno sencillo. (p. 193)

En este trabajo se ha analizado el ritmo como principal característica prosódica y se ha realizado un test de percepción. Es decir, por una parte, se ha realizado una descripción del aspecto rítmico en textos orales, y, por otra parte, se han presentado los resultados de las características rítmicas según la proximidad o lejanía de los textos orales analizados. En la introducción se ha descrito la función de la prosodia en la percepción y en el desarrollo de la competencia comunicativa oral para lograr una comunicación exitosa. La prosodia incluye la melodía, el tono, la rapidez, las pausas e inflexiones conferidas al discurso que ayudan a comprender y a dar a entender el tipo de texto producido, más allá de las emociones y los estados afectivos del hablante. Por lo tanto, se llega a la conclusión de que los elementos suprasegmentales pueden contribuir a deducir el sentido de una oración a través de la percepción de textos orales.

Los resultados obtenidos tanto en los análisis de las métricas propuestas por Ramus (1999), Dellwo (2006) y Grabe & Low (2002), como en los test de percepción, muestran que los futuros maestros y maestras identifican el tipo de texto (canción, juego de palabras y pequeñas poesías recitadas) en la mayoría de las lenguas (árabe, noruego, polaco, alemán, ruso, bereber y rumano) que desconocen. La excepción es el lituano, lengua en la que han tenido más dificultades a la hora de identificar el tipo de texto. Tal y como hemos especificado en el apartado de los resultados, los problemas han sido derivados principalmente por la confusión entre la canción y el juego de palabras en esta lengua. El texto oral que han escuchado es una pequeña poesía recitada y su estructura en cuanto al ritmo, no es compleja, puesto que su valor ΔC es de 38,83 y el rPVIC es de 93.18.

Este hecho nos lleva a deducir que las características suprasegmentales de los tipos de textos, aunque sean muy parecidos y no entendamos la lengua, pueden ayudar a reconocer lo que escuchamos; es decir, a identificar el tipo de texto producido por un hablante nativo en una lengua desconocida para el receptor. El ritmo, característica analizada en este trabajo, puede ayudar a identificar el tipo de texto aun cuando el receptor no entiende el idioma y lo desconoce por completo. En este trabajo se observa que no influyen las características de las estructuras rítmicas (estructuras silábicas simples o complejas) en la percepción del tipo de texto oral. Más allá del análisis del ritmo, habría que realizar un estudio más exhaustivo de otro tipo de características suprasegmentales para verificar la importancia que pueden tener en la percepción de textos orales de lenguas desconocidas para el oyente.

Partiendo de este estudio, se podrían realizar distintos abordajes desde el análisis de la prosodia atendiendo a la percepción. Tal y como se ha hecho en este artículo con el ritmo, se podrían realizar otros futuros trabajos atendiendo a otros componentes suprasegmentales tal y como se han realizado en varios trabajos anteriores, vinculándolos a diversos temas como la interculturalidad, afectividad (expresión de emociones) o conductas humanas (expresión de órdenes y cortesía). Por ejemplo, Gaminde et al., (2015) mostraron los resultados de una investigación en la que estudiaban la influencia de la variación diatópica y la lengua materna de jóvenes que tenían como L2 la lengua vasca en la percepción de emociones en dicha lengua. Herrero et al. (2021) investigaron si los inmigrantes chinos en España percibían la diferencia entre peticiones corteses y órdenes descorteses de la misma manera que los hablantes de español como L1, con una diferencia únicamente a nivel prosódico. Los análisis mostraron que, había diferencias en la percepción entre inmigrantes chinos y hablantes de español L1. En cuanto a la percepción audiovisual de emociones básicas, Etxebarria et al. (2020) realizaron un análisis en el que estudiantes de tercer y cuarto curso de Educación Primaria y de primero y segundo de Educación Secundaria Obligatoria, algunos de ellos hablantes de la lengua vasca como L2, debían vincular el estímulo auditivo con la emoción verbalizada y con el estímulo visual. Los resultados dieron a entender que vincular el estímulo auditivo con el tipo de emoción verbalizada no era complicado, sin embargo, generaba contradicciones y mayor dificultad para los sujetos relacionar éstos con los estímulos visuales.

Por lo tanto, la prosodia es uno de los componentes más importantes del habla, por medio de las modulaciones de la voz, las estructuras silábicas, el número y tipo de pausas, la velocidad de elocución y el ritmo muchas veces acompañados por los gestos, expresamos y percibimos emociones, actitudes y conductas humanas.

REFERENCIAS

- Alsina, M. R. (2012). La comunicación intercultural. *Anthropos*.
- Alsina, M. R. (1996). Los estudios de comunicación intercultural. *Revista de Estudios de Comunicación* 4 (1).
- Barón, L. F., Müller, O., & Galindo, O. (2014). Métodos experimentales de estudio de la percepción temprana del habla. *Revista Colombiana de Psicología*, 23(1), 73-94.
- Barquero, A. C. (2020). *Lengua, cultura, interculturalidad*. Universitätsverlag Potsdam.
- Bennett, M. J. (1998). *Basic concepts of intercultural communication: Selected readings*. Yarmouth Intercultural Press.
- Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence*. Multilingual Matters.
- Cervera, J. F., & Ygual, A. (2001). Evaluación e intervención en niños con trastornos fonológicos y riesgo de dificultad de aprendizaje de la lectura y escritura. *Cuadernos de Audición y Lenguaje*, 1, 1-41.
- Chen, G. M., & Starosta, W. (1996). Intercultural Communication Competence: A synthesis. *Communication Yearbook*, 19, 353-383.
- Cortés, M. (2001). Percepción y adquisición de la entonación española en enunciados de habla espontánea: el caso de los estudiantes taiwaneses. *Estudios de fonética experimental*, 90-119.

- Cortés, M. (2005). Análisis experimental del aprendizaje de la acentuación y la entonación españolas por parte de hablantes nativos de chino. *Phonica*, 1, 1-25.
- Dellwo, V. (2006). Rhythm and Speech Rate: A Variation Coefficient for deltaC. In P. Karnowski & I. Szigeti (Ed.), *Language and language-processing* (231–241). Peter Lang.
<https://www.zora.uzh.ch/id/eprint/111789/>
- Díaz-Campos, M., & Ronquest, R. (2007). La percepción de acentos tonales en enunciados afirmativos. *Estudios de fonética experimental*, 82-98.
- Escandell-Vidal, V. (2015). Prosodia y pragmática. *Studies in Hispanic and Lusophone Linguistics*, 4(1), 193-208.
- Etxebarria, A., Gaminde, I., Romero, A., & Iglesias, A. (2020). Las emociones básicas en un cuento: percepción audiovisual de los niños. *Círculo de Lingüística Aplicada a la Comunicación*, 83, 41-50.
<https://doi.org/10.5209/clac.70562>
- Gaminde, I., Etxebarria, A., Aurrekoetxea, G., Garay, U., & Romero, A. (2015). 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, 152-172.
- Gaminde, I., Romero, A., Garay, U., & Etxebarria, A. (2011). Los tonos de frontera de las oraciones interrogativas absolutas producidas por hablantes bilingües vasco-español. *Estudios de Lingüística Aplicada*, 29(54), 61-79.
- García, F. J., & Barragán, C. (2004). Mediación intercultural en sociedades multiculturales: Hacia una nueva conceptualización. *Revista de Trabajo Social*, 123-142. <http://hdl.handle.net/10272/219>
- Gil, E. M. (2015). INTERacción comunicativa en mensajería instantánea bajo sistemas SAACs: <<INTERSAACs>>. *Index comunicación*, 7 (1), 269-288. <https://eprints.ucm.es/id/eprint/29245/>
- Giménez, G. (2007). Estudios sobre la cultura y las identidades sociales. Iteso.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Ed.), *Laboratory Phonology* (515–546). De Gruyter Mouton.
<https://doi.org/10.1515/9783110197105.2.515>
- Gutiérrez Ordóñez, S. (2010). La pragmática en un cuento de Antonio Pereira. In J.M. Bleuca, G. Clavería & D. Poch (coor.) *Al otro lado del espejo* (321-341). Ariel.
- Herrero, C., Planelles, M., & Alhmoud, Z. (2021). Percepción de peticiones corteses y órdenes descorteses en español L2 por parte de trabajadores chinos residentes en España. *Lengua Y migración*, 12(2), 65-85.
<https://doi.org/10.37536/LYM.12.2.2020.1026>
- Hymes, D. (1971). Competence and performance in linguistic theory. *Acquisition of languages: Models and methods*, 3-23.
- Hymes, D. (1972). On communicative competence. *Sociolinguistics*, 269-293.
- Jimenez M., & Sercu, L. (2007). Challenges in Teacher Development: Learner Autonomy and Intercultural Competence. *Foreign Language Teaching in Europe*. Eschborner Landstrasse, 42-50.
- Kimura, T., Sensui, H., Takasawa, M., Toyomaru, A., & Atria, J. J. (2010). On Sentence-type Discrimination Strategies of Japanese Learners of Spanish-An Evidence from the HLH* Tone. In Poster presented in The Fourth European Conference on Tone and Intonation (TIE4), Universidad de Estocolmo, Campus Frescati.
- Mairano, P., & Romano, A. (2009). Un Confronto Tra Diverse Metriche Ritmiche Usando Correlatore. *Proceedings of the V National AISV Congress*, 79–100.
- Martínez Celdrán, E. (1993). La percepción categorial de /bp/ en español basada en las diferencias de duración. *Estudios de fonética experimental*, 223-239.
- Martínez, H., & Rojas, D. (2011). Prosodia y emociones: datos acústicos, velocidad de habla y percepción de un corpus actuado. *Lengua y Habla*, 15, 59-72.
- Ministerio de Educación y Formación Profesional e Instituto Cervantes. (2021). Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación. Volumen complementario. Secretaría General Técnica del Ministerio de Educación y Formación Profesional e Instituto Cervantes.
<https://hdl.handle.net/11162/207011>
- Muñoz, A. M. (2015). La enseñanza de la pronunciación como contribución a la mejora de la competencia oral de la lengua inglesa: una investigación sobre la percepción y la producción de la calidad vocálica en sílaba acentuada y no acentuada. Tesis doctoral. Universidad de Córdoba.
<http://hdl.handle.net/10396/12878>

- Peñalva, A., & Leiva, J. J. (2019). La interculturalidad en el contexto universitario: necesidades en la formación inicial de los futuros profesionales de la educación. *Educación*, 55(1), 141-158. <https://doi.org/10.5565/rev/educar.989>
- Pinilla, R. (2021). Hacia la competencia comunicativa intercultural. In S. Robles (eds.) *Pragmática: estrategias para comunicar* (pp. 209-223). Edelsa / Anaya.
- Ramus, F. (1999). *Rythme des langues et acquisition du langage* [PhD, Ecole des Hautes Etudes en Sciences Sociales (EHESS)]. <https://tel.archives-ouvertes.fr/tel-00242452>
- Rodríguez, A., & Souto, E. (2009). *Inmigración y resolución de conflictos: La mediación intercultural*. Dykinson.
- Romero, M. V. (1988). Sobre la formación de profesores de español como lengua extranjera. *Actas de las Primeras Jornadas Pedagógicas de ÁSELE*, 81-87.
- Ruiz, M. I. (2018). *Word stress patterns in the English of Spanish speakers: a perceptual analysis*. Tesis doctoral. Universidad de Alicante. <http://hdl.handle.net/10045/114906>
- Ruiz-Bernardo, M. P. (2012). Revisión lógica y empírica para la adaptación de un cuestionario de sensibilidad intercultural. <http://dx.doi.org/10.6035/ForumRecerca.2012.17.24>
- Saá, N. (2001). Lenguaje y hemisferio derecho. *Cuadernos de neurología*, XXV.
- Sanhueza, S. V., & Cardona, M. C. (2009). Evaluación de la Sensibilidad Intercultural en alumnado de Educación Primaria escolarizado en aulas culturalmente diversas. *Revista de Investigación Educativa*, 27 (1), 247-262. <https://revistas.um.es/rie/article/view/94391>
- Vilà, R. (2006). La dimensión afectiva de la competencia comunicativa intercultural en la educación secundaria obligatoria: Escala de Sensibilidad Intercultural. *Revista de Investigación Educativa*, 24 (2), 353-372. <https://revistas.um.es/rie/article/view/96891>
- Vilà, R. (2007). *Comunicación intercultural. Materiales para secundaria*. Narcea.
- Wharton, T. (2012). Prosody and Meaning: Theory and Practice. In J. Romero-Trillo (eds.), *Pragmatics and Prosody in English Language Teaching* (pp.97-117). Springer. https://doi.org/10.1007/978-94-007-3883-6_7
- Wichmann, A., & Blakemore, D. (2006). The prosody-pragmatics interface. *Journal of Pragmatics* 38(10), 1537-1541. <https://doi.org/10.1016/j.pragma.2006.02.009>

Artículo original

El análisis lexicométrico, el corpus y el diccionario previo: caso de la lengua Vasca¹

Lexicometric analysis, the corpus, and the previous dictionary: the case of the Basque language

Juan Abasolo

Universidad del País Vasco (UPV/EHU), España

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

e-mail: juan.abasolo@ehu.es

Recibido: 18/3/2023

Aprobado: 12/5/2023

RESUMEN

Siguiendo los lineamientos del método propuesto por Reinert (1983, 1990), podemos llevar a cabo una clasificación automática de grandes volúmenes de textos que aborden un determinado contexto semántico, con el fin de identificar los distintos campos semánticos o de sentido en base a las elecciones léxicas realizadas. Esta tarea se encuentra resuelta tanto en el software patrocinado por Max Reinert (1990), conocido como ALCESTE, como en el software de código abierto IRaMuTeQ (Ratinaud & Déjean, 2009) y sus desarrollos posteriores (Barnier, 2022). Como mencionaba Reinert (1990), es fundamental llevar a cabo un pretratamiento lingüístico que se ajuste al sentido semántico del texto, evitando así la variabilidad regida por las normas gramaticales. Para lograr este control sobre la variabilidad, se utiliza una preclasificación con un diccionario que contempla las formas plausibles de aparición en el texto, junto con sus correspondientes valores de significado asociados. En el caso de lenguas aglutinantes como el euskera, esto plantea un desafío particular. En esta comunicación, presentamos el proceso de creación de un diccionario para el uso del software IRaMuTeQ con textos en lengua vasca, su evaluación interna, mediante el análisis de autodescripciones de profesorado universitario, y externa, mediante el análisis de un corpus paralelo multilingüe vasco, castellano, inglés y francés.

Palabras clave: Método Reinert; Iramuteq; léxico; NPL; corpus multilingüe.

ABSTRACT

Following the steps of the method described by Reinert (1983, 1990), we can perform an automatic classification of large numbers of texts that address a specific semantic context in order to identify different semantic fields or meanings based on lexical choices. This task is already solved in the software sponsored by Reinert himself, known as ALCESTE, as well as in the open-source software IRaMuTeQ (Ratinaud & Déjean, 2009) and subsequent developments based on it (Barnier, 2022). As Reinert (1990) anticipated, a linguistic pre-processing is necessary to adhere to the proper semantic sense of the text, free from the variability governed by the rules of grammar. This control over variability is based on the use of a dictionary that includes plausible forms of appearance in the text along with their associated meaning values. In agglutinative languages such as Basque, this poses a challenge. In this communication, we present the process of creating a dictionary for the use of the IRaMuTeQ software with texts in the Basque language, along with an analysis of a parallel multilingual corpus.

Keywords: Reinert method; Iramuteq; dictionary; NPL; multilingual corpus.

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

INTRODUCCIÓN

En este estudio, deseamos presentar de manera concisa la adaptación que un grupo de investigadores del equipo EUDIA está llevando a cabo para hacer factible utilizar Iramuteq en lengua vasca. Iramuteq es una herramienta utilizada para el análisis de textos y cuestionarios (Ratinaud & Déjean, 2009).

Nuestra intención es compartir y transparentar el proceso de creación de un léxico adaptado al uso en esta lengua, así como brindar una visión más clara de los aspectos teóricos aplicados en el análisis de textos. Además, también queremos mostrar algunos resultados obtenidos mediante este enfoque de trabajo. Creemos que este estudio puede resultar de interés en otros entornos bilingües, especialmente aquellos con lenguas de naturaleza aglutinante.

LEXICOMETRÍA

Dentro del campo de la lingüística de corpus, el avance en la computación, el desarrollo de técnicas estadísticas más complejas y la capacidad de almacenamiento y procesamiento de grandes volúmenes de datos han permitido un rápido desarrollo de nuevas técnicas, entre las cuales se encuentran la lexicometría y la textometría.

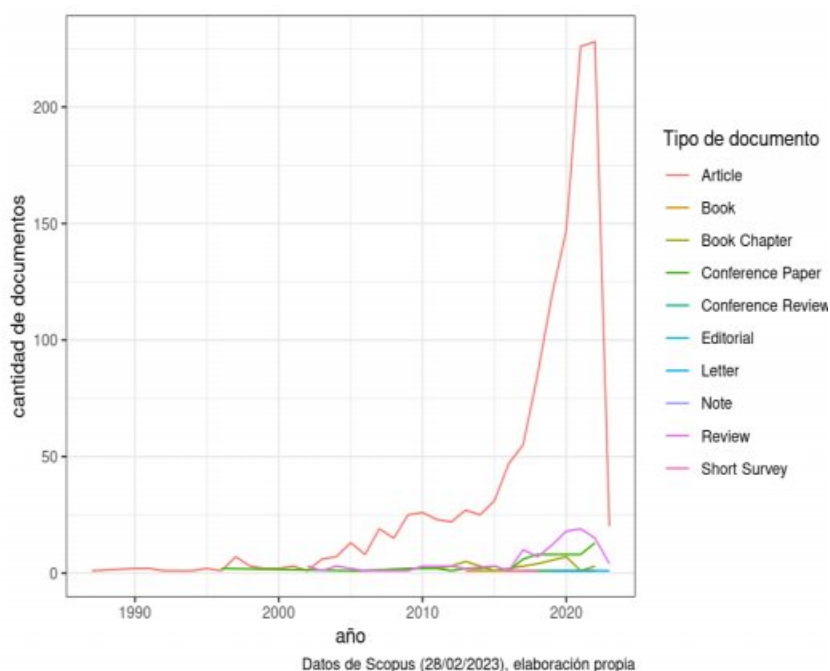
Se podría señalar como el primer paso en las técnicas que hoy se utilizan en la lingüística de corpus fue, en el siglo XIII, la indexación de la Biblia *Sancta Vulgata*, realizada a por Hugo de San Caro y 500 monjes bajo su dirección en su obra *Correctio Biblie*[†] (Hanon, 1991). Sin embargo, probablemente no fue hasta el desarrollo de la semántica interpretativa en los años 80 que se puede hablar de un desarrollo amplio en esta perspectiva. Es posible que el trabajo de François Rastier (1987) sea considerado el punto de partida de este desarrollo, al proporcionar técnicas y herramientas computacionales para la modelación del discurso.

La lexicometría, también conocida como análisis léxico cuantitativo, tiene como objetivo identificar unidades y patrones temáticos en los corpus utilizando enfoques cuantitativos. Una característica distintiva de las metodologías lexicométricas es que, si bien los análisis de corpus se basan en técnicas cuantitativas, los resultados de estos análisis no pueden ser interpretados sin una subsiguiente interpretación cualitativa que les otorgue sentido. Esto se debe a que las técnicas lexicométricas se centran en la búsqueda de patrones y coocurrencias en el corpus analizado. Una vez que se identifica la relevancia estadística de ciertas características, es fundamental que los elementos reconocidos y señalados desde el punto de vista matemático y estadístico sean reinterpretados y dotados de significado.

Entre las diversas técnicas de análisis que permite la lexicometría, posiblemente la perspectiva presentada por Reinert en 1983 (Reinert, 1983) sea la más relevante desde la experiencia académica. Esto se debe, entre otras cosas, al encuentro entre las técnicas que evalúan el grado de significancia estadística de los constructos estudiados y la interpretación cualitativa, lo cual permite identificar y reconstruir significados sociales basados en la práctica textual. Este enfoque ha sido ampliamente valorado por las editoriales científicas, como se refleja en la figura 1, que muestra la publicación de documentos científicos utilizando esta técnica. Es de gran relevancia la amplitud de editores que le dan importancia desde las políticas monetarias (Schonhardt-Bailey & Bailey, 2013) hasta la salud o la educación.

[†] Se puede consultar online la impresión de 1498: *Inkunabeln / Biblia latina: Cum postillis Hugonis de Sancto Caro*: I. [Basel]. [nach 29. X. 1498 u. nicht nach 1499]. (1498). <http://digital.ub.uni-duesseldorf.de/ink/7856461>

Figura 1. Resultados por años de la búsqueda de los descriptores "Alceste", "Reinert" e "Iramuteq" en documentos de la base de datos Scopus. Agregados según tipo de documentos.



Elaboración con base en el trabajo de investigación.

MÉTODO REINERT

A principios de la década de 1980, Reinert (Reinert, 1983) introdujo un algoritmo revolucionario que permitía clasificar los textos según su contexto, mediante un análisis multidimensional. Inicialmente, se presentó en el campo de la psicología social como una herramienta para categorizar las respuestas libres a preguntas abiertas, con el propósito de establecer análisis reproducibles. Tres años después, se presentó la primera implementación informática del algoritmo propuesto, conocida como el programa *ALCESTE - Analyse Lexicale par Contexte d'un Ensemble de Segments de Texte* (Reinert, 1986), junto con su flujo de trabajo, que se explicó en detalle en un artículo posterior en el ámbito de las ciencias sociales (Reinert, 1990).

Aunque no pretendemos abordar todos los aspectos técnicos de esta técnica en este artículo, nuestro objetivo es presentar una reconstrucción de la teoría que sustenta su enfoque técnico. Reinert, en colaboración con Benzecri (1981), propuso el concepto de que el contexto de cada palabra en un texto está formado por el texto completo, excluyendo la palabra en sí misma, tal como se establece originalmente en el concepto “con-texto”. Desde una perspectiva teórica, esto permitiría modelar los contextos de las palabras en función de las palabras que las rodean. Para lograr esto, la técnica denominada ALCESTE por Reinert propone describir los contextos mediante la coocurrencia de los términos contextuales.

Una vez que se clasifican las unidades de análisis de contexto, que idealmente deberían coincidir con unidades de sentido, se agrupan utilizando una metodología jerárquica ascendente de conglomerados (cluster). Esta agrupación permite identificar elementos con patrones regulares de aparición y calcular la presencia o ausencia específica de cada uno de los términos en todas las agrupaciones definidas.

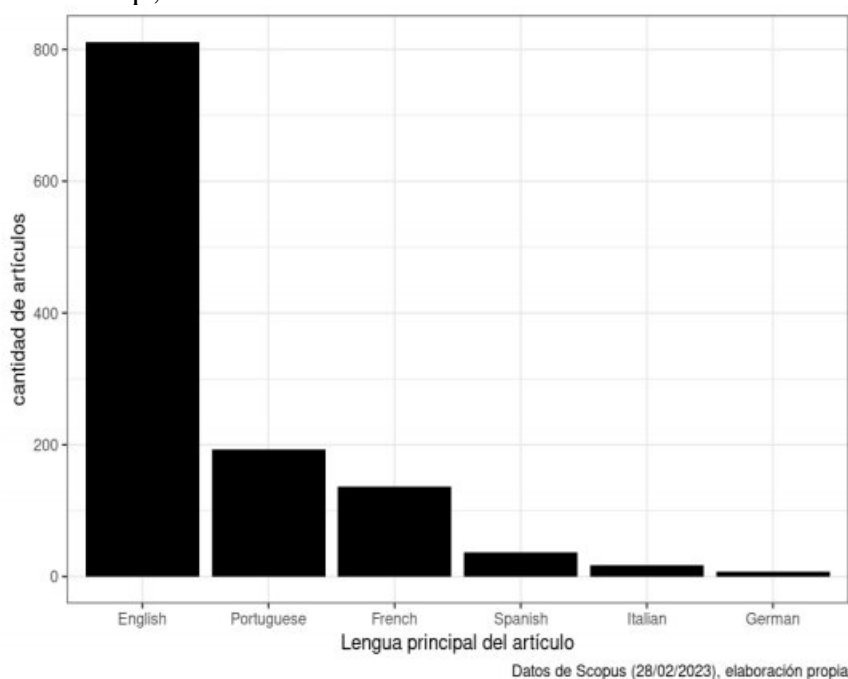
Sin embargo, esta propuesta teórica se enfrenta al desafío del uso real del lenguaje, en el que los elementos conceptuales subyacentes a las formas lingüísticas pueden ser únicos o polimórficos debido a las diversas formas que pueden tomar los lemas. Esta dificultad se agrava especialmente en lenguas aglutinantes, como el euskera.

La solución propuesta por Reinert, que actualmente se está aplicando al euskera, implica contar previamente con una lista ordenada de formas lingüísticas plausibles de aparecer, etiquetadas con sus lemas correspondientes y su valor sintáctico, antes de realizar el análisis. Esto facilita el proceso de asignar sentido a los elementos identificados y señalados mediante las técnicas lexicométricas utilizadas en el análisis.

IRAMUTEQ

Iramuteq (Ratinaud, 2014; Ratinaud & Déjean, 2009) es una adaptación del algoritmo propuesto por Reinert (1983), esta vez utilizando el lenguaje R (Ihaka & Gentleman, 1996), con una interface utilizable por el usuario final. Tanto el código como la aplicación de Iramuteq se distribuyen de forma abierta y gratuita. De hecho, Iramuteq se presenta como una interfaz de R para el análisis multidimensional de textos y cuestionarios (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*), como se indica en el sitio web de descarga[‡]. Además del análisis multidimensional, Iramuteq también permite realizar otros análisis del corpus tanto multidimensionales como unidimensionales, tales como análisis factorial, nubes de palabras, búsquedas de concordancias y análisis de similitudes. Estos análisis adicionales no serán abordados en detalle en el presente trabajo, pero pueden resultar de interés para los investigadores.

Figura 2. Lenguas de edición de artículos en la base de datos Scopus, que incluyen los descriptores "Iramuteq", "Reinerto" o "Alceste".



Elaboración con base en el trabajo de investigación.

[‡] <http://www.iramuteq.org>

Como se ha destacado en la figura 2, hasta la fecha se ha demostrado la valía de esta herramienta. En la base de datos Scopus se puede observar la relevancia de las publicaciones en varios idiomas, más allá del francés e inglés, que eran los esperados. Al revisar las bases de datos de publicaciones científicas, también se puede observar que en muchos casos las investigaciones mencionan el método Reinert (o Alceste, o Iramuteq), pero el idioma utilizado en la comunicación de la investigación no coincide con el idioma en el que se desarrolló el método. También es importante destacar la notable presencia de publicaciones científicas en Colombia, Brasil o México, además de Estados Unidos, Francia, España o Portugal, que consideran el uso de esta técnica, método y herramienta para describir diferentes representaciones sociales.

LENGUAS

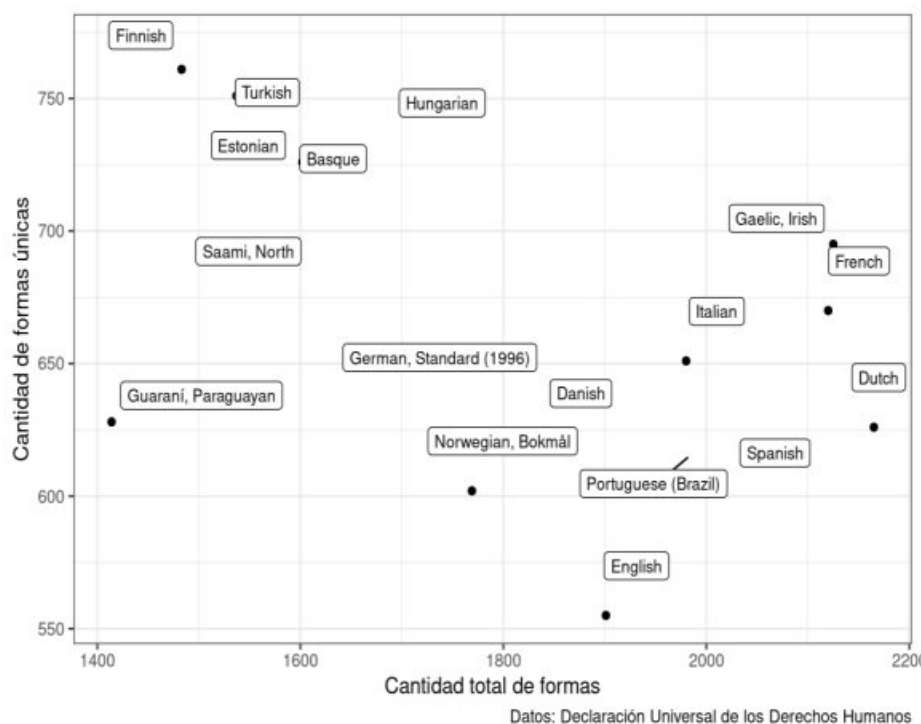
La herramienta Iramuteq, como se mencionó anteriormente, requiere de un léxico ordenado que contenga las formas lingüísticas esperadas en los textos que van a ser analizados. Es relevante destacar en qué idiomas se está utilizando la herramienta y cuáles son sus características diferenciales.

Actualmente, la distribución de Iramuteq viene con algún léxico en alemán, inglés, francés, gallego, griego, italiano, latín, portugués, rumano, español y sueco. De todos los léxicos disponibles, solo algunos se pueden acceder automáticamente desde la interfaz del programa y tienen el tamaño suficiente como para conseguir análisis efectivos. A aquellos léxicos que no ofrecen una opción de selección automática se puede acceder mediante un cuadro de diálogo, que permite asignar un archivo en particular como diccionario.

Esta posibilidad mencionada anteriormente es la que permite y justifica la existencia de numerosos léxicos preparados para su uso en diferentes idiomas y con una mayor cantidad de términos incluidos, además de los léxicos de distribución oficial mencionados anteriormente. Por ejemplo, existe un léxico en lengua castellana distribuido a través de un enlace de Twitter (Ideia [@ideiainova], 2017), o un léxico en alemán distribuido en listas de correo públicas (Loubere, 2023) propias del grupo que desarrolla el programa.

Los léxicos preparados para el método Alceste se encuentran en su mayoría en lenguas indoeuropeas y son principalmente, a excepción del latín, prepositivas. En este trabajo, estamos presentando la adaptación del método a la lengua vasca, que es una lengua no indoeuropea y se caracteriza por ser altamente aglutinante. Esto se traduce en una mayor variedad de formas escritas diferentes y una menor cantidad total de formas para la comunicación de textos equivalentes, como se puede observar en la figura 3, que se representa la cantidad de formas necesarias para la traducción oficial de la Declaración Universal de los Derechos Humanos de la UNESCO, contrastando la cantidad total de formas con la cantidad de formas diferentes.

Figura 3. Relación entre formas diferentes y total de formas empleadas en algunas lenguas.



Elaboración con base en el trabajo de investigación.

EL VASCO

Específicamente, el caso del vasco escrito se destaca como el más divergente en comparación con otras lenguas europeas en cuanto a la cuantificación de la cantidad de formas distintas y totales utilizadas en textos equivalentes. Este fenómeno se atribuye a la presencia de una morfología aglutinante que rige la producción de textos en esta lengua. Para ilustrar esta situación, podemos contrastar alguna de las diversas formas que pueden adoptar los lemas equivalentes “casa” en castellano y “etxe” en vasco:

- casa (casa singular o indefinido)
- casas (casa plural)

En vasco:

- etxe [*casa*]
- etxea [*la casa* (absolutivo)]
- etxeak [*la casa* (ergativo)]
- etxean [*en (la) casa*]
- etxera [*a (la) casa*]
- etxetik [*desde (la) casa*]
- etxerako [*para (la) casa*]
- etxeraino [*hasta (la) casa*]
- etxeko [*de (la) casa*]
- ...

Esta lista, que actualmente solo contempla las formas singulares, podría ampliarse para incluir otras formas de uso común, como plurales o indefinidos, además de la amplia variedad de combinaciones de sufijos que son posibles en el idioma, como, por ejemplo.

- etxekoa [*el de (la) casa* (absolutivo)]
- etxekoak [*el de (la) casa* (ergativo)]
- etxekoaren [*del de (la) casa*]
- etxekoarena [*el del de (la) casa* (absolutivo)]
- ...

En aras del principio de parsimonia y teniendo en cuenta la capacidad limitada de las herramientas que se utilizarán, no es apropiado trabajar con la lista de todas las formas posibles, sino más bien con aquellas formas que sean plausibles de encontrarse en los diversos textos escritos en vasco.

OBJETIVOS Y MÉTODO

El objetivo de este estudio es desarrollar y evaluar un léxico en lengua vasca para analizar textos escritos en dicho idioma utilizando la metodología propuesta por Reinert. En esta sección, se describen los métodos utilizados para la construcción de los corpus y los procedimientos aplicados posteriormente para su procesamiento.

PRIMER OBJETIVO:

CONSTRUCCIÓN DEL LÉXICO EN LENGUA VASCA

El primer objetivo de este estudio fue crear un léxico en lengua vasca. Para lograrlo, se siguieron varios pasos. En primer lugar, se construyó un corpus extenso que reflejara el uso del idioma. Luego, se llevó a cabo un análisis del corpus para identificar las relaciones entre las formas utilizadas, los lemas a los que hacen referencia y las categorías gramaticales en las que se clasifican. Por último, se redujeron estas relaciones encontradas a un conjunto único, basándose en la frecuencia de uso.

Figura 4. Captura de algunas líneas del léxico en español.

Token	Lemma	Category
21804	geográficas»geográfico»	adj
21805	geográfico» geográfico»	adj
21806	geográficos»geográfico»	adj
21807	geológica» geológico»	adj
21808	geológico» geológico»	adj
21809	geológicos» geológico»	adj
21810	geometría» geometría»	nfm
21811	geométrica» geométrico»	adj
21812	geométricas»geométrico»	adj

Elaboración con base en el trabajo de investigación.

En la imagen de la figura 4 se muestra la estructura requerida para el léxico, con una columna inicial que incluye las formas o tokens, incluyendo las flexiones de los lemas. La segunda columna muestra los lemas correspondientes, mientras que en la tercera columna se proporciona una abreviatura en francés que indica la categoría gramatical asignada.

El corpus utilizado en el análisis consistió en artículos y cartas de lectores de una revista de temática general, Argia, en formato digital, que comprendía 1.082.787 frases. Además, se incluyeron 61.567 frases de la selección de artículos en lengua vasca de Wikipedia realizada por el proyecto Common Voice, y 11.379 frases de las publicaciones del año 2020 de la revista en línea Bizkaie escrita en dialecto occidental. En total, se obtuvieron más de un millón de frases en diferentes registros y formas para su análisis.

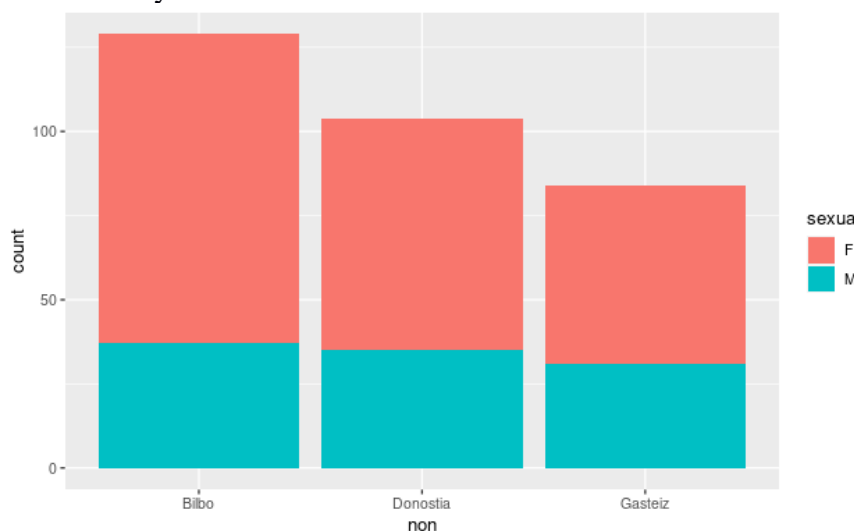
Para llevar a cabo el análisis y clasificación de las formas lingüísticas, se utilizó el lenguaje de programación R (R Core Team, 2020). Específicamente, se empleó el paquete “udpipe” (Straka et al., 2016) (Wijffels et al., 2020) en su versión 0.8.5, el cual proporciona una interfaz adaptable para aplicar modelos de Universal Dependencies (Nivre et al., 2016). Se accedió al modelo de euskera publicado por Aranzabe y colaboradores (Aranzabe et al., 2015) para llevar a cabo el modelado y clasificación de los lemas y las categorías gramaticales.

SEGUNDO OBJETIVO:

PROBAR LA CONSISTENCIA INTRALINGUAL DEL LÉXICO CONSTRUIDO

Con el fin de evaluar la eficacia del léxico construido, se llevaron a cabo dos enfoques de prueba. En primer lugar, se realizó una evaluación desde la perspectiva de la propia lengua vasca, con el objetivo de comprobar la coherencia de las respuestas obtenidas. Para este propósito, se construyó un corpus de autodescripciones redactadas por profesores de las carreras de magisterio de la UPV/EHU. Estos textos públicos, redactados en lengua vasca, se obtuvieron de las páginas web de las tres facultades donde se imparten las carreras de magisterio en educación infantil o educación primaria. Estos textos fueron procesados con Iramuteq, utilizando el léxico en lengua vasca, para realizar una clasificación jerárquica ascendente de acuerdo con el método ALCESTE propuesto por Reinert.

Figura 5. Distribución del corpus de autodescripciones según campus de pertenencia y género de los autores y autoras.



Elaboración con base en el trabajo de investigación.

El corpus mencionado anteriormente (figura 5), consta de un total de 129 autodescripciones escritas en lengua vasca, de un total de 317 docentes pertenecientes a las tres facultades donde se imparten los grados de magisterio, recopiladas en noviembre de 2021. Esto permitió obtener un total de 1279 pares de unidades de sentido y etiquetas, clasificados según el campus, el departamento al que pertenecen los docentes y su género.

TERCER OBJETIVO:**VERIFICACIÓN DE LA CONSISTENCIA INTERLINGÜÍSTICA DEL LÉXICO CONSTRUÍDO**

Por otro lado, con el propósito de comprobar la coherencia de los resultados obtenidos en lengua vasca, se construyó un corpus paralelo multilingüe para comparar los resultados en vasco con aquellos en otros idiomas con una mayor tradición en el uso de Iramuteq, como el francés, inglés y castellano. En este caso, el corpus se basa en las cartas paulinas del Nuevo Testamento, segmentadas de manera idéntica, lematizadas utilizando los léxicos proporcionados por Iramuteq en cada lengua y utilizando el léxico específico para los textos en vasco. Estos subcorpus fueron agrupados según el método ALCESTE.

La consistencia de los campos semánticos detectados en diferentes idiomas se verificó en función de las unidades contextuales utilizadas en su construcción. Para ello, se creó una matriz de similitud de los agrupamientos según el método de Jacard y se generó una clasificación jerárquica ascendente.

Para la construcción del corpus paralelo multilingüe, se utilizaron los datos disponibles en el sitio web de la Asociación Bíblica Eslovena[§], que permite la presentación emparejada de diferentes traducciones. En este sentido, se incluyó la única traducción disponible en euskera actual, la edición ecuménica Elizen Arteko Biblia de 1994; la versión en inglés de King James de 1611, la versión francesa de la traducción de Louis Segond de 1910, y la versión en castellano ecuménica de la traducción Dios Habla Hoy de 2002. Estas traducciones se utilizaron como base para el corpus multilingüe y permitieron realizar comparaciones significativas entre las diferentes lenguas.

RESULTADOS

En esta sección, se presentan los resultados obtenidos durante la construcción del léxico, así como los contrastes intra e interlingüísticos del rendimiento del mismo.

CONSTRUCCIÓN DEL LÉXICO

El análisis inicial del corpus reveló un total de 13.065.741 combinaciones diferentes de *forma-lemma-categoría gramatical*. No obstante, como se mencionó anteriormente, es necesario reducir esta cantidad a relaciones únicas entre formas y lemas, así como entre lemas y categorías gramaticales. Este proceso de reducción se basa en la frecuencia de uso, priorizando las relaciones más frecuentes. Después de aplicar esta reducción, se logró crear un léxico con 627.479 formas diferentes.

A diferencia de otros léxicos construidos para lenguas prepositivas, debido a las características aglutinantes del vasco, se incluyeron en este léxico un total de 20.093 formas clasificadas como nombres propios. Estas formas fueron etiquetadas como *pro_per*, aunque originalmente esa etiqueta pertenece a la categoría de pronombres personales del léxico francés. En este léxico en lengua vasca, no se realizó una distinción entre estos pronombres y otros, siguiendo las decisiones adoptadas en la construcción de los léxicos para castellano, inglés, portugués e italiano.

[§] <https://biblija.net>

USO EN EUSKERA

Un primer análisis del uso del léxico en lengua vasca confirma que el filtrado por categorías gramaticales y la lematización del corpus generan resultados con evidentes diferencias.

Figura 6. Nubes de palabras del corpus, sin lematizar (formas) y lematizada (lemas).



Elaboración con base en el trabajo de investigación.

En la parte izquierda de la figura 6, se muestra la nube de palabras basada en la frecuencia de uso de las formas, mientras que en la parte derecha se presenta la nube de palabras basada en la frecuencia del uso de los lemas, según lo establecido en el léxico. La forma de conjunción “eta” [y] destaca en términos de relevancia, en contraposición al lema “hezkuntza” [educación]. El lema “hezkuntza” también aparece en la nube de formas y se utiliza en un total de 487 ocasiones, en doce formas diferentes.

Al realizar un Análisis de Clasificación Descendente basado en los contextos léxicos, se puede observar la siguiente proyección (figura 7), donde se identifican cuatro campos semánticos, dos de los cuales están estrechamente relacionados.

COMPARACIÓN PARALELA

El análisis de los textos de las cartas paulinas revela un total de cuatro clasificaciones jerárquicas distintas, con diferentes cantidades de campos detectados. Estas cuatro clasificaciones primarias dieron lugar a agrupaciones basadas en similitudes resumidas en la tabla 1.

Tabla 1. Resumen de los análisis jerárquicos descendientes de los cuatro subcorpus.

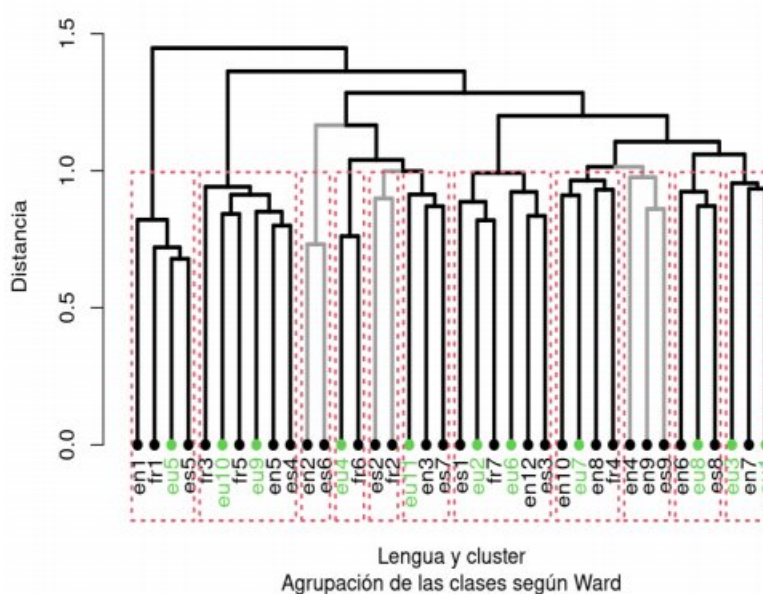
Traducción	Léxico	Ratio clasificado	clases
SEG	francés	0.9751	7
KJB	inglés	0.952	12
DHH	español	0.950	9
EAB	vasco	0.942	11

Elaboración con base en el trabajo de investigación.

El proceso de agrupamiento ascendente se llevó a cabo según la coincidencia en los pasajes asignados a cada clase, siguiendo el algoritmo de Ward. Esta clasificación determina qué grupos se consideran más similares entre los resultados de las clasificaciones. El proceso de agrupamiento basado en la similitud de los componentes de las clases se muestra en el dendrograma de la figura 8, donde se resaltan once clusters, que representan la cantidad de clases resultantes en la clasificación del subcorpus en vasco.

Las agrupaciones de las clases muestran una mayor similitud entre unidades o agrupaciones cuando están más cerca del valor 0 en el eje x. Así, podemos observar que las dos clases que muestran mayor similitud son las etiquetadas como *eu5* y *es5*, correspondientes a la quinta clase de los análisis del subcorpus en vasco y castellano, respectivamente. Estas dos clases forman una unidad con las clases *fr1* y *en1*, correspondientes a las primeras clases de los análisis del subcorpus en francés e inglés.

