

RAEL

Revista
Electrónica de
Lingüística
Aplicada

*Electronic
Journal of
Applied
Linguistics*

Volumen 23, Número 1, 2024

Volume 23, Issue 1, 2024

rael.aesla.org.es



AESLA

Asociación Española de Lingüística Aplicada

ISSN 1885-9089

Editora

Ana Ibañez Moreno, *Universidad Nacional de Educación a Distancia (UNED), España*

Editora de Reseñas

Cristina Calle Martínez, *Universidad Complutense de Madrid, España*

Comité de redacción

Clara Rodríguez Salgado, *Univesidad Camilo José Cela, España*

Miguel Lacalle Palacions, *Universidad de La Rioja, España*

Responsable de maquetación

Enrique Mateo Martínez, *Triskelion Diseño Editorial*

Consejo Asesor

Carlos Acuña Fariña, *Universidade de Santiago de Compostela, España*

Julio Borrego Nieto, *Universidad de Salamanca, España*

Christopher Butler, *Swansea University, Reino Unido*

Angela Downing, *Universidad Complutense de Madrid, España*

Daniel Everett, *Illinois State University, Estados Unidos*

María Teresa Fanego Lema, *Universidade de Santiago de Compostela*

Elisabeth Guest, *Leeds University, Reino Unido*

Rolf Kailuweit, *Albert-Ludwigs-Universität, Alemania*

María Antonia Martín Zorraquino, *Universidad de Zaragoza, España*

Jan Nuyts, *University of Antwerp, Bélgica*

Brian Nolan, *Institute of Technology Blanchardstown Dublin, Irlanda*

Ignacio Miguel Palacios Martínez, *Universidade de Santiago de Compostela, España*

José Ignacio Pérez Pascua, *Universidade da Coruña, España*

Susan Pintzuk, *University of York, Reino Unido*

Francisco J. Ruíz de Mendoza Ibáñez, *Universidad de la Rioja, España*

Cristina Sanz, *Georgetown University, Estados Unidos*

Consejo Científico

Lucía Loureiro Porto, *Universitat de les Illes Balears, España*

Cristina Suárez Gómez, *Universitat de les Illes Balears, España*

Maria Assumpció Rost Bagudanch, *Universitat Illes Balears, España*

Gloria Vázquez García, *Universitat de Lleida, España*

Stéphane Patin Patin, *Universidad de Paris Diderot Paris Sorbonne Cité, Francia* Guadalupe

Aguado de Cea, *Universidad Politécnica de Madrid, España*

María Pilar Agustín Llach, *Universidad de La Rioja, España*

José Luis Aliaga Jiménez, *Universidad de Zaragoza, España*

Inmaculada Álvarez de Mon y Rego, *Universidad Politécnica de Madrid, España*

Gemma Andújar Moreno, *Universitat Pompeu Fabra, España*

M^a Luisa Arnal Purroy, *Universidad de Zaragoza, España*

Inmaculada Concepción Báez Montero, *Universidade de Vigo, España*

Elena Bárcena Madera, *Universidad Nacional de Educación a Distancia (UNED), España*

Silvia Carmen Barreiro Bilbao, *Universidad Nacional de Educación a Distancia (UNED), España*

Asunción Barreras Gómez, *Universidad de La Rioja, España*

Julio Borrego Nieto, *Universidad de Salamanca, España*
Teresa Cadierno López, *University of Southern Denmark, Dinamarca*
José Antonio Calañas Contiente, *Universitat de Valencia, España*
María Ángeles Calero Fernández, *Universitat de Lleida, España*
Javier Calle Martín, *Universidad de Málaga, España*
Andrés Canga Alonso, *Universidad de La Rioja, España*
María Luisa Carrió Pastor, *Universitat Politècnica de València, España*
Francisco Cortés Rodríguez, *Universidad de La Laguna, España*
Alejandro Castañeda Castro, *Universidad de Granada, España*
Manuela Catalá Pérez, *Universidad San Jorge de Zaragoza, España*
Isabel de la Cruz Cabanillas, *Universidad de Alcalá, España*
Javier Díaz Vera, *Universidad de Castilla La Mancha, España*
Pilar Durán Escribano, *Universidad Politécnica de Madrid, España*
Ramiro Durán Martínez, *Universidad de Salamanca, España*
Almudena Fernández Fontecha, *Universidad de La Rioja, España*
Lydia Fernández Moreno, *Universitat Pompeu Fabra, España*
Inmaculada Fortanet Gómez, *Universitat Jaume I, España*
Luz Gil Salom, *Universitat Politècnica de València, España*
María Victoria Galloso Camacho, *Universidad de Huelva, España*
María Ángeles García Aranda, *Universidad Complutense de Madrid, España*
María del Carmen Guarddon Anelo, *Univ. Nacional de Educación a Distancia (UNED), España*
Francisco Gonzálvez García, *Universidad de Almería, España*
Juan Manuel Hernández Campoy, *Universidad de Murcia, España*
Enrique Jiménez Ríos, *Universidad de Salamanca, España*
Jeannick Yvonne Le Men Loyer, *Universidad de León, España*
Juana Marín Arrese, *Universidad Complutense de Madrid, España*
Javier Martín Arista, *Universidad de La Rioja, España*
María Antonia Martín Zorraquino, *Universidad de Zaragoza, España*
Uwe Kjær Nissen, *University of Southern Denmark, Dinamarca*
Ana Ortigosa Pastor, *City University of New York, Estados Unidos*
Alfonso Palma Reyes, *Universidad de Granada, España*
Juan Carlos Palmer Silveira, *Universitat Jaume I, España*
Florentino Paredes García, *Universidad de Alcalá, España*
Enrique Pato Maldonado, *Université de Montréal, Canadá*
Johan Pedersen, *Universidad de Copenhague, Dinamarca*
Sandra Peña Cervel, *Universidad de La Rioja, España*
Hermógenes Perdiguero Villarreal, *Universidad de Burgos, España*
Javier Pérez Guerra, *Universidade de Vigo, España*
Ulrik Sandborg-Petersen, *Aalborg University, Dinamarca*
Cristina Sanz, *Georgetown University, Estados Unidos*
Eulalia Sosa Acevedo, *Universidad de La Laguna, España*
Keith Douglas Stuart, *Universitat Politècnica de València, España*
Cristina Tejedor Martínez, *Universidad de Alcalá, España*
Carmen Valero Garcés, *Universidad de Alcalá, España*

Artículos

Compiling a Corpus of User-Generated Content Units for the Detection of Social Problems	1
<i>Rocío Jiménez Briones</i>	
Classification of Technology-based Language Learning Resources according to Pedagogical and Functional Criteria	19
<i>Rafael Seiz Ortiz y María Luisa Carrió Pastor</i>	
¿Tienen GPT-3.5 y GPT-4 un estilo de escritura diferente del estilo humano? Un estudio exploratorio para el español	34
<i>Lara Alonso Simón, Ana María Fernández-Pampillón Cesteros, Marianela Fernández Trinidad y Manuel Márquez Cruz</i>	
Anglicismos y creación léxica en el discurso de gamers españoles: el caso de Valorant	55
<i>Elena Fernández de Molina Ortés</i>	
Una primera aproximación descriptiva a la morfosintaxis de los hablantes de catalán y español con síndrome de Williams	76
<i>Elga Cremades y Lluís Barceló-Coblijn</i>	
Investigating recontextualisation processes in scientific digital practices: the SciDis database	100
<i>Daniel Pascual y Ana Eugenia Sancho-Ortiz</i>	
Expresión e interacción orales y autoevaluación: estudio descriptivo y de intervención con hablantes avanzados de español como lengua adicional	118
<i>Marta F. Nogueroles López</i>	
Etimologías del caló en el Diccionario de la Lengua Española de la Real Academia Española: comentario y análisis de algunos ejemplos	135
<i>Alba Macías Couso</i>	
La diversidad del español en Instagram. Entre entretenimiento y posibilidades didácticas	150
<i>Lisa Gozzi y Silvia Betti</i>	
Competencias lingüísticas y comunicación inclusiva a través de la mediación en inglés para fines específicos	168
<i>Adriana Bausells Espín y Diana Giner Alonso</i>	
Reseñas	
Book review: Barón, J., Celaya M. L. & Watkins, P. (2024). <i>Pragmatics in Language Teaching. From Research to Practice</i>	193
<i>Marta Carretero Lapeire</i>	



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 1-18
Artículo recibido: 14/08/2024
Artículo aceptado: 21/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/668>
DOI: <https://doi.org/10.58859/rael.v23i1.668>

Compiling a Corpus of User-Generated Content Units for the Detection of Social Problems

Creación de un corpus de contenido generado por los usuarios para la detección de problemas sociales

ROCÍO JIMÉNEZ BRIONES
UNIVERSIDAD AUTÓNOMA DE MADRID

Social media platforms like Facebook, X, and Instagram provide valuable information about daily and global social problems through their user-generated content units. Such platforms turn their users into *social sensors* capable of identifying problems such as violence against women or natural hazards. Within the field of crowdsensing, which aims to extract useful information from these social sensors, we introduce a proposal for identifying problems in ALLEGRO (<http://allegro.ucam.edu/>). Based on previous studies within this smart multimodal system, our research first explains the conceptual framework for addressing one of these problems, namely, elderly institutional abuse, within the text analysis module of ALLEGRO or DIAPASON. We also detail the methodology and challenges of compiling a subcorpus of tweets related to this problem. Such a specific subcorpus will contribute to ALLEGRO's comprehensive corpus of social problems, which is being built as a training set for deep learning models in text classification.

Keywords: *subcorpus; user-generated content; elderly institutional abuse; ALLEGRO, DIAPASON*

Plataformas como Facebook, X e Instagram proporcionan información muy valiosa sobre problemas cotidianos y globales gracias al contenido generado por sus usuarios, ya que estos se convierten en *sensores sociales* capaces de identificar problemas como la violencia contra las mujeres o los peligros naturales. Dentro del campo del *crowdsensing*, cuyo objetivo es extraer información útil de estos sensores sociales, presentamos una propuesta de identificación de problemas en ALLEGRO (<http://allegro.ucam.edu/>). Basándonos en estudios previos sobre este sistema, se explica el marco conceptual para abordar el problema del maltrato institucional a ancianos dentro del módulo de análisis de texto de ALLEGRO o DIAPASON. Se detalla también la metodología y los retos que plantea la compilación de un subcorpus de tuits para dicho problema social, el cual contribuirá al corpus integral de problemas de ALLEGRO que se está construyendo como conjunto de entrenamiento para modelos de aprendizaje profundo en la clasificación de textos.

Palabras clave: *subcorpus; contenido generado por usuarios; maltrato institucional a mayores; ALLEGRO; DIAPASON.*

Citar como: Jiménez Briones, R. (2024). Compiling a Corpus of User-Generated Content Units for the Detection of Social Problems. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 1-18. <https://doi.org/10.58859/rael.v23i1.668>

1. INTRODUCTION

Social media users have lately become rich sources of information on those issues that concern them in their daily life activities or in a more global perspective. Through text data, photos, videos, audios, or a combination of them, millions of opinions, worries, complaints, etc. that affect people's quality of life (QoL) are shared, every day, in social networks like Facebook, X (formerly Twitter), Instagram, or LinkedIn. As Ghani, Hamid, Targio Hashem, and Ahmed (2019: 418) explain:¹

All the statuses, tweets, comments, posts, and reviews are the user-generated content. User-generated content is a type of data that typically refers to images, text, and videos. This content comes from regular people and not necessarily in a standard form [...]. All these data may incorporate the users' personal opinions, behaviors, and thoughts, which makes the task of extracting high-quality information from such data becoming increasingly important.

Such a gigantic amount of information (whether structured, semi-structure and/or unstructured) that can be conveyed in various modalities (i.e. textual, visual, sound, etc.) is known as big data, which, along with the advances in computer tools, has opened a door not only to understanding human behavior but also to carrying out applied research, within Artificial Intelligence (AI) and data analytics, into areas like "health, astronomy, social network and geoscience" (Ghani et al., 2019: 417).

Companies and governments are interested in storing all the massive data generated in the various social media applications because, among other things, it allows them to collect customer feedback and reviews about their products, detect political tendencies, and locate disease outbreaks, catastrophes, traffic jams, etc. (Bazzaz, Haghi, Mahdipour & Jameii, 2021: 9). Thus, the individuals that create this user-generated content (UGC) units in the social networks are considered active social sensors (Sakaki, Okazaki & Matsuo, 2013; Musto, Semeraro, Lops & De Gemmis, 2015; Arthur, Boulton, Shotton & Williams, 2018) able to identify a great myriad of problems around them such as violence against women, pollution, bullying, floods, etc.² This has resulted in a multidisciplinary field of study known as *social sensing* or *social-media crowdsensing* (Li, Wu, Wang, Cheng, Chen, Zhou & Ding, 2017; Wang, Szymanski, Abdelzaher, Ji, & Kaplan, 2019). Although these terms are used differently in the literature (An & Weber, 2015), we understand that "social sensing through social media involves the analysis of digital communications to detect real-world events or situations" (Periñán-Pascual, 2023b: 264).

In this context, a large inter-university team of computer scientists and linguists is cooperating in the design of **ALLEGRO** (*Adaptive muLti-domain sociaL-media sEnsinG fRame-wOrk*), a smart multimodal system (text, audio, and image) for the prediction of social problems (<http://allegro.ucam.edu/>). Thus, within the multidisciplinary field of crowdsensing, whose ultimate goal is to derive content from social sensors by employing computational means, and drawing upon the works of Periñán-Pascual (2023a, 2023b, 2024a, 2024b), Alameda Hernández (2024), Felices Lago (2024, 2025), Jiménez-Briones and Felices Lago (2024), and Ureña Gómez-Moreno (2024), among others, we address a proposal for the detection of social problems related to the domain of POPULATION within ALLEGRO. In particular, zooming in on the problem of elderly institutional abuse, this research outlines the conceptual framework adopted for tackling such a problem in DIAPASON (*unifieD hybrId ApProach to microtext Analysis in*

¹ Following Ghani et al. (2019: 418), social media, social media sites and social networks are used interchangeably in this paper.

² Periñán-Pascual (2023b) details the main features that define the notion of UGC in the field. This research sides with his proposal "where (a) users play the role of citizens, and (b) the content describes events or situations perceived as problems that disrupt their QoL" (Periñán-Pascual, 2023b: 264).

Social-media crOwdseNsing), the ALLEGRO's module devoted to text analysis (<http://allegro.ucam.edu/diapason/Index.html>). Additionally, the steps followed to compile a corpus of tweets related to this problem are also detailed with a view to using the resulting subcorpus as part of the gold standard for the automatic detection of social problems by the ALLEGRO system (Periñán-Pascual, 2024b).

The article is organized as follows. Section 2 sketches the goals of ALLEGRO, delving into the ontological approach followed to address, linguistically and conceptually, social problems in DIAPASON. Section 3 presents the case study of one social problem: elderly institutional abuse. Its ontological treatment (3.1), codification in the format of a problem schema (3.2), and the compilation of a subcorpus of UGC units dealing with said problem (3.3) are fully detailed, too. Finally, the conclusions are presented in Section 4.

2. AN OVERVIEW OF ALLEGRO AND DIAPASON

As introduced in Section 1, ALLEGRO is a project for the development of real-time crowdsensing applications that can accurately depict the state of society as envisaged by the collective intelligence of the social networking platforms' users (Periñán-Pascual, 2024b). It comprises two components: Data Analysis and Data Fusion (Periñán-Pascual, 2023a: 226). The former, in turn, is built upon three modules, one for each of the data required (i.e. text, audio, and image): DIAPASON (text analysis), SOUND (*Social-media sOUNd aNalysis moDule*), and ADAGIO (*SociAl meDia imAGe analysIs moDule*), respectively. Thus, when microtexts incorporate audio and/or images, this expanded knowledge is combined in Data Fusion “by rejecting irrelevant information, minimising redundancy, resolving inconsistencies, and completing missing information” (Periñán-Pascual, 2023a: 226). Both Data Analysis and Data Fusion use a multi-modal data repository and a knowledge base. In other words, the same ontology and formalism (i.e. problem schemas) are shared in the ALLEGRO modules to model the knowledge obtained from the UGC units and, accordingly, formalize the most relevant features of the social problems expressed therein.