Figura 8. Dendrograma de agrupación de las clases calculadas en el análisis interlingual.



Elaboración con base en el trabajo de investigación.

La primera agrupación que se forma ocurre entre dos clases que utilizan léxicos diferentes (vasco y castellano) para traducciones ecuménicas de finales del siglo pasado. Luego, estas dos clases se agrupan con una clase en francés y otra en inglés. Esta primera agrupación es la primera en formarse y también la última en unirse al resto de las clasificaciones. Esta característica revela que esta agrupación de clases es la más distante de todas las demás agrupaciones calculadas.

A continuación, se presentan las nubes de palabras del primer grupo (figura 9), separadas según las clasificaciones originales de los subcorpus en cada idioma. En estas nubes de palabras se puede apreciar la coincidencia interlingual de muchos de los lemas mencionados, algunos de los cuales se resumen en la tabla 2.

Figura 9. Nubes de palabras de las clases agrupadas en el primer cluster.



Elaboración con base en el trabajo de investigación.

Es importante destacar que, en el caso de este primer cluster, la mayoría de las formas incluidas en los lemas utilizados en las clasificaciones tienen significados equivalentes.

Tabla 2. Resumen de algunas equivalencias en los lemas agrupados en el cluster 1.

eu	es	fr	en
edan	beber	boire	drink
jan	comer	manger	eat
sasijainko(ei)	idolo	idole	idol
haragi	carne	viande	meat
gose	hambre	faim	hungry
ogi	pan	pain	bread
opari	sacrificio	sacrifier	sacrifice
eskaini	ofrecer	-offre (no presente)	offer

Elaboración con base en el trabajo de investigación.

En la clase en lengua vasca, destaca la forma “sasijankoei” [*a los falsos dioses*], que se equipara a *ídolo* (español), *idol* (inglés) e *idole* (francés). La forma en vasco no está lematizada, ya que el equivalente debería haber sido *sasijainko* [falso dios].

A la luz de los resultados presentados en esta sección, se señalan a continuación algunas características sobre las cuales reflexionar.

DISCUSIÓN

En la sección de discusión, se abordan los tres objetivos planteados en el artículo.

Con relación al primer objetivo de construir un léxico para lematizar discursos escritos en lengua vasca, se presenta una primera versión que incluye un total de 473.330 formas con sus relaciones unívocas en tríadas de *forma-lema-categoría gramatical*. Esta versión inicial del léxico incorpora numerosos nombres propios de uso común, actualizados hasta el año 2020 según la prensa. Si bien esto permite detectar de manera precisa y eficiente una multitud de casos, también puede subrepresentar nombres propios relacionados con realidades que no están ampliamente cubiertas por la prensa o actores posteriores a la fecha en la que se recopiló el corpus utilizado para generar el léxico.

En cuanto al segundo objetivo, que se refiere al análisis de los textos y la detección de ámbitos de autodescripción, se considera que los resultados son satisfactorios. El análisis identificó cuatro ámbitos, dos de los cuales mostraron una relación interrelacionada. Esta interrelación entre los ámbitos de acción docente y formación académica en el contexto de las facultades de educación es esperada y coherente en un entorno de formación de docentes.

Respecto al tercer objetivo, que se centra en estudiar la coherencia externa de los resultados obtenidos con cuatro léxicos y cuatro corpus paralelos, se presentan dos aspectos a considerar. En primer lugar, se destacan las características cuantitativas de las clasificaciones, que se muestran en la tabla 1. Luego, se interpreta la comparación de los resultados de las clasificaciones.

En cuanto a las características numéricas de los análisis, se resalta que la clasificación obtenida con el léxico en lengua vasca presenta la tasa más baja de clasificación, con aproximadamente un 6% del corpus sin clasificar, mientras que en francés queda alrededor de un 2.5% sin clasificar. En las otras clasificaciones, hay más unidades sin clasificar que en francés, pero menos que en euskera. En cuanto al número de clases obtenidas, no se observa nada particularmente relevante en la oposición del léxico vasco con respecto a los demás léxicos.

El último punto de reflexión se centra en el proceso de agrupamiento de las clases de los diferentes análisis. Se destaca la relevancia de términos equivalentes en las diferentes clasificaciones de las traducciones utilizando los léxicos correspondientes. Además, se menciona la presencia de formas relevantes, pero no lematizadas, como el caso de la forma “sasijankoei” mencionada anteriormente, lo cual indica la necesidad de revisar detenidamente los resultados del análisis y ajustarlos según sea necesario.

También se observa que, en la clasificación y subdivisión en grupos, en algunos casos no se incluye ninguna clase de alguno de los subcorpus, lo que sugiere que el comportamiento entre las agrupaciones es similar.

En general, la sección de discusión aborda los resultados obtenidos y plantea puntos de reflexión sobre las características del léxico, los ámbitos detectados y el proceso de agrupamiento, resaltando la necesidad de revisiones adicionales y ajustes.

OPORTUNIDADES Y LÍMITES

En esta última sección se destacan varias áreas de enfoque para futuras investigaciones y mejoras en el uso del léxico en textos escritos en lengua vasca.

En primer lugar, es necesario continuar evaluando el desempeño del léxico en diferentes textos escritos en vasco. Esto implica realizar un análisis minucioso para identificar qué elementos lingüísticos se han lematizado correctamente y cuáles no durante el proceso de construcción de las clases. Se requiere un examen detallado para identificar posibles áreas de mejora y perfeccionar el léxico en función de los resultados obtenidos. Para lograr esto, es importante distribuir el léxico entre la comunidad interesada y solicitar su participación y retroalimentación para su uso y mejora continua.

Además, se destaca la necesidad de contar con una herramienta que permita mantener actualizado el léxico, especialmente en lo que respecta a las lematizaciones de los nombres propios. Esto garantizará que el léxico esté al día y pueda adaptarse a los textos a analizar, incluso si se encuentran nombres propios que no estén cubiertos por la prensa o sean actores posteriores a la fecha del corpus utilizado para generar el léxico.

Se subraya que la técnica lexicométrica conocida como ALCESTE o método Reinert ha demostrado ser aplicable con éxito en lenguas aglutinantes, como el vasco. Esto refuerza su utilidad y relevancia en diferentes contextos lingüísticos y culturales, lo que abre oportunidades para utilizar esta técnica en otros idiomas y en diversos campos de investigación.

En resumen, las oportunidades futuras incluyen la evaluación continua y la mejora del léxico en vasco, la participación de la comunidad en su uso y retroalimentación, la actualización constante del léxico y la exploración de la aplicabilidad de técnicas lexicométricas en otros contextos lingüísticos y culturales. Estas acciones contribuirán a fortalecer el léxico y su utilidad en el análisis de textos escritos en lengua vasca.

REFERENCIAS

- Aranzabe, M. J., Atutxa, A., Bengoetxea, K., Diaz de Ilarraza, A., Goenaga, I., Gojenola, K., & Uria, L. (2015). Automatic Conversion of the Basque Dependency Treebank to Universal Dependencies. In M. Dickinsons, E. Hinrichs, A. Patejuk, & A. Przepiórkowski (Eds.), *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14)* (pp. 233–241). Institute of Computer Science of the Polish Academy of Sciences.
- Benzécri, J. P. (1981). *Pratique de l'analyse des donnees: Linguistique et lexicologie*. Dunod.
- Hanon, S. (1991). 165. La concordance. *Wörterbücher: Ein internationales Handbuch zur Lexikographie*, 2, 1562–1567. <https://doi.org/10.1515/9783110124200.2>
- Ideia [@ideiainova]. (2017). Sharing a new version of the Spanish dictionary for #Iramuteq (+500k entries) [Tweet [Link a Archivo]]. In Twitter.
- Ihaka, R., & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314. <https://doi.org/gddc3n>
- Loubere, L. (2023). Re: [Iramuteq-users] Dictionary in german? | iramuteq.
- Nivre, J., de Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., & Zeman, D. (2016). Universal Dependencies v1: A Multilingual Treebank Collection. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 1659–1666.
- R Core Team. (2020). *R: A language and environment for statistical computing [Manual]*. R Foundation for Statistical Computing.
- Rastier, F. (1987). Représentation Du Contenu Lexical Et Formalismes De L'intelligence Artificielle. *Langages*, 87, 79–102. <https://doi.org/10.3406/lgge.1987.1964>
- Ratinaud, P. (2014). IRaMuTeQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires.
- Ratinaud, P., & Déjean, S. (2009). IRaMuTeQ: Implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre. *Modélisation Appliquée Aux Sciences Humaines Et Sociales MASHS*, 8–9.

- Reinert, A. (1983). Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte. *Les cahiers de l'analyse des données*, 8(2), 187–198.
- Reinert, M. (1986). Un logiciel d'analyse lexicale. *Les Cahiers de l'analyse Des Données*, 11(4), 471–481.
- Reinert, M. (1990). Alceste une méthodologie d'analyse des données textuelles et une application: Aurelia De Gerard De Nerval. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 26(1), 24–54. <https://doi.org/cbhfwf>
- Schonhardt-Bailey, C., & Bailey, A. (2013). *Deliberating American Monetary Policy: A Textual Analysis*. The MIT Press. <https://www.jstor.org/stable/j.ctt9qf5r7>
- Straka, M., Hajič, J., & Straková, J. (2016). UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 4290–4297.
- Wijffels, J., BNOSAC, Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University in Prague, Republic, C., Straka, M., & Straková, J. (2020). udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the «UDPipe» «NLP» Toolkit (0.8.5). <https://CRAN.R-project.org/package=udpipe>

Artículo original

Contraste de métodos y variación lingüística¹

Contrasting methods and linguistic variation

Iñaki Gaminde

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0003-3966-5002>

e-mail: inaki.gaminde@gmail.com

Naia Eguskiza

Universidad del País Vasco (UPV/EHU), España
<https://orcid.org/0000-0002-8928-7006>

e-mail: naia.egusquiza@ehu.eus

Recibido: 11/4/2023

Aprobado: 25/6/2023

RESUMEN

La variación lingüística es inherente al lenguaje y se presenta en diferentes momentos y situaciones. Se identifican cuatro niveles de variación: diatópico (geográfico), diastrático (social), diafásico (registros) y diacrónico (histórico). Este trabajo se enfoca en los niveles socio-geográficos y de género, proponiendo una investigación geo-sociolingüística. Los estudios de esta índole pueden ser cualitativos o cuantitativos, y la elección del método y análisis previo afectan a los resultados que se obtienen. Con el fin de demostrar esa afirmación en este trabajo se pretende contrastar los resultados de dos métodos cuantitativos en el análisis de la variación lingüística. En el primer método los elementos susceptibles de análisis son tratados como presencias o ausencias de los mismos en el corpus (creando matrices dicotómicas para su posterior análisis) y en el segundo, se tienen en cuenta las probabilidades de cada elemento basadas en la frecuencia de los mismos, relacionándose el número de elementos con el número de concurrencias. Al final se muestran las conclusiones obtenidas a partir del empleo de ambos métodos.

Palabras clave: variación lingüística; sociolingüística; geolingüística; métodos.

ABSTRACT

Linguistic variation is inherent to language and occurs at different times and in different situations. Four levels of variation are identified: diatopic (geographical), diastratic (social), diaphasic (registers) and diachronic (historical). This paper focuses on the socio-geographical and gender levels, proposing a geo-sociolinguistic investigation. Studies of this nature can be qualitative or quantitative, and the choice of method and previous analysis affect the results obtained. In order to demonstrate this assertion, this paper aims to contrast the results of two quantitative methods in the analysis of linguistic variation. In the first method, the elements susceptible to analysis are treated as their presence or absence in the corpus (creating dichotomous matrices for subsequent analysis) and in the second, the probabilities of each element based on their frequency are taken into account, relating the number of elements to the number of occurrences. At the end, the conclusions obtained from the use of both methods are shown.

Keywords: linguistic variation; sociolinguistics; geolinguistics; methods.

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

INTRODUCCIÓN

La variación lingüística se considera una característica inherente del propio lenguaje. Además, de variar según las características sociales o lugar de origen del hablante, la variación ocurre en distintos momentos y situaciones, por lo que hacemos nuestras las palabras de Tagliamonte: “linguistic change happens everywhere but not at the same time, not in every area, and not in the same way in places” (2012: 13).

En general, en los diferentes usos de la lengua se han definido cuatro niveles de variación: la variación a nivel diatópico, nivel diastrático, nivel diafásico y nivel diacrónico. El primer nivel se refiere a la variación lingüística que se produce según las zonas geográficas, y suele variar de un territorio a otro. El ejemplo sería el estudio de los dialectos. El segundo de ellos se refiere a las variables sociales. Al igual que en el nivel anterior hemos mencionado el dialecto, en este caso nos referimos al sociolecto para designar la variación del grupo social. Para realizar estudios de este nivel es necesario actuar de acuerdo con la organización de cada sociedad. En el tercer nivel, en el nivel diafásico, se analizan los diferentes registros que emplea y produce el hablante. Lavob (1972, 1974) lo definió como la versatilidad de estilo, incluyendo en este nivel el prestigio. A la altura del dialecto y del sociolecto, en este nivel se habla del propio registro, en referencia a los cambios que realiza el hablante. Por último, a nivel diacrónico se estudia la variación del lenguaje a lo largo del tiempo. Las formas más antiguas o arcaicas se comparan con las más recientes o modernas. Para ello, se divide el tiempo para hacer esas comparaciones. En este nivel se analiza la lengua en su contexto histórico y el cronolecto es la división del lenguaje que se hace a este nivel.

En este trabajo nos centraremos en el segundo y tercer nivel, ya que la investigación que proponemos se basa tanto en el origen geográfico de los hablantes como en el género. Por lo que se propone una investigación geo-sociolingüística. En los estudios sociolingüísticos se suelen mencionar especialmente estas variables: el género, la edad, la clase social, el nivel de estudios y el conocimiento de la lengua estándar (Eguskiza, 2019). Nosotros en este trabajo nos centraremos en la primera variable, en el género (además de en la geografía), pero cabe recalcar la importancia del resto.

Al igual que la variación ocurre en distintos niveles, los estudios también pueden ser de distinta índole. Así se diferencian los estudios de variación cualitativos y los estudios de variación cuantitativos (Aurrekoetxea, 1992, 1995, 2003, 2004, 2005, 2008, 2009, 2016; Aurrekoetxea & Ormaetxea, 2006; Gaminde, 2007; Iglesias, 2014; Ensunza, 2015; Lujanbio, 2016; Eguskiza, 2019; Etxebarria & Eguskiza, 2018; Gaminde et al., 2016a, 2016b, 2017). Suele ser el propio investigador el que se decanta por uno de ellos o decide emplear la combinación de ambos según el objetivo que pretenda alcanzar con la investigación.

Además de la elección del método a emplear para el estudio de la variación lingüística el investigador también ha de decantarse por un análisis previo u otro. Ya que no es lo mismo realizar un análisis basado en probabilidades y en frecuencias o basado en un análisis dicotómico como punto de partida. Por lo tanto, el investigador deberá elegir primeramente el tipo de estudio que quiere realizar y tras la recogida de los datos, previo al análisis de los mismos, deberá decidir cómo los tratará. Aun siendo los mismos datos en origen, según como se haga el tratamiento previo del análisis y según qué análisis se realice podrán variar los resultados obtenidos y comunicados. Dicho de otro modo, dependiendo de cuál sea el tratamiento previo de los datos, aun aplicando el mismo método los resultados pueden ser diferentes.

En muchos de los estudios se suelen clasificar los informantes. El objetivo de la clasificación de los informantes es sacar a la luz la estructuración geolingüística y/o sociolingüística que subyace en los datos desde el punto de vista de los sistemas complejos dinámicos adaptativos (Gell-Man, 1994; Beckner et al., 2011; Ellis, 2011; López Rivera, 2013; Moreno Cabrera, 2008; Léonard et al., 2015). El método más empleado para la clasificación

de los informantes ha sido el método de conglomerados jerárquicos con el que se obtiene un dendrograma como resultado. En este trabajo emplearemos este método como se especifica más adelante.

El objetivo de este trabajo es contrastar los resultados de dos métodos cuantitativos en el análisis de la variación lingüística. En el primer método los elementos susceptibles de análisis son tratados como presencias o ausencias de los mismos en el corpus, es decir, creando matrices dicotómicas para su posterior análisis, en el segundo se tienen en cuenta las probabilidades de cada elemento basadas en la frecuencia de los mismos, relacionándose el número de elementos con el número de concurrencias.

El trabajo se presenta dividido en cuatro apartados. En el segundo apartado se describe el corpus utilizado y la metodología para su procesado y posterior análisis. En el tercer apartado se presenta el análisis de los datos. Por último, en el cuarto apartado se resumen las principales conclusiones extraídas del análisis de los datos.

METODOLOGÍA

En este apartado se describe la metodología empleada para la obtención, procesado y análisis de los datos. Se han elegido 18 localidades de la comarca de Busturialdea (figura 1) en la provincia de Bizkaia, esta comarca ocupa una posición central dentro de la provincia (figura 2).

Figura 1. Localidades de la comarca de Busturialdea



Figura 2. Localización de la comarca de Busturialdea



Elaboración con base en el presente trabajo de investigación.

En cada localidad se ha contado con la colaboración de dos informantes, uno por cada género; en la tabla 1 se muestra la lista de informantes identificados por medio del nombre de la localidad añadiendo M en el caso de las mujeres y H en el de los hombres.

Tabla 1. Informantes según localidad y género

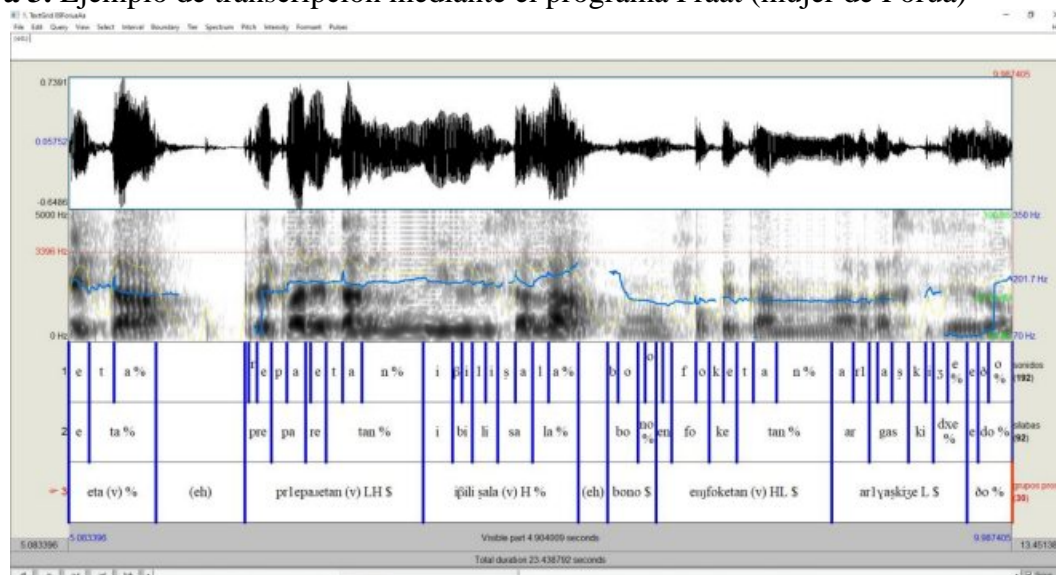
BermeoM	MuruetaM	MorgaM	GernikaM	NabarnizM	IbarangeluM
BermeoH	MuruetaH	MorgaH	GernikaH	NabarnizH	IbarangeluH
MundakaM	ForuaM	MuxikaM	AjangizM	KorteizubiM	ElantxobeM
MundakaH	ForuaH	MuxikaH	AjangizH	KorteizubiH	ElantxobeH
BusturiaM	ErrigoitiM	MendataM	ArratzuM	ArteagaM	EaM
BusturiaH	ErrigoitiH	MendataH	ArratzuH	ArteagaH	EaH

Elaboración con base en el presente trabajo de investigación.

Todos los informantes son nacidos entre los años 1981-2000 en las localidades especificadas y con estudios superiores realizados en euskera.

Para la recogida de los micro textos se utilizó un comic compuesto de seis viñetas (Gaminde, 2010) que refleja una breve historia. Después de un corto espacio de tiempo suficiente para entender la historia se le pidió que la contaran. Las grabaciones se realizaron con un grabador ZOOM y fueron procesadas con el software Praat (Boersma & Weenink, 2023) para su transcripción. Las transcripciones se hicieron en tres niveles: grupo prosódico, sílabas y sonidos (figura 3).

Figura 3. Ejemplo de transcripción mediante el programa Praat (mujer de Forua)



Elaboración con base en el presente trabajo de investigación.

En la figura 4 se muestra un ejemplo de transcripción fonética; junto con la transcripción de los sonidos aparecen algunas etiquetas correspondientes a la transcripción prosódica.

Figura 4. Transcripción de un micro texto (mujer de Arteaga)

ba (v) % (eh) aĩtfe HL \$ ta (v) \$ beian gemie (v) L % (eh) parlke βatera (v) LH \$ xup die H % (eh) (tx) okai L \$ eđo (v) \$ patui (v) LH % (eh) xaten emoten H % (eh) arlatsalde paša LH % (eh) ta (v) % (eh) eyuałdi (v) L \$ eđer2a đauelez HL \$ baiťey eiaβaki đeu (v) HL \$ (tx) arl yaški βeđ atatie L % (eh) fotoyrłafi βet LH % (eh) (tx) orletajako LH \$ βa (v) \$ gemie (v) HL % (eh) xarlťsep deų (v) H % (eh) kamaie prleparateų daų HL \$ ta (v) \$ kor2ika đqa (v) L % (eh) bejan alβora (v) HL \$ (tx) (eh) arl yaškifem plađdateko L % (eh) baja (v) \$ kor2iyađ dqała βiđien (v) HL % (eh) (eh) ar2i βateyaš itfeų deų ešťrłopošu (v) H % (eh) eta (v) \$ kamaiek arl yaškife ađajeteų deųe momentuen HL \$ (tx) (eh) aĩtfeųetaja xaušteų da H \$ eta (v) % (eh) arl yaškifen oĩ ayerłteų da L %

Elaboración con base en el presente trabajo de investigación.

Para el análisis que nos ocupa se hizo una conversión de alófonos a fonemas. De este modo se ha confeccionado el corpus de análisis. Este corpus ha servido de base para realizar las diferentes matrices, tanto las dicotómicas como las basadas en las probabilidades, que se irán describiendo en el siguiente apartado referente al análisis de datos.

Por último, reseñar que en los análisis de conglomerados se ha utilizado la distancia euclídea al cuadrado y el método de Ward; estos análisis se han realizado por medio del programa SPSS. Los índices de diversidad se han calculado por medio del programa Past (Hammer et al., 2001).

ANÁLISIS DE LOS DATOS

Con los datos de los 36 micro textos se ha formado un corpus de 12.192 fonemas; los fonemas son un total de 26, los 21 consonánticos que se muestran en la figura 5 más cinco vocálicos /i, e, a, o, u/.

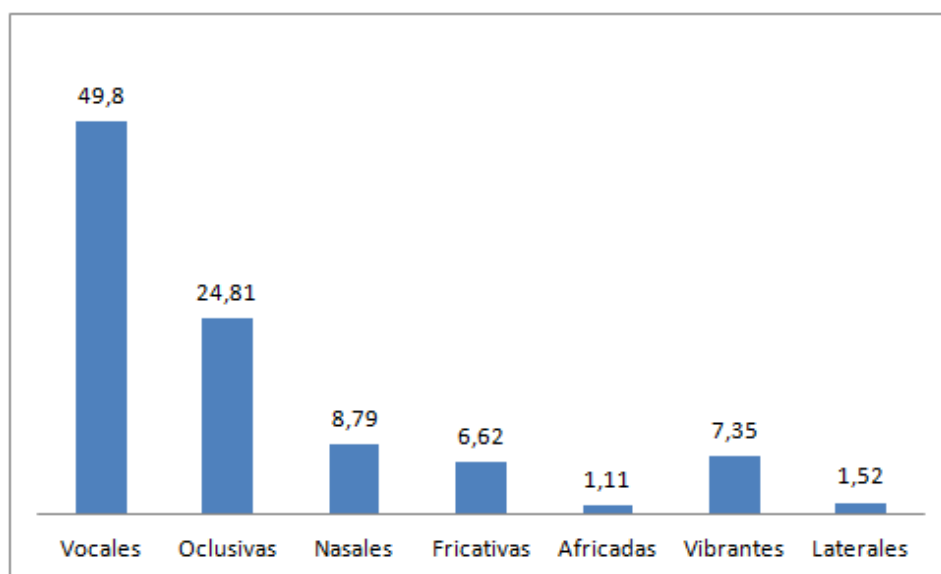
Figura 5. Ejemplo de los fonemas consonánticos

	Labiales	Labiodental	Interdental	Dental alveolar	Palatal	Velar
Oclusivas	p b			t d	ʃ ʒ	k g
Nasales	m			n	ɲ	
Fricativas		f	θ	s	ç	x
Africadas				ts	tʃ	
Vibrantes				r/r		
Laterales				l		

Elaboración con base en el presente trabajo de investigación.

Los fonemas vocálicos son los más utilizados (6072, 49,8%), la vocal más frecuente es /a/ y la menos frecuente /u/. Las oclusivas son 3025 (24,85%), la más frecuente es /t/ y la menos frecuente /j/. Los fonemas nasales son 1072 (8,79%) siendo la /n/ la más frecuente y la /ɲ/ la menos frecuente. Entre las vibrantes (896, 7,35%) la /r/ es la más frecuente. Las fricativas son 807 (6,62%), siendo las /s/ la más frecuente. La única lateral /l/ ha aparecido 185 veces (1,52%). Por último, las africadas son las que tienen una menor frecuencia en el corpus (135, 1,11%), la /ts/ tiene mayor frecuencia que la /tʃ/. En la figura 6 se muestran los datos de cada grupo de fonemas.

Figura 6. Porcentajes de cada grupo de fonemas



Elaboración con base en el presente trabajo de investigación.

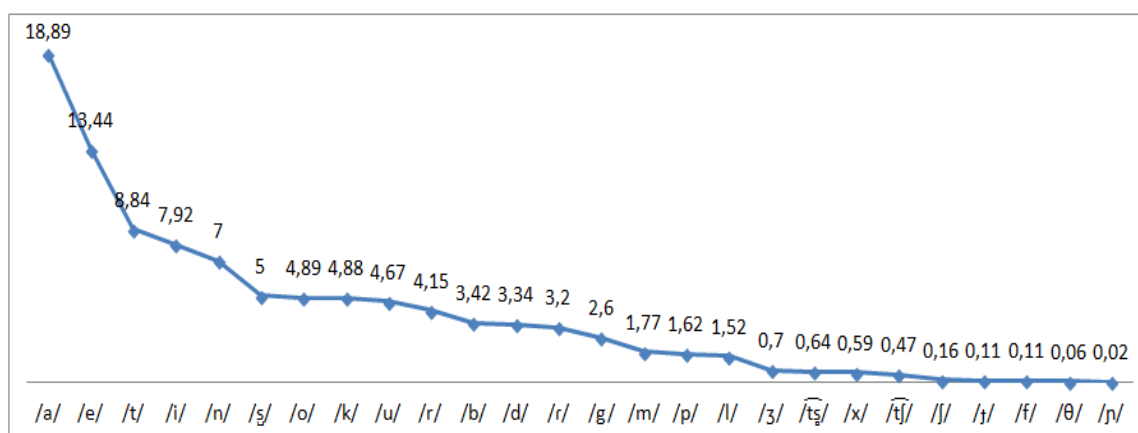
En la tabla 2 se muestran el número total y los porcentajes de cada fonema ordenados de menor a mayor (figura 7); como puede verse la mayor frecuencia corresponde a /a/ y la menor a /p/.

Tabla 2. Cantidad y porcentajes de todos los fonemas

Fonema	N	%	Fonema	N	%
/a/	2303	18,89	/g/	317	2,60
/e/	1638	13,44	/m/	216	1,77
/t/	1078	8,84	/p/	197	1,62
/i/	966	7,92	/l/	185	1,52
/n/	854	7,00	/ʒ/	85	0,70
/s/	609	5,00	/tʃ/	78	0,64
/o/	596	4,89	/x/	72	0,59
/k/	595	4,88	/tʃ/	57	0,47
/u/	569	4,67	/ʃ/	20	0,16
/r/	506	4,15	/j/	14	0,11
/b/	417	3,42	/f/	14	0,11
/d/	407	3,34	/θ/	7	0,06
/r/	390	3,20	/ɲ/	2	0,02

Elaboración con base en el presente trabajo de investigación.

Figura 7. Los porcentajes de cada fonema ordenados de menor a mayor



Elaboración con base en el presente trabajo de investigación.

En la tabla 3 se pueden ver el número de fonemas diferentes y el número total de concurrencias de los mismos por cada informante; el mínimo es 18 y el máximo 24.

Tabla 3. Número de fonemas diferentes y el número total de concurrencias según la localidad y el género de los informantes

Localidad	Mujeres		Hombres	
	Fonemas	N	Fonemas	N
Bermeo	18	259	20	279
Mundaka	20	450	21	224
Busturia	20	212	20	307
Murueta	20	585	19	282
Forua	21	194	21	350
Errigoiti	20	342	22	450
Morga	19	249	21	321
Muxika	18	263	20	287
Mendata	22	394	22	452
Gernika	20	277	20	310
Ajangiz	19	246	20	253
Arratzu	23	459	21	459
Nabarniz	21	424	21	155
Kortezubi	22	399	20	231
Arteaga	22	360	22	404
Ibarrangelu	20	265	18	202
Elantxobe	21	452	20	467
Ea	19	337	24	592

Elaboración con base en el presente trabajo de investigación.

A continuación, presentamos el análisis cuantitativo de los datos en tres secciones. En la primera sección se realiza el análisis con matrices dicotómicas, en la segunda las matrices toman en cuenta las probabilidades de los fonemas y, en la tercera, se utilizan los índices de diversidad.

MATRICES DICOTÓMICAS

Cuando analizamos los datos de cada localidad teniendo en cuenta la aparición o no aparición de los fonemas, tomamos como base una matriz dicotómica como la que se muestra en la figura 8.

Figura 8. Matriz dicotómica

	a	e	i	o	u	p	t	k	b	d	j	g	m	n	ñ	f	θ	s	ʃ	z
Bermeo	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	1
Mundaka	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1
Busturia	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1
Murueta	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	1	1
Forua	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1
Errigoiti	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	1	0	1
Morga	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	0	1
Muxika	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	1
Mendata	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1
Gernika	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0
Ajangiz	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	1
Arratzu	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1
Nabarniz	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	1	1	1
Kortezubi	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	1
Arteaga	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
Ibarrangelu	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	0	1
Elantxobe	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0	1
Ea	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1

Elaboración con base en el presente trabajo de investigación.

En la figura 9 se muestra la matriz de distancias obtenida y el dendrograma de la figura 10; las distancias de la matriz oscilan entre 1 y 7.

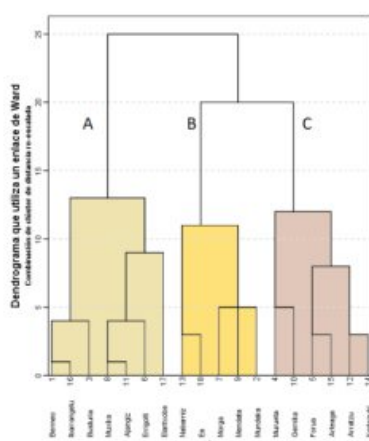
Figura 9. Matriz de distancias

	Bermeo	Mundaka	Busturia	Murueta	Forua	Errigoiti	Morga	Muxika	Mendata	Gernika	Ajangiz	Arratzu	Nabarniz	Kortezubi	Arteaga	Ibarrangelu	Elantxobe	Ea
Bermeo		4	1	3	4	2	4	1	4	5	1	3	3	2	5	0	4	
Mundaka	4		5	3	4	2	2	3	2	5	3	5	3	4	5	4	4	
Busturia	1	5		4	3	3	3	2	3	4	2	2	4	1	4	1	5	
Murueta	3	3	4		3	3	3	2	5	2	2	2	2	3	4	3	5	
Forua	4	4	3	3		2	4	3	2	3	3	1	3	2	1	4	4	
Errigoiti	2	2	3	3	2		4	1	2	5	1	3	3	2	3	2	2	
Morga	4	2	3	3	4	4		3	2	3	3	3	3	2	5	4	6	
Muxika	1	3	2	2	3	1	3		3	4	0	2	2	1	4	1	3	
Mendata	4	2	3	5	2	2	2	3		5	3	3	3	2	3	4	4	
Gernika	5	5	4	2	3	5	3	4	5		4	2	4	3	4	5	7	
Ajangiz	1	3	2	2	3	1	3	0	3	4		2	2	2	1	4	1	3
Arratzu	3	5	2	2	1	3	3	2	3	2	2		2	1	2	3	5	
Nabarniz	3	3	4	2	3	3	3	2	3	4	2	2		3	4	3	5	
Kortezubi	2	4	1	3	2	2	2	1	2	3	1	3	3		3	2	4	
Arteaga	5	5	4	4	1	3	5	4	3	4	4	2	4	3		5	3	
Ibarrangelu	0	4	1	3	4	2	4	1	4	5	1	3	3	2	5		4	
Elantxobe	4	4	5	5	4	2	6	3	4	7	3	5	5	4	3	4		
Ea	4	2	5	2	2	2	4	2	2	5	2	2	1	4	2	4		

Elaboración con base en el presente trabajo de investigación.

En el mapa de la figura 11 se muestran los resultados del dendrograma de la figura 10 proyectados sobre el plano.

Figura 10. Dendrograma de clasificación jerárquica



Elaboración con base en el presente trabajo de investigación.

Figura 11. Los resultados del dendrograma proyectados en el mapa



Al incluir los datos de todos los informantes por separado, en atención al género de los mismos, obtenemos la matriz dicotómica de la figura 12.

Figura 12. Matriz dicotómica según el género de los informantes

	a	e	i	o	u	p	t	k	b	d	f	g	m	n	ñ	l	ll	h	ch	s	z	x
BermeoM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BermeoH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MundakaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MundakaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BusturiaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BusturiaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MuruetaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MuruetaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ForuaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ForuaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ErrigoitiM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ErrigoitiH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MorgaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MorgaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MuxikaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MuxikaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MendataM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MendataH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
GernikaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
GernikaH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AjangizM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AjangizH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ArratzuM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ArratzuH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NabarnizM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NabarnizH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KortezubiM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
KortezubiH	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ArteagaM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Elaboración con base en el presente trabajo de investigación.

En la figura 13 se muestra la matriz de distancias obtenida y el dendrograma de la figura 14; las distancias de la matriz oscilan entre 1 y 8.

Figura 13. Matriz de distancias

	BermeoM	BermeoH	MundakaM	MundakaH	BusturiaM	BusturiaH	MuruetaM	MuruetaH	ForuaM	ForuaH	ErrigoitiM	ErrigoitiH	MorgaM	MorgaH	MuxikaM	MuxikaH	MendataM	MendataH
BermeoM		2	2	3	2	2	4	1	3	5	2	4	1	5	4	2	4	6
BermeoH	2		2	5	2	0	4	1	3	3	2	2	1	5	4	0	2	4
MundakaM	2	2		3	2	2	2	1	1	5	0	2	1	3	2	2	2	4
MundakaH	3	5	3		5	5	5	4	2	8	3	3	4	4	5	5	3	5
BusturiaM	2	2	2	5		2	4	1	3	3	2	4	1	3	4	2	4	4
BusturiaH	2	0	2	5	2		4	1	3	3	2	2	1	5	4	0	2	4
MuruetaM	4	4	2	5	4	4		3	3	3	2	4	3	3	2	4	4	4
MuruetaH	1	1	1	4	1	1	3		2	4	1	3	0	4	3	1	3	5
ForuaM	3	3	1	2	3	3	3	2		6	1	1	2	4	3	3	3	3
ForuaH	5	3	5	8	3	3	4	6	6		5	4	5	3	3	5	3	3
ErrigoitiM	2	2	0	3	2	2	1	1	5	5		2	1	3	2	2	2	4
ErrigoitiH	4	2	2	3	4	2	4	3	1	5	2		3	5	4	2	2	2
MorgaM	1	1	1	4	1	1	3	0	2	4	1	3		4	3	1	3	5
MorgaH	5	5	3	4	3	5	3	4	4	3	5	4	4		3	5	3	3
MuxikaM	4	4	2	5	4	4	2	3	3	5	2	4	3	3		4	4	4
MuxikaH	2	0	2	5	2	0	4	1	3	3	2	2	1	5	4		2	4
MendataM	4	2	2	3	4	2	4	3	3	5	2	2	3	3	4	2		4
MendataH	6	4	4	5	4	4	4	5	3	3	4	2	5	3	4	4	4	
GernikaM	4	4	2	5	4	4	0	3	3	3	2	4	3	3	2	4	4	4
GernikaH	4	4	2	5	2	4	2	3	3	3	2	4	3	1	2	4	4	2
AjangizM	3	3	1	4	3	3	1	2	2	4	1	3	2	2	1	3	3	3
AjangizH	2	0	2	5	2	0	4	1	3	3	2	2	1	5	4	0	2	4
ArratzuM	5	3	3	6	3	3	3	4	4	2	3	3	4	4	5	3	3	3
ArratzuH	3	1	1	4	3	1	3	2	2	4	1	1	2	4	3	1	1	3
NabarnizM	3	3	1	4	3	3	1	2	2	4	1	3	2	4	3	3	3	5
NabarnizH	3	1	3	4	3	1	5	2	4	4	3	3	2	4	5	1	1	5
KortezubiM	4	2	2	5	2	2	4	3	3	3	2	2	3	3	4	2	2	2
KortezubiH	2	0	2	5	2	0	4	1	3	3	2	2	1	5	4	0	2	4
ArteagaM	6	4	4	5	6	4	2	5	3	3	4	2	5	5	4	4	4	2
ArteagaH	4	2	4	7	2	2	6	3	5	3	4	4	3	5	6	2	4	4
IbarrangeluM	2	0	2	5	2	0	4	1	3	3	2	2	1	5	4	0	2	4
IbarrangeluH	0	2	2	3	2	2	4	1	3	5	2	4	1	5	4	2	4	6
ElantxobeM	3	3	3	4	5	3	5	4	6	3	3	4	6	5	3	3	3	5
ElantxobeH	2	2	4	3	4	2	6	3	3	5	4	2	3	7	6	2	4	4
EaM	1	3	1	2	3	3	3	2	2	6	1	3	2	4	3	3	3	5
EaH	6	4	4	3	6	4	4	5	3	5	4	2	5	5	6	4	2	4

Elaboración con base en el presente trabajo de investigación.

En el conglomerado (A) se encuentran 21 informantes (MundakaM, MundakaH, MuruetaM, ForuaM, ForuaH, ErrigoitiM, ErrigoitiH, MorgaM, MuxikaM, MendataM, MendataH, GernikaM, GernikaH, AjangizM, ArratzuM, ArratzuH, NabarnizM, KortezubiM, ArteagaM, EaM y EaH) y en el conglomerado (B) 15 (BermeoM, BermeoH, BusturiaM, BusturiaH, MuruetaH, MorgaM, MuxikaH, AjangizH, NabarnizH, KortezubiH, ArteagaH, IbarrangeluM, IbarrangeluH, ElantxobeM y ElantxobeH). En la tabla 4 pueden verse las diferencias de género en los conglomerados, estas diferencias no son estadísticamente significativas.

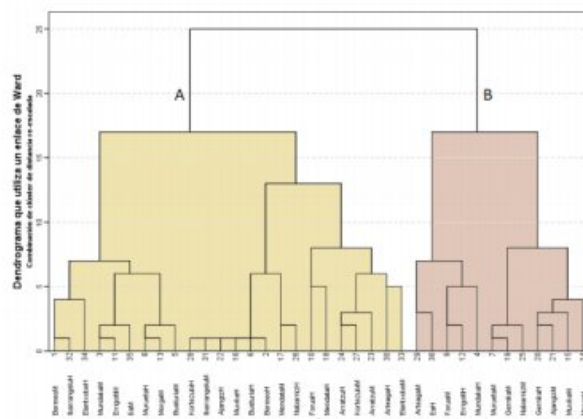
Tabla 4. Diferencias de género obtenidas en los conglomerados

	(A)	(B)	Total
Mujeres	13	5	18
Hombres	8	10	18
Total	21	15	36

Elaboración con base en el presente trabajo de investigación.

Las localidades con ambos informantes dentro del mismo conglomerado son Bermeo, Busturia, Muxika, Gernika, Arratzu, Kortezubi, Ibarangelu y Elantxobe (mapa de la figura 15).

Figura 14. Dendrograma de clasificación jerárquica



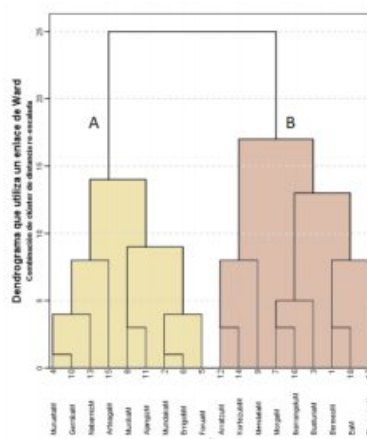
Elaboración con base en el presente trabajo de investigación.

Figura 15. Los resultados del dendrograma proyectados en el mapa



En los mapas de las figuras 16 y 17 se pueden constatar las diferencias existentes según el género de los informantes.

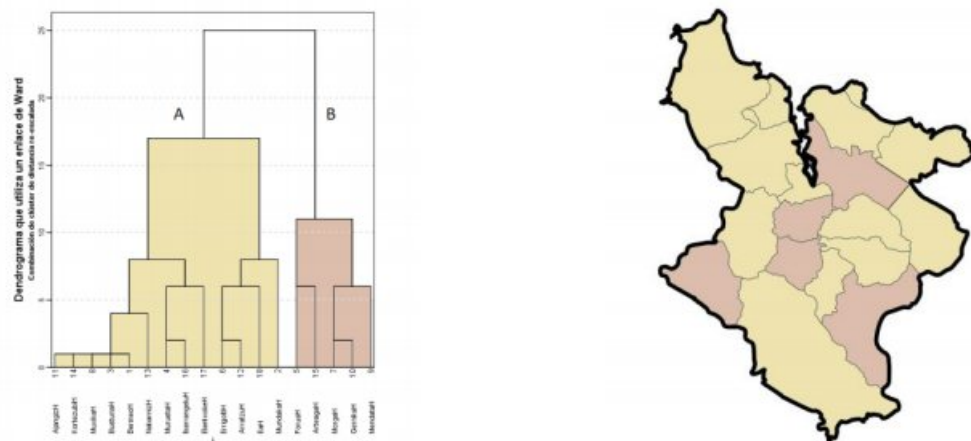
Figura 16. Dendrograma de clasificación jerárquica de las mujeres y resultados proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.



Figura 17. Dendrograma de clasificación jerárquica de los hombres y resultados proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.

ANÁLISIS DE PROBABILIDADES

Como ya hemos mencionado en la introducción otra manera de enfocar el análisis es tener en cuenta los fonemas junto con el número de ocurrencias de cada uno.

Para abordar el análisis de los datos por localidades se ha creado la matriz de la figura 18 donde se recogen los fonemas con el número de casos de cada uno en las respectivas localidades.

Figura 18. Matriz dicotómica según las localidades

	a	e	i	o	u	p	t	k	b	d	y	g	m	n
Bermeo	100	67	46	23	29	4	48	28	20	23	0	14	12	33
Mundaka	132	100	55	31	25	10	62	36	20	25	0	12	13	43
Busturia	96	72	37	25	25	8	50	28	22	24	1	11	6	39
Murueta	151	122	65	41	47	12	88	46	38	32	0	18	11	68
Forua	103	76	33	29	21	8	43	24	23	10	1	18	8	37
Errigoiti	145	104	57	49	41	12	83	36	22	30	0	18	18	52
Morga	111	78	38	31	35	12	53	21	13	20	2	14	12	31
Muxika	111	73	43	25	23	5	47	32	12	15	0	19	12	33
Mendata	161	118	61	33	42	13	83	29	39	22	1	13	14	72
Gernika	118	84	33	34	24	7	48	24	21	23	1	23	13	42
Ajangiz	93	65	46	21	22	11	47	35	12	11	0	14	4	31
Arratzu	173	117	72	49	39	15	70	60	36	19	1	27	17	55
Nabarniz	112	64	52	36	22	12	49	24	20	14	0	19	6	54
Kortezubi	127	79	62	34	22	18	58	27	23	16	5	14	5	48
Arteaga	140	118	64	30	34	15	52	36	23	30	2	22	15	54
Ibarrangelu	84	60	40	24	27	5	40	25	11	26	0	6	11	35
Elanxobe	166	120	84	39	45	16	87	50	21	30	0	25	20	54
Ea	180	121	78	42	46	14	70	34	41	37	0	30	19	73

Elaboración con base en el presente trabajo de investigación.