It is worth stressing that the ALLEGRO system needs to be understood within the field of social sensing or, more specifically, social media crowdsensing, whose goal is to infer knowledge through computational models by collecting, analyzing and interpreting UGC items (Li et al., 2017; Wang et al., 2019). This line of research views social network users as sensors of the society in which they live, as the problems that surround them are widely shared in posts, tweets, microblogs, etc. on daily basis. Consequently, the research developed within ALLEGRO contributes to the notion of *Smart City*: “a city that manages, in an intelligent way, all its associated resources with the aim to enhance the quality of the services provided to citizens and to improve their quality of life” (Espinoza-Arias, Poveda-Villalón, García-Castro & Corcho, 2019 as cited by Periñán-Pascual, 2023b: 265). With this aim, UGC units are looked into to uncover any type of problem that can negatively affect a community. Our focus, however, extends beyond the quality of urban infrastructure and services provided to citizens (e.g. water/gas/electricity supplies, waste network, healthcare, etc.) to include the sociological dimension of the city, as reflected in people's concerns about violence, ageing, crime, and education, among others. In this way, our proposal presents a more comprehensive alternative to already-existing smart city applications, which tend to focus primarily on the former aspect.³

It turns crucial then to comprehend the natural language that social sensors use to express the issues that impact their QoL and well-being, both for AI and for natural language processing (NLP). Thus, we, as linguists within the ALLEGRO project, devote ourselves to the analysis

³ The interested reader can find a detailed account of the various frameworks proposed to explore smart cities in Periñán-Pascual (2023b).

of English and Spanish microtexts in UGC through DIAPASON, “a workbench of exploration and experimentation with UGC written in English or Spanish to automatically detect a variety of problem types by integrating natural language processing (NLP), machine and deep learning, and knowledge engineering techniques” (Periñán-Pascual, 2023b: 265). As one of the DIAPASON’s goals is to assist the NLP module to accurately classify the topics and recognize the keywords of the problems posted in UGC items, let us briefly comment on its ontology, which is structured into 4 levels: problem realms, problem dimensions, problem domains and problem types (Periñán-Pascual, 2023a, 2023b):

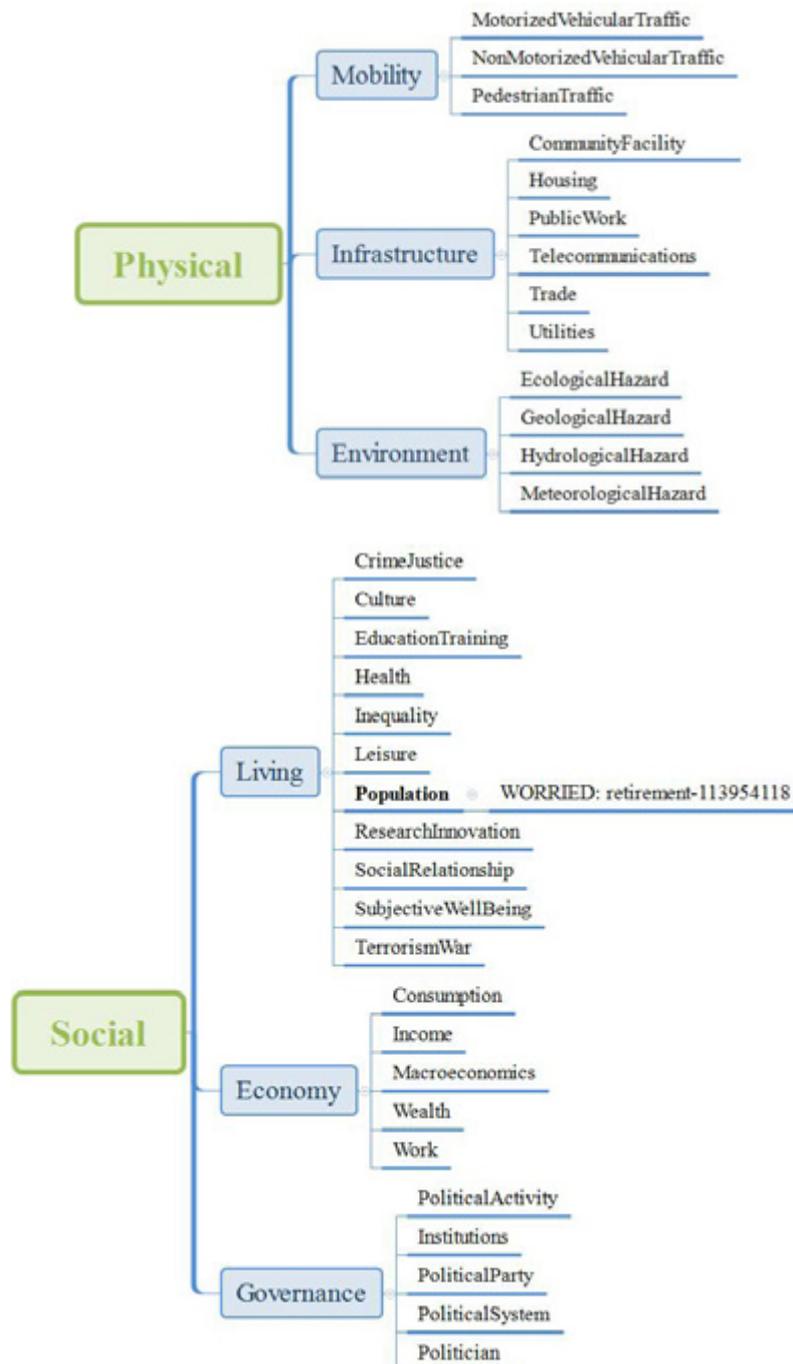


Figure 1. Architecture of the DIAPASON ontology

As Figure 1 displays, the upper ontological level comprises the PHYSICAL and Social problem realms, and six problem dimensions: MOBILITY, INFRASTRUCTURE, ENVIRONMENT, LIVING, ECONOMY, and GOVERNANCE.⁴ The intermediate ontological level includes the problem domains, such as COMMUNITYFACILITY or SUBJECTIVEWELLBEING, which store most of the areas about which citizens can complain in a community, whereas the lower ontological level refers to the specific types of problems that affect those who live in smart cities, using for that purpose a language-independent knowledge structure known as *problem schema*. For example, in the social domain POPULATION the problem *fear of retirement* is represented by means of the problem schema *WORRIED: retirement-113954118*. We turn now to explore the key role of problem schemas in DIAPASON.

As already mentioned, one of the central purposes of DIAPASON consists in automatically detecting the community problems that affect the citizens as they publish them in social media microtexts. To this end, this module first needs to have stored these problems in the ontology. Thus, the DIAPASON ontology conceptualizes the semantics of the community problems by means of problem schemas, which are the constructs employed to formalize such problems in the SOCIAL and PHYSICAL realms of the ontology. Below is the interface for editing said ontology and incorporating the various problem schemas:

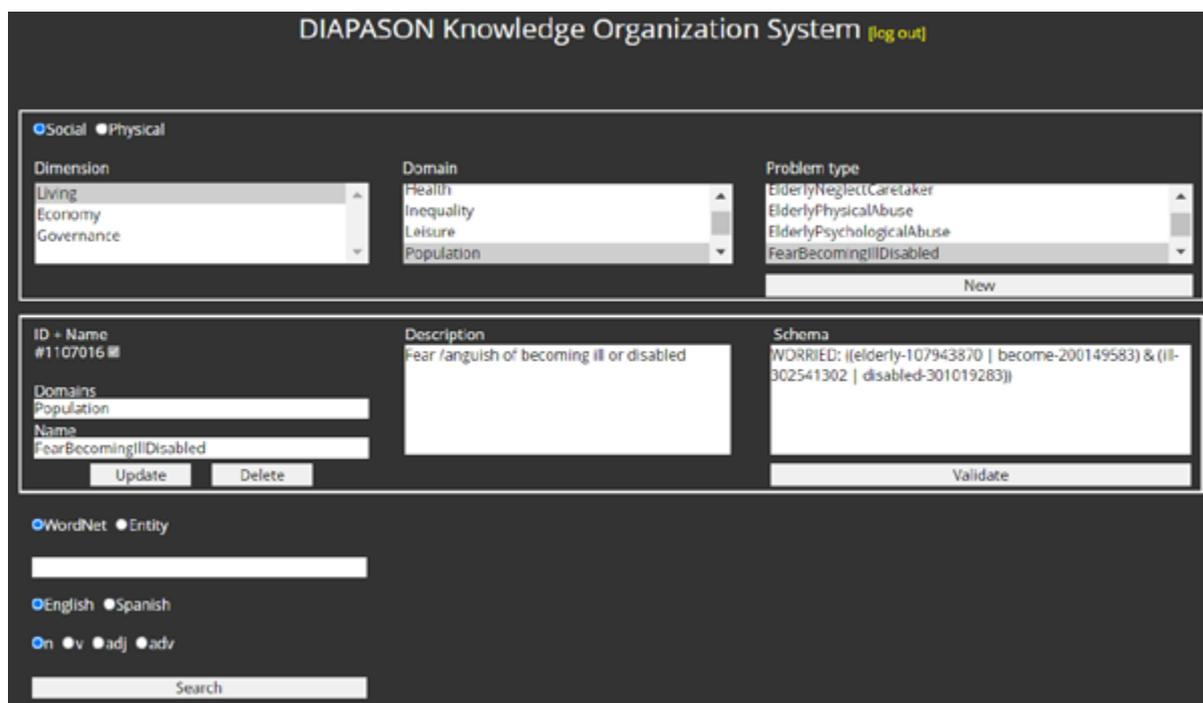


Figure 2. The DIAPASON Editor

⁴ The modeling of this ontological level stems from foundational studies on smart cities like Giffinger, Fertner, Kramar, Kalasek, Pichler-Milanovic, and Meijers (2007), Govada, Spruijt, and Rodgers (2017), and Appio, Lima, and Par (2019). A comprehensive account of how this level and the intermediate one were designed and implemented can be found in Perrián-Pascual (2023b, 2024a, 2024b). The lower level of the ontology is still being populated by the ALLEGRO linguists, although many of the community problems belonging to dimensions such as ENVIRONMENT and GOVERNANCE (Perrián-Pascual, 2024a, 2024b), LIVING (Alameda Hernández, 2024; Jiménez-Briones & Felices Lago, 2024; Ureña Gómez-Moreno, 2024), and ECONOMY (Felices Lago, 2024, 2025) have already been identified, listed and formalized.

The researchers modeling this type of knowledge in the ontology must either select or fill in the appropriate information in the DIAPASON Knowledge Organization System for each of the social problems already identified. By way of illustration, Figure 2 displays the case of the problem type *fear of becoming ill or disabled*, for which the appropriate dimension (LIVING) and domain (POPULATION) need to be selected in the interface. Likewise, an English short description of the problem must be typed in, along with its given name, which takes the shape of a list of running characters without spaces, using capital letters for the initial of each content word in the name: FEARBECOMINGILLDISABLED. Finally, the problem schema itself is provided: WORRIED: ((elderly-107943870 | become-200149583) & (ill-302541302 | disabled-301019283)). A word is needed to outline the two building blocks that make up DIAPASON problem schemas, that is, the illocutionary and locutionary components:

Table 1. Components of problem schemas

Illocutionary component	WORRIED
Locutionary component	((elderly-107943870 become-200149583) & (ill-302541302 disabled-301019283))

On the one hand, the illocutionary component expresses the non-propositional meaning of the schema, in this case, as Table 1 captures, through the communicative function WORRIED, which reflects the users' negative perception of this issue and which will activate linguistic expressions, in the microtexts, such as *I am worried that []*, *I fear that []* or *What if []?*⁵ This component, which relies upon Searle's (1969) Speech Act Theory, is optional as not all the community problems require the actual lexicalization of the users' negative perceptions.⁶ On the other hand, the locutionary component is compulsory because it contains the propositional meaning of the schema, that is, the semantic content of the problem itself. To do so, WordNet synsets (Princeton University, 2010) are employed as the conceptual units of the description, i.e. *elderly-107943870*, *become-200149583*, *ill-302541302* and *disabled-301019283*, as well as several logical and conceptual operators to connect the selected synsets and reflect notions of addition (&), inclusive/exclusive disjunction (|, ^), quantity (M, P) and negation (N). It should be noticed that direct access to the WordNet database is granted by selecting the WordNet option in the Editor and typing in, in natural language, each of the key words of the problem under study (e.g. "ill"), which will list the appropriate synsets with their original numerical code. Thus, the problem schema shown in Table 1 should be interpreted as the sensor's worry about ageing in relation to becoming sick and/or disabled.

After identifying, listing and formalizing as problem schemas the main community problems related to the domains established in Figure 1, the ALLEGRO linguists moved to the second stage in the process of ontological modelling in DIAPASON: the compilation of the subcorpora of microtexts from X. This is a crucial step in the ALLEGRO project because, to test how effective the DIAPASON module is in the detection of problems reflected in UGC units, as they are shared by citizens in the social media, these subcorpora will eventually result in a gold standard corpus for the training phase of the ALLEGRO algorithm, whose ultimate goal is to automatically sort out any microtext as either related or unrelated to a specific social problem. As Periñán-Pascual (2024b: 20) points out: "whereas the corpus is characterised by

⁵ Fernández-Martínez (2024) lists the set of communicative functions and their linguistic realizations that can be used to express the social sensors' attitude towards a topic.

⁶ Although indirect speech acts are pervasive in language—cases in which a communicative function is present in a text but without a clear systematic linguistic correspondence (e.g. Holtgraves, 1994; Stefanowitsch, 2003; Pérez, 2013; Trott & Bergen, 2017, to name a few)—, we are not yet able to address them at this stage of the project.

linguistic heterogeneity, each subcorpus displays more homogeneity. As a result, this division of labour considerably reduces the time of corpus compilation.” Section 3 illustrates this phase with the social problem of elderly institutional abuse.

3. CASE STUDY: THE SOCIAL PROBLEM OF ELDERLY INSTITUTIONAL ABUSE IN DIAPASON

Section 2 has outlined the main goals of ALLEGRO, zooming in on DIAPASON, the module in charge of textual processing. Before addressing the compilation of a subcorpus of UGC units related to elderly institutional abuse (Section 3.3), the conceptual framework employed to deal with said problem is needed, in particular, a brief introduction about the ontological modeling of POPULATION (Section 3.1), within which the problem is included, as well as an explanation of the creation of its schema (Section 3.2).

3.1. Knowledge Modeling in POPULATION

According to the sociology literature (Parrillo, 2008; Bates & Ciment, 2013; Eitzen, Baca Zinn & Eitzen Smith, 2014; Marginean, 2014; Seccombe & Kornblum, 2020, among others), complaints about elderly institutional abuse belong in the field of aging, which, along with migration and tourism, comprise the three main areas of POPULATION as it is a domain that encompasses situations that can be understood as problematic due to the cumulative processes of people and the complaints of specific groups. It is one of the QoL indicators put forward and defined by OECD_iLibrary (*Organisation for Economic Co-operation and Development*) as “all nationals present in, or temporarily absent from a country, and aliens permanently settled in a country. This indicator shows the number of people that usually live in an area” (<https://www.oecd-ilibrary.org/>). As Figure 1 lays out, POPULATION in turn is included in the dimension LIVING, whose conceptual modelling has been exhaustively accounted for in Periñán-Pascual (2023b, 2024a, 2024b).

Within this context, our proposal for knowledge modeling in POPULATION has been realized through the identification, listing, and formalization of the citizens’ complaints and concerns in this domain. To this end, our methodology has followed three stages (Jiménez-Briones & Felices Lago, 2024). First, the specific area to be modeled was selected out of the three already-mentioned fields: aging, migration, and tourism. Second, a bibliographic consultation of the chosen topic in sociology manuals and encyclopedias, as well as in specialized academic journals such as *Australian Journal on Ageing*, *Australasian Journal on Ageing*, *International Journal of Geriatric Psychiatry*, and *International Journal of Environmental Research and Public Health* was carried out. Thanks to these sources, we listed the possible issues about which social media users would express their attitude: nursing homes, pensions, health, loneliness, etc. Finally, a new bibliographic consultation was accomplished to incorporate, into the catalog of problems, more specific objections or subjective perceptions of the community about the addressed domains. As documented in the OECD report *How’s Life? Measuring Well-Being* (2020), some well-being indicators for distinct population groups and across different countries are measured considering both objective and intrinsically subjective outcomes (cf. objective and subjective measures of Income and Wealth/Health/Safety or Subjective Well-Being). In our case, the more specific consultation of surveys, questionnaires and lists made to the community on what it considers a problem within the sub-theme of aging, such as *The Attitudes to Ageing Questionnaire* (Laidlaw, Power, Schmidt & the WHOQOL-OLD Group, 2007) or *The German Ageing Survey* (Klaus, Engstler, Mahne, Wolff, Simonson, Wurm & Tesch-Römer, 2017), helped us to refine the catalog obtained in the second methodological step and incorporate more particular and subjective problems such as abuse in nursing homes, age discrimination, fear of ending up in a nursing home or of becoming a widow or widower. As a result, Jiménez-Briones

and Felices Lago (2024) present the list of problem types developed to date (e.g. AGEISM, FEAR-NURSINGHOME, LOWPENSION, etc.), along with the formalization of this knowledge on aging in DIAPASON through the development of their corresponding problem schemas. Next section approaches the formalization of the problem type ELDERLYINSTITUTIONALABUSE.

3.2. Building the problem schema of elderly institutional abuse

As explained in Section 2, the construction and formalization of the problem catalog in DIAPASON is achieved by means of the interface functionality or Editor (see Figure 2). Furthermore, the researchers in charge of this knowledge modeling task followed the methodology discussed in Jiménez-Briones and Felices Lago (2024) and Felices-Lago (2025), illustrated in Figure 3 for elderly institutional abuse:

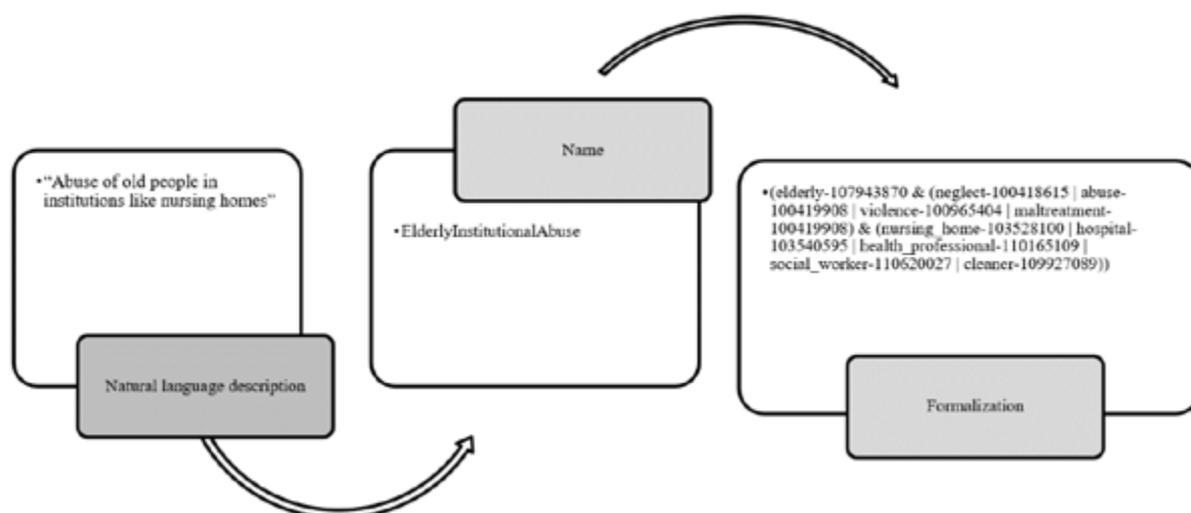


Figure 3. Problem schema methodology

Since, according to the literature (Rueda Estrada & Martín Martín, 2011; Walker, 2022), complaints about this type of abuse may involve different kinds of institutions (nursing homes, hospitals, clinics) and actors (physicians, social workers, nurses, etc.), the schema elaborated to define it, repeated in (1) for the sake of clarity, is certainly complex:

- (1) (elderly-107943870 & (neglect-100418615 | abuse-100419908 | violence-100965404 | maltreatment-100419908) & (nursing_home-103528100 | hospital-103540595 | health_professional-110165109 | social_worker-110620027 | cleaner-109927089))

First, as the synsets corresponding to *neglect*, *abuse*, *violence*, and/or *maltreatment* have negative connotations and already express a problematic situation, it was decided not to incorporate the illocutionary part into the schema. In these cases, it is better to opt for the minimal expression because, at this stage, the introduction of some communicative function in the problem schema (e.g. COMPLAIN) will force the machine to search, in the UGC units, for a linguistic expression that instantiates it (e.g. *This is unacceptable!*). That being so, the proposed schema in (1) only consists of the locutionary part, in which the following have been formalized:

- 1) Physical and psychological abuse of senior citizens. As a result, the synset *elderly-107943870* must always be present (&), followed by the synsets that conceptualize the type of abuse, that is, *neglect-100418615*, *abuse-100419908*, *violence-100965404* and/or (|) *maltreatment-100419908*.

- 2) The institutions that may be part of the abuse of the elderly. For this purpose, the synsets *nursing_home-103528100* and *hospital-103540595*, connected with the inclusive disjunction operator (\vee), have been incorporated.
- 3) People who could perpetrate the mistreatment. The presence of the inclusive disjunction between the synsets *health_professional-110165109*, *social_worker-110620027* and *cleaner-109927089* allows us to achieve this goal.

As inferred from above, the WordNet synsets turn crucial when codifying the conceptual content of each problem as a problem schema. WordNet (Princeton, 2010) is a comprehensive lexical database that organizes nouns, verbs, adjectives and adverbs into sets of cognitive synonyms, known as synsets. It also systematically documents the relationships between these synsets through semantic relations such as hyponymy, hypernymy, troponymy, holonymy, and meronymy. For this reason, if, for example, the hypernym *health_professional* is selected in the DIAPASON's knowledge organization system, no other items subordinate to it, such as *doctor*, *nurse*, *surgeon*, etc. will have to be housed in the problem schema as can be seen in the Related words box in Figure 4. In the same line, selecting the synsets *hospital-103540595*, *social_worker-110620027*, and *cleaner-109927089* already encompasses the different types of hospitals, social workers and cleaners.

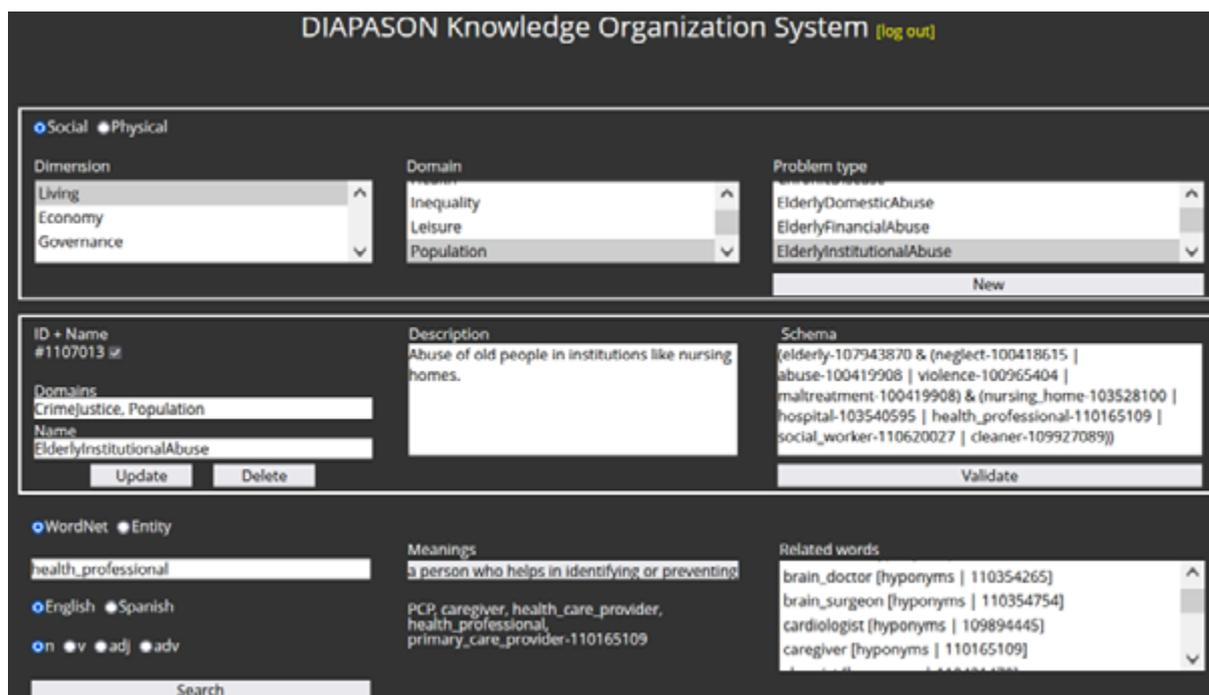


Figure 4. Elderly Institutional Abuse in DIAPASON

As made clear in Periñán-Pascual (2023a, 2023b), problem schemas are fundamental in the tasks of keyword recognition and topic categorization performed by DIAPASON. For the former, the synsets that are semantically connected to linguistic expressions in the microtexts are selected as keywords, whereas for the latter, problem schemas help to build a training set that DIAPASON will use to classify the UGC units. However, it is out of the scope of this paper to explore these two tasks in depth.

3.3. Compiling the subcorpus of UGC units on elderly institutional abuse

Employing elderly institutional abuse as a practical case, the two sections above have sketched the first stages for the automatic identification of social problems in the ALLEGRO system: a) choosing the problem domain from the DIAPASON ontology (e.g. POPULATION); b) identifying a set of socially significant problems within the domain based on the sociological literature available; c) selecting one problem and further consulting specific references (i.e. elderly institutional abuse); and, finally, d) building its problem schema with the DIAPASON Editor (Figure 4). This section adds a new and decisive step to the methodology: e) compiling a corpus of tweets on elderly institutional abuse. As mentioned in Section 2, the different subcorpora built for the DIAPASON problem schemas will contribute to the ALLEGRO gold standard corpus for the task of automatically sorting out UGC items as social problems or not. In consequence, the same attested methodology found in Periñán-Pascual (2024b), Alameda Hernández (2024) and Ureña Gómez-Moreno (2024) has been applied below.

3.3.1. Automatic extraction of microtexts

The first stage in the compilation of the subcorpus consists in gathering the necessary UGC units related to the problem under study. To this end, the application Twitter Searcher, developed by Periñán-Pascual (2024b: 20-21), was employed as it allowed us to access the Twitter API directly:

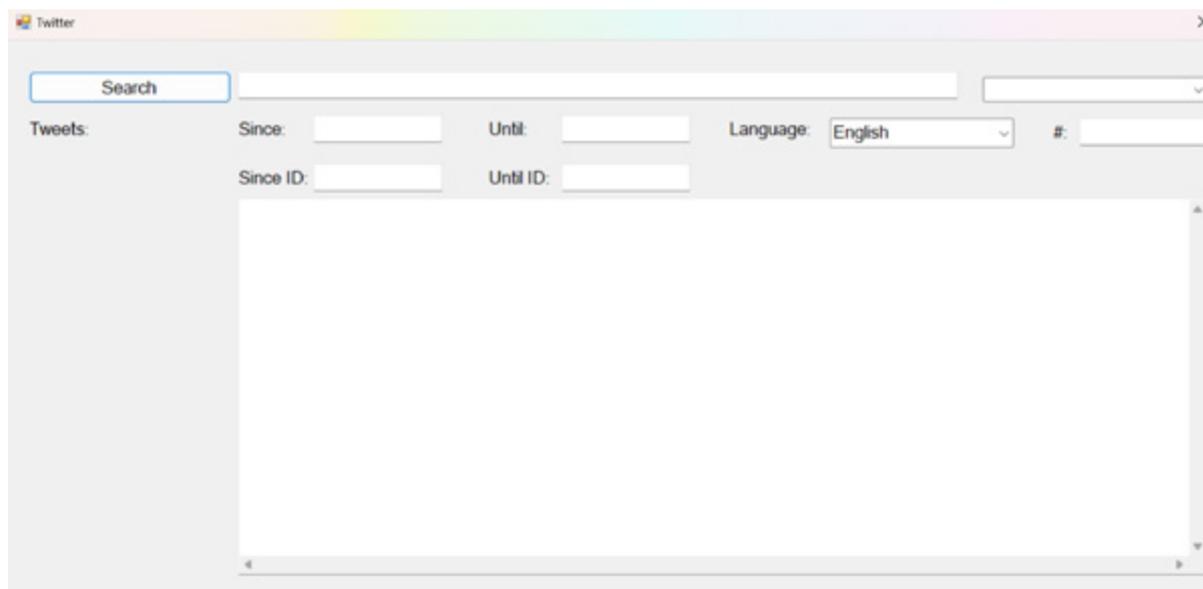


Figure 5. *Twitter Searcher* (Periñán-Pascual (2024b: 20-21))

The first query parameter in Twitter Searcher is the Search box, in which the keywords that will trigger the appropriate related tweets are typed in. As identified in the literature on building specialized corpora (Bowker & Pearson, 2002; Bathia, Sánchez Hernández & Pérez Paredes, 2011; Alayiaboozar & Hojjatpanah, 2022), this is a critical step when collecting data for a corpus because the choice of the searched words or seeds should strike a balance so that it neither biases the sample nor relies on introspection. Thus, drawing on Periñán-Pascual (2024b) and Ureña Gómez-Moreno (2024), whenever possible, keyword-based searches were avoided in favor of topic searches (e.g. words in hashtags, usernames, or named entities) so that the bias

towards using keywords from the very problem schemas for which we want to retrieve tweets could be minimized. Table 2 displays the topic words employed, from January to March 2023, to obtain tweets for our social problem:

Table 2. Search words for elderly institutional abuse

English	Spanish
elderly AND neglect AND institutions	tercera edad maltrato instituciones
elderly OR neglect OR institutions	tercera edad OR maltrato OR instituciones
seniors AND abuse AND nursing home	ancianos OR abuso OR residencia
seniors OR abuse OR nursing home	ancianos AND abuso AND residencia
seniors nursing home	ancianos residencias
seniors OR nursing home	ancianos OR residencias
seniors abuse	ancianos maltrato
seniors OR abuse	ancianos OR maltrato
elderly AND neglect	tercera edad OR maltrato
elderly OR neglect	tercera edad maltrato

In order to obtain a significant number of tweets, n-grams and various combinations with AND/OR were used, although some of the bigrams and trigrams were more productive than others. For instance, the word *institutions* was soon replaced with others because it retrieved tweets regarding universities and military or financial organizations.

Twitter Searcher also allowed us to set up the following parameters: a) Date: the time span for the searches or the tweet IDs could be established to avoid repeated microtexts; b) Language: English and Spanish tweets were retrieved; and c) Number: 500 tweets per query were introduced. Figure 6 provides a glimpse of one of the searches, whereas Figure 7 presents the results. Notice that, to filter out retweets and replies to other tweets, the instruction *filter:retweets-filter:replies* was included in the Search box.

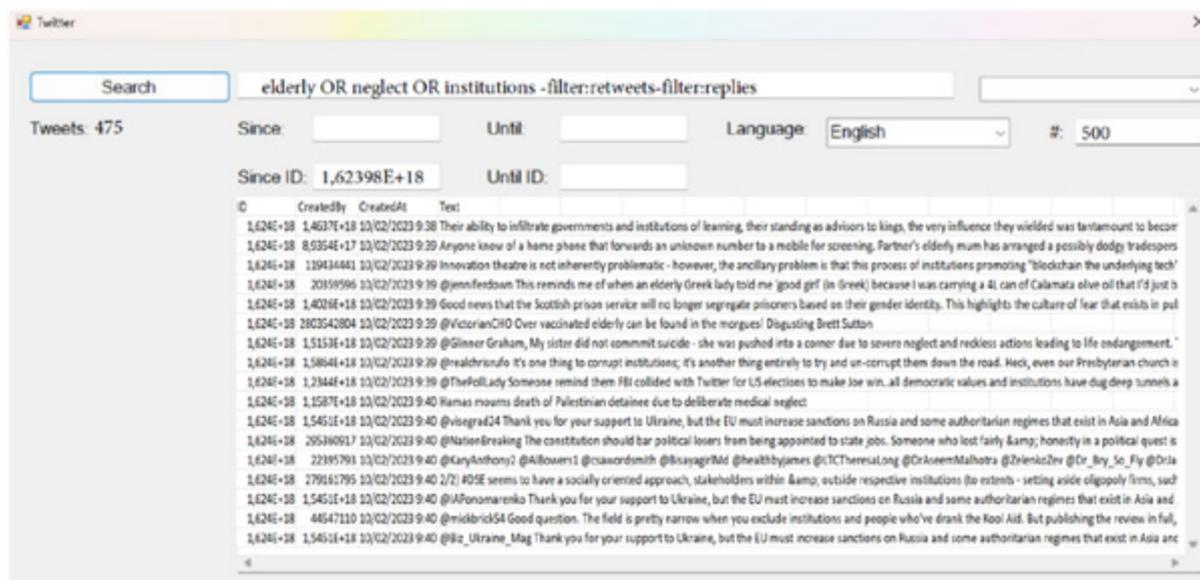


Figure 6. Twitter Searcher for elderly institutional abuse

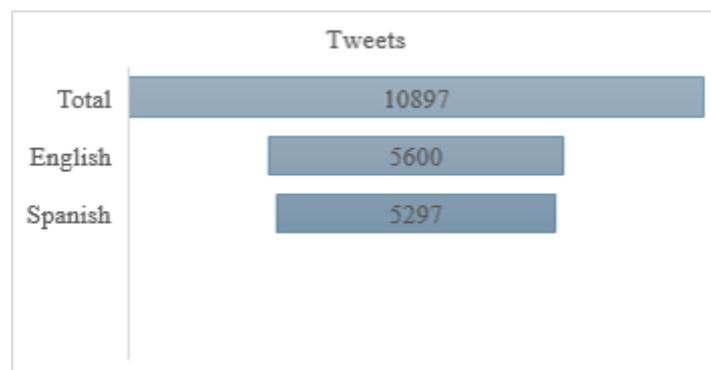


Figure 7. Number of tweets for elderly institutional abuse

Using Twitter Searcher on January 21, February 2, 10, and 28, 2023, we automatically obtained a total number of 10,897 tweets, out of which 5,600 were in English and 5,297 in Spanish. With such a vast number of tweets, we had to carry out a closer manual analysis to filter out those tweets that did not conform to our definition of social problem. Before continuing with the second step in the compilation of the subcorpus (see 3.2.2), a word is needed on the interpretation of social problems in ALLEGRO, which sides with the subjectivist approach found, among others, in Eitzen et al. (2014) and Best (1995, 2017), for which “social problems are what people view as social problems [...] No condition is a social problem until someone considers it a social problem” (Best, 1995: 4-5).⁷ In other words, a social problem comes about when society perceives it as a situation that harms or threatens the lives and well-being of some or many of the citizens and, accordingly, a solution needs to be found. Furthermore, such belief becomes a social problem only when it is collectively acknowledged and shared through social media such as X. As Periñán-Pascual states (2023b: 272), “the problem arises only when people make claims about it”. This way, such a significant number of tweets about elderly abuse indicates that this issue is viewed as a social problem by the community.