Con esta matriz se han calculado las probabilidades de cada fonema de acuerdo con las frecuencias de los mismos (figura 19).

Figura 19. Matriz creada según las probabilidades de los fonemas de acuerdo a las frecuencias

	a	e	i	o	u	p	t	k	b	d	y	g	m	n
Bermeo	0,186	0,125	0,086	0,043	0,054	0,007	0,089	0,052	0,037	0,043	0,000	0,026	0,022	0,061
Mundaka	0,196	0,148	0,082	0,046	0,037	0,015	0,092	0,053	0,030	0,037	0,000	0,018	0,019	0,064
Busturia	0,185	0,139	0,071	0,048	0,048	0,015	0,096	0,054	0,042	0,046	0,002	0,021	0,012	0,075
Murueta	0,174	0,141	0,075	0,047	0,054	0,014	0,101	0,053	0,044	0,037	0,000	0,021	0,013	0,078
Forua	0,189	0,140	0,061	0,053	0,039	0,015	0,079	0,044	0,042	0,018	0,002	0,033	0,015	0,068
Errigoiti	0,183	0,131	0,072	0,062	0,052	0,015	0,105	0,045	0,028	0,038	0,000	0,023	0,023	0,066
Morga	0,195	0,137	0,067	0,054	0,061	0,021	0,093	0,037	0,023	0,035	0,004	0,025	0,021	0,054
Muxika	0,202	0,133	0,078	0,045	0,042	0,009	0,085	0,058	0,022	0,027	0,000	0,035	0,022	0,060
Mendata	0,190	0,139	0,072	0,039	0,050	0,015	0,098	0,034	0,046	0,026	0,001	0,015	0,017	0,085
Gernika	0,201	0,143	0,056	0,058	0,041	0,012	0,082	0,041	0,036	0,039	0,002	0,039	0,022	0,072
Ajangiz	0,186	0,130	0,092	0,042	0,044	0,022	0,094	0,070	0,024	0,022	0,000	0,028	0,008	0,062
Arratzu	0,188	0,127	0,078	0,053	0,042	0,016	0,076	0,065	0,039	0,021	0,001	0,029	0,019	0,060
Nabarniz	0,193	0,111	0,090	0,062	0,038	0,021	0,085	0,041	0,035	0,024	0,000	0,033	0,010	0,093
Kortezubi	0,202	0,125	0,098	0,054	0,035	0,029	0,092	0,043	0,037	0,025	0,008	0,022	0,008	0,076
Arteaga	0,183	0,154	0,084	0,039	0,045	0,020	0,068	0,047	0,030	0,039	0,003	0,029	0,020	0,071
Ibarrangelu	0,180	0,128	0,086	0,051	0,058	0,011	0,086	0,054	0,024	0,056	0,000	0,013	0,024	0,075
Elantxobe	0,181	0,131	0,091	0,042	0,049	0,017	0,095	0,054	0,023	0,033	0,000	0,027	0,022	0,059
Ea	0,194	0,130	0,084	0,045	0,050	0,015	0,075	0,037	0,044	0,040	0,000	0,032	0,020	0,079

Elaboración con base en el presente trabajo de investigación.

En la figura 20 se muestra la matriz normalizada por medio de la raíz cuadrada de los datos.

Figura 20. Matriz normalizada empleando la raíz cuadrada

	a	e	i	o	u	p	t	k	b	d	y	g	m	n
Bermeo	0,4311	0,3529	0,2924	0,2068	0,2322	0,0862	0,2987	0,2281	0,1928	0,2068	0,0000	0,1613	0,1493	0,2477
Mundaka	0,4425	0,3852	0,2857	0,2145	0,1926	0,1218	0,3033	0,2311	0,1723	0,1926	0,0000	0,1334	0,1389	0,2526
Busturia	0,4301	0,3725	0,2670	0,2195	0,2195	0,1242	0,3104	0,2323	0,2059	0,2150	0,0439	0,1456	0,1075	0,2741
Murueta	0,4173	0,3751	0,2738	0,2175	0,2328	0,1176	0,3186	0,2303	0,2094	0,1921	0,0000	0,1441	0,1126	0,2801
Forua	0,4351	0,3738	0,2463	0,2309	0,1965	0,1213	0,2811	0,2100	0,2056	0,1356	0,0429	0,1819	0,1213	0,2608
Errigoiti	0,4279	0,3624	0,2683	0,2487	0,2275	0,1231	0,3237	0,2132	0,1667	0,1946	0,0000	0,1508	0,1508	0,2562
Morga	0,4413	0,3699	0,2582	0,2332	0,2478	0,1451	0,3049	0,1919	0,1510	0,1873	0,0592	0,1567	0,1451	0,2332
Muxika	0,4492	0,3643	0,2796	0,2132	0,2045	0,0953	0,2923	0,2412	0,1477	0,1651	0,0000	0,1859	0,1477	0,2449
Mendata	0,4362	0,3735	0,2685	0,1975	0,2228	0,1240	0,3132	0,1851	0,2147	0,1613	0,0344	0,1240	0,1286	0,2917
Gernika	0,4484	0,3783	0,2371	0,2407	0,2022	0,1092	0,2860	0,2022	0,1891	0,1979	0,0413	0,1979	0,1488	0,2675
Ajangiz	0,4317	0,3609	0,3036	0,2051	0,2100	0,1485	0,3069	0,2648	0,1551	0,1485	0,0000	0,1675	0,0895	0,2492
Arratzu	0,4341	0,3570	0,2801	0,2310	0,2061	0,1278	0,2761	0,2557	0,1980	0,1439	0,0330	0,1715	0,1361	0,2448
Nabarniz	0,4398	0,3325	0,2997	0,2494	0,1949	0,1440	0,2909	0,2036	0,1859	0,1555	0,0000	0,1811	0,1018	0,3054
Kortezubi	0,4490	0,3541	0,3137	0,2323	0,1869	0,1690	0,3034	0,2070	0,1911	0,1594	0,0891	0,1491	0,0891	0,2760
Arteaga	0,4281	0,3930	0,2894	0,1982	0,2110	0,1401	0,2609	0,2171	0,1735	0,1982	0,0512	0,1697	0,1401	0,2659
Ibarrangelu	0,4241	0,3584	0,2927	0,2267	0,2404	0,1035	0,2927	0,2314	0,1535	0,2360	0,0000	0,1133	0,1535	0,2738
Elantxobe	0,4250	0,3614	0,3023	0,2060	0,2213	0,1319	0,3077	0,2333	0,1512	0,1807	0,0000	0,1649	0,1475	0,2424
Ea	0,4402	0,3609	0,2898	0,2126	0,2225	0,1228	0,2745	0,1913	0,2101	0,1996	0,0000	0,1797	0,1430	0,2803

Elaboración con base en el presente trabajo de investigación.

El análisis de conglomerados jerárquicos nos ofrece la matriz de distancias de la figura 21 y el dendrograma correspondiente (figura 22); la distancia mayor se encuentra entre las localidades de Gernika y Elantxobe (6,98%) y la menor entre Ajangiz y Arratzu (1,28%). Como se observa en el dendrograma las localidades se aglutinan en tres conglomerados, (A) Bermeo, Ibarrangelu, Busturia y Elantxobe. (B) Murueta, Mendata, Mundaka, Errigoiti, Morga y Gernika. (C) Arteaga, Ea, Forua, Ajangiz, Arratzu, Muxika, Nabarniz y Kortezubi (mapa de la figura 23).

Figura 21. Matriz de distancias

	Bermeo	Mundaka	Busturia	Murueta	Forua	Errigoiti	Morga	Muxika	Mendata	Gernika	Ajangiz	Arratzu	Nabarniz	Kortezubi	Arteaga	Ibarrangelu	Elantxobe	Ea
Bermeo		0,0274	0,0131	0,0211	0,0393	0,0258	0,0347	0,0175	0,0343	0,0566	0,0390	0,0287	0,0439	0,0366	0,0268	0,0090	0,0340	0,0257
Mundaka	0,0274		0,0319	0,0167	0,0417	0,0159	0,0169	0,0179	0,0203	0,0332	0,0223	0,0284	0,0225	0,0371	0,0281	0,0271	0,0264	0,0263
Busturia	0,0131	0,0319		0,0203	0,0289	0,0241	0,0341	0,0242	0,0257	0,0494	0,0256	0,0380	0,0172	0,0217	0,0149	0,0255	0,0256	0,0256
Murueta	0,0211	0,0167	0,0203		0,0316	0,0215	0,0234	0,0236	0,0168	0,0239	0,0221	0,0261	0,0253	0,0338	0,0215	0,0236	0,0344	0,0206
Forua	0,0393	0,0417	0,0289	0,0316		0,0259	0,0387	0,0225	0,0233	0,0377	0,0285	0,0141	0,0311	0,0256	0,0158	0,0505	0,0406	0,0191
Errigoiti	0,0258	0,0159	0,0241	0,0215	0,0259		0,0216	0,0164	0,0182	0,0367	0,0257	0,0249	0,0292	0,0306	0,0219	0,0246	0,0230	0,0197
Morga	0,0347	0,0169	0,0341	0,0234	0,0387	0,0216		0,0218	0,0209	0,0300	0,0318	0,0333	0,0311	0,0343	0,0298	0,0347	0,0410	0,0321
Muxika	0,0175	0,0179	0,0242	0,0236	0,0225	0,0164	0,0218		0,0266	0,0387	0,0144	0,0154	0,0266	0,0298	0,0228	0,0258	0,0219	0,0259
Mendata	0,0343	0,0203	0,0257	0,0168	0,0233	0,0182	0,0209	0,0266		0,0291	0,0244	0,0222	0,0276	0,0239	0,0217	0,0359	0,0397	0,0214
Gernika	0,0566	0,0332	0,0494	0,0239	0,0377	0,0367	0,0300	0,0387	0,0291		0,0432	0,0327	0,0315	0,0511	0,0320	0,0572	0,0698	0,0342
Ajangiz	0,0300	0,0223	0,0256	0,0221	0,0285	0,0257	0,0318	0,0144	0,0244	0,0432		0,0128	0,0128	0,0248	0,0225	0,0255	0,0331	0,0246
Arratzu	0,0287	0,0284	0,0250	0,0261	0,0141	0,0249	0,0333	0,0154	0,0222	0,0327	0,0128		0,0195	0,0190	0,0380	0,0280	0,0280	0,0240
Nabarniz	0,0439	0,0225	0,0380	0,0253	0,0311	0,0292	0,0311	0,0266	0,0276	0,0315	0,0248	0,0234		0,0280	0,0347	0,0455	0,0431	0,0207
Kortezubi	0,0366	0,0371	0,0172	0,0338	0,0256	0,0306	0,0343	0,0298	0,0239	0,0511	0,0225	0,0195	0,0280		0,0244	0,0402	0,0349	0,0318
Arteaga	0,0268	0,0281	0,0217	0,0215	0,0158	0,0219	0,0298	0,0228	0,0217	0,0320	0,0255	0,0190	0,0347	0,0244		0,0301	0,0249	0,0134
Ibarrangelu	0,0090	0,0271	0,0149	0,0236	0,0505	0,0246	0,0347	0,0258	0,0359	0,0572	0,0331	0,0380	0,0455	0,0402	0,0301		0,0193	0,0305
Elantxobe	0,0140	0,0264	0,0255	0,0344	0,0406	0,0230	0,0410	0,0219	0,0397	0,0698	0,0246	0,0280	0,0431	0,0349	0,0249	0,0193		0,0280
Ea	0,0257	0,0263	0,0256	0,0206	0,0191	0,0197	0,0321	0,0259	0,0214	0,0342	0,0336	0,0240	0,0207	0,0318	0,0134	0,0305	0,0280	

Elaboración con base en el presente trabajo de investigación.

Figura 22. Dendrograma de clasificación jerárquica

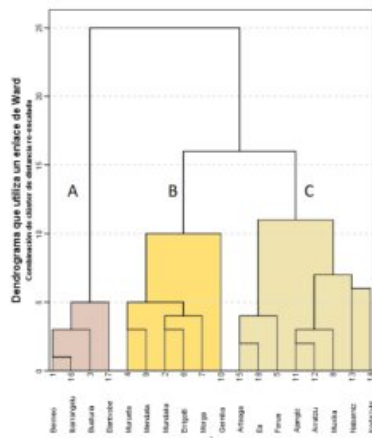


Figura 23. Los resultados del dendrograma proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.

Al introducir en el análisis los datos de todos los informantes de acuerdo al género de los mismos obtenemos la matriz de la figura 24.

Figura 24. Matriz de fonemas y el número total de casos de cada uno

	a	n	i	o	p	r	t	k	h	d	g	m	n	p	f	θ	z	l		
Bernesecht	46	11	22	12	13	3	26	10	11	11	0	0	4	17	0	0	0	15	0	
Berneseis	54	38	24	11	16	3	22	13	9	12	0	8	16	0	0	0	0	17	0	
Bernschakt	30	66	24	31	24	17	3	18	20	13	15	0	10	8	27	0	0	0	19	0
Bernschakt	42	25	24	6	1	24	9	7	8	0	0	0	0	0	0	0	1	13	0	
Bernschakt	18	27	25	8	13	2	20	11	4	11	1	0	14	0	0	0	0	13	0	
Bernschakt	17	45	17	17	12	6	27	10	16	13	0	8	2	25	0	0	0	11	0	
Bernschakt	92	90	66	26	13	8	60	29	26	23	0	11	8	47	0	0	0	20	2	
Bernschakt	59	32	15	13	14	4	28	17	14	9	0	7	3	21	0	0	0	13	0	
Bernschakt	40	23	14	10	7	3	12	12	9	1	0	12	1	0	0	0	0	13	0	
Bernschakt	80	53	19	19	14	4	31	12	12	9	1	0	12	7	22	0	0	16	3	
Bernschakt	85	48	36	17	13	5	13	18	8	12	0	11	8	21	0	0	0	0	0	
Bernschakt	80	18	27	82	28	7	16	18	14	18	0	7	10	31	0	0	0	19	0	
Bernschakt	11	35	18	11	14	6	20	13	3	11	0	5	8	0	0	0	0	16	0	
Bernschakt	18	43	26	21	6	3	13	1	8	30	9	2	4	0	0	0	0	14	0	
Bernschakt	62	13	18	15	6	2	28	14	8	7	0	0	7	13	0	0	0	13	0	
Bernschakt	49	40	25	10	17	3	19	18	4	8	0	11	5	20	0	0	0	16	0	
Bernschakt	76	55	27	12	22	7	41	14	14	14	0	5	5	37	0	0	1	21	0	
Bernschakt	85	65	34	21	23	6	40	15	23	8	1	4	9	35	0	1	0	20	0	
Bernschakt	56	40	18	11	12	1	21	12	7	11	0	14	6	0	0	0	0	13	3	
Bernschakt	62	44	15	27	12	3	24	17	52	12	1	0	7	20	0	0	0	13	0	
Bernschakt	41	14	27	19	10	4	28	17	6	7	0	5	2	18	0	0	0	8	0	
Bernschakt	52	31	19	8	12	7	19	10	6	4	0	9	2	13	0	0	0	15	0	
Bernschakt	81	62	26	18	2	9	31	32	17	11	1	7	12	7	0	0	0	0	0	
Bernschakt	82	54	37	22	13	6	39	28	13	8	0	12	10	28	0	0	0	29	0	
Bernschakt	83	48	18	23	16	6	31	21	18	4	0	14	5	43	0	0	0	23	2	
Bernschakt	29	13	14	13	6	0	18	3	2	8	0	5	3	11	0	0	1	9	0	
Bernschakt	76	49	42	19	18	10	51	20	15	9	3	9	8	8	0	0	0	19	0	
Bernschakt	11	30	26	11	4	0	11	4	7	7	0	3	17	0	0	0	0	11	0	
Bernschakt	71	57	26	17	19	7	29	17	13	14	0	0	0	18	0	2	0	12	3	
Bernschakt	89	65	18	13	21	8	23	19	10	10	14	2	14	8	36	1	0	0	22	0
Bernschakt	49	16	25	14	15	4	25	16	8	13	0	3	7	13	0	0	0	13	0	
Bernschakt	15	24	15	10	10	12	3	15	9	3	13	0	4	22	0	0	0	15	0	
Bernschakt	80	59	31	21	21	6	44	24	14	13	0	12	7	13	0	0	0	25	0	
Bernschakt	80	65	49	18	20	8	43	26	13	17	0	13	13	28	0	2	0	26	0	

Elaboración con base en el presente trabajo de investigación.

Una vez calculadas las probabilidades de cada fonema (figura 25) y normalizada la matriz por medio de la raíz cuadrada de los datos (figura 26), procedemos al análisis de conglomerados jerárquicos.

Figura 25. Matriz basada en las probabilidades de cada fonema

	a	e	i	o	u	p	t	k	h	ʃ	ʒ	g	m	n	ɲ	ɰ	ʁ	ʕ	ʔ
BermeoM	0.1776	0.1197	0.0849	0.0863	0.0902	0.0118	0.1304	0.0579	0.0425	0.0425	0.0000	0.0232	0.0154	0.0656	0.0000	0.0000	0.0000	0.0000	0.0000
BermeoH	0.1315	0.1290	0.0880	0.0394	0.0279	0.0018	0.0789	0.0486	0.0323	0.0430	0.0000	0.0287	0.0278	0.0573	0.0000	0.0000	0.0000	0.0000	0.0000
MonduM	0.2000	0.1311	0.0989	0.0513	0.0178	0.0087	0.0444	0.0422	0.0389	0.0423	0.0000	0.0222	0.0278	0.0609	0.0000	0.0000	0.0000	0.0000	0.0000
MonduH	0.1375	0.1425	0.1071	0.0313	0.0357	0.0013	0.1071	0.0337	0.0313	0.0286	0.0000	0.0089	0.0223	0.0734	0.0000	0.0040	0.0045	0.0045	0.0000
BermeoM	0.1840	0.1274	0.0943	0.0377	0.0619	0.0094	0.1385	0.0615	0.0283	0.0519	0.0047	0.0342	0.0289	0.0660	0.0000	0.0000	0.0000	0.0000	0.0000
BermeoH	0.1357	0.1466	0.0504	0.0554	0.0391	0.0195	0.0879	0.0489	0.0521	0.0421	0.0000	0.0261	0.0305	0.0834	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.1371	0.1518	0.0786	0.0476	0.0564	0.0137	0.1026	0.0496	0.0410	0.0393	0.0000	0.0188	0.0137	0.0881	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizH	0.2002	0.1335	0.0878	0.0461	0.0496	0.0142	0.0991	0.0303	0.0496	0.0319	0.0000	0.0248	0.0106	0.0745	0.0000	0.0000	0.0000	0.0000	0.0000
FornalM	0.2062	0.1188	0.0722	0.0513	0.0361	0.0208	0.0619	0.0464	0.0367	0.0052	0.0000	0.0309	0.0052	0.0773	0.0000	0.0052	0.0000	0.0000	0.0000
FornalH	0.1800	0.1534	0.0543	0.0400	0.0114	0.0886	0.0429	0.0343	0.0257	0.0029	0.0000	0.0243	0.0200	0.0629	0.0000	0.0000	0.0000	0.0000	0.0000
ElantxobeM	0.1301	0.1404	0.0877	0.0497	0.0380	0.0144	0.0960	0.0536	0.0234	0.0351	0.0000	0.0322	0.0234	0.0834	0.0000	0.0000	0.0000	0.0000	0.0000
ElantxobeH	0.1776	0.1244	0.0903	0.0791	0.0422	0.0154	0.1111	0.0489	0.0311	0.0403	0.0000	0.0154	0.0212	0.0849	0.0000	0.0000	0.0000	0.0000	0.0000
MugaM	0.1326	0.1405	0.0723	0.0462	0.0542	0.0041	0.0803	0.0522	0.0128	0.0443	0.0000	0.0201	0.0321	0.0281	0.0000	0.0000	0.0000	0.0000	0.0000
MugaH	0.1307	0.1340	0.0623	0.0623	0.0604	0.0187	0.1028	0.0249	0.0312	0.0280	0.0082	0.0280	0.0125	0.0810	0.0000	0.0000	0.0000	0.0000	0.0000
MonduM	0.2357	0.1255	0.0864	0.0570	0.0228	0.0078	0.1385	0.0532	0.0364	0.0266	0.0000	0.0304	0.0266	0.0454	0.0000	0.0000	0.0000	0.0000	0.0000
MonduH	0.1307	0.1394	0.0871	0.0348	0.0392	0.0105	0.0862	0.0627	0.0139	0.0279	0.0000	0.0383	0.0174	0.0897	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.1376	0.1394	0.0685	0.0396	0.0394	0.0178	0.1291	0.0335	0.0355	0.0096	0.0000	0.0127	0.0217	0.0939	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizH	0.1881	0.1384	0.0752	0.0465	0.0442	0.0133	0.0885	0.0382	0.0353	0.0177	0.0012	0.0177	0.0199	0.0774	0.0000	0.0000	0.0000	0.0000	0.0000
GernikaM	0.2022	0.1444	0.0804	0.0433	0.0433	0.0144	0.0794	0.0233	0.0325	0.0397	0.0000	0.0305	0.0217	0.0794	0.0000	0.0000	0.0000	0.0000	0.0000
GernikaH	0.2000	0.1413	0.0484	0.0710	0.0387	0.0097	0.0829	0.0248	0.0387	0.0302	0.0012	0.0290	0.0248	0.0645	0.0000	0.0000	0.0000	0.0000	0.0000
AlangM	0.1307	0.1382	0.0398	0.0528	0.0487	0.0163	0.1238	0.0491	0.0244	0.0285	0.0000	0.0203	0.0081	0.0712	0.0000	0.0000	0.0000	0.0000	0.0000
AlangH	0.2000	0.1225	0.0751	0.0218	0.0478	0.0277	0.0751	0.0771	0.0217	0.0226	0.0000	0.0304	0.0379	0.0514	0.0000	0.0000	0.0000	0.0000	0.0000
AratzM	0.1305	0.1373	0.0763	0.0504	0.0392	0.0194	0.0879	0.0370	0.0240	0.0612	0.0000	0.0261	0.0131	0.0832	0.0000	0.0000	0.0000	0.0000	0.0000
AratzH	0.2004	0.1378	0.0806	0.0501	0.0458	0.0133	0.0892	0.0610	0.0414	0.0133	0.0000	0.0317	0.0218	0.0746	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.1358	0.1354	0.0894	0.0547	0.0377	0.0142	0.0773	0.0493	0.0423	0.0342	0.0000	0.0330	0.0118	0.1034	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizH	0.1471	0.0969	0.0903	0.0818	0.0387	0.0187	0.1281	0.0234	0.0129	0.0518	0.0000	0.0313	0.0460	0.0702	0.0000	0.0000	0.0000	0.0000	0.0000
KortxoM	0.1305	0.1238	0.0393	0.0578	0.0453	0.0253	0.0802	0.0301	0.0378	0.0229	0.0125	0.0228	0.0075	0.0777	0.0000	0.0000	0.0000	0.0000	0.0000
KortxoH	0.2005	0.1299	0.0884	0.0476	0.0175	0.0348	0.1128	0.0303	0.0348	0.0303	0.0000	0.0218	0.0087	0.0738	0.0000	0.0000	0.0000	0.0000	0.0000
AlangM	0.1377	0.1382	0.0722	0.0472	0.0381	0.0194	0.0806	0.0472	0.0363	0.0444	0.0000	0.0212	0.0184	0.0700	0.0000	0.0000	0.0000	0.0000	0.0000
AlangH	0.1308	0.1510	0.0943	0.0322	0.0520	0.0198	0.0589	0.0470	0.0248	0.0367	0.0000	0.0447	0.0208	0.0891	0.0013	0.0000	0.0000	0.0000	0.0000
BerrangM	0.1305	0.1348	0.0443	0.0528	0.0546	0.0151	0.0943	0.0604	0.0302	0.0493	0.0000	0.0113	0.0264	0.0411	0.0000	0.0000	0.0000	0.0000	0.0000
BerrangH	0.1733	0.1180	0.0743	0.0405	0.0594	0.0240	0.0743	0.0440	0.0149	0.0244	0.0000	0.0149	0.0240	0.1089	0.0000	0.0000	0.0000	0.0000	0.0000
ElantxobeM	0.1770	0.1305	0.0774	0.0405	0.0553	0.0177	0.0973	0.0313	0.0354	0.0688	0.0000	0.0205	0.0155	0.0575	0.0012	0.0000	0.0000	0.0000	0.0000
ElantxobeH	0.1342	0.1308	0.0349	0.0305	0.0438	0.0171	0.0921	0.0557	0.0317	0.0394	0.0000	0.0278	0.0278	0.0800	0.0000	0.0000	0.0000	0.0000	0.0000
Sabi	0.2020	0.1310	0.0861	0.0217	0.0443	0.0017	0.0772	0.0336	0.0344	0.0415	0.0000	0.0297	0.0179	0.0742	0.0000	0.0000	0.0000	0.0000	0.0000
Sabi	0.1375	0.1384	0.0838	0.0376	0.0516	0.0101	0.0740	0.0388	0.0388	0.0388	0.0000	0.0318	0.0228	0.0811	0.0000	0.0000	0.0000	0.0000	0.0000

Elaboración con base en el presente trabajo de investigación.

Figura 26. Matriz normalizada empleando la raíz cuadrada

	a	e	i	o	u	p	t	k	h	ʃ	ʒ	g	m	n	ɲ	ɰ	ʁ	ʕ	ʔ
BermeoM	0.4214	0.3440	0.2914	0.2952	0.2240	0.1078	0.3108	0.2407	0.2061	0.2061	0.0000	0.1527	0.1241	0.2542	0.0000	0.0000	0.0000	0.0000	0.0000
BermeoH	0.4199	0.2952	0.2913	0.1388	0.2393	0.0599	0.2489	0.2119	0.1794	0.2078	0.0000	0.1493	0.1493	0.2393	0.0000	0.0000	0.0000	0.0000	0.0000
MonduM	0.4472	0.3887	0.2625	0.2305	0.1344	0.0818	0.2908	0.2494	0.1700	0.2055	0.0000	0.1493	0.1333	0.2449	0.0000	0.0000	0.0000	0.0000	0.0000
MonduH	0.4130	0.3780	0.3273	0.1768	0.1890	0.1768	0.3273	0.1890	0.1768	0.1837	0.0000	0.0940	0.1494	0.2673	0.0000	0.0668	0.0668	0.0668	0.0000
BermeoM	0.4439	0.3169	0.3071	0.1343	0.2478	0.0973	0.3294	0.2478	0.1882	0.2278	0.0000	0.1180	0.1374	0.2530	0.0000	0.0000	0.0000	0.0000	0.0000
BermeoH	0.4309	0.3625	0.2103	0.2315	0.1977	0.1198	0.2996	0.2239	0.2381	0.2058	0.0000	0.1814	0.0807	0.2854	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.3964	0.3749	0.2804	0.2388	0.2398	0.1275	0.3101	0.2136	0.2075	0.1983	0.0000	0.1171	0.1199	0.2834	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizH	0.4574	0.3169	0.2596	0.2347	0.2228	0.1191	0.3151	0.2435	0.2238	0.1786	0.0000	0.1576	0.1911	0.2729	0.0000	0.0000	0.0000	0.0000	0.0000
FornalM	0.4341	0.3443	0.2886	0.2270	0.1900	0.1418	0.2487	0.2134	0.2383	0.0718	0.0000	0.1799	0.0718	0.2781	0.0000	0.0718	0.0000	0.0000	0.0000
FornalH	0.4243	0.3891	0.2330	0.2330	0.2000	0.1089	0.2970	0.2870	0.1852	0.1894	0.0000	0.1852	0.1414	0.2927	0.0000	0.0000	0.0000	0.0000	0.0000
ElantxobeM	0.4298	0.3436	0.2963	0.2274	0.1860	0.1089	0.3108	0.2478	0.2075	0.1983	0.0000	0.1299	0.1299	0.2834	0.0000	0.0000	0.0000	0.0000	0.0000
ElantxobeH	0.4212	0.3520	0.2449	0.2467	0.2494	0.1247	0.3133	0.2090	0.1784	0.2089	0.0000	0.1247	0.1491	0.2625	0.0000	0.0000	0.0000	0.0000	0.0000
MugaM	0.4814	0.3749	0.2889	0.2371	0.1552	0.1084	0.2285	0.1084	0.1102	0.0000	0.0000	0.1417	0.1792	0.1417	0.0000	0.0000	0.0000	0.0000	0.0000
MugaH	0.4251	0.3660	0.2496	0.2496	0.2158	0.1367	0.3206	0.1579	0.1765	0.1874	0.0789	0.1874	0.1116	0.2846	0.0000	0.0000	0.0000	0.0000	0.0000
MonduM	0.4405	0.3542	0.2818	0.2388	0.1520	0.0872	0.3261	0.2387	0.1744	0.1451	0.0000	0.1744	0.1451	0.2213	0.0000	0.0000	0.0000	0.0000	0.0000
MonduH	0.4132	0.3713	0.2951	0.1367	0.2434	0.1022	0.2713	0.2504	0.1181	0.1878	0.0000	0.1958	0.1130	0.2647	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.4137	0.3713	0.2743	0.2355	0.2304	0.1152	0.2975	0.1852	0.2352	0.1385	0.0470	0.1330	0.1411	0.2783	0.0000	0.0470	0.0000	0.0000	0.0000
NabarnizH	0.4439	0.3800	0.2549	0.2061	0.2061	0.1282	0.2818	0.1282	0.1390	0.0390	0.0000	0.2248	0.1472	0.2818	0.0000	0.0000	0.0000	0.0000	0.0000
Sorokum	0.4472	0.3787	0.2208	0.2684	0.1961	0.0994	0.2896	0.2342	0.1967	0.2367	0.0566	0.1704	0.1301	0.2540	0.0000	0.0595	0.0000	0.0000	0.0000
ElantxobeM	0.4312	0.3718	0.2951	0.2387	0.1872	0.1028	0.3128	0.2478	0.2075	0.1983	0.0000	0.1299	0.1299	0.2834	0.0000	0.0000	0.0000	0.0000	0.0000
Ajargine	0.4134	0.3500	0.2740	0.1778	0.2178	0.1863	0.2748	0.2667	0.1540	0.1377	0.0000	0.1888	0.0800	0.3287	0.0000	0.0000	0.0000	0.0000	0.0000
AraratM	0.4201	0.3705	0.2763	0.2380	0.1980	0.1406	0.2598	0.2649	0.1925	0.1548	0.0467	0.1617	0.1225	0.2514	0.0000	0.0000	0.0000	0.0000	0.0000
AraratH	0.4477	0.3430	0.2839	0.2229	0.2139	0.1143	0.2913	0.2470	0.2015	0.1320	0.0000	0.1868	0.1476	0.2380	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.4624	0.3400	0.2594	0.2229	0.1943	0.1196	0.2794	0.2225	0.2000	0.1129	0.0000	0.1617	0.1398	0.3185	0.0000	0.0595	0.0000	0.0000	0.0000
NabarnizH	0.4312	0.3718	0.2951	0.1367	0.2434	0.1022	0.2713	0.2504	0.1181	0.1878	0.0000	0.1958	0.1130	0.2647	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobM	0.4194	0.3504	0.2844	0.2001	0.2124	0.1583	0.2851	0.2279	0.1939	0.1590	0.1119	0.1502	0.0987	0.2797	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobH	0.4499	0.3064	0.2747	0.2382	0.1318	0.1861	0.3035	0.1781	0.1861	0.1741	0.0000	0.1497	0.0902	0.2713	0.0000	0.0000	0.0000	0.0000	0.0000
AjargineM	0.4441	0.3879	0.2887	0.2173	0.1900	0.1394	0.2888	0.2173	0.1900	0.2108	0.0188	0.1481	0.1394	0.2236	0.0000	0.0188	0.0000	0.0000	0.0000
AjargineH	0.4133	0.3898	0.3067	0.1774	0.2289	0.1407	0.2388	0.2389	0.1573	0.1862	0.0734	0.1862	0.1497	0.2985	0.0488	0.0488	0.0000	0.0000	0.0000
AraratM	0.4201	0.3705	0.2763	0.2380	0.1980	0.1406	0.2598	0.2649	0.1925	0.1548	0.0467	0.1617	0.1225	0.2514	0.0000	0.0000	0.0000	0.0000	0.0000
AraratH	0.4477	0.3430	0.2839	0.2229	0.2139	0.1143	0.2913	0.2470	0.2015	0.1320	0.0000	0.1868	0.1476	0.2380	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.4624	0.3400	0.2594	0.2229	0.1943	0.1196	0.2794	0.2225	0.2000	0.1129	0.0000	0.1617	0.1398	0.3185	0.0000	0.0595	0.0000	0.0000	0.0000
NabarnizH	0.4312	0.3718	0.2951	0.1367	0.2434	0.1022	0.2713	0.2504	0.1181	0.1878	0.0000	0.1958	0.1130	0.2647	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobM	0.4194	0.3504	0.2844	0.2001	0.2124	0.1583	0.2851	0.2279	0.1939	0.1590	0.1119	0.1502	0.0987	0.2797	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobH	0.4499	0.3064	0.2747	0.2382	0.1318	0.1861	0.3035	0.1781	0.1861	0.1741	0.0000	0.1497	0.0902	0.2713	0.0000	0.0000	0.0000	0.0000	0.0000
AjargineM	0.4441	0.3879	0.2887	0.2173	0.1900	0.1394	0.2888	0.2173	0.1900	0.2108	0.0188	0.1481	0.1394	0.2236	0.0000	0.0188	0.0000	0.0000	0.0000
AjargineH	0.4133	0.3898	0.3067	0.1774	0.2289	0.1407	0.2388	0.2389	0.1573	0.1862	0.0734	0.1862	0.1497	0.2985	0.0488	0.0488	0.0000	0.0000	0.0000
AraratM	0.4201	0.3705	0.2763	0.2380	0.1980	0.1406	0.2598	0.2649	0.1925	0.1548	0.0467	0.1617	0.1225	0.2514	0.0000	0.0000	0.0000	0.0000	0.0000
AraratH	0.4477	0.3430	0.2839	0.2229	0.2139	0.1143	0.2913	0.2470	0.2015	0.1320	0.0000	0.1868	0.1476	0.2380	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.4624	0.3400	0.2594	0.2229	0.1943	0.1196	0.2794	0.2225	0.2000	0.1129	0.0000	0.1617	0.1398	0.3185	0.0000	0.0595	0.0000	0.0000	0.0000
NabarnizH	0.4312	0.3718	0.2951	0.1367	0.2434	0.1022	0.2713	0.2504	0.1181	0.1878	0.0000	0.1958	0.1130	0.2647	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobM	0.4194	0.3504	0.2844	0.2001	0.2124	0.1583	0.2851	0.2279	0.1939	0.1590	0.1119	0.1502	0.0987	0.2797	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobH	0.4499	0.3064	0.2747	0.2382	0.1318	0.1861	0.3035	0.1781	0.1861	0.1741	0.0000	0.1497	0.0902	0.2713	0.0000	0.0000	0.0000	0.0000	0.0000
AjargineM	0.4441	0.3879	0.2887	0.2173	0.1900	0.1394	0.2888	0.2173	0.1900	0.2108	0.0188	0.1481	0.1394	0.2236	0.0000	0.0188	0.0000	0.0000	0.0000
AjargineH	0.4133	0.3898	0.3067	0.1774	0.2289	0.1407	0.2388	0.2389	0.1573	0.1862	0.0734	0.1862	0.1497	0.2985	0.0488	0.0488	0.0000	0.0000	0.0000
AraratM	0.4201	0.3705	0.2763	0.2380	0.1980	0.1406	0.2598	0.2649	0.1925	0.1548	0.0467	0.1617	0.1225	0.2514	0.0000	0.0000	0.0000	0.0000	0.0000
AraratH	0.4477	0.3430	0.2839	0.2229	0.2139	0.1143	0.2913	0.2470	0.2015	0.1320	0.0000	0.1868	0.1476	0.2380	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.4624	0.3400	0.2594	0.2229	0.1943	0.1196	0.2794	0.2225	0.2000	0.1129	0.0000	0.1617	0.1398	0.3185	0.0000	0.0595	0.0000	0.0000	0.0000
NabarnizH	0.4312	0.3718	0.2951	0.1367	0.2434	0.1022	0.2713	0.2504	0.1181	0.1878	0.0000	0.1958	0.1130	0.2647	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobM	0.4194	0.3504	0.2844	0.2001	0.2124	0.1583	0.2851	0.2279	0.1939	0.1590	0.1119	0.1502	0.0987	0.2797	0.0000	0.0000	0.0000	0.0000	0.0000
KortxobH	0.4499	0.3064	0.2747	0.2382	0.1318	0.1861	0.3035	0.1781	0.1861	0.1741	0.0000	0.1497	0.0902	0.2713	0.0000	0.0000	0.0000	0.0000	0.0000
AjargineM	0.4441	0.3879	0.2887	0.2173	0.1900	0.1394	0.2888	0.2173	0.1900	0.2108	0.0188	0.1481	0.1394	0.2236	0.0000	0.0188	0.0000	0.0000	0.0000
AjargineH	0.4133	0.3898	0.3067	0.1774	0.2289	0.1407	0.2388	0.2389	0.1573	0.1862	0.0734	0.1862	0.1497	0.2985	0.0488	0.0488	0.0000	0.0000	0.0000
AraratM	0.4201	0.3705	0.2763	0.2380	0.1980	0.1406	0.2598	0.2649	0.1925	0.1548	0.0467	0.1617	0.1225	0.2514	0.0000	0.0000	0.0000	0.0000	0.0000
AraratH	0.4477	0.3430	0.2839	0.2229	0.2139	0.1143	0.2913	0.2470	0.2015	0.1320	0.0000	0.1868	0.1476	0.2380	0.0000	0.0000	0.0000	0.0000	0.0000
NabarnizM	0.4624	0.3400	0.2594	0.2229	0.1943	0.													

Figura 27. Matriz de distancias

	BermeoM	BermeoH	MundakaM	MundakaH	MuruetaM	MuruetaH	ForuaM	ForuaH	ErrigoitiM	ErrigoitiH	MorgaM	MorgaH	MuxikaM	MuxikaH	MendataM	MendataH	GernikaM	GernikaH	AjangizM	AjangizH
BermeoM	0	0.1417	0.0326	0.0479	0.0381	0.0313	0.0405	0.0381	0.0405	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
BermeoH	0.1417	0	0.0326	0.0479	0.0381	0.0313	0.0405	0.0381	0.0405	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
MundakaM	0.0326	0.0479	0	0.0326	0.0381	0.0313	0.0405	0.0381	0.0405	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
MundakaH	0.0479	0.0381	0.0326	0	0.0381	0.0313	0.0405	0.0381	0.0405	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
MuruetaM	0.0381	0.0313	0.0405	0.0381	0	0.0313	0.0405	0.0381	0.0405	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
MuruetaH	0.0313	0.0405	0.0381	0.0313	0.0405	0	0.0405	0.0381	0.0405	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
ForuaM	0.0405	0.0381	0.0405	0.0381	0.0405	0.0381	0	0.0405	0.0381	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
ForuaH	0.0405	0.0381	0.0405	0.0381	0.0405	0.0381	0.0405	0	0.0405	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
ErrigoitiM	0.0294	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
ErrigoitiH	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407
MorgaM	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473
MorgaH	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407
MuxikaM	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403
MuxikaH	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407
MendataM	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403
MendataH	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407
GernikaM	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
GernikaH	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
AjangizM	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314
AjangizH	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0411	0.0379	0.0794	0.0881	0.0414	0.0473	0.0407	0.0403	0.0407	0.0314

Elaboración con base en el presente trabajo de investigación.

En el dendrograma de la figura 28 podemos ver que en el conglomerado (A) se aglutinan 21 informantes: MundakaM, MundakaH, MuruetaM, ForuaM, ForuaH, ErrigoitiM, ErrigoitiH, MorgaH, MuxikaM, MendataM, MendataH, GernikaM, GernikaH, AjangizM, ArratzuM, ArratzuH, NabarnizM, KorteziubiM, ArteagaM, EaM y EaH y en el coglomerado (B) 15: BermeoM, BermeoH, BusturiaM, BusturiaH, MuruetaH, MorgaM, MuxikaH, AjangizH, NabarnizH, KorteziubiH, ArteagaH, IbarrangeluM, IbarrangeluH, ElantxobeM y ElantxobeH. En las localidades de Bermeo, Mundaka, Busturia, Forua, Errigoiti, Mendata, Gernika, Arratzu, Ibarrangelu, Elantxobe y Ea ambos informantes aparecen en el mismo conglomerado (mapa de la figura 29).

Figura 28. Dendrograma de clasificación jerárquica

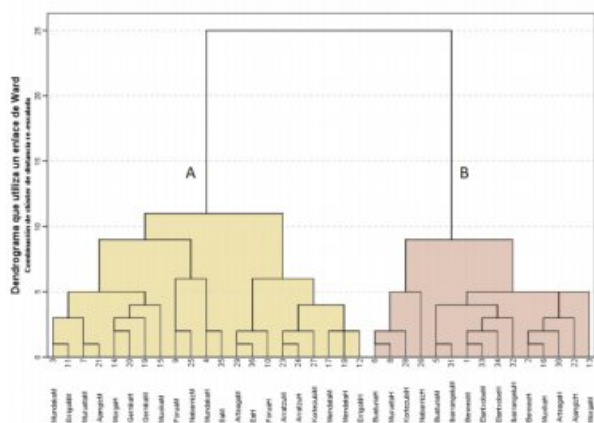


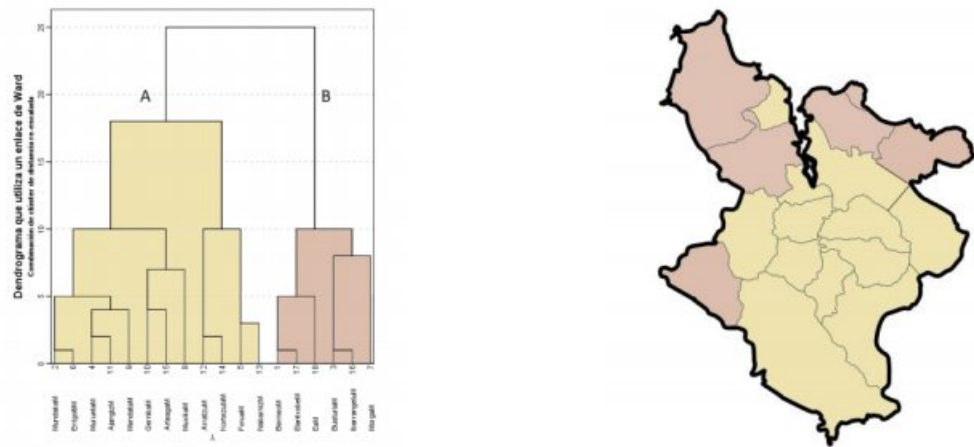
Figura 29. Los resultados del dendrograma proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.

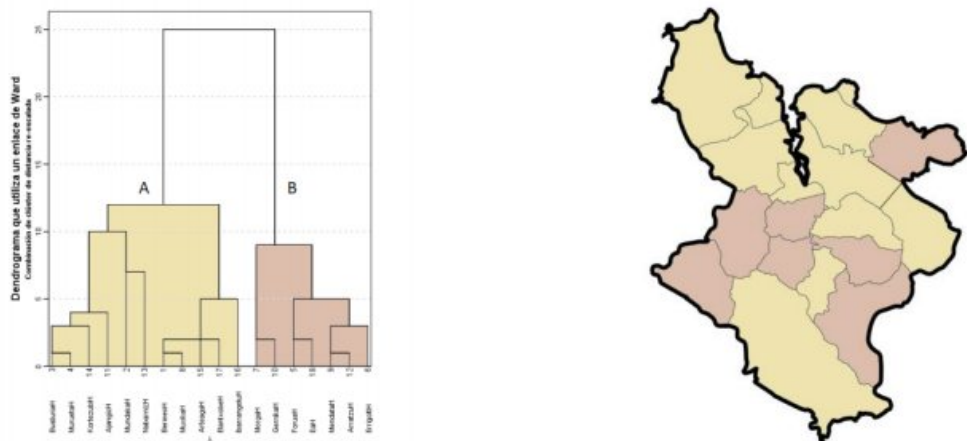
Al analizar los datos de pertenencia a uno u otro conglomerado según el género de los informantes obtenemos distribuciones diferentes, como se puede observar en las figuras 30 (mujeres) y 31 (hombres).

Figura 30. Dendrograma de clasificación jerárquica de las mujeres y resultados proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.

Figura 31. Dendrograma de clasificación jerárquica de los hombres y resultados proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.