Based on this definition of social problem, the linguists within the ALLEGRO project agreed upon three criteria to identify and include UGC items that present social problems in each subcorpus: contemporariness, self-containment, and problem content. How each criterion is applied to elderly institutional abuse tweets is detailed below.

3.3.2. Manual filtering

The second step in the collection of the specialized corpus under study comprises manually annotating each tweet retrieved as 0 (off-topic related) or 1 (topic related). To filter in and out the microtexts that present UGC related to elderly institutional abuse, the three criteria mentioned above were followed. As for contemporariness, only tweets that make claims about institutionalized elder abuse for today’s society were selected. Example (2) illustrates such a case:

- (2) 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 🇪🇸 Aged care worker, 56, charged with sexual assaults of elderly residents (25/01/2023)

As Alameda Hernández (2024) details, what citizens regard as problems varies throughout history, so there must exist a defined timeframe for identifying a specific social problem. Therefore, “the selection process excludes tweets that, although posted within the selected time span, deal with problems that have affected society in the past, even if the condition that motivated

⁷ In sociology, social problems have been addressed from two approaches: objectivism and subjectivism. The interested reader is referred to Periñán-Pascual (2023b) for a detailed explanation of each perspective.

the problem may still exist, but there is not a shared belief that perceives that as a problem at present” (Alameda Hernández, 2024: 71). Likewise, texts that deal with future or hypothetical problems like (3) were discarded:

- (3) Now u cannot protect yourself and an invalid. Now most likely she'll end up in a nursing home dead within a week from neglect, and God only knows what because services are overwhelmed in the area. And I bet you, Narcos buy the land. U 🤔 (09/02/2023)

The second criterion employed to manually filtering out off-topic microtexts is self-containment. As the subcorpus includes English and Spanish tweets, any fluent speaker of these languages must be able to grasp the content of the tweet and recognize it as addressing a social problem regardless of their cultural background or identity. For this criterion it is particularly relevant the use of names or abbreviations that refer to individuals, locations, or organizations that are understood only on a local or national level:

- (4) Casi 8k de ancianos dejo morir IDA, no Podemos, IDA, gracias a sus protocolos de exclusión y la negativa de medicalizar residencias, pero tu media neurona no puede procesar tanta información 🤔😂 (28/02/2023)

Example (4) was discarded because it includes the Spanish acronym *IDA*, which stands for Isabel Díaz Ayuso, President of the Community of Madrid, who many hold responsible for the elderly that died from COVID in nursing homes during the lockdown in Madrid. However, tweets that contain abbreviations that have made their way into the language (i.e. lol, omg, b4, ICU, ER, etc.) or are field-specific like CNAs (*certified nursing assistants*) or CdS (Spanish for *health center* or *centro de salud*) have been included in the corpus.

Finally, the filtering process also involves sorting out microtexts that, despite mentioning a problem theme, lack substantive problem-related content. In other words, the third criterion helps “to differentiate and filter the tweets that deal with the subjective belief as a problem affecting society and discard those tweets that are merely referential and deal with the social problem as a topic” (Alameda Hernández, 2024: 72). In our case, only UGC items like (5), which reflects abuse in elder care institutions, were included, discarding microtexts like (6), which announces some service for reporting cases of abuse in nursing homes. Our subcorpus, thus, collects tweets that showcase the social media users’ subjective perceptions and complaints about elderly institutional abuse.

- (5) The way seniors light up once you mention their children, its like the only thing they feel is worth remembering from life. Elderly abuse in care homes is bad. 🤖 The one's treated like crap the most is the one's that don't have family to advocate for them yes even the rich ones (28/02/2023).
- (6) Sadly, it is all too common for seniors living in nursing homes to experience abuse or neglect. Call 📞 910-839-7433 or visit 🏠 #nursinghome #nursinghomeneglect #nursinghomenegligence (03/02/2023)

3.3.3. Upcoming steps

Once the appropriate UGC units on elderly institutional abuse have been selected, the following steps need to be undertaken in the compilation of the subcorpus and in the completion of the ALLEGRO system. First, the reliability of the manual filtering explained above will be tested with another annotator. To this end, the Kappa index will be employed to evaluate the accuracy

and consistency of our tweet annotations. Second, other subcorpora need to be compiled for the rest of social problems belonging to the different dimensions of DIAPASON (see Figure 1). Thus, after the collection of the distinct subcorpora for the problem schemas in DIAPASON, linguistic studies could be pursued (e.g. word frequency lists, collocations, concordances, etc.) to research, among others, positive and negative attitudes of the social media users on topics such as institutionalized elder abuse, violence against women, unemployment, and so on. To do so, TexMiLab is the software designed by Perrián-Pascual (2024c) as a linguistic laboratory that incorporates tools for corpus building, text preprocessing, and text analysis.

It is worth highlighting that, in the light of the corpus data obtained for elderly institutional abuse, some of the already codified problem schemas for aging had to be redefined. For instance, the problem schemas `ELDERLYABUSE`, `ELDERLYDOMESTICABUSE` and `ELDERLYFINANCIAL-ABUSE` had to be reformalized differently, and even discarded in the case of the first schema, so that they did not overlap with `ELDERLYINSTITUTIONALABUSE` as finally defined in (1). Furthermore, while manually filtering out the tweets whose content was or was not related to elderly institutional abuse, we could identify UGC units for other social problems: ageism, fear of digital divide, elderly financial/domestic abuse, etc.

4. CONCLUSIONS

In the era of digital social media like X, Facebook, Instagram, etc., users have emerged as significant sources of immediate and authentic information through the content they post online or UGC. The identification and analysis of UGC can be highly valuable, particularly when such content addresses issues impacting society like elderly institutional abuse since exploring UGC can allow relevant institutions to take appropriate measures to address such specific social challenges.

In this context, the present study has focused on the compilation of a subcorpus of text messages from the social media platform X concerning elderly institutional abuse. This subcorpus has been incorporated in ALLEGRO, a multimodal intelligent system for the real-time identification and analysis of community problems published in social media platforms. After describing the ALLEGRO architecture, in particular its text-processing module or DIAPASON, as the corpus under analysis comprises written UGC, the methodological stages established for the completion of the intelligent system, as well as the challenges that lie ahead in the project, have been illustrated with the practical case of the problem of elderly institutional abuse. As one of the purposes of ALLEGRO is to build a medium-sized corpus of social problems that can serve as a training set for deep language learning models in text classification (Perrián-Pascual, 2024b), our proposed subcorpus contributes to the creation of a gold standard corpus for the automatic detection of social problems by this multimodal intelligent system.

ACKNOWLEDGEMENTS

This publication is part of the We-Collab project 2021-1-HR01-KA220-HED-000027562, co-funded by the Erasmus+ Programme and the R&D&i project PID2020-112827GB-I00, funded by MICIU/AEI/10.13039/501100011033.

REFERENCES

- Bazzaz, A. S., Haghi, K. M., Mahdipour, E., & Jameii, S. M. (2021). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and Informatics*, 57, 101517. <https://doi.org/10.1016/j.tele.2020.101517>
- Alameda Hernández, Á. (2024). Social media detection of texts on the social problem of violence against women within the multimodal intelligent system ALLEGRO. In R. Jiménez-Briones & A. Corral Esteban (Eds.), *Approaches to Knowledge Representation and Language* (pp. 63-75). Granada: Comares.
- Alayiaboozar, E., & Hojjatpanah, A. A. (2022). Steps for creating two Persian specialized corpora. *International Journal of Information Science and Management (IJISM)*, 20(4), 231-243.
- An, J., & Weber, I. (2015). Whom should we sense in “social sensing” - analysing which users work best for social media now-casting. *EPJ Data Science*, 4(22), 1-22. <https://doi.org/10.1140/epjds/s13688-015-0058-9>
- Appio F. P., Lima, M., & Paroutis, S. (2019). Understanding Smart Cities: Innovation ecosystems, technological advancements, and societal challenges. *Technological Forecasting & Social Change*, 142, 1-14. <https://doi.org/10.1016/j.techfore.2018.12.018>
- Arthur, R., Boulton, C. A., Shotton, H., & Williams, H. T. P. (2018). Social sensing of floods in the UK. *PLoS ONE*, 13(1), e0189327. <https://doi.org/10.1371/journal.pone.0189327>
- Bates, C. G., & Ciment, J. (Eds.) (2013). *Global Social Issues: An Encyclopedia*. New York: Sharpe Reference.
- Bathia, V., Sánchez Hernández, P., & Pérez Paredes, P. (2011) Specialized languages: Corpora, meta-analyses and applications. *Researching Specialized languages*, 47(1), 1-8. <https://doi.org/10.1075/sci.47.02bha>
- Best, J. (1995). Typification and social problems construction. In J. Best (Ed.), *Images of Issues: Typifying Contemporary Social Problems* (pp. 1-10). London: Routledge.
- Best, J. (2017). *Social Problems* (3rd ed.). New York City: W.W. Norton & Company.
- Bowker, L., & Pearson, J. (2002) *Working with Specialized Language: A Practical Guide to Using Corpora*. London and New York: Routledge. <https://doi.org/10.4324/9780203469255>
- Eitzen, S., Baca Zinn, M., & Eitzen Smith, K. (2014). *Social Problems* (13th ed.) Boston: Pearson.
- Espinoza-Arias, P., Poveda-Villalón, M., García-Castro, R., & Corcho, O. (2019). Ontological representation of smart city data: From devices to cities. *Applied Sciences*, 9(1), 1-23. <https://doi.org/10.3390/app9010032>
- Felices Lago, Á. (2024). Description of social problems by means of schemas related to the Income domain in the DIAPASON platform. In R. Jiménez-Briones & A. Corral Esteban (Eds.), *Approaches to Knowledge Representation and Language* (pp. 27-43). Granada: Comares.

Felices Lago, Á. (2025). Towards the characterization of WORK problem schemas in the DIAPASON ontology. *Sintagma* 37.

Fernández-Martínez, N. J. (2024). Exploring the creation of synthetic corpora of negative communicative functions for the task of communicative function identification. In R. Jiménez-Briones & A. Corral Esteban (Eds.), *Approaches to Knowledge Representation and Language* (pp. 77-93). Granada: Comares.

Gething, L. (1994). Health professional attitudes towards ageing and older people: Preliminary report of the reactions to Ageing Questionnaire. *Australian Journal on Ageing*, 13(2), 77-81. <https://doi.org/10.1111/j.1741-6612.1994.tb00646.x>

Gething, L., Fethney, J., McKee, K., Goff, M., Churchward, M., & Matthews, S. (2002). Knowledge, stereotyping and attitudes towards self-ageing. *Australasian Journal on Ageing*, 2(2), 74-79. <https://doi.org/10.1111/j.1741-6612.2002.tb00421.x>

Ghani, N. A., Hamid, S., Targio Hashem, I. A., & Ahmed, E. (2019). Social media big data analytics: A survey. *Computers in Human Behavior*, 101, 417-428. <https://doi.org/10.1016/j.chb.2018.08.039>

Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N., & Meijers, E. (2007). *Smart cities - Ranking of European medium-sized cities*. Retrieved from http://www.smart-cities.eu/download/smart_cities_final_report.pdf.

Govada, S. S., Spruijt, W., & Rodgers, T. (2017). Smart city concept and framework. In T. M. V. Kumar (Ed.), *Smart Economy in Smart Cities. Advances in 21st Century Human Settlements* (pp. 187-198). New York City: Springer. https://doi.org/10.1007/978-981-10-1610-3_7

Holtgraves, T. (1994). Communication in context: Effects of speaker status on the comprehension of indirect requests. *Journal of Experiential Psychology: Learning, Memory, and Cognition*, 20(5), 1205-1218. <https://doi.org/10.1037//0278-7393.20.5.1205>

Jiménez-Briones, R., & Felices Lago, Á. (2024). La formalización del conocimiento en DIAPASON a través de una muestra de problemas poblacionales y macroeconómicos. In F. Olmo-Cazeville (Ed.), *Investigación Lingüística en Entornos Digitales* (pp. 187-216). Granada: Tirant Lo Blanch.

Klaus, D., Engstler, H., Mahne, K., Wolff, J. K., Simonson, J., Wurm, S., & Tesch-Römer, C. (2017). Cohort Profile: The German Ageing Survey (DEAS). *International Journal of Epidemiology*, 46(4), 1105-1105. <https://doi.org/10.1093/ije/dyw326>

Laidlaw, K., Power, M. J., Schmidt, S., & the WHOQOL-OLD Group (2007). The attitudes to ageing questionnaire (AAQ): Development and psychometric properties. *International Journal of Geriatric Psychiatry*, 22, 367-379. <https://doi.org/10.1002/gps.1683>

Lee, H. J., Lee, D. K., & Song, W. (2019). Relationships between social capital, social capital satisfaction, self-esteem, and depression among elderly urban residents: Analysis of secondary survey data. *International Journal of Environmental Research and Public Health*, 16, 1445, 2-13. <https://doi.org/10.3390/ijerph16081445>

Li, W., Wu, W., Wang, H., Cheng, X., Chen, H., Zhou, Z., & Ding, R. (2017). Crowd intelligence in AI 2.0 era. *Frontiers of Information Technology & Electronic Engineering*, 18, 15-43. <https://doi.org/10.1016/j.intell.2017.04.004>

Marginean, I. (2014). Quality of Life Diagnosis (QoLD). In A. C. Michalos (Ed.), *Encyclopedia of Quality of Life and Well-Being Research* (pp. 5333-5339). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-0753-5_2358

Musto, C., Semeraro, G., Lops, P., & De Gemmis, M. (2015). CrowdPulse: A framework for real-time semantic analysis of social streams. *Information Systems*, 54, 127-146. <https://doi.org/10.1016/j.is.2015.06.007>

OECD iLibrary (Organisation for Economic Co-operation and Development). Retrieved from <https://www.oecd-ilibrary.org/>.

OECD (2020). *How's Life? 2020: Measuring Well-Being*. Paris: OECD Publishing. Retrieved from https://www.oecd-ilibrary.org/economics/how-s-life/volume-/issue-_9870c393-en. <https://doi.org/10.1787/9870c393-en>

OECD (2022). "Population" (indicator). <https://doi.org/10.1787/d434f82b-en>

Parrillo, V. N. (Ed.) (2008). *Encyclopedia of Social Problems*. Los Angeles: Sage. <https://doi.org/10.4135/9781412963930>

Pérez, L. (2013). Illocutionary constructions: (Multiple source)-in-target metonymies, illocutionary ICMs, and specification link. *Language & Communication*, 33(2), 128-149. <https://doi.org/10.1016/j.langcom.2013.02.001>

Periñán-Pascual, C. (2023a). Exploring user-generated content to detect community problems: The ontological model of ALLEGRO. In *Proceedings of the 15th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2023) - V. 2* (pp. 224-230). KEOD. <https://doi.org/10.5220/0012203300003598>

Periñán-Pascual, C. (2023b). From Smart City to Smart Society: A quality-of-life ontological model for problem detection from user-generated content. *Applied Ontology*, 18(3), 263-306. <https://doi.org/10.3233/AO-230281>

Periñán-Pascual, C. (2024a). Modelización de las quejas de los ciudadanos como artefactos digitales culturales: DIAPASON. In F. Olmo-Cazeville (Ed.), *Investigación Lingüística en Entornos Digitales* (pp. 129-156). Valencia: Tirant Lo Blanch.

Periñán-Pascual, C. (2024b). Exploring problems through social media: The case of Beach Quality. In R. Jiménez-Briones & A. Corral Esteban (Eds.), *Approaches to Knowledge Representation and Language* (pp. 11-26). Granada: Comares.

Periñán-Pascual, C. (2024c). *Minería de textos para investigadores lingüistas*. Valencia: Tirant Lo Blanch.

Princeton University (2010). *About WordNet*. Retrieved from <https://wordnet.princeton.edu/>.

Rueda Estrada, J. D., & Martín Martín, F. J. (2011). El maltrato a personas mayores. Instrumentos para la detección del maltrato institucional. *Alternativas*, 18, 7-33. <https://doi.org/10.14198/ALTERN2011.18.01>

Sakaki, T., Okazaki, M., & Matsuo, Y. (2013). Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering*, 25(4), 919-931. <https://doi.org/10.1109/TKDE.2012.29>

Searle, J.R. (1969). *Speech Acts*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139173438>

Secombe, K., & Kornblum, W. (2020). *Social Problems* (16th ed.). Boston: Pearson.

Stefanowitsch, A. (2003). A construction-based approach to indirect speech acts. In K.U. Panther & L. Thornburg (Eds.), *Metonymy and Pragmatic Inferencing* (pp. 105-26). Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.113.09ste>

Trott, S. and Bergen, B. (2017). A theoretical model of indirect request comprehension. *Proceedings of the AAAI Fall Symposium Series on Artificial Intelligence for Human-Robot Interaction* (pp. 129-132). Arlington, VA.

Ureña Gómez-Moreno, P. (2024) Integrating corpus methodology into the construction of an intelligent crowdsensing system. In R. Jiménez-Briones & A. Corral Esteban (Eds.), *Approaches to Knowledge Representation and Language* (pp. 45-62). Granada: Comares.

Walker, R. W. (2022). *Elderly Financial Abuse in New Zealand: Is the Law Sufficient?* (Master's dissertation). University of Canterbury. Retrieved from <https://ir.canterbury.ac.nz/items/5f894749-d5d1-4f47-a6c0-511e9bd299b7>.

Wang, D., Szymanski, B. K., Abdelzaher, T., Ji, H., & Kaplan, L. (2019). The age of social sensing. *IEEE Computer*, 52(1), 36-45. <https://doi.org/10.1109/MC.2018.2890173>



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 19-33
Artículo recibido: 23/08/2024
Artículo aceptado: 21/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/676>
DOI: <https://doi.org/10.58859/rael.v23i1.676>

Classification of Technology-based Language Learning Resources according to Pedagogical and Functional Criteria

Clasificación de recursos de aprendizaje de lenguas basados en tecnología según criterios pedagógicos y funcionales

RAFAEL SEIZ ORTIZ
MARÍA LUISA CARRIÓ PASTOR
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

The objective of this study is to propose a classification of ICT-based language learning resources using as organizing criteria the pedagogical functions of the language learning and teaching process. To meet this goal, first, a corpus of resources is collected, and a literature review is carried out with the aim of identifying the most important language learning pedagogical functions. Then, a two-level classification comprising fifteen types that describe key functions of the resources and a set of subcategories. This classification can be a pedagogical tool useful for language teachers and learners, researchers and instructional designers. Later, the classification is used in two studies, the aim of which is to demonstrate its usefulness to assess the real use of technology-based resources by teachers in two educational areas and to get feedback from practising teachers to improve such classification as well as to carry out an initial validation of this classification proposal.

Keywords: *Information and Communication Technologies (ICTs); Computer-Assisted Language Learning (CALL); technology-based resources and tools; pedagogical functions; language learning.*

El objetivo del presente estudio es proponer una clasificación de recursos de aprendizaje de lenguas basados en las TIC utilizando como criterio las funciones pedagógicas del proceso de enseñanza-aprendizaje de lenguas. Para ello, primeramente, se recopila un corpus de recursos y se hace una revisión bibliográfica para identificar las funciones pedagógicas más importantes en el aprendizaje de lenguas. Seguidamente, se presenta una clasificación de dos niveles que comprende quince tipos que describen las funciones de los recursos y diversas subcategorías. La clasificación puede ser una herramienta pedagógica útil para el profesorado y los aprendices de lenguas, investigadores y diseñadores instruccionales. Finalmente, la clasificación se utiliza en dos estudios cuyo objetivo es demostrar su utilidad para evaluar el uso real de recursos por parte del profesorado en dos áreas educativas y obtener retroalimentación de los docentes de lenguas, con el fin de mejorar dicha clasificación y realizar una validación inicial de esta propuesta.

Palabras clave: *Tecnologías de la Información y las Comunicaciones (TIC); Aprendizaje de Lenguas Asistido por Ordenador (ALAO); recursos y herramientas basadas en tecnología; funciones pedagógicas; aprendizaje de lenguas.*

Citar como: Seiz Ortiz, R. y Carrió Pastor, M.L. (2024). Classification of Technology-based Language Learning Resources according to Pedagogical and Functional Criteria. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 19-33. <https://doi.org/10.58859/rael.v23i1.676>

1. INTRODUCTION

Resources and tools based on Information and Communication Technologies (ICTs) have been used for a long time for the purposes of teaching or learning languages, and they increase in number, quality and functionalities at a fast pace. A large and growing body of research in areas such as Computer-Assisted Language Learning (CALL), educational technology and Second Language Acquisition (SLA) strongly suggest the pedagogical efficiency of ICTs to enhance the language learning and teaching process. To fully benefit from this technology-enhanced learning environment, language teachers and learners should know what to do with technology and how to do it in the best possible way. Thus, the pedagogical functions of technology when teaching or learning languages must be clearly identified. This is the main topic of this study, which has a double objective. First, it aims to identify the functions of the ICTs and their related resources and tools from the point of view of pedagogy in language teaching and learning. A second goal is to use these pedagogical functions to design a classification framework for ICT-based resources and tools in their use within the language teaching and learning process. Our assumption is that all ICT-based resources and tools can be used with a language learning purpose in mind, as long as the learning objectives and the related pedagogical functions, as well as the activities that may be carried out through these resources, are known by the users, i.e. teachers and learners.

The classification of technology-based language learning resources in different types and categories, therefore, should serve the purpose of assisting learners, teachers, and researchers to ascertain and identify the different varieties at hand, which will help to evaluate their potential (Rosell-Aguilar, 2017). Another advantage of this approach is that a theoretically-sound classification system of technological language learning resources implies a conceptualization of the functionalities of the technology, as well as a contextual clarification and use of specific terminology, a clear demand in the field of CALL (Levy, 1997).

To meet these objectives, first, this study delves into the functional features of the ICTs and their connection to the pedagogical functions of language learning and teaching. The classification is first presented as a tool for teachers, learners, designers and researchers. Then it is implemented, and two studies are conducted to ascertain the real use of ICT-based language learning resources by teachers, and to integrate the points of view of these teachers to get feedback on possible improvements of the classification, with the ultimate purpose of carrying out an initial validation of the proposal.

2. THEORETICAL FRAMEWORK

The term *ICT* refers to “various technologies that are used to access, collect, process, and share information” (Hiradhar & Bhattacharya, 2022: 4). A similar and complementary definition is given by the *United Nations Development Program*, which considers ITCs as “a diverse set of technological tools and resources used to communicate, and to create, disseminate, store, and manage information” (Tinio, 2003: 4). These definitions are relevant for our classification framework, because they are based on verbs, actions –i.e. functions for which the technology may be useful and feasible, in our case, within the language learning and teaching process–. If we analyse the verbs and concepts embedded in both definitions, the two fundamental components of language learning clearly emerge: communication and a set of crucial actions that can be carried out with information and content when learning a language. Therefore, these verbs and the communication factor are used to conceptualize the various pedagogical functions of the classification proposed in this study. Pedagogical functions here are conceptualized as the roles that assist in the process of achieving learning goals (Campbell et al., 2012). That said, this research focuses on the latest computer and Internet-based technologies, rather than integrating

other technologies that are also considered as ICTs, such as compact discs, voice recorders or television, since all these technologies have been integrated in practice within the Internet and the Web. The way the types of resources are mostly referred to throughout this study is “technology-based”, although it is understood that we imply “ICT-based resources”, to avoid the use of acronyms.

An initial procedure in the study was to incorporate the wide range of technologies that Godwin-Jones has comprehensively discussed since 1997. As he recognised in an account of 20 years of language learning technologies (Godwin-Jones, 2016), this approach can also provide insights into possible affordances and uses of the technologies. Over that period and beyond, many technologies, from multimedia delivery to immersive technologies and Generative Artificial Intelligence have been studied by Godwin-Jones (2024), a theoretical base that was used here to conceptualize the types and subcategories of the classification framework. Additionally, and apart from some relevant websites that present collections and lists of technological tools for learning in general or language learning in particular, there are some pertinent books that present a collection of language learning technologies, and, although their main classification criteria are not pedagogical functions, they have been used to establish some categories in our classification, such as Kenning (2007) or Farr & Murray (2016). Also relevant for this research, a recent study comprehensively reviews fifty years of technologies relevant to language learning (González-Lloret, 2023).

In the area of mobile applications, Rosell-Aguilar (2017) proposes a taxonomy that classifies apps for language learning into three categories: applications designed for language learning, those not designed for but useful for language learning, and a third group comprising dictionaries and translators, incorporating grammar, vocabulary, interaction and the four communicative skills, as the final layer of the classification. Coherent as this framework may be, the present research has merged these two categories, i.e. language-learning vs. generic, since our assumption is that, regardless of the target purpose of the technological resource, the most important consideration is the pedagogical function, and most, if not all, ICT-based resources could be used for a pedagogical function within the language learning process.

In a comprehensive study that establishes close links between language theories, CALL and ICTs, Wilkinson (2016) classifies ICT-based resources for English as a Foreign Language (EFL) into different types, although his main purpose is not classificatory: L2 skill-specific applications, Web resources, mobile apps, sites for recording, Web 1.9 (static) tools, Web 2.0 tools, and social learning platforms. With a broader scope that includes language learning and teaching resources based on ICTs, Nirmal & Mohsen (2023) also differentiate between language-specific tools and more general productivity tools and classify ICT tools in this educational context with a functional approach into the following categories: E-communication tools, E-creation tools, Assessment tools, Reading/Writing E-tools and Virtual Learning Environments.

In our classification framework, Web 2.0 technologies have been incorporated, since these are the ones that best fit the principles of Task-based Language Teaching (TBLT) and Task-based Language Learning (TBLL). They enable users -teachers and mostly learners- to create digital content and communicate with other users, so that learners can engage in doing things with language and in cooperation with other learners, through the functionalities of Web 2.0, rather than simply carrying out language-based activities. Therefore, these technologies can be integrated into language learning as an environment for interaction and learning by doing realistic, goal-oriented tasks, and getting in contact with authentic input and authentic interaction with native speakers of the L2 (González-Lloret, 2017: 236).

In a study that reviews computer technologies, and their efficiency related to the field of CALL, Golonka et al. (2014) classify these technologies under the following categories: (1) “Classroom-based technologies”, which assist classroom activity regarding material presentation and learner tracking, including *Course management systems (CMS)*, *Interactive white boards* and *ePortfolios*, with programs such as gap-filling exercises, simulations, multi-

ple-choice exercises, and the like; (2) “Individual study tools”, resources which the learner or teacher can use as a complement in an individual level language learning process, such as *Corpus tools*, *Electronic dictionaries*, *Electronic gloss or annotation tools*, *Intelligent tutoring systems*, *Grammar checkers*, *Automatic speech recognition (ASR)* or *pronunciation programs*; (3) “Network-based social computing”, including technologies such as *virtual worlds* or *serious games*, *chat-based environments*, *social networks* and *blogs*, learning environments the purpose of which is not related to language learning, but provide pedagogical advantages and value in terms of practicing language skills within real-life contexts and in a cooperative manner; and (4) “Mobile and portable devices”, which are useful for language learners and teachers beyond the conventional scope of foreign language learning and CALL, with almost endless possibilities of delivering language material, establishing social networking environments, authoring potential, and so on. Some of these categories are reconceptualized and rearranged through the lenses of language learning functions in the classification framework proposed here.

Within the field of language learning, some categories of technology resources have been proposed. In some studies, focused on language learning, the criterion for classification is often based on the four communicative skills (Dash & Kuddus, 2020). This criterion has sometimes been expanded to include the major language components, such as grammar, vocabulary, pronunciation and culture as organizing categories, for example grouping the technological resources according to their usefulness for improving these skills and components (Levy, 2009). In other cases, studies have used criteria based on the specific Computer-based technology that supports the language learning process (Krajcsó & Frimmel, 2017; Alkamel & Chouthaiwale, 2018; Bhushan, 2020; Budiman, 2020; Li & Lan, 2022; Madhavi, Sivapurapu & Kati, 2023).

All the abovementioned research studies within the field of language learning, though, do not have a classificatory purpose *per se*, and, apart from being very heterogeneous in their approach and purpose, they do not provide a comprehensive account of the pedagogical potential of ICTs in terms of the pedagogical functions of the technology (resources and tools), such as the classification framework presented in this research.

Outside the area of language learning, some studies have categorized educational ICT tools. For example, in a research work about Interactive Learning Systems (ILSs), Luo and Lei (2012) established the following four types of ICT tools: (a) *educational networking*; (b) *web-based learning*; (c) *mobile learning*, and (d) *classroom equipment*. Although learning systems are not the same as tools and resources, some of the concepts in their categorization have been incorporated into our classification, especially those to do with interactive learning technologies and functions. More informally, and without restricting the focus on language learning, the compilation of technological learning tools that Jane Hart has been building up since 2007, called *Top 100 tools for learning* (<https://toptools4learning.com>), proposes a category of general educational resources somehow resembling our research, since the tool categories are: Content (& app), Content development, Learning platform, Communication & Collaboration and Web tools.

In this section, we have been carried out a brief state-of-the-art review of earlier attempts to categorize ICT-based resources and technologies which can have a pedagogical function, not only within the field of language learning, so that our research can incorporate some key concepts that will help us to build up our classification framework. In this review process, Hart’s model has been especially relevant, although it has not been directly used. Therefore, the classification proposals discussed above have played an *inspirational*, rather than a *developmental* role, in this research.

In the most recent version of the annual *Horizon Report - Higher Education Edition* (Becker et al., 2018), which identifies and describes the higher education trends, challenges, and developments in general educational technology with an impact on learning and teaching, the latest developments in technology are considered, such as adaptive learning technologies, games and gamification, mobile learning, natural user interfaces, Artificial Intelligence,

Learning Management Systems, and mixed reality, among others. This report also deals with certain technological trends with a high potential to contribute to the field of education. These significant trends in education nowadays can serve this research by incorporating pedagogical functions that can be implemented through ICTs within the language learning and teaching processes.

3. METHODOLOGY

In this study, two research designs can be distinguished, in line with the two objectives: developing a framework for the classification of technology-based language learning resources and carrying out an initial and partial validation of the classification. Both aspects of the research are closely intertwined in such a way that they do not take place in a linear fashion throughout the research process, but they provide feedback to one another in an iterative way, as the data is collected and analysed.

This is a descriptive and analytical study which follows a mixed research methodology, including qualitative and quantitative approaches. For the elaboration of the classification, a corpus of 350 ICT-based language learning resources was generated based on the contribution of 96 teachers, including the authors, most of whom were Content and Language Integrated Learning (CLIL) teachers. Apart from this corpus, in a more informal and holistic way, the researchers also observed and analysed between 20 and 30 additional technology-based resources that could be used in language learning. At this preliminary stage of the research, a collection of specialized publications on educational technologies have been incorporated, within a process of conceptual coding. For building up the classification framework and the related taxonomy, the research methodology was, therefore, qualitative.