ÍNDICES DE DIVERSIDAD

En esta sección se clasifican las localidades primero y luego los informantes según los índices de diversidad. Para realizar las operaciones de clasificación partimos de las matrices formadas con los fonemas y el número total de casos de cada uno (sección *Análisis de probabilidades*). Por medio del programa Past (Hammer, Harper & Ryan) se han obtenido los índices de diversidad. En la figura 32 se muestran los índices de diversidad de las localidades y en la figura 33 la matriz normalizada por medio de la raíz cuadrada.

Figura 32. Matriz de los índices de diversidad según las localidades

	Dominance_D	Simpson_1-D	Shannon_H	Evenness_e^H/S	Brillouin	Menhinick	Margalef	Equitability_J	Fisher_alpha
Bermeo	0,086	0,914	2,675	0,725	2,578	0,862	3,022	0,893	4,093
Mundaka	0,093	0,907	2,655	0,646	2,569	0,847	3,224	0,859	4,359
Busturia	0,089	0,911	2,674	0,691	2,570	0,922	3,199	0,879	4,393
Murueta	0,088	0,912	2,658	0,680	2,591	0,713	2,956	0,873	3,879
Forua	0,086	0,914	2,745	0,648	2,634	1,029	3,651	0,864	5,137
Errigoiti	0,087	0,913	2,693	0,672	2,617	0,782	3,146	0,871	4,193
Morga	0,089	0,911	2,695	0,673	2,595	0,922	3,309	0,872	4,546
Muxika	0,091	0,909	2,675	0,691	2,575	0,895	3,170	0,879	4,327
Mendata	0,092	0,908	2,665	0,599	2,591	0,825	3,412	0,839	4,597
Gernika	0,091	0,909	2,683	0,697	2,588	0,867	3,137	0,881	4,256
Ajangiz	0,089	0,911	2,677	0,693	2,569	0,940	3,219	0,879	4,439
Arratzu	0,085	0,915	2,728	0,665	2,658	0,759	3,225	0,870	4,281
Nabarniz	0,088	0,912	2,699	0,646	2,598	0,956	3,458	0,861	4,788
Kortezubi	0,093	0,907	2,670	0,657	2,578	0,877	3,258	0,864	4,432
Arteaga	0,088	0,912	2,722	0,608	2,637	0,905	3,615	0,846	4,956
Ibarrangelu	0,087	0,913	2,648	0,706	2,540	0,926	3,091	0,884	4,247
Elantxobe	0,086	0,914	2,695	0,673	2,628	0,726	3,078	0,872	4,053
Ea	0,088	0,912	2,708	0,625	2,638	0,787	3,365	0,852	4,498

Elaboración con base en el presente trabajo de investigación.

Figura 33. Matriz normalizada empleando la raíz cuadrada

	Dominance_D	Simpson_1-D	Shannon_H	Evenness_e^H/S	Brillouin	Menhinick	Margalef	Equitability_J	Fisher_alpha
Bermeo	0,293	0,956	1,636	0,852	1,606	0,929	1,738	0,945	2,023
Mundaka	0,306	0,952	1,629	0,804	1,603	0,921	1,796	0,927	2,088
Busturia	0,298	0,955	1,635	0,831	1,603	0,960	1,789	0,937	2,096
Murueta	0,297	0,955	1,630	0,824	1,610	0,845	1,719	0,934	1,970
Forua	0,293	0,956	1,657	0,805	1,623	1,014	1,911	0,929	2,266
Errigoiti	0,296	0,955	1,641	0,820	1,618	0,884	1,774	0,933	2,048
Morga	0,299	0,954	1,642	0,820	1,611	0,960	1,819	0,934	2,132
Muxika	0,301	0,953	1,636	0,831	1,605	0,946	1,780	0,937	2,080
Mendata	0,303	0,953	1,632	0,774	1,610	0,908	1,847	0,916	2,144
Gernika	0,302	0,953	1,638	0,835	1,609	0,931	1,771	0,939	2,063
Ajangiz	0,298	0,955	1,636	0,832	1,603	0,970	1,794	0,938	2,107
Arratzu	0,292	0,956	1,652	0,816	1,630	0,871	1,796	0,933	2,069
Nabarniz	0,297	0,955	1,643	0,804	1,612	0,978	1,860	0,928	2,188
Kortezubi	0,304	0,953	1,634	0,810	1,606	0,936	1,805	0,929	2,105
Arteaga	0,297	0,955	1,650	0,780	1,624	0,951	1,901	0,920	2,226
Ibarrangelu	0,295	0,955	1,627	0,840	1,594	0,962	1,758	0,940	2,061
Elantxobe	0,294	0,956	1,642	0,820	1,621	0,852	1,754	0,934	2,013
Ea	0,297	0,955	1,646	0,791	1,624	0,887	1,834	0,923	2,121

Elaboración con base en el presente trabajo de investigación.

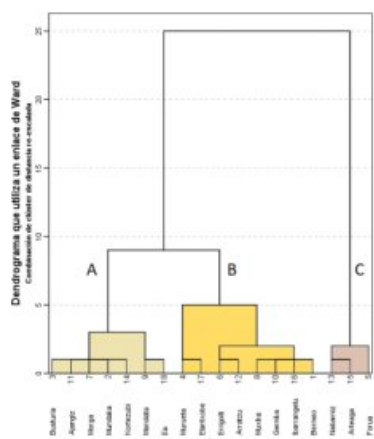
En la figura 34 se muestra una parte de la matriz de disimilaridad; la mayor similaridad la encontramos entre Muxika y Busturia (0,06%) y la menor entre Murueta y Forua (15,43%). En el dendrograma de la figura 35 se ve el resultado de la clasificación; en el conglomerado (A) aparecen siete localidades: Busturia, Ajangiz, Morga, Mundaka, Kortezubi, Mendata y Ea, en el conglomerado (B) ocho: Murueta, Elantxobe, Errigoiti, Arratzu, Muxika, Gernika, Ibarrangelu y Bermeo y en el conglomerado (C) tres: Nabarniz, Arteaga y Forua, siendo estas últimas las que más se alejan del resto (mapa de la figura 36).

Figura 34. Matriz de distancias

	Bermeo	Mundaka	Busturia	Murueta	Forua	Errigoiti	Morga	Muxika	Mendata	Gernika	Ajangiz
Bermeo		0,0105	0,0094	0,0112	0,0994	0,0053	0,0206	0,0059	0,0340	0,0031	0,0124
Mundaka	0,0105		0,0026	0,0262	0,0549	0,0042	0,0046	0,0019	0,0070	0,0026	0,0038
Busturia	0,0094	0,0026		0,0341	0,0484	0,0087	0,0024	0,0006	0,0121	0,0023	0,0002
Murueta	0,0112	0,0262	0,0341		0,1543	0,0108	0,0496	0,0262	0,0535	0,0190	0,0402
Forua	0,0994	0,0549	0,0484	0,1543		0,0837	0,0300	0,0580	0,0322	0,0693	0,0426
Errigoiti	0,0053	0,0042	0,0087	0,0108	0,0837		0,0149	0,0053	0,0177	0,0028	0,0117
Morga	0,0206	0,0046	0,0024	0,0496	0,0300	0,0149		0,0046	0,0062	0,0082	0,0016
Muxika	0,0059	0,0019	0,0006	0,0262	0,0580	0,0053	0,0046		0,0138	0,0006	0,0015
Mendata	0,0340	0,0070	0,0121	0,0535	0,0322	0,0177	0,0062	0,0138		0,0172	0,0120
Gernika	0,0031	0,0026	0,0023	0,0190	0,0693	0,0028	0,0082	0,0006	0,0172		0,0041
Ajangiz	0,0124	0,0038	0,0002	0,0402	0,0426	0,0117	0,0016	0,0015	0,0120	0,0041	
Arratzu	0,0111	0,0045	0,0100	0,0174	0,0727	0,0014	0,0130	0,0072	0,0126	0,0054	0,0126
Nabarniz	0,0472	0,0177	0,0148	0,0857	0,0103	0,0362	0,0054	0,0200	0,0082	0,0269	0,0120
Kortezubi	0,0134	0,0007	0,0015	0,0342	0,0442	0,0073	0,0017	0,0019	0,0056	0,0037	0,0019
Arteaga	0,0746	0,0326	0,0331	0,1126	0,0064	0,0542	0,0177	0,0394	0,0121	0,0477	0,0296
Ibarrangelu	0,0033	0,0055	0,0024	0,0241	0,0712	0,0077	0,0097	0,0014	0,0230	0,0016	0,0037
Elantxobe	0,0077	0,0131	0,0203	0,0034	0,1154	0,0027	0,0302	0,0145	0,0317	0,0095	0,0249
Ea	0,0252	0,0047	0,0103	0,0395	0,0435	0,0099	0,0068	0,0104	0,0019	0,0118	0,0111

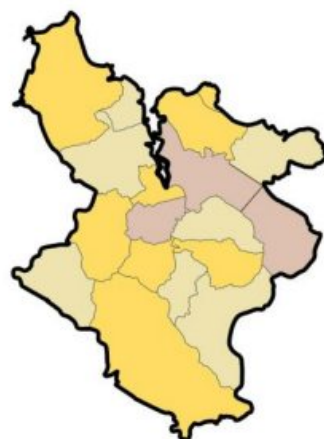
Elaboración con base en el presente trabajo de investigación.

Figura 35. Dendrograma de clasificación jerárquica



Elaboración con base en el presente trabajo de investigación.

Figura 36. Los resultados del dendrograma proyectados en el mapa



Para realizar la clasificación de los informantes hemos incluido en la matriz de datos los índices de diversidad correspondientes a todos ellos como se muestra en la matriz de la figura 37.

Figura 37. Matriz con los índices de diversidad correspondientes a todos los informantes

	Dominance	Simpson_1-l	Shannon_H	Evenness_e'	Brillouin	Menhinick	Margalef	Equitability	Fisher_alpha
BermeoM	0,083	0,917	2,676	0,807	2,509	1,118	3,059	0,926	4,398
BermeoH	0,087	0,913	2,681	0,730	2,515	1,197	3,374	0,895	4,935
MundakaM	0,094	0,906	2,646	0,705	2,533	0,943	3,110	0,883	4,289
MundakaH	0,093	0,908	2,648	0,673	2,446	1,403	3,696	0,870	5,675
BusturiaM	0,089	0,911	2,647	0,706	2,442	1,374	3,547	0,884	5,417
BusturiaH	0,087	0,913	2,690	0,737	2,534	1,141	3,318	0,898	4,789
MuruetaM	0,087	0,913	2,652	0,709	2,562	0,827	2,982	0,885	4,008
MuruetaH	0,091	0,909	2,659	0,751	2,498	1,131	3,190	0,903	4,598
ForuaM	0,087	0,913	2,714	0,718	2,483	1,508	3,797	0,891	5,984
ForuaH	0,085	0,915	2,735	0,734	2,589	1,122	3,414	0,898	4,905
ErrigoitiM	0,090	0,910	2,650	0,707	2,509	1,081	3,256	0,884	4,635
ErrigoitiH	0,085	0,915	2,722	0,691	2,599	1,037	3,437	0,881	4,843
MorgaM	0,095	0,905	2,633	0,732	2,457	1,204	3,262	0,894	4,785
MorgaH	0,087	0,913	2,694	0,705	2,539	1,172	3,465	0,885	5,035
MuxikaM	0,103	0,897	2,571	0,726	2,410	1,110	3,051	0,889	4,377
MuxikaH	0,082	0,919	2,728	0,765	2,562	1,181	3,357	0,911	4,891
MendataM	0,094	0,906	2,644	0,640	2,512	1,108	3,514	0,856	5,030
MendataH	0,089	0,911	2,684	0,666	2,565	1,035	3,435	0,869	4,837
GernikaM	0,091	0,909	2,674	0,725	2,505	1,202	3,378	0,893	4,947
GernikaH	0,090	0,910	2,688	0,735	2,533	1,136	3,312	0,897	4,775
AjangizM	0,089	0,911	2,635	0,734	2,458	1,211	3,270	0,895	4,804
AjangizH	0,088	0,912	2,695	0,740	2,512	1,257	3,434	0,900	5,096
ArratzuM	0,082	0,918	2,760	0,687	2,636	1,074	3,589	0,880	5,098
ArratzuH	0,088	0,912	2,700	0,708	2,584	0,980	3,263	0,887	4,539
NabarnizM	0,089	0,911	2,676	0,692	2,553	1,020	3,306	0,879	4,640
NabarnizH	0,085	0,915	2,706	0,713	2,433	1,687	3,966	0,889	6,552
KortezubiM	0,088	0,912	2,693	0,672	2,560	1,101	3,506	0,871	5,012
KortezubiH	0,099	0,901	2,628	0,693	2,434	1,316	3,491	0,877	5,255
ArteagaM	0,091	0,909	2,702	0,677	2,555	1,160	3,568	0,874	5,167
ArteagaH	0,086	0,914	2,714	0,686	2,581	1,095	3,499	0,878	4,994
IbarrangeluM	0,088	0,912	2,653	0,710	2,482	1,229	3,405	0,886	5,018
IbarrangeluH	0,085	0,915	2,641	0,780	2,442	1,266	3,203	0,914	4,777
ElantxobeM	0,083	0,917	2,716	0,720	2,599	0,988	3,271	0,892	4,559
ElantxobeH	0,088	0,912	2,674	0,725	2,564	0,926	3,091	0,893	4,247
EaM	0,100	0,900	2,592	0,703	2,456	1,035	3,093	0,880	4,357
EaH	0,082	0,918	2,756	0,656	2,653	0,986	3,603	0,867	5,023

Elaboración con base en el presente trabajo de investigación.

Una vez normalizada la matriz por medio de la raíz cuadrada (figura 38), se procede a la clasificación de los informantes.

Figura 38. Matriz normalizada empleando la raíz cuadrada

	Dominance	Simpson_1-d	Shannon_H	Evenness_e'	Brillouin	Menhinick	Margalef	Equitability	Fisher_alpha
BermeoM	0,288	0,957	1,636	0,898	1,584	1,057	1,749	0,962	2,097
BermeoH	0,294	0,956	1,637	0,854	1,586	1,094	1,837	0,946	2,221
MundakaM	0,306	0,952	1,627	0,840	1,592	0,971	1,764	0,940	2,071
MundakaH	0,304	0,953	1,627	0,820	1,564	1,184	1,922	0,933	2,382
BusturiaM	0,298	0,954	1,627	0,840	1,563	1,172	1,883	0,940	2,327
BusturiaH	0,296	0,955	1,640	0,858	1,592	1,068	1,822	0,948	2,188
MuruetaM	0,295	0,956	1,628	0,842	1,601	0,909	1,727	0,941	2,002
MuruetaH	0,301	0,954	1,631	0,867	1,581	1,063	1,786	0,950	2,144
ForuaM	0,295	0,956	1,647	0,848	1,576	1,228	1,949	0,944	2,446
ForuaH	0,291	0,957	1,654	0,857	1,609	1,059	1,848	0,948	2,215
ErrigoitiM	0,301	0,954	1,628	0,841	1,584	1,040	1,804	0,940	2,153
ErrigoitiH	0,291	0,957	1,650	0,831	1,612	1,018	1,854	0,938	2,201
MorgaM	0,308	0,951	1,623	0,856	1,567	1,097	1,806	0,946	2,187
MorgaH	0,294	0,956	1,641	0,839	1,593	1,083	1,861	0,941	2,244
MuxikaM	0,321	0,947	1,603	0,852	1,552	1,054	1,747	0,943	2,092
MuxikaH	0,285	0,958	1,652	0,875	1,601	1,087	1,832	0,954	2,212
MendataM	0,306	0,952	1,626	0,800	1,585	1,053	1,875	0,925	2,243
MendataH	0,298	0,955	1,638	0,816	1,602	1,017	1,853	0,932	2,199
GernikaM	0,302	0,953	1,635	0,851	1,583	1,096	1,838	0,945	2,224
GernikaH	0,299	0,954	1,640	0,857	1,592	1,066	1,820	0,947	2,185
AjangizM	0,298	0,954	1,623	0,857	1,568	1,100	1,808	0,946	2,192
AjangizH	0,297	0,955	1,642	0,860	1,585	1,121	1,853	0,949	2,257
ArratzuM	0,286	0,958	1,661	0,829	1,624	1,036	1,894	0,938	2,258
ArratzuH	0,296	0,955	1,643	0,842	1,607	0,990	1,806	0,942	2,130
NabarnizM	0,299	0,954	1,636	0,832	1,598	1,010	1,818	0,937	2,154
NabarnizH	0,291	0,957	1,645	0,844	1,560	1,299	1,991	0,943	2,560
KortezubiM	0,297	0,955	1,641	0,820	1,600	1,049	1,872	0,933	2,239
KortezubiH	0,315	0,949	1,621	0,832	1,560	1,147	1,868	0,937	2,292
ArteagaM	0,302	0,953	1,644	0,823	1,598	1,077	1,889	0,935	2,273
ArteagaH	0,293	0,956	1,647	0,828	1,607	1,046	1,871	0,937	2,235
IbarrangeluM	0,297	0,955	1,629	0,843	1,575	1,109	1,845	0,941	2,240
IbarrangeluH	0,291	0,957	1,625	0,883	1,563	1,125	1,790	0,956	2,186
ElantxobeM	0,289	0,957	1,648	0,848	1,612	0,994	1,809	0,944	2,135
ElantxobeH	0,297	0,955	1,635	0,852	1,601	0,962	1,758	0,945	2,061
EaM	0,316	0,949	1,610	0,839	1,567	1,017	1,759	0,938	2,087
EaH	0,286	0,958	1,660	0,810	1,629	0,993	1,898	0,931	2,241

Elaboración con base en el presente trabajo de investigación.

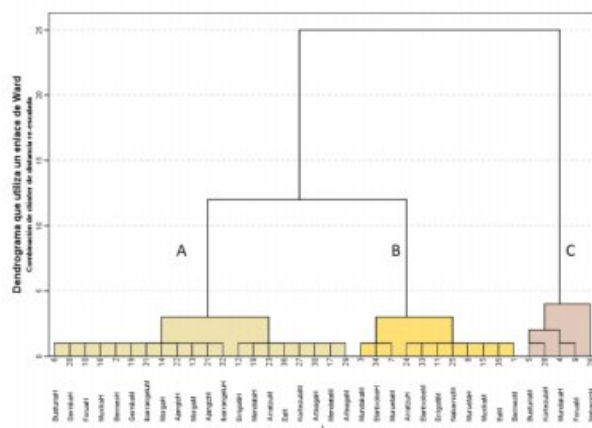
En la figura 39 se muestra una parte de la matriz de disimilaridad; la mayor similaridad la encontramos entre GernikaM y BermeoH (0,01%) y la menor entre NabarnizH y MuruetaM (53,52%). En el dendrograma de la figura 40 se ve el resultado de la clasificación; en el conglomerado (A) aparecen veinte informantes: BermeoH, BusturiaH, ForuaH, ErrigoitiH, MorgaM, MorgaH, MuxikaH, MendataM, MendataH, GernikaM, GernikaH, AjangizM, AjangizH, ArratzuM, KortezubiM, ArteagaM, ArteagaH, IbarrangeluM, IbarrangeluH y EaH.; en el conglomerado (B) once: BermeoM, MundakaM, MuruetaM, MuruetaH, ErrigoitiM, MuxikaM, ArratzuH, NabarnizM, ElantxobeM, ElantxobeH y EaM y en el conglomerado (C) cinco: MundakaH, BusturiaM, ForuaM, NabarnizH y KortezubiH, siendo estas últimas las que más se alejan del resto (mapa de la figura13). Las localidades de Bermeo, Mundaka, Busturia, Errigoiti, Forua, Mendata Gernika Arratzu, Ibarrangelu y Elantxobe tienen ambos informantes dentro del mismo conglomerado (mapa de la figura 41). Por otra parte, la pertenencia a uno u otro conglomerado según el género de los informantes no es estadísticamente significativa.

Figura 39. Matriz de distancias

	Bernesohl	Bernesohl	Mundakuhl	Mundakuhl	Buntarlakhl	Buntarlakhl	Munersakhl	Munersakhl	Forsakhl	Forsakhl	Erigakhl	Erigakhl	Mogakhl	Mogakhl	Musakhl	Musakhl	Mundakuhl	Mundakuhl	Gernakhl
Bernesohl	0.0000	0.0387	0.0128	0.1350	0.0884	0.0137	0.0154	0.0085	0.1943	0.0284	0.0104	0.0294	0.0158	0.0385	0.0058	0.0222	0.0484	0.0309	0.0283
Bernesohl	0.0387	0.0000	0.0435	0.0433	0.0203	0.0023	0.0948	0.0098	0.0814	0.0022	0.0090	0.0079	0.0026	0.0015	0.0295	0.0012	0.0072	0.0084	0.0001
Mundakuhl	0.0128	0.0435	0.0000	0.1483	0.3225	0.0275	0.0302	0.0153	0.3418	0.0372	0.0132	0.0285	0.0326	0.0523	0.0102	0.0406	0.0505	0.0274	0.0448
Mundakuhl	0.1350	0.0433	0.1483	0.0000	0.0052	0.0638	0.2802	0.0528	0.0083	0.0535	0.0080	0.0083	0.0066	0.0345	0.1138	0.0523	0.0396	0.0678	0.0412
Buntarlakhl	0.0884	0.0203	0.1270	0.0052	0.0000	0.0153	0.2004	0.0540	0.0223	0.0298	0.0544	0.0435	0.0318	0.0184	0.0891	0.0267	0.0217	0.0436	0.0196
Buntarlakhl	0.0137	0.0023	0.0275	0.0638	0.0153	0.0000	0.0948	0.0098	0.1427	0.0022	0.0090	0.0079	0.0026	0.0015	0.0295	0.0012	0.0072	0.0084	0.0001
Munersakhl	0.0154	0.0948	0.0302	0.2802	0.2004	0.0948	0.0000	0.0485	0.3482	0.0035	0.0482	0.0084	0.0774	0.1078	0.0334	0.0087	0.1031	0.0672	0.0973
Munersakhl	0.0085	0.0098	0.0153	0.0528	0.0540	0.0098	0.0485	0.0000	0.1457	0.0104	0.0017	0.0128	0.0039	0.0172	0.0067	0.0085	0.0236	0.0131	0.0105
Forsakhl	0.1943	0.0814	0.2418	0.0083	0.0223	0.1087	0.3452	0.1457	0.0000	0.0914	0.1427	0.1148	0.1054	0.1997	0.0899	0.0896	0.1167	0.0791	0.0791
Forsakhl	0.0284	0.0022	0.0372	0.0535	0.0298	0.0022	0.0635	0.0104	0.0914	0.0000	0.0079	0.0027	0.0070	0.0024	0.0323	0.0015	0.0099	0.0043	0.0028
Erigakhl	0.0104	0.0090	0.0132	0.0080	0.0544	0.0022	0.0462	0.0017	0.1427	0.0079	0.0000	0.0068	0.0050	0.0117	0.0094	0.0096	0.0153	0.0065	0.0095
Erigakhl	0.0294	0.0079	0.0285	0.0481	0.0435	0.0053	0.0684	0.0128	0.1148	0.0027	0.0068	0.0000	0.0135	0.0084	0.0319	0.0077	0.0061	0.0094	0.0085
Mogakhl	0.0158	0.0385	0.0523	0.0345	0.0384	0.0022	0.0774	0.0039	0.1054	0.0035	0.0054	0.0128	0.0000	0.0085	0.0332	0.0044	0.0157	0.0121	0.0029
Mogakhl	0.0058	0.0295	0.0522	0.0345	0.0384	0.0022	0.0774	0.0039	0.1054	0.0035	0.0054	0.0128	0.0000	0.0085	0.0332	0.0044	0.0157	0.0121	0.0029
Musakhl	0.0222	0.0484	0.0523	0.1138	0.0891	0.0188	0.0334	0.0067	0.1997	0.0123	0.0094	0.0319	0.0152	0.0410	0.0296	0.0441	0.0296	0.0131	0.0299
Musakhl	0.0309	0.0283	0.0412	0.0523	0.0274	0.0017	0.0885	0.0089	0.0899	0.0015	0.0090	0.0077	0.0044	0.0036	0.0296	0.0118	0.0099	0.0018	0.0018
Mundakuhl	0.0484	0.0072	0.0505	0.0396	0.0217	0.0303	0.0311	0.0236	0.0894	0.0089	0.0153	0.0084	0.0117	0.0033	0.0441	0.0118	0.0099	0.0018	0.0018
Mundakuhl	0.0309	0.0084	0.0274	0.0678	0.0438	0.0058	0.0672	0.0131	0.1147	0.0043	0.0062	0.0094	0.0131	0.0077	0.0296	0.0099	0.0018	0.0018	0.0018
Gernakhl	0.0283	0.0001	0.0448	0.0413	0.0196	0.0025	0.0973	0.0305	0.0793	0.0038	0.0095	0.0085	0.0029	0.0013	0.0299	0.0018	0.0007	0.0008	0.0008
Gernakhl	0.0105	0.0028	0.0236	0.0677	0.0387	0.0000	0.0673	0.0032	0.1134	0.0023	0.0025	0.0090	0.0022	0.0008	0.0177	0.0021	0.0106	0.0017	0.0029
Aljargakhl	0.0168	0.0023	0.0342	0.0578	0.0294	0.0022	0.0885	0.0045	0.1015	0.0065	0.0017	0.0124	0.0061	0.0071	0.0172	0.0036	0.0134	0.0122	0.0024
Aljargakhl	0.0422	0.0024	0.0639	0.0289	0.0099	0.0088	0.1267	0.0208	0.0567	0.0064	0.0204	0.0135	0.0085	0.0013	0.0463	0.0044	0.0096	0.0167	0.0021
Aratarkhl	0.0150	0.0188	0.0588	0.0413	0.0285	0.0134	0.1134	0.0388	0.0782	0.0057	0.0222	0.0095	0.0225	0.0095	0.0115	0.0051	0.0099	0.0111	0.0111
Aratarkhl	0.0130	0.0207	0.0682	0.1173	0.0885	0.0102	0.0295	0.0075	0.1780	0.0142	0.0088	0.0084	0.0173	0.0246	0.0183	0.0244	0.0080	0.0219	0.0219
Naharnakhl	0.0155	0.0128	0.0138	0.0945	0.0618	0.0064	0.0418	0.0054	0.1589	0.0081	0.0015	0.0040	0.0107	0.0134	0.0100	0.0121	0.0140	0.0088	0.0134
Naharnakhl	0.0154	0.0125	0.0199	0.0589	0.0825	0.0218	0.0552	0.0221	0.0201	0.1598	0.0087	0.0295	0.0152	0.0146	0.0415	0.0142	0.0170	0.0117	0.0173
Naharnakhl	0.0426	0.0051	0.0468	0.0427	0.0250	0.0072	0.0975	0.0196	0.0823	0.0013	0.0130	0.0033	0.0134	0.0018	0.0438	0.0075	0.0010	0.0030	0.0050

Elaboración con base en el presente trabajo de investigación.

Figura 40. Dendrograma de clasificación jerárquica



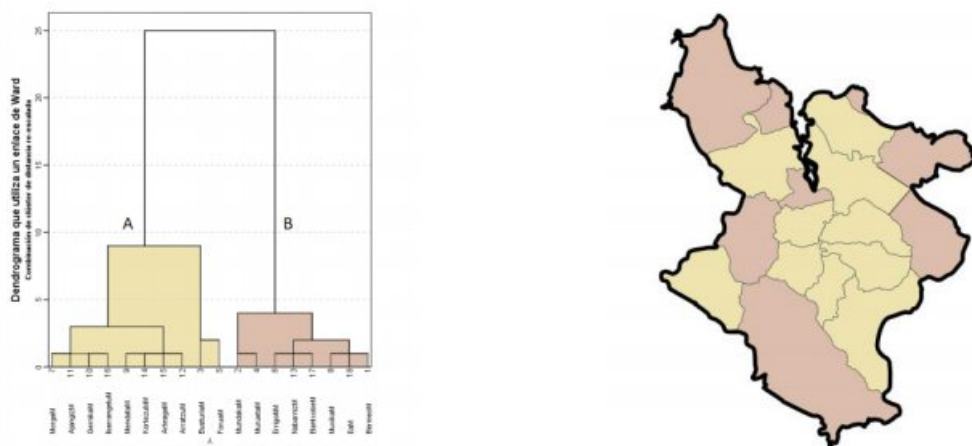
Elaboración con base en el presente trabajo de investigación.

Figura 41. Los resultados del dendrograma proyectados en el mapa



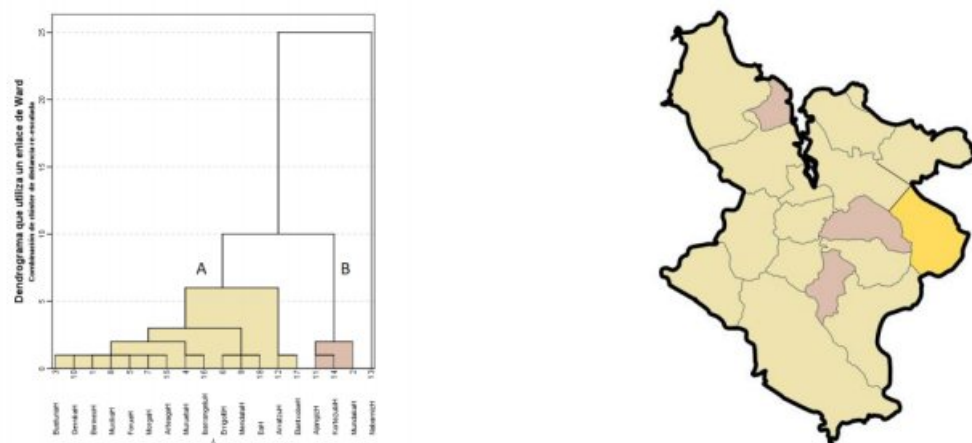
En los mapas de las figuras 42 y 43 se pueden constatar las diferencias existentes según el género de los informantes. Como bien se observa los hombres presentan mayor dispersión que las mujeres.

Figura 42. Dendrograma de clasificación jerárquica de las mujeres y resultados proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.

Figura 43. Dendrograma de clasificación jerárquica de los hombres y resultados proyectados en el mapa



Elaboración con base en el presente trabajo de investigación.

RESULTADOS Y DISCUSIÓN

En este apartado final hacemos un resumen de los resultados obtenidos en el apartado anterior y algunas sugerencias para futuros trabajos.

Como hemos podido observar, en la clasificación de las localidades elegidas inciden tanto el método de análisis como el género de los informantes. En los mapas de las figuras 44, 45 y 46 se muestran las diferentes clasificaciones de las localidades sin hacer distinción de género, sólo teniendo en cuenta el método de análisis (dicotómico “Di”, probabilidades “Pro” o índices de diversidad “Índ”). En la tabla 5 se muestra la situación de cada localidad de acuerdo con cada método de análisis. Bermeo, Ibarrangelu y Elantxobe coinciden en todos los conglomerados (a, a, b), también Mundaka, Morga y Mendata (b, b, a) y Murueta y Gernika (c, c, c).

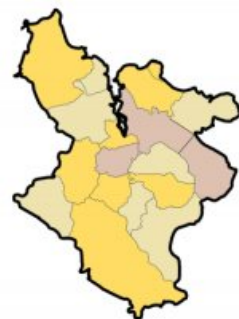
Figura 44. Mapa de la proyección del dendrograma creado a partir de matriz dicotómica



Figura 45. Mapa de la proyección del dendrograma creado a partir de matriz basada en los análisis de probabilidades



Figura 46. Mapa de la proyección del dendrograma creado a partir de los índices de diversidad



Elaboración con base en el presente trabajo de investigación.

En la tabla 5 se muestran las localidades según el conglomerado de pertenencia.

Tabla 5. Localidades según conglomerado de pertenencia

localidad	Dic	Pro	Índ	localidad	Dic	Pro	Índ
Bermeo	a	a	b	Gernika	c	b	b
Mundaka	b	b	a	Ajangiz	a	c	a
Busturia	a	a	a	Arratzu	c	c	b
Murueta	c	b	b	Nabarniz	b	c	c
Forua	c	c	c	Kortezubi	c	c	a
Errigoiti	a	b	b	Arteaga	c	c	c
Morga	b	b	a	Ibarrangelu	a	a	b
Muxika	a	c	b	Elantxobe	a	a	b
Mendata	b	b	a	Ea	b	c	a

Elaboración con base en el presente trabajo de investigación.

Cuando incluimos en el análisis la variable de género vemos que, según el método de análisis elegido, unas localidades aglutinan a ambos informantes dentro del mismo conglomerado y otras no; en la tabla 6 se muestran los datos (1 es cuando ambos informantes están en el mismo conglomerado y 0 cuando no). En Bermeo, Busturia, Gernika Arratzu, Ibarrangelu y Elantxobe los dos informantes siempre se encuentran dentro del mismo conglomerado, es decir, la diferencia de género es menor: en Mundaka, Forua, Errigoiti y Mendata los dos informantes aparecen dentro del mismo conglomerado en dos ocasiones, en Muxika, Kortezubi y Ea sólo aparecen en el mismo conglomerado una vez y en Murueta, Morga, Ajangiz, Nabarniz y Arteaga nunca aparecen ambos informantes en el mismo conglomerado, siendo la diferencia de género máxima.

Tabla 6. Localidades según género y conglomerado de pertenencia

localidad	Dic	Pro	Índ	localidad	Dic	Pro	Índ
Bermeo	1	1	1	Gernika	1	1	1
Mundaka	0	1	1	Ajangiz	0	0	0
Busturia	1	1	1	Arratzu	1	1	1
Murueta	0	0	0	Nabarniz	0	0	0
Forua	0	1	1	Kortezubi	1	0	0
Errigoiti	0	1	1	Arteaga	0	0	0
Morga	0	0	0	Ibarrangelu	1	1	1
Muxika	1	0	0	Elantxobe	1	1	1
Mendata	0	1	1	Ea	0	1	0

Elaboración con base en el presente trabajo de investigación.

Del mismo modo, si analizamos los datos según el género de los informantes también se constatan diferencias notables en la clasificación. En los mapas de las figuras 47, 48 y 49 se muestra la clasificación de las mujeres en razón del método elegido y en la tabla 7 un resumen de los datos de cada localidad.

Figura 47. Clasificación de las mujeres basada en las matrices dicotómicas



Figura 48. Clasificación de las mujeres basada en el análisis de probabilidades



Figura 49. Clasificación de las mujeres basada en los índices de diversidad



Elaboración con base en el presente trabajo de investigación.

Tabla 7. Resumen de los datos de cada localidad según las mujeres

localidad	Dic	Pro	Índ	localidad	Dic	Pro	Índ
Bermeo	b	b	b	Gernika	a	a	a
Mundaka	a	a	b	Ajangiz	a	a	a
Busturia	b	b	a	Arratzu	b	a	a
Murueta	a	a	b	Nabarniz	a	a	b
Forua	a	a	a	Kortezubi	b	a	a
Errigoiti	a	a	b	Arteaga	a	a	a
Morga	b	b	a	Ibarrangelu	b	b	a
Muxika	a	a	b	Elantxobe	b	b	b
Mendata	b	a	a	Ea	b	b	b

Con los datos de la tabla 12 vemos que las mujeres se clasifican en cinco grupos según pertenezcan a unos conglomerados o a otros:

aaa: Forua, Gernika, Ajangiz y Arteaga

aab: Mundaka, Murueta, Errigoiti, Muxika y Nabarniz

bbb: Bermeo, Elantxobe y Ea

bba: Busturia, Morga e Ibarrangelu

baa: Mendata, Arratzu y Kortezubi

En los mapas de las figuras 50, 51 y 52 se muestra la clasificación de los hombres según el método de análisis; como puede observarse los hombres presentan una situación más heterogénea que las mujeres. En la tabla 8 un resumen de los datos de cada localidad.

Figura 50. Clasificación de los hombres en las matrices dicotómicas



Figura 51. Clasificación de los hombres basada en el análisis de probabilidades



Figura 52. Clasificación de los hombres basada en los índices de diversidad



Elaboración con base en el presente trabajo de investigación.

Tabla 8. Resumen de los datos de cada localidad según los hombres

localidad	Dic	Pro	Índ	localidad	Dic	Pro	Índ
Bermeo	a	a	a	Gernika	b	b	a
Mundaka	a	a	b	Ajangiz	a	a	b
Busturia	a	a	a	Arratzu	a	b	a
Murueta	a	a	a	Nabarniz	a	a	c
Forua	b	b	a	Kortezubi	a	a	b
Errigoiti	a	b	a	Arteaga	b	a	a
Morga	b	b	a	Ibarrangelu	a	a	a
Muxika	a	a	a	Elantxobe	a	a	a
Mendata	b	b	a	Ea	a	b	a

Elaboración con base en el presente trabajo de investigación.

Observando los datos de la tabla 12 constatamos que los hombres se clasifican en más grupos que las mujeres; así según los conglomerados de pertenencia, los hombres presentan una mayor dispersión:

aaa: Bermeo, Busturia, Murueta, Muxika, Ibarrangelu y Elantxobe
aab: Mundaka, Ajangiz y Kortezubi
aba: Errigoiti, Arratzu y Ea
bba: Forua, Morga, Mendata y Gernika
baa: Arteaga
aac: Nabarniz

Creemos que llegados a este punto podemos concluir que el método elegido para el análisis de unos mismos datos influye en los resultados de la clasificación final. Por otra parte, cada método aporta una visión diferente de la realidad y más que excluyentes pensamos que pueden ser complementarios. Por último, señalar que, en el caso que nos ocupa, la variable de género es determinante en la clasificación de muchas localidades. El estudio de otras variables sociolingüísticas tales como la edad, el nivel de estudios, el conocimiento de la lengua estándar, u otras puede ser de gran importancia en el análisis de la distribución geográfica.

REFERENCIAS

- Aurrekoetxea, G., & Ormaetxea, J.L. (2006). Euskararen atlas sozio-geolinguistikoa ikerketa-proiektua. Bilbo: Mendebalde Kultura Elkartea
- Aurrekoetxea, G. (1992). Nafarroako euskara: azterketa dialektometrikoa. Uztaro 5, 59-109.
- Aurrekoetxea, G. (1995). Bizkaieraren egituraketa geolinguistikoa. [Doktoretza-tesia]. Leioa: UPV-EHUko Argitarapen Zerbitzua.
- Aurrekoetxea, G. (2003). Euskalkiak estandarren uholdepean (Arratiako kasua). Ahozketasuna aztergai, II. Jardunaldiak. Bilbo: Mendebalde Kultura Alkartea. 167-177.orr.
- Aurrekoetxea, G. (2004). Estandar eta dialektoen arteko bateratze-joerak (ikusuntu teorikotik begirada bat. Uztaro, 50, 45-57.
- Aurrekoetxea, G. (2005). Nafarroako euskararen sailkapenaz. In P. Etxeberria & H. Knörr (arg.), Nerekin yaio nun. Txillardegi omenaldia, Iker 17, EHU & Euskaltzaindia, Bilbo, 109-24.
- Aurrekoetxea, G. (2008). Bariazio soziolinguistikoa Dimako euskararen. Euskalingua, 12 17-26.
- Aurrekoetxea, G. (2009). Iparraldeko hizkeren sailkapena (1): lexikoa. ASJU XXVIII-1, 287-331.
- Aurrekoetxea, G. (2016). Distantzia geografikoaren eta hizkuntza distantziaren arteko korrelazioa. In Aurrekoetxea, G., Makazaga, J. M. eta Salaberri, P. (arg.). Hire Bordatxoan: Txipi Ormaetxea Omenduz. Bilbo: UPV/EHU.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen; Freeman, D., & Schoenemann, T. (2011). La lengua es un sistema adaptativo complejo. Lingüística en la Red. http://www.linred.es/articulos_pdf/LR_articulo_04092011.pdf
- Boersma, P., & Weenik, D. (2023). Praat: doing phonetics by computer. [Computer program].
- Eguskiza, N. (2019). hizkuntza aldakortasun geo-soziolinguistikoa arratian. [Doktore-tesia]. UPV7EHU.
- Ellis, N. C. (2011). The emergence of Language as a Complex Adaptative System. In J. Simpson (ed.), Routledge Handbook of Applied Linguistics. London: Routledge/Taylor Francis.
- Ensunza, A. (2015): Busturialdeko euskararen hizkuntza-aldakortasuna denboran eta espazioan. [Doktore-tesia]. UPV/EHU.
- Etxebarria, A., & Eguskiza, N. (2018). Bariazioa esaldien intonazioan. Bilbo: Euskal Herriko Unibertsitateko Argitalpen Zerbitzua.
- Gaminde, I., Olalde, A., Etxebarria, A., Eguskiza, E., & Gaminde, U. (2017). Hizkuntza aldakortasuna Larrabetzun. Larrabetzuko udala.
- Gaminde, I., Etxebarria, A., Eguskiza, N., Romero, A. & Unamuno, L. (2016b). Lexikoaren bariazioa eta multzokatze-azterketa. Euskalingua 28, 19-3
- Gaminde, I., Romero, A., Etxebarria, A., & Eguskiza, N. (2016a). Bizkaiko aditz laguntzaileen bilakaeraren azterketaz. In A. Iglesias, A. Romero, & A. Ensunza, (ed.), Linguistic variation in the Basque language and education – II / Euskararen bariazioa eta bariazioaren irakaskuntza – II. Bilbo: UPV/EHU.

- Gaminde, I. (2007). Bizkaian zehar: Euskararen Ikuspegi Orokorra, Mendebalde Kultura Alkartea eta Bizkaiko Foru Aldundia, Bilbo
- Gell-Mann, M. (1994). El quark y el jaguar. Aventuras en lo simple y lo complejo. Tusquets.
- Hammer, Ø., Harper, D., & Ryan, P. (2001). PAST: paquete de programas de estadística paleontológica para enseñanza y análisis de datos. [Programa estadístico].
- Iglesias, A. (2014). Igorreko hizkeraren azterketa dialektologikoa. [Doktore-tesia]. UPV/EH.
- Labov, W. (1972). Sociolinguistic Patterns, Philadelphia: University of Pennsylvania Press. [Berrarg. [1983], Modelos sociolingüísticos. Cátedra.
- Labov, W. (1974). Language change as a form of communication, In A. Silverstein (ed.), Human Communication.
- Leónard, J.L., Heinsalu, E., Patriarca, M., & Darlu, P. (2015). Modeling regional variation from eas: complexity and communal aggregates. In G. aurekoetxea, A. Romero, A. Etxebarria (ed.), Linguistic Variation in the Basque and Education-I/Euskararen Bariazioa eta Bariazioaren Irakaskuntza-I, 50-58. Bilbo: UPV/EHU.
- López Rivera, J. J. (2013). Aplicación del marco de los sistemas complejos adaptativos a un modelo de variación lingüística. Moenia 19, 5-24.
- Lujanbio, O. (2016). Hizkuntza-aldakortasuna euskarán. Nafarroa ipar-mendebaldeko bi udalerritan egindako azterketa. [Doktoretza-tesia]. UPV/EHU.
- Moreno Cabrera, J. C. (2008). El nacionalismo lingüístico: Una ideología destructiva. Ediciones Península.
- Tagliamonte, S. (2012). Variationist sociolinguistics: change, observation, interpretation. Malden, MA: Wiley-Blackwell.

Artículo original

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

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

Ariane Ensunza

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

e-mail: ariane.ensunza@ehu.eus

Lorea Unamuno

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

e-mail: lorea.unamuno@ehu.eus

Gotzon Aurrekoetxea

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

e-mail: gotzon.aurrekoetxea@ehu.eus

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

RESUMEN

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

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

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

ABSTRACT

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

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

INTRODUCCIÓN

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

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

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

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

CLASIFICACIONES BASADAS EN DIFERENTES MARCOS TEÓRICOS

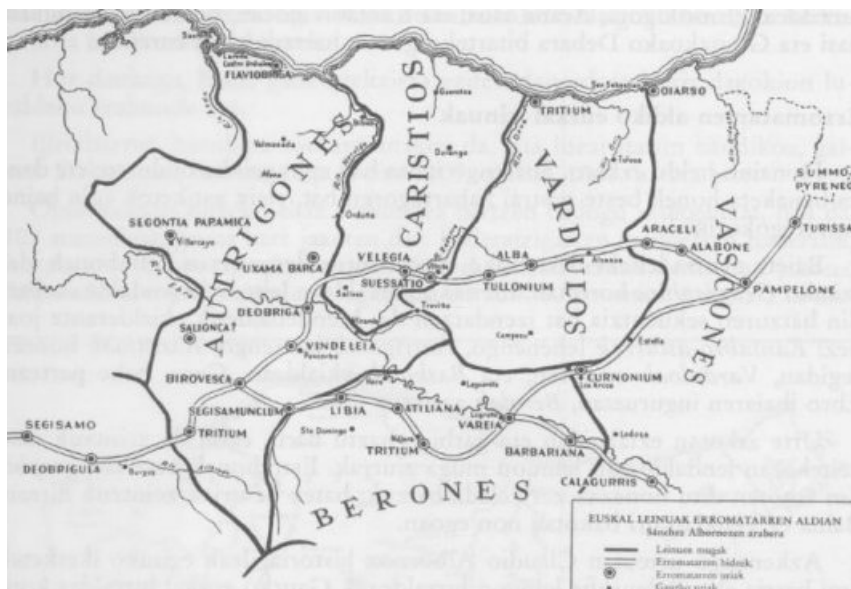
Siendo conscientes de que se trata de un tema muy controvertido, en el que no hay anclajes asumidos por una extensa gama de dialectólogos; a rasgos generales se podrían establecer diversos marcos teóricos en los que se basan las clasificaciones dialectales llevadas a cabo en la historia de la dialectología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

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

CLASIFICACIONES EXTRALINGÜÍSTICAS

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

Figura 1. Dialectos vascos atendiendo a criterios extralingüísticos



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

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

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

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

CLASIFICACIONES ISOGLÓTICAS

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

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

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

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

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

Figura 2. Clasificación de los dialectos vascos



Zuazo 2019.

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

CLASIFICACIONES PERCEPTUALES

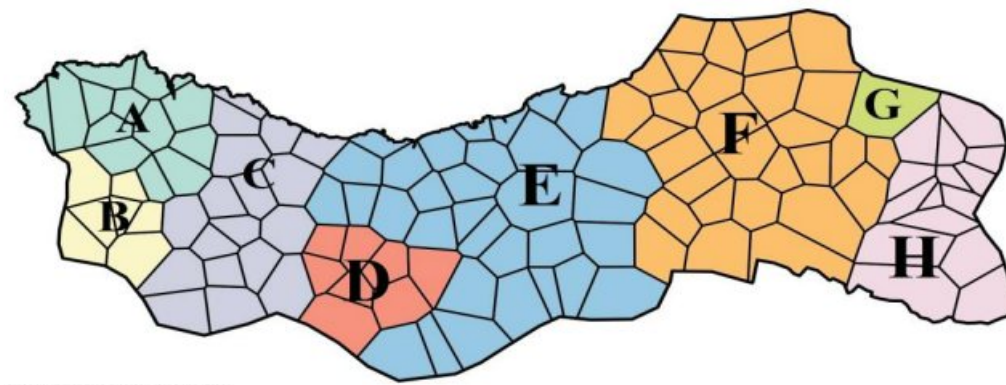
Las clasificaciones perceptuales son aquellas que se basan en las creencias, percepciones u opiniones de hablantes no-lingüistas; dicho de otro modo, clasificaciones llevadas a cabo mediante técnicas usadas en la dialectología perceptual. El máximo impulsor y referente de la dialectología perceptual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

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

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

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

Figura 3. Clasificación subjetiva de los dialectos vascos



Aurrekoetxea et al. (2017).

CLASIFICACIONES AGLOMERATIVAS

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

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

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

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

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

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

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

¿QUÉ DIFERENCIA HAY ENTRE ESTOS TIPOS DE CLASIFICACIÓN?

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

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

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

Figura 4. Mapa de los dialectos y subdialectos



Aurrekoetxea et al. 2019.

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

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

Elaboración con base al trabajo de investigación.

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

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

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

METODOLOGÍA EN LA CLASIFICACIÓN DIALECTAL

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

METODOLOGÍA DE LAS CLASIFICACIONES ISOGLÓTICAS

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

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

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

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

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

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

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

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

METODOLOGÍA DE LAS CLASIFICACIONES AGLOMERATIVAS

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

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

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

deben cumplir ciertos requisitos, como:

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

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

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

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

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

Como asevera Francis (1983:235)

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

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

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

En resumen, la dialectología cuantitativa necesita de:

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

CONCLUSIONES

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

REFERENCIAS

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

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

Artículo original

Notas sobre el Atlas Lingüístico Vasco Notes on the Basque Linguistic Atlas

Xarles Videgain

Universidad de Pau y Paises del Adour & Euskaltzaindia, Francia
<https://orcid.org/0000-0002-7053-2286>

e-mail: charles.videgain@univ-pau.fr

Recibido: 17/4/2023
Aprobado: 30/6/2023

RESUMEN

El presente trabajo aborda la importancia y el potencial del Atlas Lingüístico Vasco en el estudio de la variación dialectal del euskera. Se destaca que el atlas proporciona una gran cantidad de datos lingüísticos detallados que pueden ser utilizados para investigaciones dialectométricas y consultas específicas. A través de ejemplos tomados de diferentes mapas del atlas, se muestran y se señalan diversas divisiones y características dialectales en las distintas regiones. Se mencionan ejemplos de variaciones léxicas y fonéticas, préstamos lingüísticos y formas endémicas. En el trabajo se quiere subrayar que el atlas ofrece un amplio conjunto de datos que pueden ser explorados de manera más amplia y utilizados en investigaciones futuras. Se destaca la necesidad de una explotación organizada de los datos del atlas y se invita a los investigadores a aprovechar este recurso valioso en el estudio de la variación dialectal del euskera.

Palabras clave: Atlas Lingüístico Vasco; euskera; variación; dialectología.

ABSTRACT

This paper discusses the importance and potential of the Basque Linguistic Atlas in the study of Basque dialectal variation. It is emphasized that the atlas provides a large amount of detailed linguistic data that can be used for dialectometric research and specific queries. Through examples taken from different maps in the atlas, various dialectal divisions and characteristics in the different regions are shown and pointed out. Examples of lexical and phonetic variations, linguistic borrowings and endemic forms are mentioned. The paper wishes to emphasize that the atlas offers a large body of data that can be further explored and used in future research. It highlights the need for an organized exploitation of the atlas data and invites researchers to take advantage of this valuable resource in the study of Basque dialectal variation.

Keywords: Basque Linguistic Atlas; Basque; variation; dialectology.

DESARROLLO

El Atlas Vasco ha sido dirigido por los co-directores técnicos Gotzon Aurrekoetxea y Xarles Videgain, quienes han trabajado en conjunto con un grupo de investigadores. Este proyecto es conocido como EHHA, que originalmente se refería a *Euskal Herriko Hizkuntz-Atlasa*, es decir, 'Atlas Lingüístico del País Vasco'. Sin embargo, debido a presiones políticas, su denominación fue modificada a *Euskal Hizkeren Herrien Atlasa*, que significa 'Atlas de las hablas vernaculares vascas', eliminando así la referencia a un territorio específico.

El proyecto ha recibido el respaldo de Euskaltzaindia, la Academia de la lengua vasca, y ha contado con la colaboración intelectual del CNRS (Centre National de la Recherche Scientifique) y de expertos previamente involucrados en la elaboración de un atlas lingüístico.

Como es ampliamente reconocido, el objetivo principal de un atlas lingüístico es proporcionar información sobre la variación espacial y datos diacrónicos. En este sentido, el conjunto de datos recopilados para el EHHA (*Euskal Hizkeren Herrien Atlasa*) es de gran importancia tanto en términos de cantidad como, en nuestra opinión, de calidad. Se ha realizado un esfuerzo notable para garantizar la homogeneidad metodológica en la recopilación de estos datos, que incluyen tanto registros orales como transcripciones escritas.

Una parte significativa de estos datos se encuentra disponible en 11 volúmenes publicados, que están accesibles para su consulta en línea*. Cabe mencionar que el euskera ha sido la lengua de trabajo utilizada por los autores en todo el proyecto. Hasta el momento, no se ha elaborado una versión en otro idioma distinto al euskera.

Tuvimos la oportunidad de llevar a cabo el Atlas mediante el uso de un cuestionario exhaustivo, una red de puntos de encuesta meticulosamente seleccionados y numerosas horas de grabaciones. El cuestionario, desarrollado por Aurrekoetxea y Videgain en 1993, consta de un total de 2857 preguntas.

El léxico, como ocurre en muchos atlas lingüísticos, representa la parte más extensa del cuestionario, con un total de 2163 preguntas organizadas por campos semánticos, siguiendo una práctica habitual. Además, se otorgó una importancia significativa a la morfología nominal y verbal (595 preguntas), la sintaxis (62 preguntas) y la fonética (37 preguntas). En el caso de la morfología, se tuvo en cuenta la influencia de la declinación (o conjunto de posposiciones) en el euskera, así como las formas verbales que indican el sujeto, objeto, dativo y, en algunos casos, el alocutivo. Esto, naturalmente, ha alargado la duración de las encuestas. Cabe destacar que, al utilizar mapas relacionados con el léxico y la morfología, será posible tratar un número mayor de datos en fonética y fonología, superando los 37 inicialmente previstos.

La red de puntos de encuesta abarca todo el territorio donde se habla la lengua vasca, incluyendo tres zonas administrativas distintas: la Comunidad Autónoma Vasca, la Comunidad Foral de Navarra y una parte del departamento de los Pirineos Atlánticos, situada por encima de la frontera política entre Francia y España. Consideramos que lograr esta unidad de tratamiento de la lengua en su estatus dialectal es un logro destacable. En concreto, el EHHA ha realizado encuestas en un total de 145 puntos, con una densidad de red más amplia que muchos atlas lingüísticos vecinos. Estos puntos de encuesta se distribuyen de la siguiente manera: 36 en Vizcaya, 1 en Álava, 36 en Guipúzcoa, 27 en Navarra, 15 en Labourd, 18 en Baja Navarra y 12 en Soule. Esta cobertura exhaustiva refleja un esfuerzo significativo para capturar la variación dialectal y lingüística en toda la región del euskera.

Las encuestas realizadas en el marco de este proyecto se han llevado a cabo en lengua vasca, a excepción de la sección de morfología, donde hemos utilizado la traducción de estímulos en castellano o francés. Además, hemos aplicado el método de datos negativos, presentando respuestas plausibles, de las cuales hemos utilizado solo una parte mínima en la elaboración de los mapas. El enfoque metodológico que hemos seguido nos ha permitido

* https://www.euskaltzaindia.eus/index.php?option=com_ehha&Itemid=466&lang=eu&layout=sarrera&view=frontpage

recopilar más de 4.000 horas de grabaciones en cintas magnéticas, las cuales posteriormente se han digitalizado y almacenado en formato digital. Una selección de estas respuestas se encuentra disponible para su audición en el sitio web del atlas.

Durante la fase de recopilación de datos, se obtuvieron un total de 830,000 respuestas, siendo la mayoría de ellas espontáneas. Frecuentemente, se acompañaron de comentarios metalingüísticos relacionados con la respuesta proporcionada. Por ejemplo, los testigos pueden informar sobre:

- Localización geográfica: En la pregunta 02550, después de proporcionar la respuesta *atun* ('atún'), se destacó la diferencia regional con el comentario *Gipuskoarrak egalusi; guk geuk estu esaten* [Los guipuzcoanos dicen "egalusi", nosotros no].
- Distinción semántica entre las respuestas: En la pregunta 08100, se exploró la diferencia entre *oroï* y *lañu*. Se proporcionó el siguiente comentario: *Oroixak goikuk, lañuk menditarteetan egoten direnak* ["oroixak" es la niebla en las alturas, "lañuk" son las nubes que se quedan abajo en los montes].
- Uso de palabras obsoletas: En la pregunta 10030, se mencionó el cambio en el uso de ciertas palabras con el comentario: *Guk geure denpora gustin beti martisena; oin esate u (asteartea)* [En mis tiempos siempre hemos dicho "martisena", pero ahora se dice "asteartea"].
- Conocimientos enciclopédicos: En la pregunta 45050, se hizo referencia a un conocimiento enciclopédico con el siguiente comentario: *Gio thanbûa, bahia bezalako zen ber üngüia, bena pia eginik larriüz, ardilarriü edo hua zerbüttxatzen zen* [El tamiz que se llamaba "thanbûa" se parecía al "bahe", tenía el mismo círculo pero estaba hecho de piel, piel de carnero, y se utilizaba para...].
- Formas fijas y fraseología: En la pregunta 12010, se mencionó una expresión fija o fraseología con el comentario: *Urte barri txarri belarri / daukanak estekonari emoti* [Nuevo año, oreja de cerdo, que el que tiene le dé al que no tiene, canción de Año Nuevo].

Además de las observaciones aisladas y observaciones breves mencionadas anteriormente, también hemos recopilado una gran cantidad de textos libres. Algunos de estos textos tienen transcripciones asociadas, aunque no pueden ser incluidos en los mapas del atlas. Estos textos se encuentran almacenados en una base de datos específica para su posterior análisis y consulta.

Las encuestas realizadas se han transcrito minuciosamente en cuadernos manuscritos. A continuación, se muestra un ejemplo de la reproducción de una pregunta en la figura 1. En esta reproducción se incluyen el número de la pregunta, el número del testigo, el número de la cinta magnética, el contexto fónico de la respuesta, la respuesta transcrita en alfabeto fonético, los comentarios adicionales sobre la fiabilidad de la respuesta, la categoría gramatical y otros aspectos relevantes relacionados con la elocución, entre otros detalles.

Figura 1. Cuaderno manuscrito de encuesta de EHHA

Zenb.	Lek.	Zint.	Zm.	Ezeoko datuak																		
519		11710A	275																			
1. Testuing. <u>bust</u> Alf. Fon. <u>oʃ</u>				1. _____																		
2. Testuing. Alf. Fon.				2. _____																		
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 5%;">1.</td> <td style="width: 10%;"></td> <td style="width: 10%;"></td> <td style="width: 10%;"></td> <td style="width: 10%;"></td> <td style="width: 10%;"></td> <td style="width: 10%;">J</td> <td style="width: 10%;">+</td> <td style="width: 10%;">t</td> </tr> <tr> <td>2.</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>				1.						J	+	t	2.									3. _____
1.						J	+	t														
2.																						
lek. oharrak				4. _____																		
+ <u>bust</u> [pamiɛ]				5. _____																		
+ <u>bust</u> [ɛtɛmɛ]: urune aratg, da kontzeon yakeam sɛtɛmɛ da a susio kende da getatan sɛtɛmɛ fɛlɛn- gu mɛn ɛmɛ sɛtɛ, da... kɛrɛtɛ dɛmɛ tɛmɛmɛ				6. _____																		
Transkribatzailearen oharrak				7. _____																		
+ <u>bust</u> [ɛtɛkɔɪpɪ] <i>Emteko otonotik berizteko</i>				8. _____																		
Zenb.	Lek.	Zint.	Zm.	Ezeoko datuak																		
5110		11710A	296																			
1. Testuing. <u>bost</u> Alf. Fon. <u>ɔʃtɔ</u>				1. _____ β ≠																		
2. Testuing. Alf. Fon.				2. _____ δ																		
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 5%;">1.</td> <td style="width: 10%;"></td> <td style="width: 10%;">?</td> <td style="width: 10%;"></td> <td style="width: 10%;"></td> <td style="width: 10%;"></td> <td style="width: 10%;">J</td> <td style="width: 10%;">+</td> <td style="width: 10%;">t</td> </tr> <tr> <td>2.</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>				1.		?				J	+	t	2.									3. _____ v
1.		?				J	+	t														
2.																						
lek. oharrak				4. _____ γ																		
				5. _____ φ																		
Transkribatzailearen oharrak				6. _____																		
+ <u>ɛtɛkɔɪpɪ</u> <i>garioruna nahastuz</i>				7. _____																		
				8. _____																		

Aurrekoetxea, 2008.

La información recopilada en los cuadernos manuscritos ha sido posteriormente integrada en la base de datos, tal como se mencionó anteriormente hemos optado por el enfoque sistemático de informatización del atlas. En la figura 2, se puede observar un ejemplo que ilustra un punto de encuesta (Dima), su número correspondiente (117), la respuesta transcrita en alfabeto fonético (ógi) o en la ortografía habitual (ogi) que se obtiene de manera automática, el contexto fónico (*bost*: un numeral que afecta la aparición de posposiciones y modifica la forma del lexema nudo), la referencia al testigo (D. Aurrekoetxea Manzarraga), las indicaciones sobre la facilidad o dificultad del testigo para responder, la categoría gramatical, la elocución (normal, rápida, trabada) y la fiabilidad de la transcripción del locutor. Todos estos

aspectos están relacionados con la respuesta en sí. Además, en el ejemplo se observa que el hablante menciona una palabra adicional como *pamitxe*, que no era el objetivo de la pregunta, y cuya realización fonética se proporciona al mismo tiempo que su significado, aportado por el hablante o por la persona encargada de transcribir la encuesta. De esta manera, se han obtenido numerosos elementos lingüísticos complementarios que no se han incluido en la cartografía del atlas.

Figura 2. Respuesta a la pregunta “pan” en el banco de datos del EHHA

The screenshot shows the 'EHHA-Datu Sarrera' application window. The main form is titled '51090 - hogaza redonda' and '1 - óyi / ógi'. It contains several fields for data entry:

- Lekukoa:** 1 - Domingo AURREKOETXEA MANZARRAGA
- Testuingurua:** bost -
- Fonetikoki:** óyi
- Ortografikoki:** ógi
- Ezagutza:** -: Arazorik gabe
- Kategoria gramatikala:** izena
- Elokuzioa:** Erantzuna normal eman du
- Lekukoaren fiabilitatea:** Arazorik gabe
- Transkrip. fiabilitatea:** Arazorik gabe
- Egiatzaketa:** Egiaztatuta
- Oharra:** Oharra gehitu
- Testua:** Ogi txiki bat, biribila, lebadurarik gabea: bost #01 pámitxe
- Buttons:** Utzi, Ados

The left sidebar shows a list of categories, with '51090 - hogaza redonda' selected. Below it, there are several 'Oharra' (Note) icons and a list of categories: 1 - Oharra, 2 - etzekógi, 1 - Oharra, 51100 - pan de comuña, 1 - árto, 1 - Oharra, 2 - trémes, 1 - Oharra, 2 - Oharra, 1 - Oharra, 2 - Oharra, 3 - Oharra, 4 - Oharra.

Fuente: Aurrekoetxea, 2008

Una vez completada la encuesta, procedimos a la cartografía de los datos. Para ello, utilizamos un fondo de mapa basado en polígonos de Thiessen. Esto significa que la delimitación de los puntos de encuesta no coincide exactamente con las fronteras geográficas detalladas de cada municipio. El mapa se modelizó de la siguiente manera: cada punto dentro de un polígono de Thiessen está siempre más cerca del centro de su polígono que del centro de cualquier otro polígono. Los pueblos donde no se realizó encuesta no aparecen en el mapa. En su momento, el enfoque utilizado para el mapa fue bastante novedoso y también generó críticas. Se le acusaba de forzar la interpretación en función de las áreas presentes en el mapa. Además, la utilización de la lematización en lugar de la semiología tradicional de los mapas punto por punto también era una novedad. Sin embargo, la representación por lemas tiene la ventaja, que hoy en día se discute mucho menos, de permitir una visualización más clara de la distribución geográfica de los datos en el mapa. Por supuesto, el detalle de cada respuesta también se muestra, pero no en el mismo mapa. Todas las respuestas se presentan en una lista, en alfabeto fonético, pueblo por pueblo, indicando las respuestas solicitadas.

La lematización es un paso crucial previo a la construcción del mapa. Hay varios elementos que pueden ayudar en el proceso de lematización de los datos. Se utiliza la etimología y se pueden distinguir diferentes tipos de elementos lingüísticos: autóctonos, hereditarios, préstamos, históricos o recientes, y crudos. Además, se pueden tener en cuenta la presencia o ausencia de un fonema, una regla fonológica, un morfema, entre otros. Durante la lematización, se realiza un agrupamiento de los lemas. Sin embargo, debido a limitaciones técnicas, el número total de lemas no puede exceder los 16. Esto se debe, por ejemplo, a la necesidad de diferenciar distintos colores en el mapa.

El mapa 137 (figura 3) muestra las designaciones de la palabra 'codorniz'. En el mapa se pueden observar tres zonas importantes: (1) Zona oriental (en verde): En esta zona, la respuesta predominante es *kalla*, que es un claro préstamo del francés `caille` u occitano. (2) Zona central (en marrón): Aquí, la respuesta más común es *galeper*, que es un compuesto de los términos *gari* y *eper*, que significa 'perdiz del trigo'. Esta designación está relacionada con el hábitat de la codorniz. También se menciona que una canción tradicional amorosa de la zona oriental cita a la codorniz cantando debajo del trigo, aunque en esta canción el trigo no se llama *gari*. (3) Zona occidental: En esta área, la respuesta más común es *poxxpolin*, que originalmente se refería al canto de la codorniz, pero con el tiempo pasó a designar al pájaro en sí. Sin embargo, la forma *galeper* todavía se encuentra en esta región, formando una especie de isla, probablemente debido a su antigüedad en comparación con *poxxpolin*.

El uso de colores y lemas en el mapa proporciona información general sobre la distribución geográfica de las respuestas. Además, los círculos de colores pequeños en el mapa, llamados "datos negativos", brindan información adicional. Por ejemplo, en la zona oriental, se muestran dos puntos de encuesta que tienen un área en blanco, lo que indica la ausencia de respuestas espontáneas. Sin embargo, los círculos verdes indican que los testigos aceptaron la palabra *kalla* como respuesta exacta después de que se les hubiera propuesto esa palabra.

En la zona occidental se observa un fenómeno similar con los círculos rojos y marrones. Por supuesto, el mapa no puede proporcionar toda la información sobre la transición entre *galeper* y *poxxpolin*, y se brinda más información en la leyenda (contribución de Bergara). Según lo especificado por las personas encuestadas en este pueblo y en otros, «se dice *poxxpolin* cuando se escucha cantar, pero si un cazador lo mata, se le llama *galeper*». La diferencia entre cómo se llama el canto y cómo se llama al pájaro es evidente. El mapa muestra cómo se elige una u otra de las dos soluciones (en marrón para *galeper* y en rojo para *poxxpolin*). Es importante admitir que la variación léxica no se explica únicamente a través del mapa. Es evidente que muchos de estos mapas abren el camino a monografías con el respaldo de las respuestas y comentarios de los testigos (sin mencionar los recursos de la filología). La distribución en tres áreas, con los dialectos orientales de Labourd, Baja Navarra y Soule, la zona central con Navarra y Guipúzcoa, y la zona occidental en Vizcaya, se corresponde más o menos con la fragmentación dialectal que hemos expuesto en la dialectometría (Aurrekoetxea et al., 2020).

Figura 3. Mapa de la respuesta a la pregunta “codorniz”



Fuente: Aurrekoetxea & Videgain (2008).

Aunque hemos llevado a cabo investigaciones de dialectometría utilizando conjuntos de datos y técnicas estadísticas, existe otra manera de explorar el atlas, analizando mapa por mapa, lo que permite observar los patrones destacados en la distribución geográfica de los datos. A continuación, presentaremos distintos tipos de mapas, que van desde aquellos con un solo lema (monónimos, una única respuesta) hasta aquellos con más de 20 lemas diferentes.

Por ejemplo, en el mapa 191 sobre la palabra 'liebre', se observa que la única respuesta registrada es erbi. En otros casos, la variación se limita a un solo fonema, como se puede apreciar en el mapa 114 que trata sobre la palabra 'serpiente', donde se contrastan los términos *suge* y *sube*, que son términos hereditarios.

Es importante mencionar que el atlas en sí no explica las razones detrás de las diferencias en el número de significantes o términos utilizados para describir una misma entidad, como en el caso de la zoología. Sin embargo, proporciona un material valioso para llevar a cabo estudios e investigaciones que puedan analizar y explicar dichas variaciones.

La variación fonémica puede tener un impacto en los préstamos, como en el caso de la palabra 'fagus' del latín, donde se observan variantes en la inicial, como *fago*, *bago*, *pago* y *phago*. Por otro lado, en el mapa 18 'hormiga', se aprecian múltiples variantes fonológicas en un término hereditario siempre trisilábico, como *txingurri*, *txiniurri*, *txinurri*, *txindurri*, *xinurri*, *zínurri*, *zínhurri*, *zinaurri*, *zínhurri*, *xingurri*, *xinaurri*, *iñurri*, *inhurri*, *üñhurri*, *üñhürri*, *intxorri*, *ipurri* y *zimurri*.

En ocasiones, la distribución de las formas en tres lemas indica que un lema está compuesto por los otros dos lemas. Por ejemplo, en el mapa 30 'pulga', se puede observar que el lema compuesto de la zona central, *arkakuso*, se origina a partir de dos lemas: *ardi* occidental

y (*ku*)*kuso* oriental, siguiendo una estructura nominal compuesta de *ardi+kukuso* que evoluciona a *arkukuso*. Otro fenómeno interesante abordado por la etnolingüística es el uso de designaciones con motivaciones formuladas en una narración etiológica (muy conocida fuera del País Vasco), que a menudo presentan una desconexión entre la narración y la designación en sí.

Tomemos como ejemplo el mapa 121 'lución', donde en las formas occidentales, como *sugitsu*, se encuentra la motivación de '*serpiente ciega*'. Sin embargo, los testimonios de los hablantes proporcionan narraciones similares pero utilizan palabras como *ziraun*, que no están relacionadas con la ceguera. Según la narración en la que confían algunos testigos, Dios castigó a una serpiente que deseaba hacer daño a todo lo relacionado con ella, dejándola ciega por voluntad divina. Tanto *sugitsu* como *ziraun* se aplican a esta narración. Es importante destacar que el animal llamado *suge itsu*, literalmente '*serpiente ciega*', no es una serpiente ni tiene ceguera.

Existen otras razones para llevar a cabo la lematización. Además de los préstamos, que ya hemos mencionado como posible criterio, en el mapa 121 'laurel', se contrastan las formas *erramu* como préstamo y *ereiñotz* como término hereditario. También se puede identificar la presencia de un calco. En el mapa 180 'comadreja', se encuentra la forma occidental *ogigaztaia*, que significa 'pan y queso' y tiene paralelismo semántico con el término románico 'paniquesa'. Por supuesto, pueden coexistir geosinónimos, unos motivados y otros, no. Por ejemplo, en el mapa 167 'cabrito', el lexema *ahuntzkume* está motivado como 'pequeño de cabra' y contrasta con *pittika*, forma aparentemente no motivada donde no se aprecia un elemento correspondiente a una designación de la 'cabra'.

A menudo, las motivaciones culturales, enciclopédicas y los tabús en las respuestas son compartidos con las lenguas vecinas, aunque no sea evidente para el lector que no conoce el euskera. Es que las diferencias entre los significantes que separen el vasco de las lenguas románicas hacen difícil darse cuenta de la unidad de motivación detrás de unos significantes tan alejados.

La dialectometría busca sintetizar y condensar la información proporcionada por cada mapa del atlas, pero no está prohibido examinar cómo una zona dialectal se aísla o se conecta con otros territorios, pregunta por pregunta. Aquí proponemos examinar el caso de Navarra. Por ejemplo, en el mapa 843 'hiel', se indica que el término *beazun* en Navarra es común con la mitad occidental de Guipúzcoa, en contraste con el término *birestun* más occidental de Vizcaya al este y el término (*gibel*)*min* oriental. En el mismo sentido, en el mapa 1017 'apagar', el término *itzeli* en Navarra coincide con el término *itzali* en Guipúzcoa, en oposición a los términos occidentales *amatau* y los orientales *hil* o *eho*. En el mapa 838 'ala', Navarra se alinea con *egal* en acuerdo con los dialectos occidentales que producen *egal* y *hegal*, mientras que Guipúzcoa y Vizcaya solo conocen una forma ligeramente diferente, *ego*.

En el mapa 935 'cal', el término navarro *kisu* es común con *gisu* y *kisu* ('gypsum') al otro lado de la frontera política en Labourd y Baja Navarra, en contraste con el término *kare* en el occidente y *latsiün* en el oriente. Sin embargo, es importante mencionar la presencia en el sur de Navarra, donde se habla vasco, de una forma endémica llamada *galtziñ*. En el mapa 785, 'el cerdo más pequeño de la camada', el término navarro *kadentxerri* es común con el término bajonavarro y suletino *kadantxerri*, en contraposición a los términos occidentales *erreus*, *nano* y *erkin*. Los términos comunes entre la forma navarra y otra zona pueden encontrarse en espacios discontinuos. Por ejemplo, en el mapa 813 'esófago', el término navarro *iranoi* guarda relación a distancia con el término suletino oriental *giranoi*.

Continuando con el análisis, en el mapa 781 'verraco', se observa que en todo el territorio navarro se utiliza la forma *aketz*, que se extiende también sobre Labourd y solo parte meridional de Baja Navarra. En contraste, en el suletino oriental se utiliza el préstamo *berrat*, mientras que en el occidental se emplean formas como *apote* o el motivado *txarriar* (puerco macho). En

otros mapas, Navarra aparece dividida en dos partes. En el mapa 827 'intestino', un cambio fonético (rotacismo) separa el término *ertze* en la Navarra oriental y los dialectos de Labourd y Baja Navarra de la palabra *este*, que cubre la mitad de Navarra. Esta división se repite en el mapa 831 'pringue', con *urin* en Navarra al este, al igual que en los dialectos orientales, y *manteka* común con Guipúzcoa.

En ocasiones, se produce una división en tres zonas, como en el mapa 805 'jamón', donde encontramos *azpi* al norte de Navarra, *pernil* al sur y *magra*, endémica en una zona intermedia. El término *pernil* no es común con el vecino guipuzcoano *urdaiazpiko*, pero muestra un efecto espejo con el término oriental de Vizcaya *pernil*. En lugar de ser horizontal y basarse en la latitud, la longitud puede influir en algunos mapas. Navarra puede revelar un espacio lingüístico con muchas variaciones, como se ve en las seis respuestas *arraila*, *artesi*, *zirriku*, *lertu*, *edeki*, *arteka* en el mapa 34, que se dedica precisamente a la 'grieta de la pared'. Por supuesto, Navarra (al igual que otros territorios) presenta formas endémicas, es decir, que no se encuentran en otras partes.

A nivel fonético, en el mapa 958 'contrapar' se utiliza la forma *kaprio* en Navarra, con anaptyx, en contraposición a los vecinos *kapiro*, *gapirio*, *kaperio*. En el mapa 1006 'cama', Navarra utiliza la forma *goatze* con una oclusiva inicial desconocida en las formas mayoritarias *ohe*, *oe*, *oi*. La distinción también puede ser léxica, como se ve en el mapa 956 'alfarda' con el término *astazaldi*, especialmente utilizado en Navarra, o en el mapa 968 'puerta' con *atare*, *atari*, así como en el mapa 975 'portal' con *larrain*.

Efectivamente, el recorrido realizado nos ha permitido explorar algunos datos específicos de Navarra, pero sin llegar a conclusiones generales. La dialectometría, por su parte, adopta un enfoque más global en el análisis de los dialectos. Sin embargo, existen aún muchos datos en el Atlas Lingüístico Vasco que no han sido completamente explotados y que merecen ser consultados de manera más detallada, abordando preguntas específicas. Esta tarea queda abierta para los investigadores interesados en utilizar el valioso recurso que ofrece el atlas lingüístico vasco, sin excluir otras fuentes de información.

REFERENCIAS

- Aurrekoetxea, G., Gaminde, I., Ormaetxea, J. L., & Videgain X. (2020). Euskalkien sailkapen berria. UPV/EHU.
- Aurrekoetxea, G. (2008). Basque Linguistics Atlas EHHA. From speech to automatic maps. *Dialectologia*, 1, 117-119.
- Aurrekoetxea, G., & Videgain, X. (1993). Euskal Herriko hizkuntza atlasa: galdesorta / cuestionario / questionnaire. *Euskera* 38, 529-647.
- Aurrekoetxea, G., & Videgain, X. (2008-2019). Euskal Herriko hizkuntza atlasa (11 volumen). Euskaltzaindia.

ARTICULOS

Artículo original

Aspecto socio-cultural de la situación lingüística del Paraguay

Socio-cultural aspect of the linguistic situation in Paraguay

Ígor Protsenko

Universidad del Norte, Paraguay

<https://orcid.org/0000-0002-6195-636X>

e-mail: protsent2002@mail.ru

Recibido: 4/3/2023
Aprobado: 25/5/2023

RESUMEN

En el artículo Aspecto socio-cultural de la situación lingüística del Paraguay se explica la noción socio-cultural, se destacan los componentes del fenómeno mencionado. Se revela lo que es potencia demográfica y etnodemográfica de la situación lingüística del Paraguay. Presta atención a la lengua guaraní como uno de los componentes potentes en el sistema lingüístico del Paraguay. Se subraya que el fenómeno de jopará o “tercera lengua” por una parte, es una categoría cultural, culturologema, sirve de medio de comunicación. Por la otra, es factor socializante, un símbolo potente de la solidaridad de los hablantes en la sociedad única. Se confirma que en Paraguay pasó el mestizaje cultural y no biológico. La “tercera lengua” es la esencia de la formación de la nueva sociedad paraguaya. El artículo explica que de las cuatro principales estrategias de aculturación (asimilación, separación, marginalización, integración) al Paraguay corresponde la integración: identificación una cultura con la otra, igualdad de los componentes la situación lingüística del Paraguay.

Palabras clave: situación lingüística; socio-cultural; potencia demográfica y etnodemográfica; adaptación psicológica y sociocultural; jopará.

ABSTRACT

The article “Aspecto socio-cultural de la situación lingüística del Paraguay” (Socio-cultural aspect of the linguistic situation in Paraguay) explains about the socio-cultural notion, the components of the mentioned phenomenon are highlighted. It shows what is the demographic and ethnodemographic power of the linguistic situation of Paraguay. Pay attention to the Guaraní language as one of the powerful components in the linguistic system of Paraguay. It is highlighted the phenomenon of jopará or “third language” on the one hand, is a cultural category, culturologem, it works as a way of communication. On the other hand, it is a social factor, a powerful symbol of the solidarity of the native speakers in the unique society. It is confirmed that the mixing was cultural and not biological. The “third language” is the essence of the formation of the new Paraguayan society. The article explains the four main strategies of acculturation (assimilation, separation, marginalization, integration) Paraguay matches with integration: identification of one culture with another, equality of the components on the linguistic situation of Paraguay.

Keywords: linguistic; socio-cultural situation; demographic; ethnodemographic; power; psychological and sociocultural adaptation; jopará.

LA INVESTIGACION

La situación lingüística es un fenómeno de una parte, lingüístico, de la otra, socialmente condicionado y se representa como un concepto complejo, difícilmente inteligible que a lo largo de la historia fue interpretado por varios lingüistas en las formas diferentes.

Actualmente la trayectoria de las investigaciones se concentra en los factores extralingüales y en el aspecto funcional de los sistemas lingüísticos (idiom's) en el contexto sociocultural (Nikolskiy, L. V., 1976; Avrorin V.A., 1975; Shveytser A.G., Nikolskiy L.V., 1978; C. Ferguson C., 1959; Haugen E., 1975; Catcart M.C., 2009; etc.).

Sin embargo, los factores relevantes de la situación lingüística son estatuto político-administrativo del país, nivel del desarrollo de las infraestructuras socioculturales, potencia demográfica y etnodemográfica, procesos históricos y migratorios que a su vez complica la definición del concepto analizado.

Basándose en los resultados de las investigaciones de V. A. Avrorin, la situación lingüística (SL) nosotros entendemos como un sistema jerárquico de las lenguas (sistemas lingüísticos) y las formas de sus existencias, formada para el tiempo determinado en un territorio bajo la influencia de los factores: geográficos, demográficos, sociales, socio-históricos, culturales, políticos, económicos, migratorios, etc., que a su vez lleva a la creación del ambiente lingüístico (AL), una variedad de la comunidad de los seres humanos unidos entre sí con los lazos de parentesco, sociales, territoriales dentro de la cual se realiza la comunicación (V.A. Avrorin, *op. cit.*).

La finalidad de nuestra investigación consiste en determinar la noción sociocultural o socioculturalismo como la base de nuestra investigación. Revelar potencia demográfica y etnodemográfica de la situación lingüística del Paraguay y de cada uno de sus componentes principales. Especificar el rol de la situación lingüística en la creación y desarrollo de la sociedad paraguaya.

Aprovechando el legado clásico (Émile Durkheim, Karl Marx, Max Weber, Talcott Parsons, Pitirim Sorokin, Nikolay Lápin, etc.) entendemos la noción *sociocultural*, *socioicaturalidad* o *aspecto sociocultural* (SC) como la trinidad indisoluble de tres componentes: personalidad, cultura y sociedad.

Personalidad (homo activus) realiza las acciones e interacciones sociales, cambiando y desarrollando los procesos de Influencia mutua entre culturalidad y sociedad.

Cultura se entiende como el conjunto de los métodos y resultados de la actividad humana, material y espiritual: ideas, valores, necesidades. Revela aspecto funcional del fenómeno analizado.

Sociedad es el conjunto de relaciones de una persona u otro sujeto social con otros sujetos: político-económicos, ideológicos, etc., que surgen como resultado de actividades destinadas a crear, desarrollar y garantizar condiciones de coexistencia en la interrelación de necesidades, intereses y valores.

Así que el enfoque sociocultural significa comprender la sociedad como la unidad de la cultura y la socialidad formada por la actividad humana (Lapin),

La ciencia del lenguaje, en la parte que se ocupa del aspecto social y funcional (cultural), está constantemente en peligro de disolverse en la sociología del lenguaje, cuyo punto de partida es la sociedad y no el lenguaje, por lo que el objeto de estudio de esta última no debe considerarse el lenguaje en la sociedad, sino al revés, la sociedad en la lengua, según la teoría de Wilhelm von Humboldt.

Las lenguas, al ser parte notoria de la cultura, crean una imagen del mundo para el individuo, es decir, una imagen conceptual del mundo, y forman alrededor de las personas a las que pertenecen, un espacio especial con sus características sociales, culturales o funcionales.

Como sistema de visión del mundo, el lenguaje tiene un efecto regulador en el comportamiento humano, en las relaciones entre sí y con la sociedad. Un individuo, como parte de la sociedad, ve y se comunica con la realidad, ya que le presenta el lenguaje. Pero cuando se trata de una sociedad multilingüe, la situación se complica. Lo que conduce tanto a contradicciones entre lo social y lo cultural, o viceversa, complementándose mutuamente desarrollan y en algunos casos forman una nueva sociedad. A lo que nos enfrentamos en Paraguay.

Hablando de la SL del Paraguay, hay que mencionar que es un país pluricultural, pero bilingüe: español y guaraní, según el artículo 40 de la Constitución Nacional del 1992.

Las demás lenguas indígenas, así como de las otras minoritarias forman parte del patrimonio cultural de la Nación.

Aunque en las regiones fronterizas con Brasil, en Pedro Juan Caballero, Ciudad del Este, por ejemplo, se observa la situación siguiente.

Un paraguayo se acerca a un bar pidiendo algo de comer. Lógico que los empleados le atienden en castellano. Enseguida entra un luso-parlante, preguntando algo, y los mozos le contestan en portugués. Despachados los clientes, los camareros se ponen a hablar en guaraní. En unos minutos utilizaron tres lenguas diferentes con la misma fluidez y pasando de una a otra con total naturalidad. Resulta que son perfectos trilingües.

En la comunidad de Santa Rita se puede observar la situación cuando una persona guaraní hablada mezcla esa lengua con portuñol.

No hace falta olvidar que en el territorio del Paraguay se asentaron casi veinte comunidades, agrupadas según su cultura y familia lingüística, se trata de los inmigrantes de diferentes olas: alemanes, italianos, coreanos, japoneses, ucranianos, etc. que, además del castellano y guaraní, dominan idiomas de sus etnias (sobre todo los asiáticos: coreanos y japoneses).

Por supuesto, existen culturas y lenguas de las etnias indígenas minoritarias de tupí-guaraní que en la vida cotidiana usan sus lenguas nativas entre los que están *aché*, *mbyā*, *avā*, *chripā*, *pā'i tavyterā* etc. Estas lenguas se están perdiendo, pero, según de las confirmaciones de algunos investigadores, hoy día son casi la mitad de la población del país.

Tabla 1. División de los pueblos indígenas por las familias lingüísticas

POBLACIONES INDÍGENAS POR FAMILIAS LINGÜÍSTICAS Y PUEBLOS

FAMILIA LINGÜÍSTICA	PUEBLO
Guaraní	Pa'i Tavyterā Mbya Ava Aché Guaraní Occidental Guaraní Ñandéva
Mataco	Nivacle Manjui Maká
Zamuco	Ayoreo Ybytoso Tomarāho
Maskoy	Toba Maskoy Enlhet Norte Enxet Sur Angaité Guaná
Guaicurú	Qom

Secretaría de políticas lingüísticas del Gobierno del Paraguay.

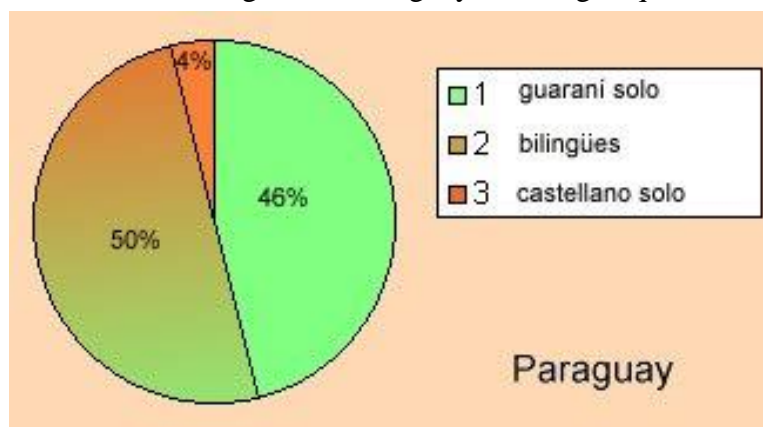
Según información de la tabla las diecinueve lenguas indígenas del Paraguay se agrupan en cinco familias lingüísticas. La más potente de todas es guaraní. ¿Por qué ni otros idiomas, ni las familias lingüísticas no evolucionaron igual que guaraní-ete? Es porque fueron menos “gramaticalizados”. Gracias a los jesuitas, la lengua guaraní si no fortalece su posición en la sociedad durante el periodo colonial y posterior, al menos no la pierde y más con el tiempo se convirtió en lo que ahora llamamos guaraní paraguayo (*jehe’a*) o criollo. Es el guaraní, con pocas influencias del castellano que por una parte deriva del cario-guaraní y, según nuestras observaciones, por la otra, tiene la base de *ava ñe’ê*. Actualmente casi 80% de los paraguayos habla o entiende el guaraní (ya sea como idioma materno o segunda lengua). Lo que toca la población indígena que son en total 117.000 aproximadamente, casi la mitad, 48,9%, en la vida cotidiana hablan el guaraní paraguayo. Además, esa lengua sirve como herramienta en la comunicación intertribal. Por ejemplo, las tribus vecinas de Mbya y Maká en Itapúa, el Sur del Paraguay, tienen sus lenguas, una de las cuales pertenece a la familia guaraní, la otra a la Mataco respectivamente. Para mantener las relaciones económicas, culturales, etc. usan guaraní.

Del día, al día, gracias a la política gubernamental el uso del guaraní se amplía, ganando un punto de apoyo en varias áreas de la vida social: educación, economía, política, cultura, etc. Lo que es muy importante mencionar que existe y se desarrolla la literatura en guaraní: Jorge Canese, David Galeano Olivera, Julio Correa, etc. son algunos escritores de nuestra contemporaneidad.

Español, a su vez, siendo lengua de la comunicación internacional, cambiándose poco a poco bajo la influencia del guaraní, no pierde sus posiciones, en primer lugar, en Asunción, la capital del país, que se explica por la necesidad de llevar los negocios internacionales, es la lengua de ONU, del Parlamento de los tratados internacionales.

Así, la potencia demográfica de la situación lingüística del Paraguay, basada en guaraní, la lengua de confianza de los paraguayos (Gobierno Nacional) y español, la lengua de la comunicación internacional, actualmente se considera como estable, con el predominio del castellano. Influyen a la situación los procesos migratorios de los países del mundo hispanohablante que sobreviven actualmente crisis político-económica. Los migrantes se quedan en Paraguay sin sufrir barreras lingüísticas para adaptarse a las nuevas realidades de la vida cotidiana, a la cultura del país nuevo. Al asociarse con las nuevas realidades, los descendientes de ellos absorben las tradiciones del español paraguayo y guaraní, en primer lugar, el idioma, que forma parte principal de la cultura del Paraguay. Por otra parte, vale la pena mencionar que el español paraguayo es leal a otros dialectos del castellano, lo que crea ambiente agradable para los migrantes.

Diagrama 1. distribución de las lenguas del Paraguay en el hogar, que es la base de la sociedad.



Fuente: <https://sites.google.com/site/paraguaycorazonsudamericano/informacion-general/lenguas>

Y si la potencia demográfica revela en la adaptación de la lengua a las condiciones socio-económicas, potencia etnodemográfica se caracteriza por la posibilidad de servir a las necesidades de la vida cotidiana, revelar los fenómenos culturales, describir los valores nacionales, reflejar el espíritu del pueblo.