Thus, the resources and tools were analysed qualitatively applying principles of the grounded theory research methodology, in this case, coding the information emerging from the analysis of the data, i.e. the ICT-based learning resources and tools, from the point of view of pedagogical functions. The resources were coded using the concepts of pedagogical functions and purposes, i.e. asking the following questions: (1) *What can the resource be used for within the language learning process?* (2) *What is the resource's primary function within the language learning process?* and (3) *What technology is behind the resource and allows it to have this function?* As a result of this coding, certain categories emerged from the data (i.e. the resources and tools) in the form of phrases starting with “for”, which made up the first level of the classification framework, with the addition of further details about the technology or the way or format in which the pedagogical function is enabled in the resource, which made up the second level or subcategories of the classification. Consequently, the first level of the classification framework (labelled as phrases with “for”) refers to pedagogical functions, which, in turn, may belong to 2 subfamilies: those focused on teaching (carried out by teachers) and those focused on learning (carried out by learners). And the second level of the framework refers to subcategories which correspond to practical results or uses of those pedagogical functions of the first level. This analytical procedure finally resulted in fifteen types of technology-based language learning resources, with a variable number of subcategories. The categorization process experienced a series of iterations aiming at refining the classification, including feedback from questionnaires completed by language teachers.

In the second stage of the study, a quantitative methodology was implemented, through the implementation of descriptive statistics. Two almost identical questionnaires, one in Spanish, called *Uso de herramientas digitales en la enseñanza de Español como Lengua Extranjera* (available online at <https://forms.gle/naHohQjFER4oqGgN8>) and the other in English, named *Use of digital resources for learning English as a Foreign Language* (available at

<https://forms.gle/BES8xLjUVPABzCrMA>), were designed using *GoogleForms* and administered to teachers of Spanish as a Foreign Language (ELE) and English as a Second Language (ESL), respectively.

The structure of both questionnaires was the same, comprising four parts. The first part included 7 questions covering ethnographic information about the participants, such as age, nationality, country of work, years of experience, academic qualification, etc. The second part had 17 questions about the use of technology-based language learning resources for specific functions. This part started with 3 questions dealing with the general use of these resources in the participants' teaching practices: (1) *Do you use digital tools/resources (ICT, Information and Communication Technologies) in your EFL teaching practice?* (2) *What is your favourite digital tool/resource (ICT) for teaching/learning purposes? (Please write the full name in capital letters)*, and (3) *For which teaching activities or functions do you use digital tools/resources (ICT)?* This last question included the following options: *Assessment, Development and preparation of learning materials, Adaptation of learning materials, Search for learning materials, Game-based language learning, Presentation of didactic content, Synchronous communication, Asynchronous communication, Simulations and Other*. These categories roughly correspond to key functions that came into view during the coding process. Then, the following questions of this second part of the questionnaire integrated the types and subcategories of the classification, corresponding to its first version (for the ELE questionnaire) and its second version (for the ESL questionnaire). This set of questions, based on the classification proposal, asked teachers whether they used the types and subcategories of language learning resources. The third part of the questionnaire comprised 2 questions about the classification of technology-based language learning resources. The first Yes/No question was: *The typology (questions 1 to 14) describes digital tools and resources (ICT) for the teaching-learning of EFL. Do you consider that they include all the pedagogical possibilities regarding the types of tools that can be found in ICT today for the teaching-learning of EFL?* The second question should only be answered if the participant marked "No" in the previous question, and was an open-ended question the objective of which was for the participant to write any possibility of resource type that was not included in the classification: *What type(s) of tools would you include that have not been described in this questionnaire?* This question was used to provide feedback to refine and improve the final version of the classification. The fourth and final part of the questionnaire also comprised 3 questions the purpose of which was to prompt a general reflection about the importance and pedagogical significance of ICTs and their resources within the language learning process, from the point of view of the teachers.

These questionnaires were part of an iterative process and provided the research with both qualitative and quantitative data, which were processed with the grounded theory methodology (conceptual coding) and descriptive statistics, respectively. The purpose of the questionnaires was two-fold. Firstly, and mainly, they provided data and feedback from teachers that was used to refine the classification framework, after a second process of data coding, also with concept of pedagogical functions in mind in order to construct the categories and subcategories. Secondly, the questionnaires also served as an initial and partial validation of the classification framework from the perspective of teachers. Using the grounded theory approach, and after a series of iterations in a non-linear process of analysis of the data (using codes to identify pedagogical functions and uses of the ICT-based resources), the questionnaires served, together with the observation of the resources, to the main objective of the research, that is, the building up of the final proposal of the classification framework.

The reason why two questionnaires for teachers of two target languages (Spanish and English) were used is to show how the classification framework may be valid and may be used in different target languages, and, possibly, in the future, it could also be used to carry out research studies about the real use of the different categories of ICT-based resources, comparing different target languages. Nevertheless, this was beyond the scope and purpose of the present

study, partly because the sample of teachers and the demographic data the questionnaires developed for this study was not sufficient to carry out such a comprehensive analysis. Additionally, it must be noted that the present research is only focused on teachers and, therefore, it cannot draw any conclusion about the point of view of language learners or the usefulness of the framework for learners. This validation is also beyond the scope of this study.

4. A CLASSIFICATION OF TECHNOLOGY-BASED LANGUAGE LEARNING RESOURCES

The classification is organized around two levels. The first level, represented in bold letters in Table 1, describes the pedagogical function of the resources, i.e. the action that is facilitated by the resource within the language learning and teaching process. At this first level, the ICT-based resources may belong to 2 different types: those mainly aimed and used by teachers and those mostly useful for learners. This difference is illustrated in the framework with the letters T (for teachers) and L (for learners). The second level in the classification framework incorporates a characterization of the technology, mechanics, or a more detailed account of the results or use of the pedagogical functions when they are implemented in the language learning and teaching process. A second purpose of the second level, more specifically, is to open the door for new subcategories, since the classification, at this level, can integrate new type descriptions to account for potential new uses and applications of the resources, as technology evolves. An overview of the categories is shown in Table 1:

Table 1: Types of Technology-based Language Learning resources

RESOURCES FOR...
T-1. the development of didactic content: Elaboration of general language learning didactic materials / Creation of game-based didactic materials / Development of questionnaires / AI-based learning materials tool
T-2. obtaining didactic content: Language learning materials download / Access to linguistic input in the L2 / Games and gamified materials download
L-3. language practice and use: Content-based online practice / Content-based practice (through mobile app) / Online language practice / Language practice (through mobile app) / Game-based online language practice / Game-based language practice (through mobile app)
L-4. creativity purposes: Tool for general creative tasks / Tool for language-based creative tasks / AI-based creative tool
L-5. communication and collaboration: Communication tool / Collaborative platforms
T-6. learning management: Learning Management System (LMS) / Learning Experience Platform (LXP) / Self-assessment tool / Language proficiency assessment tool
L-7. notetaking: Note-taking application / Mind map tools
L-8. vocabulary management: Online dictionary / Dictionary mobile app / Concordancer / Searchable lexical database
L-9 translation management: Online translation tool / Translation mobile application
T-10. teaching professional development: Teaching information tool / Teaching training tool / Teaching network/association
L-11. Natural Language Processing: Text-to-speech tool / Speech recognition tool / Speech synthesis tool / Text analysis tool
L-12. social and cultural projects: Social networks / Online communities of practice / Social and cultural project
L-13. ICT-based exploration and basic research: WebQuest / Online resource for basic research purposes
L-14. immersive learning: Virtual world / Virtual Reality tool / Augmented Reality tool / Mixed Reality tool
L-15. text review purposes: Grammar checker / AI-based textual review tool

At the first level, the classification comprises fifteen types of learning resources and tools the identification of which starts with the preposition *for*, reflecting a functional and practical approach in the analysis and description, followed by a statement of the function that plays a role within the language learning process and is enabled or enhanced by the digital resource or tool. The actions used to build up the classification refer to pedagogical functions not necessarily restricted to the area of language learning, but they could also be regarded as relevant to the learning and teaching of other subjects through a target language. The 15 categories of the first-level classification are discussed below.

- 1) 1st category: “*Resources for the development of didactic content*” includes tools that allow teachers to develop all kinds of learning material, from individual Learning Objects (LOs) to complete didactic units and courses, for the purpose of language teaching and learning. These resources are particularly useful for teaching professionals and instructional designers who want to create language learning materials to be used in different learning settings. An especially relevant group of resources within this category are those known as Authoring Tools. Currently, the advent of Artificial Intelligence (AI) has contributed to expanding the possibilities of accessing specific L2 linguistic input that can be used in the development and adaptation of language learning materials. Subcategories within this type are: (1) Elaboration of general language learning didactic materials, (2) Creation of game-based didactic materials; (3) Development of questionnaires, and (4) AI-based learning materials tool.
- 2) 2nd category: “*Resources for obtaining didactic content*” comprises those tools that permit teachers and, in this case, also learners, to access, obtain and download language learning materials. The pedagogical material thus obtained may be both language learning materials or LOs that have been designed to learn a given language, and any relevant digital material that can be useful to teach or learn an L2 because it is expressed in that target language, even though it has not been devised with language learning in mind, such as online encyclopaedias, digital magazines, information dissemination platforms, multimedia repositories, and so on. Practically everything written or spoken in the L2 that is accessible through ICTs belongs to this category and can therefore be a source of linguistic input. The most important subcategories of this second type of resources are: (1) Language learning materials download; (2) Access to linguistic input in the L2, and (3) Games and gamified materials download.
- 3) 3rd category: “*Resources for language practice and use*” covers all the ICT-based tools for learners to practice and improve their linguistic and communicative knowledge and skills. This type of language learning resources implies some kind of interactivity between the learner and the digital resource, which usually includes tasks and exercises whereby the learner uses the interface of the technology to provide some kind of linguistic output, which can be oral or written, although at times the interaction can be more mechanical, for instance in drag-and-drop exercises. Another usual implication of this type of resources is the provision of some sort of feedback or evaluation of the learner’s linguistic production or mechanical action. A distinction must be made between interactive resources of this type and text files (e.g. PDF documents with ready-to-print language exercises) that can be downloaded to be completed by learners without any interaction with the technology, which would be classified within the previous category. Another distinction is made between practising the L2 for general language use and carrying out this practice with the integration of specific content, as well as between online and mobile learning. Some subcategories of this type include: (1) Content-based online practice; (2) Content-based practice (through mobile app);

- (3) Online language practice; (4) Language practice (through mobile app); (5) Game-based online language practice, and (6) Game-based language practice (through mobile app).
- 4) 4th category: “*Resources for creativity purposes*” includes resources and tools that allow learners to develop their creative skills, in a very broad sense, by using the target language. In this type, learners practise the L2, either for generating something that is not directly related to the language under study, or for creating a language-based product or output. AI technology is currently very relevant for this type of resources. The major subcategories of this family are: (1) Tool for general creative tasks; (2) Tool for language-based creative tasks, and (3) AI-based creative tool.
 - 5) 5th category: “*Resources for communication and collaboration*” enables communication between the various participants in the language teaching-learning process and facilitate collaborative work and learning. These resources are especially relevant for language learning, since the main purpose and function of language is communication, as well as collaboration in a process of mutual understanding. ICTs have brought about a wealth of tools directly related not only to communicative functions, like the Social Web, but also to an unprecedented potential for collaboration between humans. Within this group, taking these components into consideration, two subcategories may be identified: (1) Communication tool, and (2) Collaborative platforms.
 - 6) 6th category: “*Resources for learning management*” includes tools, resources and platforms that allow teachers and learners to administer and manage the language teaching and learning processes and tasks. The most popular resources of this kind at present are those referred to as Learning Management Systems (LMS), defined as server-based learning technologies for the planning, creation, management and delivery of course material (Turnbull et al. 2021). A distinction should be made between these LMS, which usually require learners to follow a program as designed by the course provider, and the more modern Learner Experience Platforms (LXPs), which use advanced technologies, such as AI, to make the learning experience better adapted to the learner’s specific needs (Feffer, 2024; Kirvan & Brush, 2024). Tools and resources used for language proficiency assessment (including testing), as well as those which can be used for self-assessment purposes, are also included here, because language assessment and testing may be considered as a fundamental part of learning and teaching management. Thus, this group of resources comprises the following subcategories: (1) Learning Management Systems (LMS); (2) Learning Experience Platforms (LXPs); (3) Self-assessment tool, and (4) Language proficiency assessment tool.
 - 7) 7th category; “*Resources for notetaking*” facilitates the task of collecting information and content from texts, mostly -yet not necessarily- oral, for study, analysis or interpretation purposes. Notetaking is a major academic skill which may be ideally developed with the help of these digital resources. The two most common subcategories here are: (1) Note-taking application, and (2) Mind map tools.
 - 8) 8th category: “*Resources for vocabulary management*” enables learners to get to know, learn and practice the vocabulary of the language under study. There is a further classification into some subcategories, such as (1) Online dictionary; (2) Dictionary mobile app; (3) Concordancer, and (4) Searchable lexical database.
 - 9) 9th category: “*Resources for translation management*” includes tools that assist the translation process and help in learning the target language. Despite certain criticism, translation plays a major role at all levels of language learning and teaching. Current

- ICTs, especially with the advent of AI, have significantly expanded the potential of assistance in translation. The subcategories of this type are: (1) Online translation tool, and (2) Translation mobile application.
- 10) 10th category: “*Resources for teaching professional development*” allows teachers to improve their teaching and pedagogical skills. ICTs provide an ideal environment for the professional development of teachers, with plenty of specific publications, repositories, collaborative platforms, computer-mediated communication tools, associations and networks. The most important subcategories of this family of resources are: (1) Teaching information tool; (2) Teaching training tool, and (3) Teaching network/association.
 - 11) 11th category: “*Resources for Natural Language Processing*” includes a wide range of tools integrating technologies that allow the synthesis, analysis and simulation of natural language. Currently, Natural Language Processing (NLP) technologies have evolved and improved at a very fast pace, partly due to the development of AI. The use of NLP to handle and treat passages of linguistic input and output in the L2 has many possibilities in language learning and teaching. The technical functionalities of NLP are considered to establish the subcategories of this group of resources: (1) Text-to-speech tool; (2) Speech recognition tool; (3) Speech synthesis tool, and (4) Text analysis tool.
 - 12) 12th category: “*Resources for social and cultural projects*” enables teachers and students to carry out tasks for socialization and for the promotion, acknowledgment, and understanding of the culture and society of the target language. Like in other areas, ICTs, especially the Social Web, are efficient environments to facilitate unparalleled ways of human interaction, engagement in social networks and access to cultural projects and events. Learners can also get involved in communities of practice and a wide range of social and cultural activities. The subcategories within this family of resources are: (1) Social networks; (2) Online communities of practice, and (3) Social and cultural project.
 - 13) 13th category: “*Resources for ICT-based exploration and basic research*” comprises those tools that facilitate conscious and efficient use of the Web and related ICTs with the aim of improving knowledge on a specific topic, through the target language. ICTs are ideal tools to access huge amounts of information and content about virtually any topic, which makes them suitable to carry out basic research tasks by learners, who can improve not only their communicative skills in the target language, but also and simultaneously, a series of relevant digital literacies. It must be noted that the term “research” here does not refer to higher-order research carried out by teachers and researchers, but to more basic tasks related to knowledge building. This is the fundamental principle behind certain pedagogical activities, such as WebQuests and other Web-based tasks. Thus, several relevant subcategories may be identified within this group: (1) WebQuest, and (2) Online resource for basic research purposes.
 - 14) 14th category: “*Resources for immersive learning*” creates virtual environments where language learning and teaching can take place in an immersive manner. ICTs are used to simulate environments where language learners can interact in the target language with the aim of using that language in a natural and authentic, i.e. contextualized, way. Some subcategories of this type are: (1) Virtual world; (2) Virtual Reality tool; (3) Augmented Reality tool, and (4) Mixed Reality tool.
 - 15) 15th category: “*Resources for text review purposes*”, i.e. tools that use specific technologies to analyse and evaluate oral or written texts to provide the user with correction feedback and suggest changes at the levels of form or meaning for improve-

ment or error detection and rectification. These tools promote the learner's language awareness. The performance and usability of these resources have been significantly improved with the advent of AI. Some subcategories of this group are: (1) Grammar checker, and (2) AI-based textual review tool.

As can be observed in this classification framework, the most important criteria are related to pedagogical functions, rather than to the technology, although the various technologies are sometimes integrated in the definitions. This classification is an open-ended system in two ways. Firstly, a given digital language learning resource might present features and functions from different types, and, consequently, belong to more than one type in the framework. Secondly, the subcategories included in the second level of the classification may not fully cover or exhaust all the technological possibilities in terms of pedagogical functions, since ICTs evolve at a swift pace, and, in the future, new subcategories may arise to describe the new potential.

5. IMPLEMENTATION AND INITIAL VALIDATION OF THE CLASSIFICATION PROPOSAL

The next step in this research study was to put the classification framework into practice, mainly with three objectives: firstly, to illustrate one possible application of the classification, i.e. conducting educational research; secondly, to incorporate feedback from language teachers with the aim of refining and improving the classification, and, thirdly, to carry out an initial evaluation and partial validation of the classification. To meet these objectives, two questionnaires were used, at different times within the research process.