Potencia etnodemográfica esta enlazada y depende de la interacción de las culturas y, como consecuencia, de las lenguas como sistemas lingüísticos, la función de los cuales es la transmisión de la información semántica valorativa no tanto de un grupo social al otro, sino de una cultura a la otra.

El nivel de la potencia etnodemográfica de la SL se determina por el nivel de cada uno de los componentes de sistema y en Paraguay, sin duda, tiene carácter de un monolito estable y potente. Así que el español era y sigue siendo la lengua del gobierno, justicia, parlamento, sistema educativo, relaciones internacionales y el guaraní, por su parte, siendo lenguas de las imágenes sigue relegando a la intimidad, confianza en las relaciones humanas y entre miembro de la sociedad y la sociedad misma, amistad, ira y amor, es la lengua del hogar que a su vez es componente básico de la formación de la sociedad. Hablando sobre el tema de la potencia etnodemográfica del guaraní, numerosos investigadores buscan en la historia del Paraguay: “cómo el guaraní ha logrado sobrevivir al pueblo que lo creó, hecho que se registra en ninguna otra parte de América. El indio muere ... En el Paraguay queda el idioma, pero no el indio, y aquí reside la originalidad del problema: el legado de una lengua a la raza dominante...”. (Insfran, 1942, 60).

De hecho, durante la Guerra del Chaco, el 28 de mayo de 1933 consiente importancia de mantener la máxima garantía posible en las comunicaciones, el Comandante de Ejército en Campaña José Félix Estigarribia declara al guaraní como idioma oficial de la guerra ordenado que conversaciones y claves secretas se hagan exclusivamente en idioma guaraní. Paraguay obtuvo la victoria. Desde aquel momento puede decirse que el guaraní fue el idioma de la victoria que fortalecían el patriotismo de los paraguayos y orgullo por su lengua. En la lucha no habían ido solo fusiles, cañones, armas de soldado. La lengua guaraní fue un arma más, un arma imbatible al que solo tenían acceso los paraguayos. Por eso los paraguayos hacen la guerra y la ganan en guaraní (Guaraní durante la Guerra del Chaco, 2016). Así la lengua guaraní se convirtió en orgullo, símbolo de la nación, símbolo del país y sigue viviendo en todas esferas de la sociedad actual.

Interacción de las lenguas en Paraguay llevó a aparecer un fenómeno único para todo el territorio de América Latina que en la forma metafórica se llama “tercera lengua” del país: jopará: “Que tal pío si los paragua celebramos día de la galopera con cambuchi o la chiperita con canastas de chupas y nuestro bello typoi de Ñanduti...mbae día de brujas...importa si vienen de los celtas o sajones o yankis o de donde sea valoremos lo nuestro elevemos nuestro autoestima Latinos mío somos en fin .cada loco con su tema ..Manada besotes” (Facebook. 31.10.2022. Alicia Pesoa Tomberli)

En el proceso de la influencia mutua de las culturas e idiomas aparece nueva cultura de la sociedad paraguaya con su individual y único SL que Bartomeo Melià se llama bilingüismo folklórico (Melià, 2004). Desde nuestro punto de vista sería más correcto caracterizar el SL del Paraguay como bilingüismo equitativo con elementos de diglosia. Ambas lenguas, español y guaraní, tienen igual poder, ambas se absorben con la leche de la madre, es decir, las ambas son nativas

Actualmente en la mayoría de los casos un bilingüe paraguayo que habla en español reconoce la importancia del guaraní para todo el país. El que prefiere hablar guaraní no deja de reconocer el poder, necesidad y la importancia internacional del español tanto en el campo político-económico, como en el cultural.

Esta interpretación de la SL paraguayo no es nada más que la reflexión de la opinión sociocultural que es la consecuencia de la coexistencia de dos diferentes creencias, acepciones del mundo, mentalidades. Hablamos del sentimiento de ambivalencia con respeto al idioma para los paraguayos. Muy importante mencionar que el bilingüismo paraguayo (fenómeno de jopará) apareció de la relación comunicativa de los adultos, dejando aparte la razón racial, como sería el mestizaje (en su entendimiento tradicional). Lo que, a su vez, nos convence confirmar que en Paraguay observamos el fenómeno del mestizaje cultural y no biológico. La “tercera lengua” es la esencia de la formación de la nueva sociedad paraguaya, y ahora sirve para las necesidades de nueva sociedad y no de una nueva raza.

Fenómeno de jopará, de hecho, por una parte, es categoría cultural, culturologema, sirve de medio de la comunicación. Por la otra, es factor socializante, un símbolo potente de la solidaridad de los hablantes en la sociedad única. Además, es símbolo de la pertenencia de los parlantes a la misma sociedad, factor de integración y desintegración de los paraguayos, herramienta para guardar su identidad. Refleja el modo de la vida de la gente del país, su carácter nacional, tradiciones, costumbres, aceptación conceptual del mundo con el propio sistema de valores.

RESUMEN

La comunicación de los pueblos, formantes de la sociedad paraguaya, llevó a la aparición de las capacitaciones comunicativas únicas, creó su propio modelo lingo-cultural, su propia imagen etno-cultural.

Actual SL del Paraguay, siendo la base de la formación de la sociedad, se forma en el resultado de aculturación: procesos y resultados de interacción de las culturas diferentes, en los que todos o la parte de los representantes de una cultura adoptan las normas, valores y tradiciones del otro. De las cuatro principales estrategias de aculturación (asimilación, separación, marginalización, integración) al Paraguay corresponde integración: identificación una cultura con la otra, igualdad de los componentes SL. Es decir, identificación étnica y tolerancia étnica son positivas. Así, en el territorio del Paraguay aculturación logró su finalidad principal: observamos es proceso de adaptación completa de las culturas de los pueblos del país y aparición la sociedad nueva con su propia forma de hablar, su “lengua” jopará.

Adaptación se caracteriza por aspecto psicológico y sociocultural.

Adaptación psicológica positiva en Paraguay se revela en satisfacción, tranquilidad, lealtad psicológica de los portavoces de varias culturas, sin diferenciar las culturas dominantes o donantes y marginales; identificación cultural individual no se opone a los valores nacionales de la sociedad, que se formaron a lo largo de historia, siendo el resultado de los procesos de interacción, influencia mutua y enriquecimiento mutuo de las culturas.

Adaptación sociocultural es la capacidad de adaptarse a la nueva sociedad, aceptar otras culturas sin otorgar ni un elemento de ellos. En caso del Paraguay, como estaba mencionado más arriba, interacción de las culturas y lenguas llevaron a la creación de la nueva sociedad, y no la raza, en la que cada cultura se acepta con respeto. Español y guaraní tienen estatuto de las lenguas oficiales y las lenguas de los minoritarios son Patrimonio Cultural del país.

SL del Paraguay, según la opinión de Bartomeo Malià se desarrolla junto con la sociedad. Cada uno de los componentes del sistema reacciona a su manera en los cambios de la vida social. Ninguna realidad social nueva (o cambio de valores entre la gente) no se escapa de la atención de la lengua viva, la función principal de la cual es entrelazar las relaciones entre un individuo y la sociedad.

W. von Humboldt en sus obras mencionó que cada lengua es la capacidad de la transformación de la realidad a las palabras, a las lenguas. Cada lengua vale la pena aceptar como la intención satisfacer esa necesidad interior. Hablando de la SL multilingüe hay que

minuciosamente examinar las herramientas disponibles a cada una de las lenguas para realizar su función principal para crear la imagen común de cualquier situación lingüística, en nuestro caso, del Paraguay. Así seguiremos en nuestras investigaciones futuras.

REFERENCIAS

- Nikolskiy L.B. (1976). Sociolingüística sincrónica (teoría y problemas), Moscú, Nauka.
- Avrorin V.A. (1975). Problemas de la investigación de la parte funcional de las lenguas: desde el punto de sociolingüística. Leningrado, Nauka.
- Shveytser A.G., Nikolskiy L.B. (1978). Introducción a sociolingüística. Moscú, Escuela Superior.
- Ferguson, C. (1959). "Diglosia". World 15: 325 – 340.
- Haugen, E. (1966). Language Conflict and Planning: the Case of Modern Norwegian, Cambridge, Mass. Recuperado de: <http://www.philology.ru/linguistics1/haugen-75.htm>
- Novedades en lingüística. Vol. VII. Sociolingüística. (1975): 441-472.
- Causse Cathcart, M. (2009). El concepto de comunidad desde el punto de vista socio-histórico-cultural y lingüístico. Departamento de Letras, Facultad de Humanidades Universidad de Oriente, Santiago de Cuba. Disponible: <https://www.redalyc.org/pdf/1813/181321553002.pdf>
- Secretaría de Políticas Lingüísticas del Gobierno del Paraguay. Disponible: <https://www.spl.gov.py/es/index.php/misional/lenguas-indigenas> y <https://www.spl.gov.py/es/index.php/noticias/cuantas-lenguas-indigenas-hay-en-paraguay>.
- Gobierno Nacional. Disponible: <https://www.mre.gov.py/la/index.php/noticias/el-guarani-el-idioma-de-confianza-de-los-paraguayos>,
- Guaraní durante la Guerra del Chaco. (2016). Disponible: <http://www.chacosinfronteras.com/2016/06/02/idioma-oficial-guerra-del-chaco/>
- Melià, B. (2004). La tercera lengua del Paraguay. Asunción, Servilibro.
- Insfran, P.M. (1942). El Paraguay país bilingüe Revista de Ateneo Paraguayo 1, 5º6: 59-61. Paraguay, 1942.
- Lapin, N.I. Transformación socio-cultural en Rusia: liberalización versus tradicionalización. Disponible: http://www.jourssa.ru/sites/all/files/volumes/2000_3/Lapin_2000_3.pdf

Artículo original

A memória discursiva em quadrinhos da Marvel sobre Thor

Marvel's comic discursive memory about Thor

Bruno Aguinaldo Feitosa

Universidade Estadual de Mato Grosso do Sul (UEMS), Brasil

<https://orcid.org/0000-0001-9821-9171>

e-mail: bruno.feitosa@outlook.com

Rosemere de Almeida Agüero

Universidade Estadual de Mato Grosso do Sul (UEMS), Brasil

<https://orcid.org/0000-0001-7250-4206>

e-mail: rosemere@uems.br

Recibido: 19/12/2023

Aprobado: 13/3/2023

RESUMO

Este estudo intitulado “A memória discursiva em quadrinhos da Marvel sobre Thor” tem o objetivo de analisar as práticas discursivas emergentes na sociedade americana, no século XX, de modo a descrever como o articulista, Stan Lee, instaura a memória discursiva em quadrinhos recortados da Marvel sobre o sujeito/personagem Thor, por meio dos discursos imagéticos em circulação nas HQs, no âmbito do corpus recortado. Busca-se verificar se os discursos recortados evidenciam a memória discursiva no aspecto apenas da estabilização de sentidos, pelo viés do interdiscurso, ou se também é possível observar os deslizamentos de sentidos. O corpus é constituído por discursos imagéticos recortados das HQs das histórias em quadrinhos do “Thor”, de Stan Lee, e revista Journey Into Mystery (1962), no Brasil publicada pela editora Panini como Biblioteca Histórica Marvel, Volume I. As análises são realizadas com base nas teorias da Análise do Discurso (AD), a partir da voz teórica de Michel Pêcheux. A metodologia utilizada inclui o fechamento do espaço discursivo e o exame das condições de produção e da memória discursiva. As análises evidenciaram o duplo viés da memória discursiva, que de um lado se apresenta como um espaço de repetição de sentidos e, de outro, é um espaço de deslizamentos de sentidos, que irrompem na perspectiva de um acontecimento novo que vem perturbar a aparente estabilidade presente nos discursos imagéticos analisados.

Palavras-chave: histórias em Quadrinhos; sujeito Thor; memória discursiva.

ABSTRACT

This study entitled “Marvel’s discursive memory in comics about Thor” aims to analyze the discursive practices emerging in American society in the 20th century, in order to describe how the writer, Stan Lee, establishes the discursive memory in cut-out comics from the 20th century. Marvel on the subject/character Thor, through the imagery discourses in circulation in the comics, within the scope of the cut corpus. It seeks to verify if the clipped speeches evidence discursive memory in terms of stabilization of meanings, through the interdiscourse bias, or if it is also possible to observe the slippage of meanings. The corpus consists of imagery discourses cut from the comics of the comics of Stan Lee's "Thor", and the magazine Journey Into Mystery (1962), in Brazil published by the publisher Panini as Biblioteca Histórica Marvel, Volume I. The analyzes are carried out based on Discourse Analysis (DA) theories, based on the theoretical voice of Michel Pêcheux. The methodology used includes closing the discursive space and examining the conditions of production and discursive memory. The analyzes showed the double bias of discursive memory, which, on the one hand, presents itself as a space of repetition of meanings and, on the other hand, is a space for slipping of meanings, which erupt in the perspective of a new event that comes to disturb the apparent stability. present in the imagery discourses analyzed.

Keywords: comics; subject Thor; discursive memory.

INTRODUÇÃO

A aparição de *Thor* nos quadrinhos foi na revista *Journey Into Mystery* n° 83, em agosto de 1962, intitulada “O poderoso *Thor* e os homens de pedra de Saturno”. A ideia de Stan Lee era criar um super-herói que fosse metade humano e metade deus, dando-lhe o nome de *Thor*, o poderoso deus do trovão da mitologia nórdica.

No período em que a revista foi lançada, o mundo vivia o advento da Guerra Fria. Hobsbawn (1995) narra que a denominação Guerra Fria é utilizada para designar uma situação de conflito geopolítico que envolveu duas grandes superpotências mundiais (os EUA e a URSS) e seus aliados, após o término da Segunda Guerra Mundial. Nesse período, os Estados Unidos (EUA), país capitalista, e a União das Repúblicas Socialistas Soviéticas (URSS), nação socialista, centralizaram as relações políticas internacionais buscando mostrar seu poderio bélico e influência sobre o mundo.

Embora não tenha acontecido um conflito armado de fato, havia um permanente estado de tensão mundial, pois as potências mencionadas detinham a tecnologia da bomba atômica, arma que poderia destruir o planeta caso uma guerra viesse a ser deflagrada. Cada um desses países buscava provar a sua hegemonia ao resto do planeta e, para isso, usavam propagandas, investimentos em armas, tecnologias, lançando-se até mesmo à conquista do espaço por meio da *corrida espacial*.

Foram nessas condições de produção histórica que Stan Lee criou *Thor*, partindo da ideia de um herói que pudesse combater ataques comunistas. Tomando por base essa ideologia anticomunista, Stan Lee cria o médico Donald Blake, um personagem manco, fraco e que aparentemente não mostra causar qualquer tipo de perigo, condição física que facilita a sua entrada nas barreiras inimigas. Quando esse personagem, Donald Blake, se vê em perigo ele se transforma no poderoso *Thor*, o deus da mitologia nórdica, um herói tão forte que é temido pelos seus inimigos, pois sua força é incomparável e seu martelo destrói tudo que fica em seu caminho. Ao assumir a personalidade de *Thor*, Blake passa a agir e ter as mesmas emoções do deus nórdico, tornando-se forte, corajoso e impiedoso com seus inimigos.

Essas condições de produção históricas influenciaram as primeiras aventuras do deus asgardiano, pois suas histórias eram impregnadas de ideologias políticas que perpassavam a sociedade na década de 1960. Deste modo, os quadrinhos desse período acompanhavam o que estava acontecendo no momento, apresentando gírias modernas, problemas sociais e questões do mundo.

AS HISTÓRIAS EM QUADRINHOS (HQs) NA PERSPECTIVA DA ANÁLISE DO DISCURSO (AD)

As histórias em quadrinhos (HQs) estão incorporadas ao campo da literatura popular executando uma leitura do cotidiano, pois devido ao seu formato simplificado e dinâmico estão ao alcance da compreensão de um grande número de leitores. Este fato só é possível devido às características fundamentais dos quadrinhos, que trabalham o discurso imagético por meio de uma sucessão de imagens, da sequência dos personagens e do diálogo incluso no quadro, por meio do formato próprio. Desta forma, os quadrinhos instauram um método de narrativa fluída, baseada nos desenhos, que são formas universais de linguagem e, por isso, vêm ganhando espaço nas universidades e nas escolas de ensino fundamental e médio.

As temáticas discutidas nas HQs e, principalmente nas histórias sobre *Thor* abordadas neste artigo, são bastante relevantes para estudos em Análise do Discurso (AD), pois a materialidade analisada, recortada de histórias em quadrinhos, deve ser considerada não apenas pelo seu aspecto cultural, mas principalmente porque veicula ideologias e sentidos influenciando os modos de pensar da sociedade. Essas discursividades refletem valores presentes na sociedade, instaurando modelos a serem seguidos pelos sujeitos que compram as HQs. Daí a importância de serem elucidadas.

Thor nasce como um herói humanizado pelo seu criador, Stan Lee, proveniente de uma cultura dos povos escandinavos, historicamente hostil, com características do bom cidadão americano, que combate os crimes em Nova Iorque e defende os EUA de todas as ameaças. Paradoxalmente, Thor em sua origem nórdica não tem nada de humano, pelo contrário, é um deus que busca suas próprias glórias, pois quanto mais ele vence batalhas e desafios mais forte ele fica. É esse sujeito projetado parcialmente frágil, em parte heroico e também transgressivo que se transforma em sujeito neste artigo, na perspectiva da Análise do Discurso (AD), por meio da ideologia que atravessa a sociedade americana e que o interpela a ser sujeito, ou seja, o chama à existência e a enunciar discursos, identificado a uma formação ideológica (FI).

A MEMÓRIA DISCURSIVA NOS QUADRINHOS SOBRE THOR. O INTERDISCURSO E O DESLIZAMENTO DE SENTIDOS

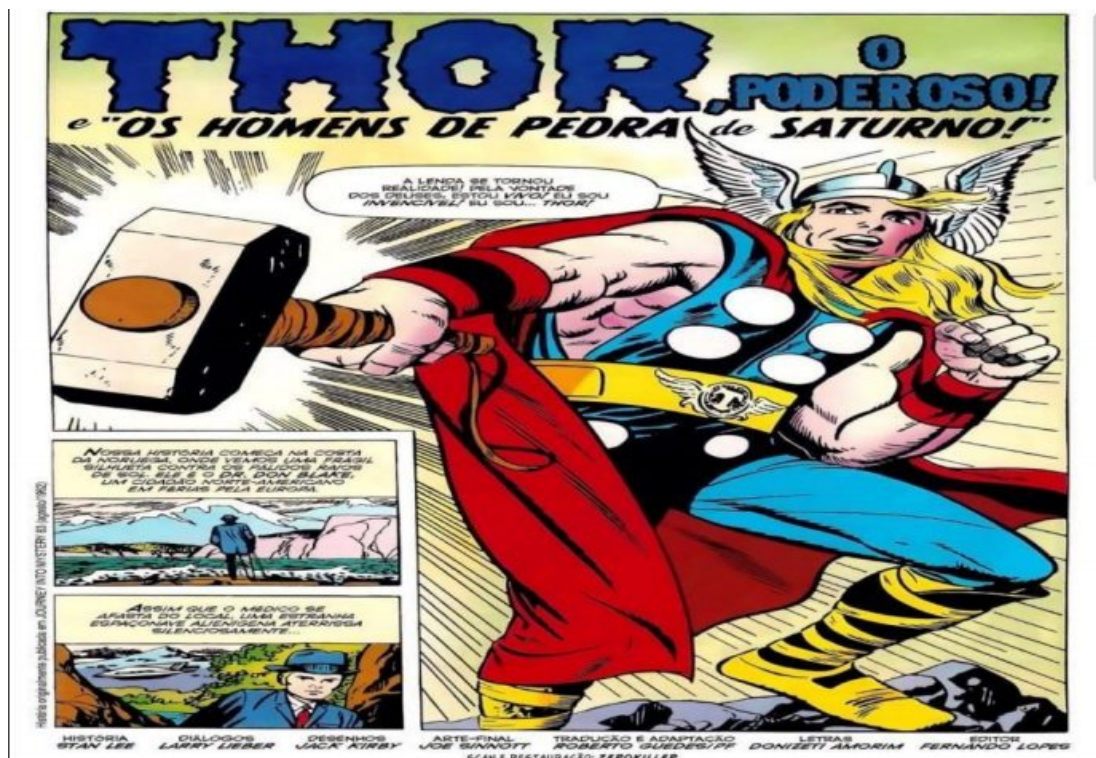
As noções da AD foram aperfeiçoadas durante toda a trajetória teórica de Michel Pêcheux, seu criador. O conceito de memória discursiva, uma das noções teóricas chave da AD, também decorre de evoluções na teoria. No artigo intitulado “Papel de Memória”, do ano de 1983, Pêcheux trata da noção mencionando que a memória discursiva “[...] seria aquilo que, face a um texto que surge como acontecimento a ler, vem restabelecer os implícitos ([...] os pré-construídos [...] discursos transversos) de que sua leitura necessita” (Pêcheux, 2010, p. 52). No mesmo texto, Pêcheux alerta, entretanto, que apesar dessa rede de implícitos parecer que irrompe de forma estável e sedimentada no acontecimento, na perspectiva de uma repetição ou regularização (sob forma de interdiscurso, paráfrase, etc), é sempre suscetível de ruir ante o peso de um acontecimento discursivo novo que vem perturbar a aparente regularização de sentidos. Segundo o teórico:

A memória tende a observar o acontecimento, como uma série matemática prolonga-se conjecturando o termo seguinte em vista do começo da série, mas o acontecimento discursivo, provocando interrupções, pode desmanchar essa “regularização” e produzir retrospectivamente uma outra série sob a primeira, desmarcar o aparecimento de uma nova série que não estava constituída enquanto tal e que é assim o produto do acontecimento; o acontecimento, no caso, desloca e desregula os implícitos associados ao sistema de regularização anterior. (Pêcheux, 2010, p. 49).

Assim, de acordo com Pêcheux, a memória discursiva possui um duplo viés: de um lado, ela se apresenta como um espaço de repetição de sentidos (já-ditos, implícitos) e, de outro, é um espaço de deslizamentos de sentidos, que irrompem na perspectiva de um acontecimento novo que vem perturbar a aparente estabilidade (Pêcheux, 2010, p. 52-3).

Considerando a memória discursiva como espaço de repetição e deslizamento de sentidos passaremos a observar a figura 1, a seguir:

Figura 1: A lenda Thor



Fonte: [http://www.guiadosquadrinhos.com/personagem/Thor-\(Thor-odinson-\(donald-blake-jake-olson\)\)](http://www.guiadosquadrinhos.com/personagem/Thor-(Thor-odinson-(donald-blake-jake-olson)))

Ao analisar as condições de produção do aparecimento de Thor pode-se identificar um viés da memória discursiva que, de acordo com o Pêcheux, coloca em jogo uma imagem que evoca um discurso enunciado em outro lugar (Pêcheux, 2010, p. 51-2).

A memória discursiva, que atravessa as discursividades da figura 1, acima, remete ao Dr. Blake que, durante a sua transformação, de médico americano frágil em Thor, consegue rememorar o seu passado de deus poderoso das terras nórdicas, quando era temido por sua força e arrogância.

Neste aspecto, pode-se identificar na figura 1 a memória discursiva em seu aspecto de “repetição e de reconhecimento que faz da imagem como que a recitação de um mito” (Pêcheux, 2010, p. 51). O viés da memória discursiva mobilizado na figura 1 é o interdiscurso em seu aspecto de “remissões, de retomadas e de efeitos de paráfrase” (Pêcheux, 2010, p. 51) que mantém a regularização do discurso enunciado em outras condições de produção. Nesse aspecto, o aparecimento dele nas HQ nos leva a pensar que o domínio associado de memória se refere às associações que um enunciado estabelece com outros enunciados no interior de uma memória. Na análise, esse domínio fornece tudo o que se relaciona com aquilo que o sujeito diz.

Na figura 2, a seguir, pode-se observar mais um conjunto de enunciados que mobiliza a memória discursiva em seu aspecto de remissões e retomadas, característica do interdiscurso.

Figura 2. Transformação do *Thor* em deus do trovão¹



Fonte: [http://www.guiadosquadrinhos.com/personagem/Thor-\(Thor-odinon-\(donald-blake-jake-olson\)\)](http://www.guiadosquadrinhos.com/personagem/Thor-(Thor-odinon-(donald-blake-jake-olson)))

Pode-se afirmar que a figura 2, acima, representa o ápice da HQ. É o momento em que o *Dr. Donald Blake* se transforma em *Thor*. No momento de sua transformação em deus do trovão, automaticamente, retoma a memória e se lembra de quem ele realmente é. Ao empunhar o martelo, diz: “*Thor!* O lendário Deus do Trovão! O mais poderoso guerreiro da mitologia! Este é o martelo dele! E eu... eu sou *Thor!*” (Biblioteca Histórica Marvel, 2008, p. 17).

³Momento da história em que o Dr. Donald Blake se transforma no poderoso *Thor*. E consegue se lembrar de quem ele realmente é. Imagem disponível em:

<[http://www.guiadosquadrinhos.com/personagem/Thor-\(Thor-odinon-\(donald-blake-jake-olson\)\)](http://www.guiadosquadrinhos.com/personagem/Thor-(Thor-odinon-(donald-blake-jake-olson)))/14>. Acessado em 20 de janeiro de 2021.

Os enunciados da figura 2 - “Thor! O lendário deus do trovão” e “O mais poderoso guerreiro da mitologia” e “Este é o martelo dele!” – também mobilizam o interdiscurso, em seu aspecto de remissão, de retomada de um já-dito produzido em outras condições de produção históricas. Os itens lexicais “lendário”, “deus do trovão”, “mais poderoso” e “guerreiro da mitologia” instauram efeitos do sujeito como um ser mítico, lendário, fabuloso e heroico, cuja imagem retoma um passado mitológico. No enunciado “E eu... eu sou Thor!” as reticências marcam um estado de suspensão, de pausa, de lapso, um intervalo entre o esquecimento e a lembrança, instante em que o sujeito se recorda de quem realmente é.

É importante mencionar que não apenas nas discursividades 1 e 2, mas em muitos discursos imagéticos presentes nas HQs de Thor, o articulista, Stan Lee, retoma o passado de deus da mitologia nórdica associado ao sujeito Thor, projetando-o com características de grandeza, força física, orgulho, altivez, atrevimento, ousadia, imodéstia, impulsividade e arrogância. Em outras condições de produção, quando retrata o sujeito como o herói americano humanizado, que luta contra o Comunismo, Stan Lee o instaura com sentidos de salvador, defensor da humanidade, protetor dos fracos e oprimidos, apresentando características de humanização, sociabilidade e civilidade.

Deste modo, a memória discursiva, na construção identitária do sujeito Thor exibe o duplo aspecto teorizado por Pêcheux (2010, p. 52-3) conforme se pode observar a seguir:

Quadro 1. O duplo viés da memória discursiva nas HQs de Thor

MEMÓRIA DISCURSIVA (INTERDISCURSO E DESLIZAMENTOS DE SENTIDOS) NA CONSTRUÇÃO IDENTITÁRIA DO SUJEITO THOR	
THOR como herói mitológico: características de grandeza, força física, orgulho, altivez, atrevimento, ousadia, imodéstia, impulsividade e arrogância.	THOR como o herói humanizado americano: características de humanização, sociabilidade e civilidade. Thor combate o Comunismo e é projetado como salvador, defensor da humanidade, protetor dos fracos e oprimidos.

Fonte: O autor

O quadro anterior (quadro 1) mostra, portanto, que os sentidos deslizam nas representações identitárias do sujeito Thor.

Neste aspecto, pode-se afirmar que a maioria das HQs de Thor instaura uma memória discursiva, tanto de algo que já foi dito em determinado momento histórico, mas que deslizam

para novos sentidos em outras condições de produção. As representações de Thor, por Stan Lee, instauram, portanto, sentidos de repetição e de deslizamentos dependendo das condições históricas em que são produzidos.

Santos (2014) também nos lembra que os sentidos de um discurso se modificam de acordo com o acontecimento e as condições históricas. Segundo o autor:

Os sentidos produzidos no aparecimento de um enunciado fazem parte do acontecimento discursivo, de modo que entra no fio de discursos possibilitados pela história. Assim, a circulação é marcada pelo imbricamento entre linguagem e história e reafirma o discurso como estrutura e acontecimento, sendo que a inscrição dos enunciados na história expõe os diferentes discursos que circularam em determinado momento e em determinada materialidade. (Santos, 2014, p. 32).

Nesta perspectiva, os enunciados recortados nos discursos imagéticos analisados, do *Thor*, permitem-nos identificar determinados dizeres que fazem parte de uma rede de memória do personagem (interdiscurso), fazendo assim uma relação com algo que já foi dito. *Thor* carrega consigo uma existência passada, que não sabe ao certo como era. As primeiras revistas deste personagem foram apenas um esboço de quem realmente ele seria e para que veio ao mundo dos quadrinhos. De fato, Stan Lee queria um herói que tivesse características da mitologia nórdica.

Nos quadrinhos, os discursos articulados por *Thor* instauram sentidos de uma personalidade múltipla. Quando está fora de perigo é apenas o *Dr. Donald*, quando está em apuros ou precisa salvar a vida de alguém ele se transforma em *Thor*. Na representação do sujeito, como herói, essa imagem também pode ser projetada como o herói transgressor da mitologia nórdica e como o herói humanizado americano. A construção identitária do sujeito, portanto, é instaurada de maneira complexa por Stan Lee.

Figura 3. Ao assumir a identidade de *Thor*, o personagem passa a redescobrir seus poderes.



Fonte: [http://www.guiadosquadrinhos.com/personagem/Thor-\(Thor-odinon-donald-blake-jake-olson\)](http://www.guiadosquadrinhos.com/personagem/Thor-(Thor-odinon-donald-blake-jake-olson))

Analisando a figura 3 observamos que o sujeito rememora que o cajado que encontrou é mágico e que, ao batê-lo no chão, o sujeito se transforma no poderoso Thor e o cajado no martelo *mijölnir*, que só *Thor* pode empunhar. O enunciado “De acordo com a lenda”, presente no discurso imagético 3, anterior, instaura efeitos de sentido de passado, de crença ou mito popular resgatado de condições de produção remotas. Isso se dá pelo fato do sujeito ter escutado histórias sobre a mitologia.

Figura 4. Dr. Donald Blake assumindo a forma de Thor.



Fonte: [http://www.guiadosquadrinhos.com/personagem/Thor-\(Thor-odinson-donald-blake-jake-olson\)](http://www.guiadosquadrinhos.com/personagem/Thor-(Thor-odinson-donald-blake-jake-olson))

Nas duas imagens acima podemos observar a transformação do personagem. Na primeira figura o cajado está ao lado dele, no chão e ele está preso em uma cela, correndo perigo de ser fuzilado por comunistas. Quando empunha o cajado, ele recebe os poderes de Thor. Podemos observá-lo enunciando o *Dr. Donald Blake*, que ele irá assumir a personalidade de *Thor* e, dali para frente, não será mais humilhado, pois os sujeitos que fizeram isso com ele irão sentir sua fúria. Este discurso é atravessado pelo sentido de relação de forças, sendo as principais características dessa formação imaginária a imposição pela força de sentidos que irrompem pelo viés da memória discursiva do sujeito.

Orlandi (2015) observa que o mecanismo das formações imaginárias compreende a antecipação, as relações de força e de sentido.

Por relações de sentido entende-se que todo discurso se relaciona com outros e os sentidos do discurso derivam dessa relação. No que tange à antecipação, temos a figura de um sujeito que “[...] antecipa-se assim a seu interlocutor quanto ao sentido que suas palavras produzem” (Orlandi, 2015, p.39). Neste aspecto, a argumentação é alocada ao discurso de acordo com o efeito que se propõe instaurar, baseando-se na compreensão de seu ouvinte. Quanto à relação de forças, deriva da idéia de que o sentido do discurso é instaurado a partir do lugar no qual o sujeito que enuncia está inserido. A autora dispõe essas três perspectivas como “formações imaginárias”, validando assim as projeções discursivas provenientes dos sujeitos e dos locais de fala destes.

Na imagem 3 temos a transformação completa do deus do trovão, na qual ele deixa de segurar um cajado frágil e passa a segurar o *mjolnir*, a arma mais poderosa da mitologia nórdica. Segundo os contos nórdicos, apenas *Thor* teria a força para empunhar este martelo. Quando o *Dr. Donald Blake* cede lugar a *Thor*, a transformação aciona a memória discursiva do sujeito que emerge pelo viés do interdiscurso, mobilizando os sentidos do deus nórdico, ou seja, daquele que pune e que age pela força. Nesse aspecto, a memória discursiva se instaura na conduta deste deus, que enuncia a partir do lugar de um deus pagão e não da fragilidade instaurada pela figura do *Dr. Blake*.

Pêcheux, (2002, p. 53) escreve que “Todo enunciado, toda sequência de enunciados é, pois, linguisticamente descritível como uma série (léxico-sintaticamente determinada) de pontos de deriva possíveis, oferecendo lugar à interpretação”. O teórico ensina, ainda, que “*as coisas a saber* [...] não são jamais visíveis [...] mas sempre tomadas em redes de memória dando lugar a filiações identificadoras [...]” (Pêcheux, 2002, p. 54). Assim sendo, a interpretação desses enunciados e discursos imagéticos passa pelo domínio de memória e por uma rede de filiações que mostram os movimentos dos sentidos.

Deste modo, todo enunciado retoma outro - seja para repeti-lo, refutá-lo e transformá-lo - ou desliza para sentidos diferentes em condições de produção distintas ou mediante o aparecimento de um novo acontecimento discursivo. Nessa rede, quando o enunciado é repetido por determinada posição-sujeito em uma conjuntura histórica diferente estabelece-se a atualização, ou seja, os sentidos deslizam e instauram-se sentidos novos.

CONSIDERAÇÕES FINAIS

Este artigo analisou um *corpus* constituído por alguns recortes da *HQ, Biblioteca histórica Marvel, O Poderoso Thor Volume 1* com o objetivo de verificar regularidades que evidenciassem aspectos da memória discursiva atravessando as discursividades.

No que se refere às condições de produção em torno do aparecimento de Thor nas histórias em quadrinhos (HQs) foram expostos brevemente os processos históricos e ideológicos em torno de sua criação nos anos da Guerra Fria, acontecimento geopolítico que deixou o mundo em estado de tensão, uma vez que os países mentores (EUA e URSS) detinham a tecnologia da bomba atômica, podendo destruir o mundo caso entrassem em guerra.

As análises dos discursos imagéticos recortados apontaram o funcionamento da memória discursiva em seu duplo viés, irrompendo em condições de produção diferentes. Deste modo, é possível observar a memória discursiva em seu aspecto de “repetição e de reconhecimento que faz da imagem como que a recitação de um mito” (Pêcheux, 2010, p. 51) em seu aspecto de “remissões, de retomadas e de efeitos de paráfrase” (Pêcheux, 2010, p. 51) quando o *Dr. Donald Blake* cede lugar a *Thor*, pois a transformação aciona a memória discursiva do sujeito que emerge pelo viés do interdiscurso, mobilizando os sentidos do deus nórdico, ou seja, daquele que pune e que age pela força. Nesse aspecto, a memória discursiva se instaura na conduta desse deus, que enuncia a partir do lugar de um deus pagão e não da fragilidade instaurada pela figura do *Dr. Blake*.

Em outras condições de produção, quando projeta a imagem do sujeito Thor como o herói americano humanizado, que luta contra o Comunismo, Stan Lee o instaura com sentidos de salvador, defensor da humanidade, protetor dos fracos e oprimidos, apresentando características de humanização, sociabilidade e civilidade. Ou seja, os efeitos de sentido deslizam tornando-se outros.

Deste modo, a memória discursiva, na construção identitária do sujeito Thor exibe o duplo aspecto teorizado por Pêcheux (2010, p. 52-3): de um lado, há um espaço de repetições, retomadas e paráfrases, mantidas graças à força do interdiscurso e, de outro, há

o deslizamento de sentidos, em condições de produção diferentes e mediante um acontecimento novo.

Trata-se de uma construção identitária bastante complexa, que instaura Thor como um sujeito dividido, heterogêneo e descentrado. Essa identificação é característica dos sujeitos modernos clivados, divididos entre consciente e inconsciente, que não apresentam unidade como se pensava no Iluminismo, mas que mantêm a ilusão subjetiva de ser o centro de si mesmos e daquilo que dizem.

REFERÊNCIAS

- Biblioteca Histórica Marvel (2008). O Poderoso Thor - Volume I. Panini Comics,
- Hobsbawm. E. J. E. (1995). Era dos Extremos: o breve século XX: 1914-1991. Trad. Marcos Santarrita. São Paulo: Companhia das Letras.
- Orlandi, E. P. (2015). Análise de Discurso: princípios e procedimentos. 12. ed. Campinas, SP: Pontes Editores,
- Pêcheux, M. (2010). Papel da memória. In.: ACHARD, P. (org.), Papel da memória. Trad.
- Nunes, J.H., 3ª. ed., Campinas - SP: Pontes Editores.
- Orlandi, E. P. (2002). O discurso: estrutura ou acontecimento. Trad. ORLANDI, E. P. 3ª ed., Campinas, SP: Pontes.
- Santos, J. J. (2014). Produções discursivas no horror: Materialidade fílmica e memória na trilogia de Zé do Caixão. São Paulo: UNESP, Disponível em: <<https://repositorio.unesp.br/handle/11449/123253>>. Acesso em 1º de junho de 2020. DOI: <http://doi.org/10.17616/R31NJN>

Artículo original

Educación bilingüe hacia la dignidad: Implicaciones para países polilingües como EE. UU. y Paraguay

Bilingual Education Towards Dignity: Implications for Multilingual Countries like the US and Paraguay

Christian Faltis

Texas A&M International University, EE.UU.
<https://orcid.org/0000-0002-1486-231X>

e-mail: christian.faltis@gmail.com

Recibido: 11/5/2023
Aprobado: 16/8/2023

RESUMEN

El artículo aborda la educación bilingüe en los Estados Unidos, destacando cómo la política lingüística ha subestimado la riqueza cultural y lingüística de las comunidades bilingües, asociándolas con ideologías de lenguaje puro y jerarquías sociales. Según Luis Poza, se requiere un enfoque que promueva la dignidad y el empoderamiento de estudiantes y comunidades minoritarias. Esto implica un compromiso colectivo por la justicia educativa y la vigilancia constante de las políticas lingüísticas. Se propone una pedagogía crítica que desafíe las prácticas que devalúan los idiomas de las poblaciones colonizadas. El artículo ofrece una breve historia de la educación bilingüe en los Estados Unidos, resaltando la importancia de cuestionar y resistir las políticas gubernamentales y las leyes relacionadas con la educación bilingüe en las escuelas públicas. Luego, establece conexiones entre la situación en los Estados Unidos y Paraguay, y concluye con sugerencias para mejorar la educación bilingüe, especialmente para aquellos cuya lengua y cultura minoritaria se ve marginada. En resumen, el artículo aboga por un enfoque que celebre la diversidad lingüística y cultural, promoviendo la dignidad y la justicia en la educación bilingüe.

Palabras clave: educación bilingüe; política lingüística; dignidad; justicia educativa.

ABSTRACT

The article addresses bilingual education in the United States, highlighting how language policy has underestimated the cultural and linguistic richness of bilingual communities, associating them with ideologies of pure language and social hierarchies. According to Luis Poza, an approach that promotes the dignity and empowerment of minority students and communities is required. This entails a collective commitment to educational justice and constant vigilance of language policies. A critical pedagogy is proposed to challenge practices that devalue the languages of colonized populations. The article provides a brief history of bilingual education in the United States, emphasizing the importance of questioning and resisting governmental policies and laws related to bilingual education in public schools. It then draws connections between the situation in the United States and Paraguay, concluding with suggestions to enhance bilingual education, particularly for those whose minority language and culture are marginalized. In summary, the article advocates for an approach that celebrates linguistic and cultural diversity, promoting dignity and justice in bilingual education.

Keywords: bilingual education; language policy; dignity; educational justice.

PRESENTACIÓN

Al hablar de la educación bilingüe, para en los EEUU, es transcendental decir que la política de lenguaje creado por académicos de la clase dominante siempre ha ignorado la riqueza tanto lingüística como cultural de las comunidades bilingües, que se han retratado en términos negativos aliñados con creencias basadas en ideologías de lenguaje pura y perfecta que insisten en definir el bilingüismo de comunidades minoritarias de tener una deficiencia en su uso popular porque éste para las clases dominantes está relacionado con jerarquías de clase y de raza y etnicidad. Según el trabajo de Luis Poza (2021), lo que nos hace falta en la educación bilingüe es la **afirmación completa hacia la dignidad** y actualización humana de los estudiantes y comunidades minorizados por grupos e individuos dominantes. Hacia la dignidad educativa en la educación bilingüe requiere que todos seamos compañeros en la lucha por la justicia (Valdés, 2021). Para ser compañero en la lucha hacia la dignidad y por justicia educacional es sumamente importante que siempre estemos vigilando las políticas de lenguaje y apoyando a los estudiantes que usan dos lenguas o que llegarán a ser bilingües en sus vidas cotidianas. Además, hay que proponer pedagógicas críticas en las cuales se problematiza las ideas que vienen de ideologías y prácticas dominantes que siempre van a menospreciar el idioma y el bilingüismo de las poblaciones que han sido colonizados.

En lo que sigue, presento una breve historia de educación bilingüe en los EE. UU. para que tengan una idea de porqué es necesario tener dudas y aún resistir las políticas gubernamentales y las leyes sobre las variedades de educación bilingües que se permiten usar en las escuelas y colegios públicos. Después, elaboro y enlazo lo que ha sucedido en los EE. UU. a lo que está pasando en Paraguay, y termino con unas sugerencias para mejorar la educación bilingüe para estudiantes minoritizadas por su lengua de casa y de identidad.

LA EDUCACIÓN BILINGÜE EN LOS ESTADOS UNIDOS

La educación bilingüe apoyado por el gobierno federal y estatal tiene más de 50 años, para precisar, la formación de educación bilingüe se inició en 1968, como el acta de la educación bilingüe, que permitió que las escuelas públicas crearan programas en dónde se podría usar el español como mediante de enseñanza para estudiantes que usan un lenguaje que no sea inglés en su casa y que según un examen de dominio de inglés (oral y escrita) tuvieran un nivel de inglés designado por un examen de proficiencia como “limitado” o “sin proficiencia ninguna” en el inglés. Estas designaciones privilegian a la lengua dominante y estándar, desvalorizan su capacidad en comparación con normas monolingüistas (Valdés y Georffrion-Vinci, 1998). Además las designaciones reclaman que los de no habla inglés no saben lo que se llama “lengua académica” en su propia lengua, sin clarificar lo que es lengua académica (Faltis, 2013; Valdés, 2004), y borran completamente la idea de que estos estudiantes ya dominan oralmente muy bien su lengua de casa. Para ellos quienes mantienen estas ideas basadas en ideologías de una lengua perfecta y académica, ser bilingüe y usar una lengua imperfectamente y/o una lengua no marcada como académica significa que la capacidad lingüística (y por extensión la capacidad intelectual) es deficiente (Faltis, 2022).

“NO SABEN HABLAR ACADÉMICAMENTE”

Hoy en día muchos en la educación bilingüe suelen usar la designación “bilingüe emergente” en lugar de “estudiante con inglés limitado.” Bilingüe emergente por lo menos reconoce con los estudiantes son bilingües. Sin embargo, para algunos, el concepto de emergente es problemático, porque ¿cuándo llegarán a ser bilingües? El problema se mantiene porque siguen usando exámenes de proficiencia contruidos con la norma de dos monolingües y con lenguaje orientado a usos en contextos académicos, en donde hay que contestar en oraciones completas. O sea, los exámenes buscan medir lo que un monolingüe que sabe sobre

el uso de lenguaje estandarizado para contextos académicos en donde la lengua escrita y la lengua oral se usan para la enseñanza. No se puede comparar un estudiante bilingüe con un monolingüe en un examen de proficiencia y un examen creado por monolingües no puede medir la segunda lengua de un bilingüe (Valdés & Figueroa (1994). El resultado de usar exámenes de proficiencia de inglés para estudiantes bilingües es una gran discriminación. Muchos de estos estudiantes bilingües enfrentarán bajas expectativas en sus clases y muy pocas oportunidades para desarrollar su bilingüismo y su alfabetización en su lengua de casa.

Como mencioné arriba, en los EE. UU. tenemos un gran problema en contextos escolares de que muchos maestras/os y educadores tanto monolingües como bilingües aceptan sin problematizar la idea de “lengua académica” y aceptan erróneamente que la mayoría de los estudiantes que entran la escuela primaria sin hablar inglés tienen muy poca experiencia con “lengua académica.” Se basan sus creencias en lo que promovió el psicólogo educativo, Jim Cummins (1979; 1981; 2000) quien, hace más de 40 años, inventó la distinción entre Basic Interpersonal Communicative Skills (BICS) (destrezas comunicativas interpersonales básicas) y Cognitive Academic Language Proficiency (CALP) (dominio de lengua académica y cognitiva). BICS implica la capacidad de comunicarse socialmente sin usar vocabulario especial y sin la capacidad de elaborar pensamientos e ideas que se encuentran en textos escolares y que usa la gente educada con lengua estandarizada. Se supone que los estudiantes hispanoparlantes que son principiantes de verdad o saben muy poco inglés al entrar a la escuela también sufren de un déficit intelectual, por no tener experiencia con lengua académica. Al contrario, CALP implica que la “lengua académica” contiene algo cognitivo, más allá de lo social o sea, una capacidad intelectual que no se encuentra en BICS, pero que sí se nota en bilingües élites que hayan tenido la oportunidad de formarse en un idioma, y luego estudiar en otro idioma o sea un bilingüismo con alfabetismo en dos lenguas que refleja dos monolingües bien educados en una sola persona.

El bilingüismo élite es poco común, siendo que menos de 5 por ciento de los bilingües y biliterados en el mundo tienen los recursos y la manera de asistir a un colegio durante la primaria y la secundaria en un idioma, y luego asistir a la preparatoria y la universidad en otra lengua. La mayoría de los bilingües no son la suma de dos lenguas distintas, pero sí tienen una gran capacidad de moverse dinámicamente a través de múltiples contextos (Flores & Schissel, 2014; García Y Kleifman, 2010).

En los EE. UU., los niños que lleguen a ser bilingües bajo la circunstancia (Valdés, 1992) de haberse inscrito en un programa bilingüe o un programa de inglés como segunda lengua usarán sus dos lenguas oralmente en una manera que hoy en día se designa como *translanguar* (García y Li Wei, 2014; MacSwan, 2022; MacSwan & Faltis, 2019), en donde las dos lenguas provienen de un sistema fluida y dinámicamente y así usar su bilingüismo sin poner límites entre las dos lenguas. La mayoría de bilingües en el mundo translanguan, cambiando de una lengua a la otra cuando hablan con otros bilingües y polilingües de las mismas lenguas en varios contextos sociales. La pregunta que sigue es: ¿Por qué el bilingüismo de los estudiantes en los EE. UU. que hablan un idioma que no sea inglés al inscribirse en programas bilingües lleguen a ser bilingües que aprenden a translanguar y que tienen muy poca experiencia con leer y escribir en su lengua de casa? La respuesta es sencilla y a la vez complicada.

La mayoría de los programas bilingües en los EE. UU. son basados en la idea de “dejar de usar español cuanto antes” o sea educación bilingüe transicional en donde los estudiantes tienen acceso a maestras bilingües por 2 años o menos. Luego, recibirán clases de inglés como segunda lengua, muchas veces usando programas de lenguaje en computadores o tabletas, y colocados en clases con maestras que tienen poco conocimiento de las teorías de lengua o de pedagogía especialmente para estudiantes bilingües. Lo malo de esto es que los estudiantes bilingües en la mayoría de casos están separados para sus clases de inglés, y cuando se

encuentran en clases regulares, sus maestras no tienen suficiente conocimiento de las prácticas buenas para que estos estudiantes puedan participar y beneficiar plenamente en las actividades de la clase, y no hay apoyo ni oral ni escrito para sus lengua de casa.

También en varios estados, como California, Tejas, y Utah hay programas de lenguaje dual, o sea, programas en donde hay 2 maestras o dos tiempos o alternación de días para el español y el inglés, modelado en la idea de que es importante no mezclar las dos lenguas y así respetar una ideología monolingüe y preferida por la clase en control político que insiste en que los estudiantes aprendan cada lengua bajo contextos y condiciones de lenguajes separados, sin mezclar o translenguar dinámicamente los dos lenguas. Además, en estos programas de lengua dual, la mayoría de las materias y literatura apoyan al inglés y los monolingües blancos y dominantes (vea Przymus & Lindo, 2021; Valdés, 2020)

Más de 60 millones de personas de los EE. UU. hablan cotidianamente un idioma que no sea inglés. Al contraste, unos 230 millones hablan inglés en sus casas y no saben comunicarse en un otra lengua. Respectivamente, tanto de los políticos como muchos de los inglés parlantes creen en la ideología de **un país, un lenguaje**, lo que se ha convertido en menos apoyo para la educación bilingüe y más atención al inglés para los niños o sea aprender el inglés lo más pronto posible. Sin embargo, entre los estadounidenses que desean aprender una segunda lengua, hay familias blancas que registran sus niños en programas de lenguaje dual, para que sus niños obtengan más capital en un mundo globalizado, y así usan a los estudiantes que hablan el otro idioma para apoyar su mejoramiento en la otra lengua. Hablando de “la otra lengua”, la mayoría de los programas bilingües y duales, menosprecian a las variedades de español u otro lenguaje que se encuentra en los EE. UU. (Dorner, Cervantes-Soon, Heiman, & Palmer, 2021). En vez de valorar su lengua y su bilingüismo, usan **el español de otro lugar**, (Mena, 2022) creyendo que el español que usamos en los EE. UU. es deficiente, y lleno de pochismos, y mezclado con el inglés, no es un español bueno, académico, etc., etc. Menosprecian el español de los bilingües, sin tomar en cuenta que las políticas e ideologías de una lengua perfecta son responsables por sus creencias equivocadas.

LA FALTA DE DOCENCIA BILINGÜE Y ALFABETIZACIÓN EN ESPAÑOL

Otra gran problema que tenemos en los EE. UU., es la escasez de programas de docencia en educación bilingüe y la alfabetización en español. Es cierto que cuando empezó el Acta de Educación Bilingüe en 1968, o Título VII del Acto de Primario y Secundario, había becas para los que estudiaran educación bilingüe al nivel de doctorado. Sin embargo, había muy poco apoyo para el mejoramiento de alfabetización en el español. Y en ese tiempo, no existía una diferencia de enseñanza del español para los que ya eran bien bilingües, o sea los que hablaban bien las dos lenguas. En muchos de los programas de docencia para maestros bilingües, todas las clases les daban en inglés, y no había recursos en español y materiales que representaban culturalmente bien a las familias México Americanos, Hispanos, y Latinos que vivían en los EE. UU. La gran mayoría de departamentos de español en universidades no reconocen la presencia de estudiantes bilingües de origen Hispánica. El español se enseña como una lengua extranjera, con un currículo basado en vocabulario, gramática, cultura, composición, literatura y civilización peninsular. (Valdés, 1997). Sabían muy poco sobre la enseñanza de español para los estudiantes bilingües o lo que hoy en día se llaman estudiantes de herencia (Parra & Faltis, in press; Valdés, 2005; 2017). Al principio, las clases de español para estudiantes de herencia tenía las siguientes metas diferentes y a veces contradictorias (Valdés, Fishman, Chávez & Pérez, 2008):

1. Mantener el uso del español durante toda la vida;
2. Expandir el rango de bilingüismo;
3. Aplicar las prácticas de alfabetización aprendidas en una lengua para la otra;

4. Desarrollar prácticas generalmente académicas; y
5. Crecer la comunidad bilingüe y alfabetizada y desarrollar el autoestima.

Actualmente, las metas principales en la enseñanza de español a los estudiantes de herencia tienen agendas más amplias, como las necesidades y capacidades de los estudiantes hacía la comunicación oral y escrito; reconocer la superdotación de ser bilingüe; promover el orgullo y agencia hacia un cambio social, y estar consciente clara y críticamente de las ideologías de lenguaje, raza, y clase social que menosprecian el hablar de los bilingües (Parra, 2016; Parra, 2020) y que figuran en las conceptualizaciones de lenguaje y la teorías de adquisición de una segunda lengua y un dialecto segundo (Valdés, 2017). Las ideologías de lenguaje juntas con políticas del estado, en términos de apoyar o no a la enseñanza de lenguas en escuelas de primaria, secundaria, superiores, y en programas de docencia, determinan la curricularización (Valdés, 2015; 2017), o sea, el currículo, las materiales, enfoques en la enseñanza, la evaluaciones requeridas, y las características de los maestro/as, especialmente en programas que dependen del estado, los que deciden todo el proceso de la curricularización. Hasta la fecha, en los EE. UU., no hemos logrado a convencer ni a la nación o al estado que la educación y la alfabetización bilingües para estudiantes de herencia merecen metas completamente diferentes a las que hay por lo general en las clases bilingües y en los departamentos de español.

LA EDUCACIÓN BILINGÜE EN PARAGUAY ¿HAY PARALELAS?

Según lo que he aprendido de la situación del bilingüismo y la alfabetización bilingüe para guaraní hablantes refleja mucho a lo que ha pasado en los EE. UU. en términos del apoyo o sea, falta de apoyo en educar docentes, promover libros para el estudio de varias disciplinas en lenguas que no sea los dominantes del país, y exámenes de alfabetización y proficiencia bilingüe. Claro que hay diferencias entre la dos naciones, pero en las dos naciones los grupos minoritarios sufren mucho por no tener una educación bilingüe de alta calidad y que promueve alfabetización bilingüe. A la misma vez, hay importantes diferencias en términos del número y la origen de los hablantes la lengua minorizada. En Los EE. UU., la mayoría de hispanoparlantes viven en los estados de California, Arizona, Nuevo México, Texas, Illinois, y Nueva York y Florida. La mayoría de los hispanoparlantes del suroeste y Illinois son de origen México, mientras muchos de Nueva York son de origen Puertorriqueño, y los de la Florida han inmigrado de Cuba, Puerto Rico, y Venezuela. El grupo hispano que tiene más residencia en lo que ahora son los EE. UU. son los México Americanos, porque el suroeste era parte de México hasta la Guerra Civil entre EE. UU. y México (1846-1848).

Al contrario, la lengua que ha coexistido con el español (castellano) en Paraguay es guaraní en nombre, para la verdad es que hay muchos guaraníes. Según Riveros Zalaya (2020), casi 8 % de la población Paraguayo habla una versión de guaraní según su región; menos de 4% de la población habla solamente castellano, y casi 65% de los habitantes de Paraguay son bilingües que usan ambas lenguas, pero dominantes en guaraní (Nickson, 2009). Como sucede en todos los lugares donde hay contacto entre hablantes de diferentes lenguas con el tiempo siempre hay una mezcla entre las lenguas. En los EE. UU., hay muchos que lo llaman el Spanglish, y en Paraguay, tiene la etiqueta de yopará o jopara o sea guaraní paraguayo, “una variedad rotundamente distinta de la que se practicaba en las Misiones (Blestel, 2021, p. 74) y los élites guaraní hablantes. Jopara, como el español de los México Americanos y otros grupos hispanos y bilingües se ve con una evaluación muy negativa o sea, como un déficit lingüístico. En los dos casos, el Spanglish y jopara se notan al nivel léxico, en donde se mezclan afijos y otras estructuras léxicas (Blestel, 2021; Riveros Zelaya, 2020). También comparten hablantes bilingües de los dos países una manera dinámica y vibrante de hablar con otros bilingües que

se nombra codeswitching o sea alternación de lenguas. Hoy en día en los EE. UU. se ha nacido el concepto de translenguar (translanguaging), que incluye codeswitching, pero es mucho más. El acto sociolingüístico de translenguar no es aceptado por muchos de los educadores y políticos quienes conceptualizan español y guaraní como lenguas puras y perfectas, lenguas que jamás se deben mezclarse con otras lenguas o que se debe de o hablar en una completamente o en la otra completamente.

En la historia de Paraguay desde su independencia de España en 1811, la lengua guaraní ha tenido poco de prestigio. Bajo el gobierno de Carlos Antonio López (1840-1962 y su hijo Francisco Solano López (1862-1870) se prohibió el uso de guaraní en las pocas escuelas que existían en esa época (Nickson, 2009). A fines del siglo, los élites de Paraguay continuaron la prohibición del uso de guaraní en contextos escolares y de negocios. En los 1920s, había un cambio ideológico sobre el valor de guaraní debido a un nuevo interés nacionalista que deseaba que guaraní formara parte de la identidad nacionalista de Paraguay que duró varias décadas. Este nuevo interés cambió la actitud general de los políticos hacia el valor y la enseñanza de guaraní en todo el país. Hasta muchos en el círculo alto de poder en Paraguay exaltaron el guaraní. Y es más, como Rona (1966, p. 286) argumenta, Paraguay es un país de hablantes guaraní, en donde solamente en los altos niveles de administración, educación, y comercio por mayor, se usa el castellano por necesidad.

En la Constitución de 1967 se reconoció a guaraní como lengua nacional, mientras que el castellano siguió como la lengua de todo relacionado oficialmente con el gobierno y para los negocios internacionales. Hasta se utilizó el guaraní como un símbolo de unidad nacional, pero en realidad, la larga historia de discriminación y la creencia que quienes usan guaraní representan un atraso.

En la década de los 80, el Ministerio de Educación promovió un programa de educación bilingüe en todo solamente se enseñaba guaraní oral, solamente para enseñar el español. No había enseñanza de lecto-escritura en guaraní. Fue un fracaso. En 1992, el Ley 28, dictó que la enseñanza de guaraní fuera obligatorio en todos los niveles del sistema educativo público. Dos años después, el plan de educación bilingüe de 25 años se inició para enseñar lectura y escritura en guaraní con castellano, por 15% en el primer grado y aumentando el uso de castellano un 5% cada año académico con la meta de que después de 9 años de escuela cada alumno llegara a ser bilingüe y biliterado. Sin embargo, el plan de PEB no funcionó como planeaban: en 2006, solamente 280 escuelas de las 1,400 que existían, seguían el PEB, mientras que la gran mayoría (como 95%) nada más enseñaban en castellano (Nickson, 2009).

Hay varias explicaciones porqué el PEB ha fallado en gran parte. Una de las razones importantes es porque no había suficientes maestros de guaraní, y el guaraní que enseñan en los programas de docencia no es el guaraní que usa los pueblos en la mayor parte de Paraguay (Elias, Rodas & Rodas, 2022; Ministerio de Educación y Ciencias, 2021). Hay muchos guaraníes y determinar cuál de las variaciones de guaraní sea la que más entiende y habla no es un tema menor. Y también se comunica muchos de los guaraníes usando las dos lenguas, o sea translenguan. Como menciona Demellenne, (2008, p. 149)

- El docente en el aula: enseña en guaraní, escribe en español, evalúa en español y se relaciona en guaraní y español.
- Los miembros de las familias: se comunican en guaraní, y se relacionan en la comunidad en guaraní y español. Los padres educan a sus hijos en español.

Al fin y al cabo, el castellano sobrevive, mientras que el guaraní lenta y verdaderamente se menosprecia como una lengua que merece mucho más atención al nivel nacional y dignidad hacia las poblaciones que comunican todos los días en guaraní.

CONCLUSIONES

En esta comparación entre lo que ha sucedido en los EEUU y el Paraguay en términos de los esfuerzos o carencia de esfuerzos que se han tomado en defensa del derecho a la educación bilingüe y la alfabetización bilingüe en las lenguas minoritarias revela que las acciones de los que tienen el poder en cada país dejan al margen del sistema educativo las políticas de lenguaje apoyados por las clases dominantes con su hegemonía de mantener la lengua de los colonizadores sin apoyar las lenguas de los minoridades lingüísticas que son los verdaderos dueños de la tierra.

Si queremos crear un sistema educativo hacia la dignidad de las comunidades indígenas en Paraguay y de las comunidades hispanas que viven en los EE. UU., es menester que sigamos la lucha hacia la dignidad y así luchar para la recuperación, la valoración y la potencia de tener fuertes programas de educación y alfabetización bilingües para los que hablan lenguas de sus raíces culturales y sus contribuciones diarios a la humanidad de todos.

REFERENCIAS

- Blestel, É. (2021) Entramados lingüísticos e ideológicos a prueba de las prácticas: Español y guaraní en Paraguay. Sánchez Moreano, S. and Blestel, É.(Eds). Prácticas lingüísticas heterogéneas: Nuevas perspectivas para el estudio del español en contacto con lenguas amerindias.(pp.69-86). Berlin: Language Science Press.
- Demellenne, D. (2008). Educación bilingüe en Paraguay como ejemplo de transformación de las prácticas de enseñanza en un contexto pluricultural. *Educación en foco*, 12 (2), 139-152.
- Dorner, L. M., Cervantes-Soon, C. G., Heiman, D., & Palmer, D. (2021). "Now it's all upper-class parents who are checking out schools": gentrification as coloniality in the enactment of two-way bilingual education policies. *Language Policy*, 20(3), 1-27.
- Elías, R. Rodas, S., & Rodas C. D. (2022). Reconocimiento, retrocesos y resistencias en defensa del derecho a la educación en lengua guaraní en Paraguay. *Revista Educación, Política y Sociedad*, 7(2), 77-102.
- Flores, N., & Schissel, J. L. (2014). Dynamic bilingualism as the norm: Envisioning a heteroglossic approach to standards-based reform. *TESOL Quarterly*, 48(3), 454-479.
- García, O. & Li Wei (2014). *Translanguaging: Language, bilingualism and education* (Palgrave Pivot). Palgrave Macmillan.
- García, O., & Kleifgen, J. A. (2010). *Educating emergent bilinguals: Policies, programs, and practices for English language learners*. Teachers College Press.
- MacSwan, J. (Ed.). (2022). *Multilingual perspectives on translanguaging* (Vol. 1). Channel View Publications.
- Nickson, R. A. (2009). Governance and the revitalization of the Guaraní language in Paraguay. *Latin American Research Review*, 44(3), 3-26.
- MacSwan, J., & Faltis, C. J. (Eds.). (2019). *Codeswitching in the classroom: Critical perspectives on teaching, learning, policy, and ideology*. Routledge.
- Mena, M. (2022). The Language-Elsewhere: A Friendlier Linguistic Terrorism. In *The Spanish Language in the United States* (pp. 80-95). Routledge.
- Ministerio de Educación y Ciencias (2021). *Análisis del Sistema Educativo Nacional. Diagnóstico técnico para contribuir al debate sobre educación en Paraguay, como parte del proceso de Transformación Educativa 2030. Eje desarrollo profesional docente*. Disponible en <https://bit.ly/3wExPfA>
- Parra, M. L. (2020). Working with diversity in the Spanish heritage language classroom: A critical perspective. *Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras*, (34), 1-24.
- Parra, M. L. (2016). Understanding identity among Spanish heritage learners. *Advances in Spanish as a heritage language*, 49, 177-204.
- Rona, J. P. (1966). The social and cultural status of Guaraní in Paraguay. In W. Bright (Ed.), *Sociolinguistics: Proceeding of the UCLA Sociolinguistics Conference* (pp. 277-298). The Hague: Mouton.
- Valdés, G. (2017). From language maintenance and intergenerational transmission to language survival: Will "heritage language" education help or hinder? *International Journal of the Sociology of Language*, 2017(243), 67-95.

- Valdés, G. (2015). Latin@s and the intergenerational continuity of Spanish: The challenges of curricularizing language. *International Multilingual Research Journal*, 9(4), 253-273.
- Valdés, G. (2005). Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The modern language journal*, 89(3), 410-426.
- Valdés, G. (2004). Between support and marginalization: The development of academic language in linguistic minority children. *International Journal of Bilingual Education and Bilingualism*, 7(2-3), 102-132.
- Valdés, G. (1997). The teaching of Spanish to bilingual Spanish-speaking students: Outstanding issues and unanswered questions. *La enseñanza del español a hispanohablantes: Praxis y teoría*, 8-44.
- Valdés, G. (1992). Bilingual minorities and language issues in writing: Toward professionwide responses to a new challenge. *Written communication*, 9(1), 85-136.
- Valdés, G., Fishman, J. A., Chávez, R., & Pérez, W. (2008). Maintaining Spanish in the United States: Steps toward the effective practice of heritage language re-acquisition/development. *Hispania*, 4-24.
- Valdés, G., & Geoffrion-Vinci, M. (1998). Chicano Spanish: The problem of the “underdeveloped” code in bilingual repertoires. *The Modern Language Journal*, 82(4), 473-501.
- Valdés, G., & R. A. Figueroa. (1994). Bilingualism and testing: A special case of bias. *Ablex*.

Artículo original

Writing Skill Training of tertiary students by using Letter Writing on All Topics App

Entrenamiento de la escritura en estudiantes universitarios mediante la aplicación Letter Writing on All Topics

Oraily Madruga Rios

*Universidad de La Habana,
Instituto Superior de Tecnologías y Ciencias Aplicadas, Cuba*
<https://orcid.org/0000-0002-1571-0246>

e-mail: oraily@instec.cu

Recibido: 20/5/2023
Aprobado: 15/8/2023

ABSTRACT

Writing in English is a productive skill, more difficult for students to manage than perceptive skills as reading and listening. Therefore, the more resources teachers provide students with to strengthen writing the better. The paper's aim is to initially analyze the training of pre-intermediate level tertiary students by using the Letter Writing on All Topics application. A poll, Iadov technique, documental analysis, descriptive statistics, and MapALL analytical approach are used as tools and methods. The study conducted achieves positive results, showing greater improvement of students' writing skill by using the app. Hence, mobile learning has turned into an outstanding and efficacious teaching resource for learning foreign languages.

Keywords: writing; skill; mobile; learning; students.

RESUMEN

La escritura en inglés es una habilidad productiva, más difícil de dominar para los estudiantes que las habilidades perceptivas como la lectura y la comprensión auditiva. Por ende, cuantos más recursos proporcionen los profesores a los estudiantes para fortalecer la escritura, mejor. El objetivo del trabajo es analizar preliminarmente el entrenamiento de los estudiantes universitarios de nivel pre-intermedio con el empleo de la aplicación móvil Letter Writing on All Topics. Se utilizan como herramientas y métodos: la encuesta, la técnica Iadov, el análisis documental, la estadística descriptiva y el enfoque analítico de MapALL. El estudio realizado obtiene resultados positivos, se percibe una mejoría notable de dicha habilidad en los estudiantes al utilizar la aplicación. Por consiguiente, el aprendizaje móvil se ha convertido en un recurso didáctico excepcional y eficaz para el aprendizaje de lenguas extranjeras.

Palabras clave: expresión; escrita; aprendizaje; móvil; estudiantes.

INTRODUCTION

Mobile learning or m-learning has been gaining followers all over the world. That trend occurs not only because of the tremendous breakthrough of technology in the 21st century, but also grounded in the current global Covid-19 crisis. Nowadays, various countries have suspended classes due to the novel coronavirus, so students are likely to use educational mobile applications for not losing track of their school training. However, educational apps, particularly those teaching foreign languages have appeared before the coronavirus outbreak. According to Rosman (2008), the emergence of m-learning within tertiary education may have started with the innovation of mobile technology and web learning platforms, which paved the way for delivering educational content and learning experiences via mobile devices such as mobile phones, personal digital assistants (PDAs), and iPods. Many authors support the potential apps have on language learning, so are the cases of: (Alexiadou, 2022; Andersen, 2019; Fouz-Gonzalez, 2020).

Most researchers have conducted studies based on the Mobile Assistance Language Learning (MALL) approach to find out apps' advantages and shortcomings from the pedagogical point of view. Nevertheless, MALL approach does not include set patterns to follow. Hence, experts can indistinctively measure the quality and features of mobile applications for language learning using different indicators, mainly focusing on student-centered learning. "More general and theoretical frameworks for MALL have predominantly been focused on student-centered learning" (Lindaman & Nolan, 2015, 4). By using MALL, students can learn anytime and anywhere without pressure.

Students using mobile language apps can learn foreign languages everywhere and every time as assured by these authors "[...] both fluid learning contexts (eg, formal, informal, in-class and out of class settings) and mobile technologies influence each other and continuously alter each other" (Kukulska-Hulme & Viberg, 2018, 215).

Nonetheless, m-learning still shows some limitations as small screen size and a general lack of pedagogically tested apps. Despite those difficulties, MALL enhances students self-confidence and ownership as some authors wisely noted, "It permits students to decide their own learning processes" (Banister, 2010:121), and "Students are increasingly taking ownership of their own education" (Mitra, 2005, 71).

Before making decisions on what educational mobile applications are viable to include as language learning resources of English course programs, teachers should examine them focusing on the aforementioned approach. Currently, in post Covid-19 pandemic times, the University of Havana including the Higher Institute of Applied Sciences and Technologies (InSTEC) has endorsed the use of ITCs to strengthen the teaching-learning process. Therefore, teachers from Languages Department at InSTEC have studied several apps to find out the most adequate for students to practice out of classes the English contents that are going to be assessed in their certification exams.

In addition, the most concerning issue for teachers is that students gain all skills required to approve B1 examinations, which are compulsory for their final graduation. Regarding the contents tested, students face most difficulties in writing, particularly letters. Thus, the paper's aim is to initially analyze the training of pre-intermediate level tertiary students by using the Letter Writing on All Topics app.

THEORETICAL REFERENCE

The research is based on Guo (2014)'s new perspective of MALL approach. This author made some changes to Kim et al.'s (2012) analytical MALL framework. He aimed at broadening its scope, integrating pedagogical characteristics, and learning theories and paradigms, naming his as analytical framework of Mobile-apps Assisted Language Learning

(MapALL). Guo's approach comprises four main categories: Target Learners, Content and Focus, Theoretical and Pedagogical Features, and Reputation. The first category seeks:

[...] to identify intended users' profile including age, interest, and English language proficiency level. Content and Focus examine each app's learning materials and its structure in addition to whether it focuses on pronunciation, speaking skills, or sociolinguistic aspects. Theoretical and Pedagogic Features cover three elements, instructional activities, individual exercises, and learning theories and paradigms. Reputation includes download count, users' average rating, and their comments. (Guo, 2014, 32)

Meanwhile, for students to gain B1 proficiency level, the English tertiary program encompassed these goals concerning writing skills:

- Write well- structured and coherent texts, which can be also concise, using some wide vocabulary and a limited range of complex grammar structures.
- Write letters using simple information of immediate relevance for friends, service personnel, teachers and others, communicating in a comprehensible way.
- Write letters accepting apologies and apologizing for, requesting, and giving advice.

The aforementioned aims were consistent with the contents tackled in the Writing Letter on All Topics app as it is further seen in this article.

METHODOLOGY

Approach

The current research has been based on a mixed approach since it relies on analyzing pedagogical features and averaging results from a poll out. The paper reaches a descriptive scope since it details the characteristics of the app, and it depicts students' reactions to the mobile application.

Study Population

The study population was 300 pre-intermediate level tertiary students from Meteorology, Nuclear Physics, Radiochemistry and Nuclear Engineering majors. The paper's author requested their consent to publish their opinions.

Recollection techniques

First, the Letter Writing on All Topics app had been shared to the 300 students. Weeks later, a poll containing three closed and two open questions was carried out to measure students' satisfaction with the use of the Letter Writing app. Its outcomes were assessed by means of the Iadov technique, which comprises six levels of satisfaction, from one to six, depending on individual answers. Furthermore, the Group Satisfaction Index was calculated, providing five values whose meanings are displayed below:

Table 1. Five values of the Iadov technique

Scale	Meaning
+1	Clear satisfaction
+0,5	More satisfied than dissatisfied
0	Not defined or contradictory
-0,5	More dissatisfied than satisfied
-1	Clear dissatisfaction

Source: (Batista, 2018)

In addition, it was analyzed the app's features through the guidance on the four categories of Guo (2014)'s MapALL approach, which was pivotal to deeper examine it from a pedagogical point of view.

RESULTS

After having shared the Letter Writing on All Topics app, 282 students out of 300 used it, representing 94 percent. The outcomes of the first three poll questions using Iadov technique revealed that 50 percent of the study population feels a clear satisfaction with the Letter Writing on All Topics app and 40 percent said to be more satisfied than dissatisfied. Just 6.6 percent reflected to feel more dissatisfied than satisfied, and the answers of the 3.3 percent turned out not defined. As the ensuing Table 2 shows:

Table 2. Poll's results in Iadov quadrant

3. Do you think that the Letter Writing on All Topics app satisfies your linguistic needs for learning English?	1. How often do you study by yourself with the Letter Writing on All Topics app?								
	NEVER			SOMETIMES			ALWAYS		
	2. How would you value the mobile application's quality?								
	Good	Acceptable	Bad	Good	Acceptable	Bad	Good	Acceptable	Bad
It satisfies me a lot	2	2	6	1	2	6	1	1	6
It satisfies me more than what it dissatisfies me	2	2	6	2	2	6	2	2	6
It is indifferent for me	3	3	3	3	3	3	3	3	3
It dissatisfies me more than what it satisfies me	6	4	5	6	4	5	6	4	4
It does not satisfies me at all	6	5	5	6	6	4	6	5	4
I do not know what to say	3	3	4	2	2	3	2	2	3

Source: (Batista, 2018; own creation).

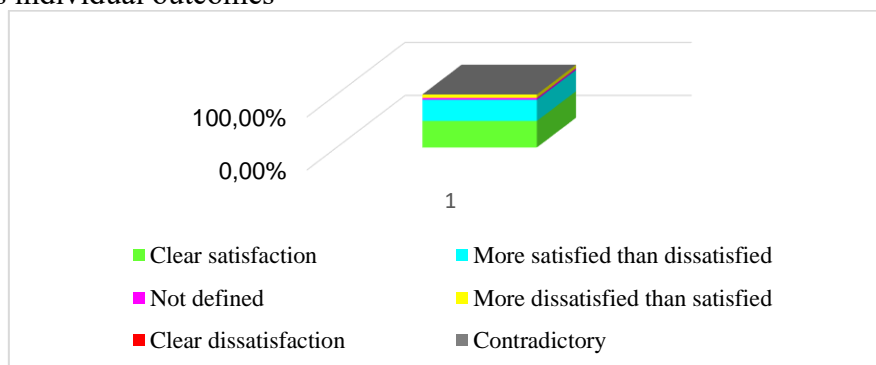
The colored numbers were the ones chosen by the study population.

Table 3. Results of the study population's individual satisfaction with the Letter Writing on All Topics app:

Results	Amount	%
Clear satisfaction	150	50%
More satisfied than dissatisfied	120	40%
Not defined	10	3,3%
More dissatisfied than satisfied	20	6,6%
Clear dissatisfaction	-	-
Contradictory	-	-

Source: (Own creation)

Fig.1. Poll's individual outcomes



Source: (Own creation)

Taking into account the previous data, the study population's different levels of satisfaction are associated with a numerical scale oscillating between +1 and -1 to calculate the Group Satisfaction Index, as it is explained in the Table 4 below:

Table 4. Data to calculate the Group Satisfaction Index

Scale	Meaning	Individual Satisfaction
+1	Clear satisfaction	150
+0,5	More satisfied than dissatisfied	120
0	Not defined or contradictory	10
-0,5	More dissatisfied than satisfied	20
-1	Clear dissatisfaction	-

Source: (Own creation)

The Group Satisfaction Index is calculated following the formula:

$$ISG = \frac{A(+1)+B(+0,5)+C(0)+D(-0,5)+E(-1)}{N}$$

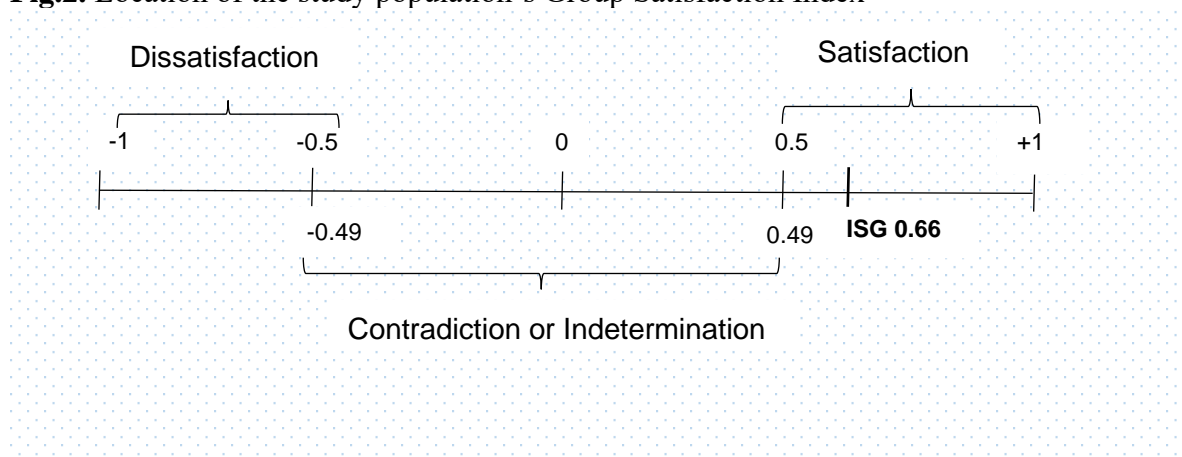
In this formula A, B, C, D, E represent the number of subjects with the individual index 1; 2; 3 o 6; 4; 5 and N represents the total of the study population, thus:

$$ISG = \frac{150(+1)+120(+0,5)+10(0)+20(-0,5)}{300}$$

$$ISG = \frac{150 + 60 + 0 - 10}{300} = 0,66$$

The study population's Group Satisfaction Index concerning the use of the mobile app was calculated, which value was GSI=0,66, a number that is within the satisfaction range (>0,5) as it can be observed in Fig. 2:

Fig.2. Location of the study population's Group Satisfaction Index



Source: (Own creation)

Regarding the fourth question, most of the study population deemed that the app addressed a vast number of noteworthy issues. However, some considered that the Letter Writing on All Topics app could include letter templates discussing on requesting awards or supporting the approval of innovations.

The majority responded to the poll's fifth question that using the mobile application helps them to improve their writing letter proficiency. A few answered that their learning was not that fast as they wish because they did not count on teachers to check their work.

MapALL ANALYSIS OF THE LETTER WRITING ON ALL TOPICS APP

Target Learners

Age: Teenagers, youngsters, and adults, in general.

Interest: this app is intended for students' independent study and self-preparation in the English subject. It is particularly aimed at helping users know the different types of letter writing formats and the correct structure of a letter in English.

Proficiency level: the app is designed for pre-intermediate, intermediate, upper intermediate and advanced students. They are more independent and proficient users; noticing the app is written in English, including the explanation on how to write a letter.

Content and Focus

Topic/Theme: Letter writing

Structure:

Letter Templates

Letter Writing Guide

Letter Sample

Banking Letter Samples

Thank you Letter Samples

Content size: The contents are grouped in an arranged and predictive way. The explanation on how to write letters is very clear and easy to understand, supported by a considerable amount of templates and samples.

Letter Writing Templates: in this section, users will find templates containing the structures used in each type of letter as follows:

Business Letter Template

Cease and Desist Letter Template

Character Reference Letter Template

Complaint Letter Template

Condolence Letter Template

Cover Letter Template

Employee Termination Letter Template

Employment Verification Letter Template

Friendly Letter Template

Job Interview Thank You Letter Template

Letter of Intent Template for a job

Letter of Intent Template for Graduate School

Letter of Recommendation Template

Reference Letter Template

Resignation Letter Template

Thank You Note Template for a Gift

Letter Writing Guide: This section will help users know important elements when writing each type of letter. It is divided into the following:

How to write a letter

Types of letters

Business Email Writing

Common Mistakes in Letter Writing

Cover Letter Writing

Differences between Formal and Informal Letters

How to Write a Business Letter
 How to Write a Cancellation Letter
 How to Write a Complaint Letter
 How to Write a Condolence Letter
 How to Write a Farewell Letter to your Co-workers
 How to Write a Formal Letter
 How to Write a Friendly Letter
 How to Write a Letter of Interest
 How to Write a Resignation Letter
 How to Write a Thank You Letter After an Interview
 How to Write an Address Change Letter
 How to Write an Apology Letter
 How to Write an Appeal Letter
 How to Write an Invitation Letter
 Letter Writing Rules
 Letter Writing Tips
 Requesting a Letter of Recommendation
 Writing a Letter of Recommendation

Letter Samples: this section constitutes an extension of the Templates Section in which users will find the structures of the letters already written in a social context as real samples. It comprises the following issues:

Address Change Letter to a Bank
 Business Condolence Letter
 Company Address Change Letter
 Effective Job Application Cover Letter
 Examples of Formal Letters
 Examples of Informal Letters
 Insurance Cancellation Letters
 Letter of Application – different versions
 Military Condolence Letter
 Request Letter
 Sample Apology Letter (Business)
 Sample Apology Letter to Professor
 Sample Apology Letter to Teacher About Homework
 Sample Business Letter
 Sample Class Activity Suggestion Letter
 Sample Complaint Letter
 Sample Cover Letter
 Sample Critical Letter to Classmate
 Sample Critical Letter to Coworker
 Sample Farewell Letter
 Sample Interview Thank You Letter
 Sample Invitation Letter (Business)
 Sample Letter Asking Employer for Extension
 Sample Letter Asking for Extension After Retirement
 Sample Letter Asking Teacher for Extension
 Sample Letter from Female to Male
 Sample Letter from Male to Female
 Sample Letter of Appeal

Sample Letter of Interest
 Sample Letter of Recommendation
 Sample Letter to Publisher
 Sample Letter to Teacher About Problem
 Sample Protest Letter to Manager
 Sample Resignation Letter
 Sample Semi Formal Letter
 Sample Teacher Apology Letter
 Sample Teacher Appreciation Letter
 Sample Welcome Letter to Parents for Back-To-School
 Teacher's Letter to Parents
 Wedding Cancellation Letter

Banking Letter Samples: they are related to the banking area, as their titles described:

Bank Authorization Letter for Travel
 Enclosing Cheques for Insurance
 Write a Letter requesting changing the beneficiary of an existing policy
 Write a Letter to the Regional Manager of Bank requesting for a loan for further studies.
 Write to the Bank requesting a copy of your Bank Statement

Thank you Letter Samples: it only contains one sample:

Sample Thank You Note for a Job Offer

Focus: the app is goal-centered on writing skill and sociolinguistic aspects. However, in an indirect way, when analyzing the letters, users will also be strengthening the reading skill through the understanding of what they read.

Theoretical and Pedagogic Features

Learning theories and Paradigms

The Letter Writing on All Topics app is likely to follow a *Behaviorist learning* theory, which outlines as a learning mechanism the association of stimuli and answers by means of conditioning. The fact that society is humanly computerized mastering the ICTs is an essential element of this learning theory. Many people find it easier and more amusing to study with an app they have at hand, which constitutes in itself, a stimulus when learning.

Pedagogic features of the mobile application are focused on the student-centered approach since all examples and exercises are directed to students' needs. It is also based on the communicative approach because users learned how to enhance their writing skill by practicing different language functions.

As one of the outstanding benefits of m-learning is ubiquity, that characteristic triggers informal and lifelong learning because students can study everywhere at any time. Nonetheless, students can reinforce their curriculum contents but without having pressure in the case teachers do not give instructions in their classes for using the Letter Writing on All Topics application.

The mobile application's creators took into account Constructivist Learning grounded in the app helps to generate and enlarge learning experiences that stimulate users to the construction of knowledge with autonomy and responsibility. The acquisition of knowledge is reflected as an authentic construction operated by users themselves.

It does not possess instructional activities, such as tutorials on how to use different sections, but it is not strictly necessary thanks to its predictive and easy-to-use interface. Even though the app does not have typical individual exercises, it does have some activities because its templates include some gaps that students can fill in using their own information, making samples and templates points of departure for users to start writing their own letters.

Reputation

India's MBDevApps Company launched the Letter Writing on All Topics app on August 20, 2020. Nowadays, it is part of All Indian Apps Company, and it was last updated on June 18, 2021, reaching 3.6 average rating and counting 258 users' reviews in Google App Store.

Download count: 100 K+

CONCLUSIONS

The current paper has been an initial stage to prove that the Letter Writing on All Topics application helps train students, so they gain further writing proficiency and self-independence. In addition, the research paved the way for teachers to rely more on mobile learning, which is one of the leading strategies nowadays because of an upwards trend in computerization of society, in general.

The poll unfolded that most users were happy with the app, according to the Group Satisfaction Index value, which was 0.66, representing clear satisfaction with the Letter Writing on All Topics app. It also revealed that the app's creators could improve templates and samples in a near future by adding other communicative functions, such as writing letters aimed at requesting awards or supporting innovations.

Meanwhile, the MapALL analytical approach aids to initially examine the Letter Writing on All Topics app, considering pedagogical criteria as learning theories and paradigms, target learners, content, and focus. Furthermore, teachers learned significant aspects to take into account when choosing an app as a resource material for students to learn English.

REFERENCES

- Alexiadou, E. (2022). Analysis and Evaluation of Interaction Applications for Effective Mobile-Assisted Language Learning [Master Thesis, Aristotle University of Thessaloniki]. IKEE Institutional Repository of Scientific Publications.
- Andersen, K. (2019). Perceived User Experience Associated with M-Learning: An Exploratory Case Study [Doctoral Project, Ashford University].
- Banister, S. (2010). Integrating the iPod Touch in K-12 education: Vision and vices. *Computers in the Schools*, vol. 2: 121.
- Batista, N. Valcárcel, N., Leyva-Vázquez, M., Smarandache, F. (2018). Validation of the pedagogical strategy for the formation of the competence entrepreneurship in high education through the use of neutrosophic logic and IADOV technique. *Neutrosophic Sets and Systems*, 22.
- Fouz-Gonzalez, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation App. *Language Learning & Technology*, 24, 62-85.
- Guo, H. (2014). Analysing and Evaluating Current Mobile Applications for Learning English Speaking. [Master Dissertation, University of London] TeachingEnglish.
- Kim, H., & Kwon, Y. (2012). Exploring smartphone applications for effective Mobile-Assisted Language Learning. *Multimedia-Assisted Language Learning*, 16, 31-57.

- Kukulska-Hulme, A., & Viberg, O. (2018). Mobile collaborative language learning: State of the art. *British Journal of Educational Technology*, 49(2), 207-218.
- Lindaman, D., & Nolan, D. (2015). Mobile-Assisted Language Learning: Application Development projects within reach for language teachers. *IALLT Journal of Language Learning Technologies*, 45(1), 1-22.
- Mitra, S. (2005). Self-organising systems for mass computer literacy: Findings from the 'hole in the wall' experiments. *International Journal of Development Issues*, 4, 71.
- Rosman, P. (2008). M-learning- as a paradigm of new forms in education. *E + M Ekonomie a Management*, 1, 119-125. Retrieved from ProQuest ID: 1569056991.

Artículo original

Architecture and Research Corpus Indexation of Texts by Yuri Shevelyov

Indexación de textos del Corpus de Arquitectura e Investigación
por Yuri Shevelyov

Illia Danyliuk

National Commission for State Language Standards, Ukraine
<https://orcid.org/0000-0001-8681-9321>

e-mail: i.g.danyluk@gmail.com

Anatolii Zahnitko

Vasyl Stus Donetsk National University, Ukraine
<https://orcid.org/0000-0001-7398-6091>

e-mail: a.zahnitko@donnu.edu.ua

Hanna Sytar

Vasyl Stus Donetsk National University, Ukraine
<https://orcid.org/0000-0001-8806-8322>

e-mail: h.v.sytar@donnu.edu.ua

Recibido: 15/2/2023
Aprobado: 25/5/2023

ABSTRACT

The article is devoted to the problem of creating a corpus of texts by Yuri Shevelyov. The main structural components of the text's corpus have been determined, the functions have been singled out, the features of its filling have been characterized, extralinguistic and linguistic marking has been explained.

Keywords: concordance; a corpus of texts; linguo-person; linguo-personology; corpus marking; frequency analysis; Yuri Shevelyov.

RESUMEN

El artículo está dedicado al problema de la creación de un corpus de textos de Yuri Shevelyov. Se han determinado los principales componentes estructurales del corpus de textos, se han señalado las funciones, se han caracterizado las características de su llenado, se ha explicado el marcado extralingüístico y lingüístico.

Palabras clave: concordancia; corpus de textos; linguo-persona; linguo-personología; marcado del corpus; análisis de frecuencia; Yuri Shevelyov.

INTRODUCTION

Linguo-personology is based on philosophical understanding of personalism (B. P. Bowne, R. H. Lotze, J. H. Howison, W. James, J. Royce, M. W. Calkins, W. E. Hocking, G. W. Allport, E. Mounier and others), one of the manifestations of which is the thesis of the spiritual renewal of the individual, the longevity of self-improvement. In modern linguistics, the issues of a linguistic portrait of the personality, establishment of structural and functional components of the personality are actualized. The stated and a number of other questions appear to be essential for modern applied linguistics, partly for psycholinguistics, linguo-psychiatrics, neurolinguistic programming, suggestive linguistics, and others. Linguo-personology as a science of linguistic personality studies the individual (monolinguo-person) or collective (polylinguo-person) – linguo-person in functional-communicative, verbal-semantic, linguo-psychological, speech-behavioral, discursive-situational, motivational-suggestive, and other dimensions.