The first study took place in December 2023, using the questionnaire, called *Uso de herramientas digitales en la enseñanza de Español como Lengua Extranjera (ELE)*. It was completed by 36 teachers of ELE. They came from different geographical locations and most of them had more than 10 years of teaching experience. Regarding the use of technology in their language teaching, the most widely used pedagogical functions and teaching tasks for which teachers preferred to implement technology-based resources were: developing/preparing learning materials, presentation of didactic content, assessment and adaptation of learning materials. According to these results, teachers still seem to prefer using new ICTs for rather "traditional" purposes, instead of incorporating other technology-assisted functionalities, such as course management, simulations or gamification.

The specific functions and categories that were used the least were the following: note-taking, online translation tools, NLP-based tools, Web-based exploration and research tools, resources for immersive learning and social networks. This, together with the fact that the most widely mentioned tools were *Canva*, *Paddlet* and *Kahoot* (with a long tradition in the market), could lead to the conclusion that teachers (at least the participants in the study) do not take full advantage of the many possibilities of ICTs in terms of pedagogical functions within the language learning process. In the last general reflection question, the answers were coded and summarized, and the teachers believe that the most outstanding aspects that provide an added value to ICT-based resources in language learning are dynamism and variety in the learning process, as well as interaction, communication and collaboration possibilities.

The second study, which used the questionnaire in English called *Use of digital resources for learning English as a Foreign Language (EFL)*, was conducted in April 2024 and was completed by 35 teachers of English. The size, gender, age range and years of experience of this group were equivalent to those in the first ELE study. Among most teachers, the preferred ICT-based tools were, like in the first study, resources that are well-known and have been available for a long time: *Kahoot*, *Quizlet*, *Moodle* and *YouTube*, which may lead to the same conclusion. The same is true of the functions of ICTs for language teaching that were more significantly

mentioned by teachers in the study, i.e. development and preparation of learning materials, presentation of didactic content, search for learning materials and, unlike in the first study, game-based learning.

The results showed that the functions, types and subcategories that were used the least by teachers were the following: translation apps, note-taking, NLP-based tools, teaching professional development tools, social networks, Web-based research and exploratory activities and tools for immersive learning (VR and AR). These results, consistent with the ELE study, suggest that language teachers tend to use ICTs and their related tools in a rather traditional manner still. The studies showed that the classification can be used to design surveys and draw conclusions on the actual use of ICT-based language learning resources and on whether teachers or learners take full advantage of the pedagogical potential of these resources.

The two studies carried out using the questionnaires were also aimed at carrying out an initial validation of the classification, by incorporating the teachers' point of view. To meet this objective, the questionnaires incorporated two questions about the teachers' opinions and views on whether the classification of resources is adequate and comprehensive. In both studies, as shown in Figure 1, the results were equally positive on this issue, and most teachers (except for 3 participants in each case) considered that the classification framework is comprehensive and adequate to account for such potential.



Figure 1: **Question - Do you consider that they include all the pedagogical possibilities regarding the types of tools that can be found in ICT today**

The answers given by these 6 teachers were analysed and coded, and they coincided in that AI should be integrated into the classification. Consequently, this feedback was provided and incorporated into the classification framework to refine and improve it, as part of the iterative research process. Thus, AI appears in different types and subcategories in the final version of the classification.

6. CONCLUSIONS AND FUTURE WORK

If language teachers and learners are to take full advantage of the great potential of ICTs within the teaching and learning process in terms of pedagogical functions, it is necessary to know exactly not only what relevant resources and tools are available and how they work, but also to ascertain how they can be used in the most pedagogically efficient way when learning or teaching a language, which is the main purpose of the classification proposal presented here.

Constructing a coherent and comprehensive classification of technology-based language learning resources is a very complex endeavour, for different reasons. First, language teaching and learning is a dynamic and multifaceted area in continuous evolution, and the classification criteria and purposes often emerge from heterogeneous sources. Second, the field is subject to

diverse terminology, since the terms used by different researchers to refer to the categories are often diverging or conflicting. Third, the establishment of a typology involves setting boundaries between functions and concepts that are usually overlapping.

In this study, a comprehensive classification of technology-based language learning resources has been proposed and initially validated. All the information from the data, previous educational research, the teachers participating in the two studies and the researchers have been incorporated into the types and categories of the classification framework, in an iterative process of improvement. The result is a classification framework that has great potential from a pedagogical standpoint, both in theoretical and practical terms. Additionally, it can take different formats depending on the context of use: database, software, application, spreadsheet, online platform, repository, template, checklist, and form, among others. It can be useful for instructional designers mostly at two levels: it may be employed to consider which function(s) should be integrated when designing specific language learning materials or courseware, and it can be used to clearly establish the pedagogical function(s) of the tools or to incorporate different didactic possibilities to their development.

Furthermore, it holds distinct possibilities of implementation for language teachers, at least in four areas of their professional practice: teaching management, learning materials development, learning tools use, and professional development. Also, language learners may use the classification for several fundamental purposes in their language learning process. Language learners, like teachers, can implement the classification for two tasks: organizing their learning and ensuring variety in their contact with the L2. Another key purpose of the classification is to adapt the language learning tasks to their individual preferences. At a more general level, the classification framework may serve for creating an organized and searchable library or database of technology-based language learning resources. This would enable a pedagogically coherent integration of technology within that process.

We are conscious that this research has some limitations, especially in terms of the number of teachers that participated in the initial and partial validation and regarding the method to carry out the qualitative coding of the information used in the construction of the classification. Although this coding was systematic, at times some intuitive or more subjective views may have resulted in biased interpretations. Thus, two future research work possibilities may be suggested: carrying out a more systematic validation of the classification and incorporating AI tools and sophisticated big data processing methods to make the classification even more comprehensive and representative of the potential of ICTs in language learning.

REFERENCES

Alkamel, M. A. A. & Chouthaiwale, S. S. (2018). The Use of ICT Tools in English Language Teaching and Learning: A Literature Review. *Veda's Journal of English Language and Literature-JOELL*, 5(2), 29-33.

Becker, S. A., Brown, M., Dahlstrom, E., Davis, A., DePaul, K., Diaz, V., & Pomerantz, J. (2018). *NMC horizon report: 2018 higher education edition*. Educause. Retrieved from <https://eric.ed.gov/?id=ED594367>

Bhushan, R. (2020). ICT and ELT: The Pedagogical Perspective. In S. Satapathy, V. Bhateja, J. Mohanty & S. Udgata (Eds.), *Smart Intelligent Computing and Applications. Smart Innovation, Systems and Technologies* (pp. 13-20). Springer. https://doi.org/10.1007/978-981-32-9690-9_20

- Budiman, A. (2020). ICT and foreign language learning: An overview. *Tarling: Journal of Language Education*, 3(2), 245-267. <https://doi.org/10.24090/tarling.v3i2.3913>
- Campbell, T., Oh, P. S., & Neilson, D. (2012). Discursive modes and their pedagogical functions in model-based inquiry (MBI) classrooms. *International Journal of Science Education*, 34(15), 2393-2419. <https://doi.org/10.1080/09500693.2012.704552>
- Dash, A. & Kuddus, K. (2020). Leveraging the Benefits of ICT Usage in Teaching of English Language and Literature. In S. Satapathy, V. Bhateja, J. Mohanty, S. Udgata (Eds.), *Smart Intelligent Computing and Applications. Smart Innovation, Systems and Technologies* (pp. 225-232). Springer. https://doi.org/10.1007/978-981-32-9690-9_22
- Farr, F. & Murray, L. (Eds.). (2016). *The Routledge handbook of language learning and technology*. Routledge. <https://doi.org/10.4324/9781315657899>
- Feffer, M. (2024). LXP vs. LMS: What are the differences? *Workplace learning: A complete guide for businesses*. <https://www.techtarget.com/searchhrsoftware/tip/LXP-vs-LMS-What-are-the-differences>
- Godwin-Jones, R. (2016). Looking back and ahead: 20 years of technologies for language learning. *Language Learning & Technology*, 20(2), 5–12. <http://lt.msu.edu/issues/june2016/emerging.pdf>
- Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–31. <https://hdl.handle.net/10125/73570>
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L. & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer assisted language learning*, 27(1), 70-105. <https://doi.org/10.1080/09588221.2012.700315>
- González-Lloret, M. (2017). Technology for Task-based Language Teaching. In C. A. Chapelle & S. Sauro (Eds.) *The Handbook of Technology and Second Language Teaching and Learning* (pp. 234-247). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118914069.ch16>
- González-Lloret, M. (2023). The road System travelled: Five decades of technology in language education. *System*, 118, 103-124. <https://doi.org/10.1016/j.system.2023.103124>
- Gruba, P. & Hinkelman, D. (2012). *Blending technologies in Second Language classrooms*. Palgrave Macmillan. <https://doi.org/10.1057/9780230356825>
- Hiradhar, P. & Bhattacharya, A. (2022). *ICT in English Language Education: Bridging the Teaching-Learning Divide in South Asia*. Springer Nature. <https://doi.org/10.1007/978-981-16-9005-1>
- Kenning, M. M. (2007). *ICT and language learning: From print to the mobile phone*. Springer. <https://doi.org/10.1057/9780230591325>
- Kirvan, P. & Brush, K. (2024). What is a learning management system (LMS)? *Workplace learning: A complete guide for businesses*. Retrieved from <https://www.techtarget.com/searchcio/definition/learning-management-system>

Krajcsó, Z. & Frimmel, U. (2017). Retrieving Online Language Learning Resources: Classification and Quality. *Universal Journal of Educational Research*, 5(1), 11-22. <https://doi.org/10.13189/ujer.2017.050102>

Levy, M. (1997) *CALL: context and conceptualisation*. Oxford University Press.

Levy, M. (2009). Technologies in use for Second Language Learning. *The Modern Language Journal*, 93, 769-782. <https://doi.org/10.1111/j.1540-4781.2009.00972.x>

Li, P. & Lan, Y. J. (2022). Digital Language Learning (DLL): Insights from Behavior, Cognition, and the Brain. *Bilingualism: Language and Cognition*, 25, 361–378. <https://doi.org/10.1017/S1366728921000353>

Luo, H. & Lei, J. (2012). Emerging technologies for interactive learning in the ICT age. In J. Jia (Ed.). *Educational stages and interactive learning: From kindergarten to workplace training* (pp. 73-91). IGI Global. <https://doi.org/10.4018/978-1-4666-0137-6.ch005>

Madhavi, E., Sivapurapu, L. & Kati, P. (2023). Value Role of ICT Tools in English Language Teaching and Learning-Emphasis on Covid-19 Pandemic. *World Journal of English Language*, 13(2), 324-324. <https://doi.org/10.5430/wjel.v13n2p324>

Nirmal, P. B. & Mohsen, B. H. (2023), The contributions of technologies and ICT tools for teaching and learning languages. In S. Fong, N. Dey & A. Joshi (Eds.), *ICT Analysis and Applications. Proceedings of ICT4SD 2023. Lecture Notes in Networks and Systems*, vol 782 (pp. 403-409). Springer, Singapore.

Rosell-Aguilar, F. (2017). State of the App: A Taxonomy and Framework for Evaluating Language Learning Mobile Applications. *CALICO Journal*, 34(2), 243–258.

Tinio, V. L. (2003). *ICT in Education*. United Nations Development Programme-Asia Pacific Development Information Programme (pp. 37-41). Retrieved from <https://digitallibrary.un.org/record/524544?ln=en&v=pdf>

Turnbull, D., Chugh, R. & Luck, J. (2021) Learning management systems: a review of the research methodology literature in Australia and China. *International Journal of Research & Method in Education*, 44(2), 164-178. <https://doi.org/10.1080/1743727X.2020.1737002>

Wilkinson, M. (2016), Language Learning with ICT. In W. A. Renandya & H. P. Widodo (Eds.) *English Language Teaching Today. Linking theory and practice* (pp. 257-276). Springer Nature. https://doi.org/10.1007/978-3-319-38834-2_18



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1

Enero-diciembre 2024

Páginas: 34-54

Artículo recibido: 14/08/2024

Artículo aceptado: 01/12/2024

Artículo publicado: 31/01/2025

Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/666>

DOI: <https://doi.org/10.58859/rael.v23i1.666>

¿Tienen GPT-3.5 y GPT-4 un estilo de escritura diferente del estilo humano? Un estudio exploratorio para el español

Do GPT-3.5 and GPT-4 Have a Writing Style Different from Human Style? An Exploratory Study for Spanish

LARA ALONSO SIMÓN

ANA MARÍA FERNÁNDEZ-PAMPILLÓN CESTEROS

MARIANELA FERNÁNDEZ TRINIDAD

MANUEL MÁRQUEZ CRUZ

UNIVERSIDAD COMPLUTENSE DE MADRID

La cuestión que se aborda en este trabajo de investigación es la comprobación, mediante técnicas estadísticas, de que los modelos generativos de lenguaje GPT-3.5 (versión gratuita) y GPT-4 (versión de pago) de ChatGPT tienen un estilo de escritura distinto al de los humanos, y que pueden diferenciarse, al menos, por tres tipos de rasgos: léxicos, signos de puntuación y estructura sintáctica de las oraciones. Determinar si los grandes modelos de lenguaje tienen un estilo propio es relevante de cara a poder detectar la autoría automática de los textos. En trabajos anteriores se construyó un corpus comparable de textos humanos y automáticos en español y, mediante un estudio cualitativo, se localizó un conjunto de rasgos lingüísticos y estilísticos propios de cada autor. En este trabajo se ha podido comprobar cuantitativamente que 17 variables lingüísticas presentan diferencias estadísticamente significativas entre autores humanos y los modelos GPT-3.5 y GPT-4.

Palabras clave: *estilo de escritura; grandes modelos de lenguaje; GPT-3.5; GPT-4; lingüística de corpus*

The aim of this research is to verify, using statistical methods, that the generative language models GPT-3.5 (free version) and GPT-4 (paid version) of ChatGPT have their own writing style distinct from that of humans and that they can be distinguished by at least three types of features: lexical features, punctuation marks and syntactic sentence structure. Determining whether large language models have their own style is relevant in order to detect automatic authorship of texts. In previous work, a comparable corpus of human and automatic texts in Spanish was constructed and, through a qualitative study, a set of linguistic and stylistic features specific to each author was identified. In this work, it has been quantitatively demonstrated that 17 identified linguistic variables show statistically significant differences between human authors and the GPT-3.5 and GPT-4 models.

Keywords: *writing style; large language models; GPT-3.5; GPT-4; corpus linguistics*

Citar como: Alonso Simón, L., Fernández-Pampillón Cesteros, A.M., Fernández Trinidad, M. y Márquez Cruz, M. (2024). ¿Tienen GPT-3.5 y GPT-4 un estilo de escritura diferente del estilo humano? Un estudio exploratorio para el español. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 34-54. <https://doi.org/10.58859/rael.v23i1.666>

1. INTRODUCCIÓN

Los grandes modelos de lenguaje (en inglés LLM – *Large Language Models*) son modelos estadísticos de lenguaje creados mediante el entrenamiento de grandes redes neuronales, con enormes cantidades de texto, y utilizando potentes recursos computacionales (Hadi, Al-Tashi, Qureshi, Shah, Muneer, Irfan, Zafar, Shaikh, Akhtar, Wu y Mirjalili, 2023). Actualmente, estas redes neuronales de varias decenas o centenas de miles de conexiones están creadas utilizando una arquitectura (o variaciones de ésta) denominada Transformador (en inglés *Transformer*) diseñada por Google (Vaswani, Brain, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser y Polosukhin, 2017). Cada una de estas conexiones tiene un peso, representado por un valor numérico de tipo real, que se calcula mediante el entrenamiento. El peso de cada conexión indica la importancia de esa conexión para entender o producir algún componente del lenguaje natural. Así, un gran modelo del lenguaje no es más que un conjunto de cientos de miles de parámetros numéricos aprendidos durante el entrenamiento que contienen implícitamente los patrones y las estructuras lingüísticas de los textos del entrenamiento.

Si bien un mismo LLM puede aplicarse a múltiples tareas de Procesamiento del Lenguaje Natural (en adelante PLN) como la clasificación, la generación de texto o la traducción automática, podemos distinguir entre los LLM creados para entender el lenguaje y aplicados, preferentemente, a tareas de clasificación –p. ej. el modelo del español BETO (Cañete, Chaperon, Fuentes, Ho, Kang y Pérez, 2020)– y los LLM creados para generar lenguaje –p. ej. los modelos de la familia GPT (Radford, Narasimhan, Salimans y Sutskever, 2018)–. En este trabajo nos centramos en los LLM generativos y, en particular, en dos modelos de la familia GPT, GPT-3.5 y GPT-4. Estos modelos subyacentes al sistema ChatGPT pueden generar texto multilingüe y, en el caso del español, con una calidad muy alta.