The questions of the present interest are the formation of speech needs of the person, determination of their duration, diagnosing changes in language and speech needs in the living space of linguo-persons. The authors pay close attention to internal differentiation of status and functional load of speech and speech transitions, coding, and recoding, formation of sets of linguo-individualizations, ethical, emotional and situational and other factors. Especially important are the studies of prominent linguists, to whom Yuri Shevelyov – a prominent linguist, literary historian, critic belongs. His speech-discursive practices correlate different language elements, unequal discursive practices. He intensifies territorial, social, age, aesthetic, emotional, and other motivators of individual changes. The scientific task of establishing the linguistic-corpus structure of discursive practices of Yuri Shevelyov is highly motivating. The statement correlates with the practical task – the introduction into active use not only of the scientific and creative heritage of the outstanding scientist but also the use of language models and constructions of linguo-individualizations in modern language-codified practice developed by Yuri Shevelyov.

THEORETICAL BASIS

For the first time the concept of linguo-person was used in the early twentieth century by M. Trubetskoy. It was based on the idea of personalism (E. Mounier and others). In the 2nd half of the XX century in connection with the intensification of the study of linguistic personality, issues of active, associative grammar (J. Karaulov, 1987), the concept of linguistic person has acquired a different dimension and status.

Activation of linguo-personological studies was observed in the late XX – early XXI century. Theoretical and practical developments of communicative strategies and tactics, discourse, text linguistics, attempts to study speech-behavioral stereotypes, expansion of studies in applied linguistics, outlining the theoretical foundations of objective grammar, etc. became significant. Attempts at the linguistic portrayal of the individual and the creation of authorial corpora of texts are no less significant.

Yuri Shevelyov's linguistic personality was considered by R. Trifonov. He clarifies the linguistic and cultural components of the scientist's individual speech on the basis of essays and letters. M. Moser reveals the main stages of the dynamics of Yu. Shevelyov's creative personality on the way from Sovietism (Trifonov, 2009; Trifonov, 2015; Moser 2014). The studies of A. Danylenko, M. Kotsyubynska, L. Masenko, L. Tarnovetskaya, and others are no less relevant. Today, most of Yu. Shevelyov's works are posted on the Internet (http://movahistory.org.ua/wiki/%D0%A8%D0%B5%D0%B2%D0%B5%D0%BB%D1%8C%D0%BE%D0%B2_%D0%AE_%D0%9F%D1%80%D0%B0%D1%86%D1%96_%D0%BD%D0%B0%D1%8F%D0%B2%D0%BD%D1%96_%D0%B2_%D1%96%D0%BD%D1%82%D0%B5%D1%80%D0%BD%D0%B5%D1%82%D1%96)

(09.08.2021)) , which in general does not solve the problem of a holistic vision of the creative linguistic persona of Yuri Shevelyov. It is possible on the condition of integral corpus structuring of his heritage, the development of linguo-computer technologies of free navigation within hypertext. Therefore, the purpose of the study is an experimental-research model of text corpus of the linguistic persona of Yuri Shevelyov with the definition of the main components of the corpus, the laws of free linguistic-textual navigation.

METHODOLOGICAL BASIS OF THE STUDY

The integral corpus of the linguistic persona of Yuri Shevelyov, as well as the consideration of the peculiarities of the linguistic portrait of the scientist, requires the use of appropriate methods. Among them are the following methods: the method of discursive analysis using the methodology of corpus content analysis to establish all possible fixations of the studied word contexts, as well as identifying non-random combinations of words-collocations, comparing the latter with regular and quantitatively dominant ones. The use of corpus content analysis can serve as a basis for creating a sketch of a word with intra-corpus differentiation of usual and occasional-authorial (individualized) word combinations. It is also essential for determining their load within the communicative registers, the manifestation of communicative intentions.

INDIVIDUAL-AUTHOR EXPERIMENTAL RESEARCH CORPORA: FUNCTIONAL LOAD IN THE STUDY OF A LINGUO-PERSON

The complete linguistic portrayal is possible provided that the whole set of personal-generated texts is covered with the diagnosis of territorial, social, generational, and other motivations for changes in discursive practices. The set of emphasized questions appears as one of the sides of modern linguo-personology. Its directions are actively studied at the Department of General and Applied Linguistics and Slavic Philology of Vasyl'Stus Donetsk National University (A. Zahnitko, I. Danyliuk, J. Krasnobayeva-Chorna, H. Sytaretc.). Elaboration of levels, categories, and aspects of linguo-personology will make it possible to create an objective and subjective linguistic-societal grammar, at the center of which is a mono-and/or polylinguo-person in an indefinite number of his/her discursive practices. Diagnosing changes in 1) the structure of language personality, 2) in levels of its implementation (verbal-semantic, linguocognitive, motivational (according to Yu. Karaulov), functional (V. Konetska), situational-psychological), 3) the ratio of components (formation of communication skills, the need for communication and proper competence, the formation of language consciousness and language and/or speech behavior) based on: a) the natural need for communication in the native language; b) the communicative need to communicate in a foreign language; c) discursive-motivated practice (pedagogical, medical, etc.) of communication in a non-native language, possible under the condition of creating an integral body of texts of the linguistic personality.

Individual author's experimental corpora make it possible to establish the levels and aspects of a linguo-person with reliance on his/her personal texts. Through the lens of the total size of the text corpus, it's possible to see the identification of certain thematic and key elements of disclosing the duration of linguo-persons, the formation of gaps motivated by territorial, social and other factors. The appropriate linguistic landscape of linguo-persons, reconstruction of creative potential and mechanisms for implementation become visible. It is noteworthy that the results of learning, dissemination, and self-realization of linguo-persons, etc. can be reconstructed. The text corpus will make it possible to trace the patterns of linguistic and social conditioning of the linguo-person and the status of linguo-individualizations as individual self-expression: attraction, desire, inclination, worldview, beliefs. A Linguo-person manifests through relationships with others, where relationships with other individuals are the

experience and awareness, and beliefs outline the direction of the individual, his/her language values. Subsequently, knowledge, skills, abilities, and habits to use the language, to differentiate it in different corporate groups, to be able to use different language codes depending on situations, become significant. This confirms the language space of Yuri Shevelyov in different periods of his life – childhood, student years, Kharkiv period, years of emigration in Munich, teaching at Lund University, and later at Harvard and Columbia Universities.

Consistent use of corpus content analysis provides the definition of the emotional background of discursive practices, the expression of feelings and sensations, memory levels, etc. In its entirety, the author's experimental research corpus of texts provides an understanding of the properties of the temperament of the *linguo-person*. It tracks the change of its speech-behavioral realizations, establishing the typological properties of a *linguo-person*. In general, the discourse of a *linguo-person* is his/her living space with different functionally loaded sets of self-realizations in language and in language.

INDIVIDUAL-AUTHOR'S EXPERIMENTAL RESEARCH CORPUS OF A LINGUO-PERSON YURI SHEVELYOV: STRUCTURE, TECHNOLOGIES

Yuri Shevelyov is an outstanding personality who is realized in numerous texts, various discursive practices, unequal linguistic-landscape spaces.

To study the peculiarities of Yuri Shevelyov's speech, it was decided to compose a research corpus of texts. This corpus of texts was designed and implemented by A. Zahnitko, I. Danyliuk, and H. Sytar, teachers of the Department of General and Applied Linguistics and Slavic Philology of Vasyl'Stus Donetsk National University. Within the framework of the internship, 2nd-year students of the specialty "Applied Linguistics" took an active part in the preparation of texts for the corpus.

Yuri Shevelyov's text corpus was created with the help of the free corpus manager NoSketch Engine (<https://www.sketchengine.eu/>), developed at the University of Masaryk (Brno, Czech Republic) (RychlŮ, SmrŮ, 2004). This corpus, along with others, is available on the server of the Department of General and Applied Linguistics and Slavic Philology of Vasyl'StusDonNU at corpora.donnu.edu.ua.

According to the authors' plan, the corpus contains all the texts of Yuri Shevelyov published by today, in particular:

Shevelyov, Yuri. From the history of the unfinished war. (Compilers Oksana Zabuzhko, LarysaMasenko). Kyiv: Ed. Kyiv-Mohyla Academy, 2009. 471 p. ISBN 978-966-518-519-2; Shevelyov, Yuri. From the history of the unfinished war. (Compilers Oksana Zabuzhko, LarysaMasenko). Kyiv: Ed. Kyiv-Mohyla Academy, 2009. 471 p. ISBN 978-966-518-519-2; Sherekh, Yuri. Outside of books and from books. Kyiv: Chas Publishing House, 1998. 456 p. ISBN 966-95238-3-4; Sherekh, Yuri. Another rug. "Library of Prologue and Modernity Part 130", 1978. 393 p.; Sherekh, Yuri. Not for children. New York: PROLOGUE Publishing House, 1964. 416 p.; Shevelyov, Yuri. Selected works: in 2 books. Book II. Literary Studies. Kyiv: Kyiv-Mohyla Academy Publishing House, 2009. 1151 p. ISBN 978-966-518-496-6; Shevelyov, Yuri. Selected works: in 2 books. Book I. Linguistics. Kyiv: Kyiv-Mohyla Academy Publishing House, 2009. 583 p. ISBN 978-966-518-494-2; Sherekh, Yuri. The third watchman. Baltimore-Toronto: Smoloskyp, 1991. 454 p.; Shevelyov, Yuri. "I, me, me... (and around)". Memoirs. In two volumes. Berezil Magazine Publishing House, MP Kots Publishing House. Kharkiv– New York, 2001; Zabuzhko, Oksana, Shevelyov, Yuri. Selected correspondence against the background of the era: 1992-2002: with appendices, works, comments, reasons for biographies and other documents. Kyiv: VysokaPolitsa, VD Fakt, 2011.

The Ukrainian-language part of the corpus currently covers 104 documents in the Ukrainian language, including 1346424 tokens.

Extralinguistic and linguistic markup is used in the created case. **Extralinguistic corpus markup** combines:

- a) metatext data. They include:
- *area* with possible meanings of literary studies, linguistics, general works;
 - *author* - in this building only Yuri Shevelyov, created for the possibility of combining with other buildings;
 - *genre*: essay, article, monograph, interview, preface, speech, report, memoirs, article, speech, article, report, letter, introductory word;
 - *name* (title of the work);
 - *source*;
 - *style*: journalistic, scientific and epistolary;
 - *type*: in the original language, translation from English, translation from German, translation from French;
 - *year*.
- b) *structural markup*. The corpus contains data:
- about text borders in <doc> - </doc>tags;
 - about paragraph boundaries in <p> - </p>tags;
 - sentence boundaries in <s> - </s>tags;
 - special tag <g /> indicates punctuation marks that are not separated by a space from the previous token.

Linguistic marking of the analyzed corpus of Yuri Shevelyov's texts today is the result of automatic morphological analysis and lemmatization performed with the help of the author's tools. The structure of the standard tag for each token is as follows: in the first position is the mark of the grammatical class of the word, then the mark of subclasses, all marks – single characters in Latin or numbers, each subclass is assigned a position that does not change for different classes.

For example, for a *conference* word, the tag looks like Izzooi1m (Fig. 1).

Scheme 1. Deciphering the tag for the word form “*conference*”

<i>Noun</i>		<i>feminine</i>		<i>Ablative case</i>		<i>inanimate</i>		<i>Soft group</i>
I	Z	Z	o	O	i	n	1	m
<i>General name</i>		<i>Singular</i>		<i>Noun type of declension</i>		<i>First declension</i>		

The system of tags is described in detail in the study (Zahnitko, Danyliuk, 2013). In particular, the classification of grammar classes has the following form (Table 1):

Table 1. Classes of words in the corpus of texts by Yuri Shevelyov

Noun	I.*
Verb	D.*
Participle	Dk.*
Adverb	Ds.*
Adjective	K.*
Pronoun	Z.*
Adverb	S.*
Preposition	J.*
Conjunction	P.*
Numeral	C.*
Particle	T.*
Interjection	W.*
Abbreviation	A.*
The rest	R.*

The gender category is described by the following tags (Table 2):

Table 2. Grammar gender in the corpus of texts of Yuri Shevelyov

masculine	..c.*
feminine	..z.*
neutral	..s.*

The quantitative parameters of the created corpus are as follows: 1,346,424 tokens were collected in 104 documents, of which 1,037,949 were words in 66,039 sentences. The general lexicon includes 138,187 different word forms and punctuation marks described by 157 original tags, and 32,172 lemmas.

Typical functions are available in the corpus manager, such as *building a concordance* based on a simple search, searching in lemmas, searching for a phrase, word form, symbol, or a specific template built using a regular expression.

Figure 1. Interface for selecting the type of query in the case

The query can be based on an additional search in the context of filtering the desired lemmas or word forms at a distance of up to 15 tokens to the right or left of the main word.

Figure 2. Interface for selecting the context of the query in the case

Finally, the search can be limited to the different types of texts provided by extralinguistic markup.

In the constructed concordance different types of data sorting and filtering, frequency analysis of morphological symbols or word forms for the lemma, etc. are possible.

Another typical function of the corpus manager is frequency analysis with the ability to select the minimum or maximum frequency, part-of-speech filters, N-grams.

Figure 3. Interface for building a frequency dictionary in the case

Параметри словника частот ?

Корпус: Корпус текстів Юрія Шевельова (Шереха) ▼

Підкорпус: Жодного (цілий корпус) ▼ [відомості створити новий](#) ?

Шукати атрибут: word ▼

☐ вживати N-грами. Значення N: від 2 ▼ до 2 ▼ ?

☐ сховати/вкласти під-N-грами

Параметри фільтрування:

Фільтрувати словник за: Регулярний вираз: ?

Мінімальна частота: 5

Максимальна частота: 0 (0 = без обмежень зверху)

Білий список: Файл не вибрано

Чорний список: Файл не вибрано [формат](#)

☐ Занести не-слова

Параметри виводу:

Показувані частоти: ☒ Кількість знайдень ☐ Кількість документів ☐ ARF

Тип виводу: ☒ Простий

☐ Ключові слова

Опорний (під)корпус: Корпус текстів Юрія Шевельова (Шереха) ▼ (цілий корпус) ▼

Давати перевагу: рідким словам частим словам 1 ?

☐ Змінити показувані атрибут(и)

--- ▼ --- ▼ --- ▼

Можна вибрати один або декілька показуваних атрибутів. Ця опція може бути часовитратною.

The frequency analysis tool also allows you to select all lemmas and all word forms.

Figure 4. Fragment of the frequency of word forms from the corpus by Yuri Shevelyov

Титульна	Словник частот	
Пошук	Корпус: Корпус текстів Юрія Шевельова (Шереха)	
Словник частот	Total number of items: 22,218	
Про корпус	Сторінка 1	<input type="button" value="Перейти"/> Наступна >
Серверні задачі	<u>word</u>	<u>частота</u>
Інструкція ↗	і	26,971
	в	22,811
	не	20,855
	з	14,831
	що	13,703
	на	13,464
	й	10,224
	до	9,251
	у	8,249
	а	7,040
	як	6,548
	це	5,777
	його	5,114
	про	5,075
	він	4,950
	але	4,659
	я	4,074
	від	3,920
	було	3,840
	тільки	3,444

CONCLUSIONS

Thus, the created by Yuri Shevelyov corpus is research, full-text and dynamic, and it has extralinguistic and linguistic markup. Among the important functions are a) building a concordance based on a simple search, search in lemmas, search for a phrase, word form, symbol, or pattern created using a regular expression, b) frequency analysis for word forms, lemmas, and tags.

Now the team of authors is working to eliminate errors of automatic morphological analysis and filling the English-language part of the corpus. The prospect of the study is to create a comprehensive body of texts by Yuri Shevelyov, which will cover documents in Ukrainian and English and will be a reliable basis for studying the features of the speech of this prominent scholar.

REFERENCES

- Zahnitko, A., Danyliuk, I. (2013). Corpus of texts of grammatical service. *Applied Linguistics and Linguistic Technologies: Mega Ling 2012* (pp. 102-112). Kyiv: UMIF.
- Karaulov, Y. (1987). *Russian language and linguistic personality*. Moscow: URSS.
- Moser, M. (2014). The Language Behavior of Galician Russophiles during the Interwar Period. *Russian Linguistics* 38, 2014, 315–339
- Moser, M. (2014a). Yuri Shevelyov on the road of reviva
<http://www.historians.in.ua/index.php/en/ukrayinska-mova/1066-mikhael-mozer-yurii-shevelov-na-dorozi-vidradianshchennia>
- Трифонов, Р.А. (2009). Metalanguage fragments of Yuri Shevelyov's memoirs are representatives of the individual picture of the linguist's world. *Bulletin of Kharkiv National University named after V.N. Karazina*, 843, 55, 19-26.
- Trifonov, R. (2015). *Linguo-cultural components of Yuri Shevelyov's individual speech (based on essays and letters)*
<http://www.historians.in.ua/index.php/en/ukrayinska-mova/1697-roman-tryfonov-linhvokulturni-skladnyky-indyvidualnoho-movlennya-yuriya-shevelova-na-materiali-ese-ta-lystiv>
- Rychlý, P., Smrž, P. (2004). Manatee, Bonito and Word Sketches for Czech. *Proceedings of the Second International Conference on Corpus Linguistics* (pp. 124-132). Saint-Petersburg: Saint-Petersburg State University Press.

Reseña

**Reseña de la novela de Miguel Zapata Ros:
"CARADOC" (I y II parte)**
Review of Miguel Zapata Ros' novel:
"CARADOC" (Part I and Part II)

José Antonio Alonso Navarro
Universidad del Norte, Paraguay
<https://orcid.org/0000-0002-7041-8803>

e-mail: meildeja@yahoo.com

Recibido: 20/3/2023
Aprobado: 13/4/2023

El profesor Miguel Zapata Ros es de esos hombres que destacan a primera vista porque habiendo dedicado toda una vida al estudio de la Ingeniería Informática, a la Informática Pluridisciplinar, a las Tecnologías de la Comunicación, a las Ciencias Experimentales y a la investigación científica en general, un día decidió alternar su coqueteo con la razón epistemológica de la ciencia y su lenguaje con el alma platónica de la literatura y su rica dimensión semántica y cultural tras casarse con la novela. Fruto de esa etapa tan vital como proyectiva, cristalizaron algunas obras literarias, entre ellas, las dos partes de su novela Caradoc. Parece, por lo tanto, que el profesor Zapata abandonó temporalmente los números, los algoritmos, las aplicaciones informáticas y los dispositivos computacionales para aventurarse en la creación literaria alumbrando una novela bien hecha, coherente, unitaria, llena de personajes reales y ficticios, y forjando una trama tan interesante como penetrante. La primera parte de la novela está acompañada de un epígrafe que reza así: "Hacer cualquier cosa para conseguir el efecto deseado", y la segunda parte de la misma, en su deseo de mantener una misma línea de coherencia con relación al título, está acompañada de otro no menos intrigante o sugerente título que dice así: "Jugando con los sentimientos de las mujeres".

La primera parte de Caradoc fue publicada en el año 2019, un año de cierta resonancia histórica e intrahistórica puesto que con él vino la pandemia del COVID-19 con tanto misterio arcano como vino también entre los años 1347 y 1353 la peste bubónica o peste negra; y la segunda parte de la misma vio la luz en el año 2021, un año algo convulso, sin duda alguna, por los cambios tan radicales que trajo en su esquivo zurrón la susodicha pandemia global. Ambas entregas literarias están disponibles en la plataforma Amazon- Kindle junto con una interesante biografía del autor de Caradoc que podemos resumir así: "Miguel Zapata Ros es profesor en las universidades de Alcalá de Henares y de Murcia e investigador de estas universidades y del Instituto Interuniversitario de Economía Internacional. Recibió el título de doctor en Ciencias de la Computación en la Universidad de Alcalá de Henares y el de licenciado en Matemáticas, especialidad en Probabilidad, Estadística e Investigación Operativa, en la Universidad de Valencia. Desde 1984 a 1994 ha trabajado para el Ministerio de Educación español, coordinando programas de integración de las tecnologías digitales. Ha participado en programas institucionales de la Unión Europea de evaluación de la calidad en estudios de grado y de postgrado en línea.

Es profesor y coordinador desde 2006 del Máster Universitario en Informática Pluridisciplinar de la Universidad de Alcalá de Henares en la especialidad en enseñanza y aprendizaje electrónicos. Ha sido profesor visitante y asesor de universidades en Latinoamérica, Europa y Asia, y en su momento fue el fundador de RED, Revista de Educación a Distancia, así como el editor de la Revista Española de Docencia Universitaria (RED-U). Por último, ha escrito más de 200 publicaciones arbitradas sobre enseñanza, aprendizaje en entornos digitales, docencia universitaria y cambio de paradigma”.

Alternándose con las entregas anteriores encontramos otra novela relacionada con el mundo de Caradoc que lleva por título *Joaquina Plana, Lord Howden y la torre Caradoc de Murcia*, disponible igualmente en Amazon. Fue publicada en 2018, y la misma retoma algunos de los personajes que aparecen en las dos partes de Caradoc y a los que vamos a referirnos a continuación. Como filólogo, doctor en Filología Inglesa y especialista en literatura medieval inglesa e irlandesa, me llamó de manera especial el título: Caradoc, y me vino a la mente un personaje legendario emparentado con los reyes de Gwent de nombre Caradoc Vreichvras (“Caradoc Brazofuerte”). En la Edad Media Gwent fue un reinado galés que estaba situado entre los ríos Wye y Usk. Su aparición tuvo lugar desde el fin de la dominación romana en Britania en el siglo V aproximadamente hasta la invasión normanda de Gales en el siglo XI.

Es posible que Caradoc Vreichvras viviera durante los siglos V o VI. En el Mito artúrico o en la vasta colección de historias relacionadas con Britania y sus reyes y héroes legendarios, como el rey Arturo, aparece como un caballero de la Tabla Redonda conocido como Carados Briefbras (“Carados Brazocorto”). Sin embargo, a lo largo de la historia han existido muchos otros personajes que ostentaron ese mismo nombre o nombres similares, como el héroe britano Carataco (en latín Caradoc), hijo del rey Cunobelino, que actuó como caudillo de los catuvellaunos, una tribu celta-belga del sureste de Britania antes de la conquista romana, y líder de la resistencia contra los romanos invasores; o Cerdic de Wessex (en latín Cerdicus), el cual es descrito en la Crónica Anglosajona como el caudillo del asentamiento anglosajón de Britania y como el fundador y primer rey de Wessex, el cual reinó desde el año 519 hasta el 534 d.C.

Lo interesante también de este nombre propio, de Cerdic, es que es posible que proceda del nombre Ceretic, que sea de origen britónico, no germánico, y que se derive del nombre Caratīcos o Corotīcos. En la obra de Geoffrey de Monmouth, *Historia Regum Britanniae*, Caradocus es un gobernante de Cornualles bajo el poder de Octavio que murió durante el reinado del emperador Magnus Maximus. Y adentrándonos en el mundo del fabliaux medieval europeo o de sus bourdes (o historias cómicas), sería igualmente oportuno destacar la aparición de un singular personaje de nombre Caradoc que aparece en la obra del escritor anglonormando del siglo XII Robert Biket: *Lay du cor*. Esta obra de menos de 600 versos se escribió en Inglaterra hacia 1170-1180, y parodia o satiriza los ideales de los caballeros medievales al contar la historia de un cuerno que se rehúsa a servir su contenido a los maridos cornudos, todos ellos caballeros de la tabla redonda, entre los que se encuentra el propio rey Arturo. Solamente hay un caballero que no ha sido engañado por su esposa: el caballero Caradoc. Esta obra se conserva en un manuscrito de finales del siglo XII que se encuentra en la Bodleian Library de la Universidad de Oxford: Digby MS (86). Por otro lado, un dato curioso es que, en el Barrio el Progreso perteneciente a la pedanía de San Benito, en el Municipio de Murcia, llamado en un tiempo Barrios de las Ranas, existe una torre conocida como la Torre Caradoc, un antiguo caserón del siglo XIX situado en una huerta.

En las dos partes de la novela Caradoc del profesor Zapata Ros, Caradoc, sin embargo, hace referencia a uno de sus personajes principales: John Caradoc, que, curiosamente, tuvo un papel destacado en la historia porque fue un personaje real que se alterna o actúa a la par con otros personajes reales con los que va a interactuar vívidamente a lo largo de las dos partes de la novela, como la murciana Joaquina Plana Riquelme. Pero, ¿quién fue este John Caradoc que

copa ambas entregas, y en torno al cual giran una serie de situaciones tan intrigantes como azarosas? La historia nos cuenta que el irlandés John Hobart Caradoc (Dublín 1799-Bayona 1873), segundo barón Howden, fue ministro plenipotenciario británico en la embajada británica de Madrid desde 1850 hasta 1858. John Caradoc fue, además, un par británico, es decir, un alto miembro de la nobleza británica, desde 1819. John Caradoc se destacó como político y soldado en la batalla de Vinegar Hill (1798), Enniscorthy, condado de Wexford, en el marco de la rebelión irlandesa. En 1830 contrajo matrimonio con la princesa Catherine Bagration (condesa Skavronskaya). Su padre, el general John Francis Cradock, primer barón Howden (1759-1839), fue también un par británico (ingresó en la Cámara de los Comunes irlandesa para Clogher en 1785), además de político y soldado, el cual comandó el décimo tercer regimiento en las Indias Occidentales (Antillas y Bahamas) en 1790, y en 1793 un batallón de granaderos. Cradock. Su abuelo, John Cradock (¿1708?-1778), fue arzobispo de Dublín desde 1772.

John Hobart Caradoc intervino en acontecimientos históricos de enorme significación histórica para Gran Bretaña. Mencionamos el bloqueo del estuario del Río de la Plata en Argentina, la revolución de 1854 en España, las guerras napoleónicas, las guerras del Norte, la restauración de la monarquía en Francia después de Napoleón, y el Congreso de Viena, entre otros. El bloqueo francés al Río de la Plata, por ejemplo, comenzó entre el 28 de marzo de 1838 y el 29 de octubre de 1840. Durante este tiempo la escuadra francesa impidió cualquier relación comercial con Buenos Aires y los puertos fluviales de la denominada Confederación Argentina. El motivo fue la negativa del gobernador de Buenos Aires Juan Manuel de Rosas (1793-1877) a aceptar eximir a los súbditos franceses residentes en Buenos Aires y con propiedades allí, de realizar el servicio militar, y que Francia recibiera como nación un trato de favor por parte de la Confederación Argentina. Juan Manuel de Rosas se amparó en un decreto de abril de 1821 en virtud del cual el gobierno de la provincia de Buenos Aires extendía la obligación del servicio militar a los extranjeros “propietarios de bienes raíces o tiendas, que tuviesen profesión liberal o que ejerciesen arte mecánica”, y a todos aquellos que hubiesen residido más de dos años consecutivos en la provincia.

De manera adicional, otra ley de 1823 obligaba a los extranjeros en tránsito en las milicias en caso de inminente peligro a realizar el servicio militar. No obstante, debido al apoyo que los británicos habían concedido a la independencia de Argentina, los súbditos británicos estaban exentos de cumplir el servicio militar en Buenos Aires. Todas estas circunstancias provocaron una reacción inmediata por parte de Francia que trató de fomentar numerosos movimientos revolucionarios en el Uruguay en la provincia de Buenos Aires, entre otras acciones hostiles. En cuanto a la revolución de 1854, esta se inició cerca del pueblo madrileño de Vicálvaro tras el enfrentamiento entre las tropas del sublevado general Leopoldo O'Donnell y las tropas gubernamentales. Este pronunciamiento militar fue seguido también por un levantamiento popular acaecido entre el 28 de junio y el 28 de julio de 1854 bajo el reinado de Isabel II.

John Hobart Caradoc aparece de esta manera como uno de los protagonistas centrales de ambas novelas en torno al cual se concitan, como dijimos más arriba, otros personajes no menos importantes, como el de Joaquina Plana Riquelme, un personaje muy importante en la vida de Caradoc, en especial, en los momentos más difíciles de su vida. Joaquina fue una humilde murciana (“huertana” en palabras del autor) que por circunstancias de la vida terminó trabajando como actriz y cantante en un café de Cartagena. El autor no se demora mucho en presentar a estos dos personajes, y lo hace al comienzo mismo de la primera parte de una novela que sabe a “crónica” y a “trabajo de investigación”:

Baste decir que a través de distintas circunstancias me he encontrado con una serie de datos, con retazos de hechos, descritos y valorados desde distintas perspectivas, que me han puesto delante de unas vidas apasionadas: las de Joaquina Plana, una humilde huertana, y John Caradoc, un noble inglés con gran influencia en el mundo de su época y aún en el actual. He

contrastado las informaciones obtenidas, verificando con la información de que he dispuesto fechas, personajes y hechos. He aceptado aquellos que son consistentes, o que coinciden en dos o más documentos. Con todo ese material he escrito otro libro. El resto, lo que no está documentado hasta ahora, lo he completado de manera compatible con lo que es crónica, y constituye la fabulación del autor, lo que humildemente podemos considerar la creación literaria.

La historia y las dos partes de Caradoc nos dicen que Joaquina tuvo con él una relación sentimental intensa, y se mantuvo fiel y unida a él cuando este cayó en desgracia después de la revolución de 1854 y en sus últimos momentos en Murcia y Bayona, momentos en los que el aristócrata inglés carecía de dinero y estaba enfermo. Caradoc había sido expulsado por Narváez por inmiscuirse más de lo que hubiera sido deseable en los asuntos de estado concernientes a España. Esto ocurrió en circunstancias adversas en las que el segundo barón Howden contaba con problemas de salud, y se había retirado de la diplomacia sin derecho a pensión. Joaquina no llegó a casarse oficialmente con él, pero si existen testimonios históricos fehacientes de que Caradoc se divorció de su esposa mencionada más arriba: Ekaterina Pavlovna Skavronskya, princesa de Bagration e hija del conde Pavel Vassilievich Skavronsky. La convivencia de Caradoc con su esposa fue compleja y poco duradera. La princesa de Bagration fue una mujer refinada, exquisita, libertina, extravagante (en especial en los últimos momentos de su vida) y muy bien relacionada con algunos de los personajes políticos, intelectuales, militares y sociales más destacados y notables de su época, como, por ejemplo, Goethe, Balzac, Víctor Hugo, Mijaíl Bulgákov, el arquitecto Metternich, Catalina de Rusia o Gregorio Potemkin.

Ekaterina fue, además, una mujer poco común para la época que no se dejó amilanar por las estrictas convenciones sociales de la época, y se mostró siempre emancipada e independiente, en especial en las lides del amor y en su relación con los hombres, a los que escogía o bien como amigos o bien como amantes. Parece que Víctor Hugo se dejó inspirar por ella, pues aparece en algunas de sus obras bajo nombres y personajes diferentes. Asimismo, Ekaterina recibió una esmerada educación en la corte de la emperatriz Catalina II y de la emperatriz María Feodorovna, respectivamente. En los últimos momentos de su vida la princesa de Bagration se destacó por mostrar públicamente una conducta inusual para una dama de su condición y clase, lo que llamó la atención de la sociedad de su época y provocó comentarios desopilantes y no exentos de malicia, en especial cuando llegó a aparecer despampanante ante notorios personajes públicos vestida de ninfa. La princesa de Bagration murió en Venecia el 11 de junio de 1857. John Hobart Caradoc no fue feliz con Ekaterina, pero sí lo fue, y mucho, con Joaquina Plana Riquelme, según atestigua la historia y las dos novelas de Caradoc en boca de los amigos británicos del propio Caradoc.

Los amigos de Caradoc pudieron comprobar de primera mano esta radiante felicidad en la pareja en un lugar mágico y reverberante como la Torre Caradoc. Asimismo, me consta que el famoso periodista de ABC y Blanco y Negro, dibujante y académico de la Real Academia Bellas Artes de San Fernando nacido hacia 1884 en la localidad lucense de Monforte de Lemos, don Antonio Méndez Casal, dio cuenta de la dicha entre Caradoc y Joaquina en un artículo publicado en ABC. Parece, además, que John Hobart Caradoc dejó en herencia a Joaquina los bienes materiales que poseía, entre ellos la Torre Caradoc, torre que mandó construir, si no estoy errado, después de conocerla, y que estaba situada cerca de la Ermita del Rosario. Sin embargo, en las dos partes de Caradoc, suceden muchos eventos de importancia más, como la desaparición en misteriosas circunstancias de un cuadro del primer barón Howden, padre de John Hobart Caradoc, pintado por Thomas Lawrence (1769-1830). Para entender la importancia de este hecho, hay que considerar la fama de la que gozaba este pintor inglés nacido en Bristol, y que desde niño había sobresalido como un magnífico pintor, especialmente

como retratista. En 1794 Lawrence fue nombrado miembro de la Royal Academy, institución en la que había estudiado algún tiempo, y en 1815 el rey Jorge III le concedió el título de Sir. Lawrence fue un pintor de estilo romántico, elegante y de buen gusto muy cotizado en su época. No es de extrañar que la desaparición de uno de sus cuadros, el retrato del primer barón Howden, causase cierto revuelo. Lawrence fue el autor de La señora Peel, John Fane, X conde de Westmoreland, o El papa Pío VII y el archodique Carlos de Austria.

En alternancia con los personajes principales, en las dos partes de Caradoc aparecen otros personajes secundarios, como el de Isabel II que sirve fundamentalmente para contextualizar la primera parte de Caradoc y otorgar al conjunto de la novela cierto sabor y sazón histórico. Al fin y al cabo, los personajes principales viven durante parte de su mandato. Isabel II (1830-1904) fue reina de España entre 1833 y 1868. Su aparición en la escena pública como monarca provocó que Carlos María Isidro, hermano de Fernando VII y tío de Isabel II, se alzara contra ella apoyado por los llamados “carlistas”. Isabel II fue la hija del rey Fernando VII y de María Cristina de Borbón-Dos Sicilias (su cuarta esposa). Isabel II ascendió al trono el 29 de septiembre de 1833 después de morir su padre Fernando VII, pero como contaba tan solo tres años de edad su madre María Cristina de Borbón fue nombrada regente del reino. Al cumplir la mayoría de edad, el 10 de noviembre de 1843 Isabel II juró la Constitución ante las cortes. Parece ser que el hecho de querer postularse como presidenta del gobierno derivó en la Revolución de 1868. Sin embargo, es de destacar que España sufrió una etapa de modernización importante, como la creación de la red ferroviaria, la construcción del Canal de Isabel II, el trazado de carreteras y puentes, la señalización de costas y la apertura de las universidades que Fernando VII había cerrado. Tras la Revolución de 1868 Isabel II abandonó España y se exilió en Francia donde recibió el apoyo de Napoleón III y de Eugenia de Montijo.

La primera parte de la novela se divide en ocho (8) capítulos seguidos de un prefacio. Los 5 primeros están dispuestos por fechas claves destinadas a presentar en parte a algunos de los personajes y sus circunstancias personales históricas e intrahistóricas (como el de Joaquina Plana Riquelme). El primer capítulo se centra en diciembre de 1972; el segundo, en noviembre de 1977; el tercero, en febrero de 1859; el cuarto, el 30 de julio de 1847; y el quinto, el 16 de agosto de 1854. El capítulo sexto se ocupa de Isabel II; el séptimo presenta a Joaquina Plana Riquelme; y el octavo presenta y describe el café de cante de Cartagena. La segunda parte de Caradoc contiene también un prefacio en el que se lleva a cabo una detallada descripción del militar y diplomático británico a cargo de Alejandro Alonso y consta de siete (7) capítulos.

El primer capítulo se centra en un personaje llamado Pepa la Malagueña (Josefa Montenegro); el segundo se ocupa de una fecha como marzo de 1924; el tercero lleva por título El encuentro y se centra en julio de 1858; el cuarto lleva por título La enfermedad y en agosto de 1858; el quinto lleva por título La recuperación; el capítulo sexto lleva por título El tren a Murcia; y el séptimo El statu quo. Otros personajes que se concitan en las dos partes de Caradoc (algunos sirven como entrada dentro del menú de ambos relatos) son Perico el conejo, Anita (una secretaria), Juan Pérez el Coscorrónes, Zacarías, Rosa Lozano Pagán, don Juan Pérez Belando, Juan García Torres, Eugenia (Genny), Christian Franzen, Ángel, Lady Theodosia Meade, Naopoleón, Arthur Wellesley, Sir Thomas Maitland, Lord Cunning, Jean Hanoteau, Karl Marx, Juan Manuel de Rosas y su hija, Manuelita Rosas, Josefa Montenegro (ya mencionada), el Duque de Infantado, Otway, Doña María Cristina, etc.

En cuanto al estilo de los dos partes de Caradoc destacamos las siguientes características:

- Ambos relatos amalgaman historia con ficción histórica, o viceversa.
- Se recurre a los diálogos para dinamizar la narrativa y como reflejo de la personalidad y condición social y cultural de los personajes.
- Los diálogos pueden ser coloquiales o informales o cercanos a la jerga o “antilinguaje”.
- El estilo de ambas entregas de Caradoc es propio de un libro o manual de historia, una enciclopedia, una crónica histórica o un trabajo de investigación de campo. Una primera prueba de ello podría ser el uso de notas a pie de página que se despliegan en la primera parte de Caradoc: 1 El 19 de diciembre de 1831 su nombre fue cambiado legalmente a John Hobart Caradoc, por un afán de su padre de enlazar su nombre y su linaje con el héroe celta Caractaco, caudillo de los catuvellani. Su padre, que recibió el título de nobleza del Reino Unido en los honores de la coronación, cambió el nombre de la familia, ridículamente, a Caradoc en enero de 1832, habiéndose convencido de que descendía de Caractaco, también llamado Caradawc; una segunda prueba, las citas a las que recurre el autor de Caradoc como prueba de autoridad o testimonio histórico: Sydney Smith observó que "hay algo en él, de lo cual no es consciente y por tanto no controla" y Lady Holland, pasado el deslumbramiento inicial, dejó caer que “deslumbra, pero no ilumina”; y una tercera prueba, la recurrencia a fuentes primarias como cartas: Alors qu'il n'était que ministre à Dresde, M. de Metternich s'était pris de passion pour une belle russe, la princesse Catherine Pavlovna Bagration, femme du général qui, à la tête de l'une des armées moscovites, devait périr en 1812 d'une blessure reçue à la bataille de Borodino, u opúsculos como Los salones del pasado. Memorias íntimas: «Qu'on se figure un jenne visage, blanc comme l'albâtre, légèrement coloré de rose, des traits mignons, une physionomie douce, expressive et pleine de sensibilité, un regard auquel sa vue basse donnait quelque chose de timide et d'in certain, une taille moyenne mais parfaitement prise, dans toute sa personne une mollesse orientale unie à la grâce andalouse».
- El estilo, además, es claro y ameno, y recoge elementos culturales o culturemas que ayudan a entender el componente histórico, social, político y cultural de ambos relatos. Tales culturemas recogen personajes históricos de las diferentes épocas que aparecen en ambos relatos, intelectuales y escritores (como el poeta Percy Bysshe Shelley, Byron, Godwin, Hunt, Mary Shelley-Harriet Westbrook), obras literarias (como Prometheus Unbound), obras artísticas, instituciones políticas, sindicatos, marcas de coche, palacios (como el Palacio de Guilhou), pueblos (como Chamartín de las Rosas), barrios, lugares o zonas geográficas (en España: Murcia, Cartagena, castillo de San Julián; Inglaterra: Londres, Eton; Portugal-Lisboa; París; Malta; El Cabo; Egipto; Grecia; Turquía; San Petersburgo; Buenos Aires), calles (del Osario), carreteras, autovías, bares, y otros referentes culturales en un marco temporal sincrónico (transversal) y diacrónico (longitudinal).
- Llama la atención la relación entre los cronotopos y el léxico empleado.

- La primera parte de la novela está escrita en tercera persona gramatical (muy cercana, creemos, al narrador heterodiegético).
- Este narrador que hace uso de la tercera persona gramatical salta de un personaje a otro. Tan pronto nos cuenta algunas anécdotas o vicisitudes propias de Lord Howden (amante de Joaquina Plana Riquelme) como pinta y describe un cuadro en el que está enmarcado Juan García Torres.
- Un aspecto muy dinamizador de la primera parte de Caradoc es su frecuente salto en el tiempo: de la etapa contemporánea al pasado (flashback-analepsis) fundamentalmente.
- El autor de Caradoc gusta de trazar y delinear muy bien sus personajes abundando en su contexto cultural, histórico, social y cultural.
- Asimismo, el autor de Caradoc le complace incorporar en su narración numerosos extranjerismos, bien sean latinismos del tipo rara avis bien anglicismos del tipo blend, jet set, y lo hace siguiendo las normas y convenciones del DRAE a tal efecto.
- En la narración de las dos partes de Caradoc descollan abiertamente dos tipos de descripciones: la effictio o prosopografía y la etopeya. Estos dos tipos de descripciones se aplican muy bien en las descripciones de los personajes principales: de John Hobart Caradoc y de Joaquina Plana Riquelme, y de otros personajes secundarios, pero no menos relevantes, como Ekaterina Pavlovna Skavronskya, Josefa Montenegro, etc.

Ejemplo 1:

John Hobart Cradock, cuando utilizaba ese apellido todavía, mostró una gran habilidad lingüística en el periodo de su primer aprendizaje, cuando sus padres lo internaron en Eton. Esta habilidad, junto con las relaciones adquiridas en este periodo, le sería de gran utilidad y se manifestaría en su carrera, no solo el dominio de idiomas, particularmente el español y el francés que utilizaría como suyos, sino en una más que corriente habilidad para la persuasión y para la diplomacia, pero también en el cortejo a las mujeres. Recuerda, sonriendo para sus adentros, las leyendas y comidillas que circulaban sobre este tema en su estancia en París. Realmente sobre este talento que Dios le había concedido no se podía quejar. De hecho, como demuestran las misiones que le encomendaron, el aspecto galán era un factor fundamental en la resolución de los problemas políticos o diplomáticos.

Ejemplo 2:

En este capítulo tenemos varios ejemplos de ese hecho. Sus protagonistas encarnan esos conflictos morales. Uno de estos personajes es una mujer: Josefa Montenegro. Es un personaje perverso en extremo, no sólo al uso de la época, según las convenciones, sino incluso según criterios que las trascienden. Se podría considerar pues un ejemplo de vicio, de maldad y de depravación en cualquier época. Sin embargo, no sólo ella misma y sus más próximos, sino la sociedad y la historia han considerado, en buena parte, su obra, ya que no a su persona, loable y ejemplar, y así está atestiguado por muchos y escrito en documentos y en relatos. El personaje al que nos referimos quizá en algunos sitios puedan localizarla por su sobrenombre: Pepa la Malagueña.

Ejemplo 3:

Por contrapartida, este capítulo tiene otro protagonista. Es el Duque de Infantado. Un personaje de estricta moral, y de una inmensa responsabilidad en la historia y en el mundo de las ideas, de las ideas tradicionales y conservadoras en nuestro país, todo ello a fuer de ilustrado. Pues bien, este personaje conservador, de una estricta moral que defiende con vigorosos argumentos en los foros y en los ámbitos públicos, es capaz como veremos de someterse a las más grandes sevicias morales, carentes de razón y de lógica.

Y ahora, algunas palabras del prefacio de la primera parte que explica la motivación de Caradoc en sus dos partes como obra literaria:

Muchas veces las fachadas de las casas, los imafrentes, las esculturas que los adornan o la decoración de espacios comunes que los llenan, han sido la manifestación más o menos explícita de un amor, de una ambición, o la ostentación de un poder personal. Cuando abrimos las páginas de una novela sabemos que hay una historia de una pasión. Pero ¿cuántas veces no habremos pasado frente a un edificio singular ignorando lo que hay detrás? Uno de los propósitos de este libro es desentrañar las pasiones que llevaron a escribirlas en una de estas mansiones, los sentimientos que hay escritos detrás de su fachada y en ésta. La infancia y la juventud de quien les escribe han estado vinculadas a una dirección, mejor a una palabra en una dirección: “caradoc” ¿Quién o qué cosa es “caradoc” o “cáradoc”? Cuando se escribían cartas en papel, era una parte, una línea, en la dirección del sobre de la carta o en el remite, pero era también una bella mansión en ruinas en lo más profundo de la huerta, con su carga de misterio e intriga.

Finalmente, para concluir, podría añadirse que Miguel Zapata Ros es, en cierta manera, un rescatador de la historia como disciplina, historia a la que hace pasar por un filtro, quizá periplo de fabulación literaria, con el objeto de ofrecer al lector una obra entretenida, didáctica, histórica y moral, y digo moral porque en la segunda parte de la novela Joaquina Plana Riquelme da lo mejor de sí misma como ser humano, de manera especial en la última etapa de la vida del segundo barón Howden. Además, en las dos entregas de la novela Caradoc el autor nos hace saber de qué pasta o materia está forjada la naturaleza humana, mostrando sus virtudes y sus pecados junto con sus anhelos e inquietudes, sus sueños y ambiciones, sus actos y sus consecuencias, en definitiva, su vida y su muerte.

INSTITUTO SUPERIOR DE LENGUAS
Facultad de Filosofía
Universidad Nacional de Asunción
PARAGUAY
2023

REVISTA



ISSN: 2707-1642 (en línea)
ISSN: 2707-1634 (impreso)