Los LLM generativos son capaces de producir textos tan coherentes y contextualmente relevantes que es difícil distinguirlos de los textos escritos por humanos (Uchendu, Le y Lee, 2023). De hecho, la investigación de Casal y Kessler (2023) revela que la tasa global de identificación positiva para distinguir textos generados por ChatGPT (GPT-4 Plus versión de pago) de los textos generados por humanos fue solo del 38,9 % en una evaluación realizada por 72 revisores de las 30 principales revistas de Lingüística Aplicada.¹ Así, la alta calidad del lenguaje que generan ha supuesto una mejora sin precedentes de las aplicaciones de PLN (Hadi et al., 2023).

Ocurre, sin embargo, que los LLM generativos también se utilizan con fines maliciosos, como la generación de noticias falsas (desinformación) y reseñas sobre productos y servicios (desprestigio), el plagio en el ámbito académico (suplantación) o el envío de correos electrónicos de amenaza o engaño (Pizarro, 2019; Pavlyshenko, 2022; Cardenuto, Yang, Padilha, Wan, Moreira, Li, Wang, Andaló, Marcel y Rocha, 2023; Crothers, Japkowicz y Viktor, 2023). Es, en estos casos, cuando resulta necesario encontrar métodos y herramientas que ayuden a detectar el uso malicioso de la generación automática de textos (Maloyan, Nutfullin y Ilyushin, 2022). La detección automática o semiautomática de estos textos generados por LLM es un problema difícil, y aún no resuelto, que tiene gran impacto en todos los sectores: político, económico, académico y social (Uchendu et al., 2023).

Nuestro objetivo es contribuir a mejorar la identificación de los textos generados automáticamente. La hipótesis de partida se basa en considerar que cada LLM tiene su propio estilo de escritura y que existen diferencias significativas en cómo utilizan el lenguaje las personas y los LLM. En este trabajo exploramos la viabilidad de la hipótesis evaluando dos LLM utilizados por ChatGPT para generar texto en español. No hemos encontrado trabajos que respondan a esta cuestión, a excepción de nuestro trabajo previo en el que observamos que parecen exis-

¹ El experimento consistía en la identificación, sin entrenamiento previo, de cuatro resúmenes de artículos de investigación lingüística en inglés.

tir, al menos, seis tipos de rasgos lingüísticos potencialmente discriminativos entre los textos humanos y los automáticos: el uso del léxico, la utilización de expresiones idiomáticas, el uso del orden sintáctico canónico SVO, de las estructuras comparativas y superlativas, el manejo de los marcadores discursivos y el uso de la puntuación (Alonso Simón, Gonzalo Gimeno, Fernández-Pampillón Cesteros, Fernández Trinidad y Escandell Vidal, 2023). El objetivo del presente estudio es verificar estadísticamente el carácter discriminativo del léxico, de los signos de puntuación y del orden sintáctico canónico entre textos humanos y textos generados por los modelos GPT-3.5 y GPT-4.

El uso del español como lengua de trabajo es relevante, fundamentalmente, porque es una de las lenguas más habladas (Fernández Vítóres, 2023). Sin embargo, se dispone todavía de escasos recursos y aplicaciones de PLN. En este sentido, este estudio puede contribuir a aumentar el conocimiento y los recursos de PLN para el español.

Hemos organizado el artículo en siete secciones. En la sección 2 presentamos una síntesis de los trabajos previos, junto con sus principales conclusiones. En la sección 3 explicitamos las dos preguntas de investigación y las hipótesis de partida. En la sección 4 describimos el conjunto de datos utilizado para demostrar las hipótesis, el corpus ROBOT-TALK. En la sección 5 describimos la metodología. En la sección 6 discutimos los resultados obtenidos y la forma en que estos responden a las preguntas de investigación planteadas. Finalmente, en la sección 7 presentamos las conclusiones y futuras líneas de trabajo.

2. ESTADO DE LA CUESTIÓN

Hasta ahora, el problema de la distinción entre un texto generado por un LLM y otro escrito por un humano se ha formulado como una tarea de clasificación automática de autoría, habitualmente binaria *humano vs. automático*. Para crear estos clasificadores se han empleado aproximaciones estadísticas basadas en aprendizaje automático tradicional o en aprendizaje profundo (Jawahar, Abdul-Mageed y Lakshmanan, 2020). Los clasificadores basados en aprendizaje automático tradicional necesitan conocer cuáles son las características o rasgos discriminatorios que les permiten realizar dicha clasificación (Jurafsky y Martin, 2024). Estos rasgos deben ser definidos previamente por un especialista.

Ocurre que en los trabajos que hemos revisado, la selección de los rasgos discriminatorios se realiza de forma empírica, sin un estudio lingüístico sistemático previo de los textos (Berber Sardinha, 2024). Sencillamente, se realiza basándose en la intuición de los creadores de los clasificadores (que no son lingüistas), o en experimentos de identificación de textos humanos-automáticos por parte de personas que tampoco son especialistas en lingüística, o, en el mejor de los casos, en trabajos estilométricos² previos sobre análisis de autoría de textos humanos, como en Savoy (2020). Solo conocemos un caso en el que los humanos que participaron en la distinción entre textos humanos y automáticos eran lingüistas, aunque en dicho estudio no se buscaba la creación de un clasificador (Casal y Kessler, 2023).

Una vez definido el conjunto de rasgos, se construye el clasificador, se evalúa su eficacia mediante métricas como la precisión, la cobertura o el valor-F1 y, de forma iterativa, se va ajustando el conjunto de rasgos hasta obtener mejores valores de eficacia (Jawahar *et al.*, 2020; Savoy, 2020). Entre los trabajos que siguen esta estrategia de construcción empírica del vector de rasgos se encuentran Pizarro (2019), Savoy (2020), Nguyen, Hatua y Sung (2023), Shijaku y Canhasi (2023), Zaitzu y Jin (2023), He, Mao y Liu (2024) y Berber Sardinha (2024).

Una segunda estrategia para la clasificación es la que se basa en clasificadores creados con grandes redes neuronales (aprendizaje profundo). En este caso, la red neuronal *aprende* por sí misma durante el entrenamiento cuáles son los rasgos característicos de los textos. El proble-

² La estilometría se basa en el análisis cuantitativo de los textos para tratar de identificar los patrones y características lingüísticas y paralingüísticas propios del estilo de cada autor.

ma es que los rasgos aprendidos están implícitamente incluidos en la configuración de la red neuronal (los pesos de las conexiones entre neuronas) y solo de forma indirecta es posible, en algunos casos, deducir estos rasgos. Esta aproximación es la que se ha seguido en los trabajos de Desaire, Chua, Isom, Jarosova y Hua (2023), Guo, Zhang, Wang, Jiang, Nie, Ding, Yue y Wu (2023), Mitrović, Andreoletti y Ayoub (2023) y Yu, Chen, Feng y Xia (2024).

En otros estudios, se opta por utilizar ambas estrategias (aprendizaje automático y aprendizaje profundo) para construir y evaluar varios clasificadores y encontrar, empíricamente, la mejor opción. Este es el caso de Uchendu, Le, Shu y Lee (2020), Fröhling y Zubiaga (2021), Desaire et al. (2023), Liao, Liu, Dai, Xu, Wu, Zhang, Huang, Zhu, Cai, Li, Liu y Li (2023), Ma, Liu, Yi, Cheng, Huang, Lu y Liu (2023) y Corizzo y Leal-Arenas (2023).

Respecto a la lengua de los textos estudiados, se observa que, de los dieciséis trabajos revisados, todos, excepto uno, se centran en el inglés. La excepción es el trabajo de Zaitzu y Jin (2023) que trabaja con el japonés, aunque en los trabajos de Pizarro (2019) y Corizzo y Leal-Arenas (2023), además del inglés, se incluye el español, y el chino en Guo et al. (2023).

Los LLM generativos utilizados en los trabajos revisados son, mayoritariamente, los modelos GPT en sus versiones 2, 3, 3.5 y 4. Uchendu et al. (2020) y Fröhling y Zubiaga (2021) eligieron GPT-2, GROVER (además de otros modelos); Zaitzu y Jin (2023) utilizaron GPT-3.5 y GPT-4, y el resto usó ChatGPT, el modelo GPT-3.5, o bien no especifican el modelo concreto, como en Corizzo y Leal-Arenas (2023), Desaire et al. (2023), Guo et al. (2023), Liao et al. (2023), Ma et al. (2023), Mitrović et al. (2023), Nguyen et al. (2023), Shijaku y Canhasi (2023), He et al. (2024), Yu et al. (2024) y Berber Sardinha (2024). Finalmente, Pizarro (2019) y Savoy (2020) eligieron los modelos generativos de los bots de Twitter.

La eficacia de los clasificadores construidos en estos trabajos presenta una métrica de evaluación macro-F1 de entre un 70% y un 98%. Sin embargo, es importante señalar que los valores altos se obtienen solamente cuando los corpus de evaluación pertenecen a la misma tipología textual que los textos utilizados en el entrenamiento. Cuando se aplica el clasificador a textos de tipología diferente de la utilizada para el entrenamiento, los valores se quedan en torno al 70%.

Finalmente, en la Tabla 1 se muestra una síntesis de los rasgos utilizados para la clasificación, junto con los autores que los han utilizado. Se ha querido separar los rasgos de puntuación del resto de los rasgos lingüísticos por su frecuente uso en los trabajos revisados y su relevancia para organizar bloques de información y sus relaciones en el texto escrito. Los rasgos de tipo lingüístico han sido utilizados por todos los autores; los rasgos de puntuación, por la mayoría de ellos y, otros tipos de características, paralingüísticas y psicolingüísticas, se han considerado solo de forma puntual.

Los rasgos de puntuación se utilizan en nueve de los dieciséis trabajos. La mayoría no especifica exactamente qué símbolos ha considerado; no obstante, es relevante que tres estudios coinciden en la importancia diferenciadora del estilo que ofrece el uso de la coma y del punto (Fröhling y Zubiaga, 2021; Corizzo y Leal-Arenas, 2023; Ma et al., 2023). En el estudio de Zaitzu y Jin (2023) se utiliza el posicionamiento de las comas como rasgo discriminatorio entre el estilo humano y el automático.

A partir de los valores de estos rasgos, algunos autores extraen conclusiones, no siempre coincidentes, sobre el diferente estilo de escritura en inglés de ChatGPT y el supuesto estilo de los humanos. Así, a nivel léxico, He et al. (2024), Liao et al. (2023), Guo et al. (2023) y Shijaku y Canhasi (2023) concluyen que la densidad de palabras de los humanos es mayor que la de ChatGPT. Por el contrario, Yu et al. (2024) y Berber Sardinha (2024) indican que ChatGPT puede proporcionar un vocabulario más diverso y hacer los textos más informativos. Corizzo y Leal-Arenas (2023) indican que las máquinas utilizan palabras muy frecuentes en los corpus mientras que, en sentido opuesto, Mitrović et al. (2023) sostienen que los textos de ChatGPT tienden a usar palabras poco comunes.

Tabla 1: Síntesis de rasgos distintivos de los textos humanos frente a automáticos

Tipos de rasgos	Autores
1) Rasgos lingüísticos:	Todos
recuento de caracteres y palabras. Incluye: unigramas de caracteres o palabras, TTR (tokens únicos/total-tokens), LD (palabras léxicas únicas/total-palabras), espacios en blanco, saltos de línea, longitud palabras, palabras únicas, palabras por oración, longitud media del texto, números, oraciones por texto, raíces únicas de palabras, frecuencia de palabras funcionales, palabras con mayúscula, errores gramaticales o de tipeo y número de párrafos.	Pizarro (2019), Savoy (2020), Uchendu et al. (2020), Fröhling y Zubiaga (2021), Corizzo y Leal-Arenas (2023), Desaire et al. (2023), Guo et al. (2023), Liao et al. (2023), Ma et al. (2023), Nguyen et al. (2023), Shijaku y Canhasi (2023), He et al. (2024) y Berber Sardinha (2024)
recuento de combinaciones de caracteres. Incluye: n-gramas de diferentes rangos.	Pizarro (2019)
recuento de combinaciones de palabras. Incluye: n-gramas de diferentes rangos.	Pizarro (2019), Fröhling y Zubiaga (2021), Corizzo y Leal-Arenas (2023), Nguyen et al. (2023), Shijaku y Canhasi (2023), He et al. (2024)
recuento de categorías léxicas. Incluye: entidades nombradas, adjetivos, adverbios, conjunciones, sustantivos, nombres propios, números, pronombres, verbos y preposiciones.	Fröhling y Zubiaga (2021), Guo et al. (2023), Liao et al. (2023), Ma et al. (2023), Nguyen et al. (2023), Shijaku y Canhasi (2023), Zaitzu y Jin (2023), He et al. (2024), Yu et al. (2024), Berber Sardinha (2024)
sintáctico-semánticos. Incluye: patrones argumentales y análisis de dependencias.	Guo et al. (2023), Liao et al. (2023), Yu et al. (2024), Berber Sardinha (2024)
semántico-pragmáticos. Incluye: opinión, enfáticos, atenuadores, intensificadores y coherencia semántica.	Corizzo y Leal-Arenas (2023), Guo et al. (2023), Liao et al. (2023), Ma et al. (2023), He et al. (2024), Berber Sardinha (2024)
pragmático-discursivos. Incluye: complejidad del párrafo, lógica de argumentación, partículas discursivas y legibilidad del texto.	Uchendu et al. (2020), Fröhling y Zubiaga (2021), Desaire et al. (2023), Liao et al. (2023), Ma et al. (2023), Mitrović et al. (2023), Nguyen et al. (2023), Berber Sardinha (2024)
2) Rasgos de puntuación. Incluye: paréntesis, guion corto, punto y coma, dos puntos, interrogación, comillas, coma y punto, entre otros.	Fröhling y Zubiaga (2021), Corizzo y Leal-Arenas (2023), Desaire et al. (2023), Guo et al. (2023), Ma et al. (2023), Nguyen et al. (2023), Shijaku y Canhasi (2023), Zaitzu y Jin (2023), Yu et al. (2024)
3) Rasgos paralingüísticos. Incluye: hipervínculos, menciones, <i>hashtags</i> y <i>emojis</i> .	Savoy (2020)
4) Rasgos psicolingüísticos. Incluye: procesos psicológicos y preocupaciones personales.	Uchendu et al. (2020)

A nivel sintáctico, se concluye que los textos humanos tienden a aplicar más modificadores (adjetivales, preposicionales y compuestos) para definir las palabras con precisión (Yu et al., 2024). También sugieren que las proporciones del objeto directo, el sujeto nominal y las palabras raíz³ en los textos humanos son menores que las de los textos de ChatGPT (Yu et al., 2024). Finalmente, se ha observado que ChatGPT tiende a usar combinaciones sistemáticas de palabras, mientras que los humanos utilizan una mayor variedad de palabras individuales en contraposición a combinaciones de palabras (He et al., 2024).

A nivel pragmático-discursivo, Mitrović et al. (2023) concluyeron que ChatGPT tiende a expresarse de forma más impersonal y formal, y que carece de expresiones emocionales, frecuentemente utilizadas por los humanos. Yu et al. (2024) observaron que los humanos suelen prestar más atención a la estructura, la consistencia y la lógica del texto. He et al. (2024) concluyeron que los textos de ChatGPT parecen ser más sistemáticos y menos específicos en contenido que los textos humanos. Guo et al. (2023) determinaron que la frecuente coocurrencia de conjunciones, junto con sustantivos, verbos y preposiciones, indica que la estructura del texto y las relaciones de causa-efecto, progresión o contraste son claras. Finalmente, Berber Sardinha (2024) llevó a cabo un análisis lingüístico sistemático, considerando diversos dominios textuales, y observó que, en la conversación, ChatGPT utiliza aproximadamente la mitad de las características relacionadas con el grado de implicación del emisor en los textos que los humanos. Además, concluyó que ChatGPT no es capaz de captar la complejidad de la gramática del lenguaje hablado y que se basa en patrones del lenguaje escrito para generar los diálogos. También encontró que los textos generados automáticamente tienen 1,4 veces menos probabilidades de emplear referencias dependientes del contexto. Finalmente, observó que los textos generados por ChatGPT no suelen adoptar un tono persuasivo a diferencia de los textos humanos.

3. PREGUNTAS DE INVESTIGACIÓN E HIPÓTESIS

En consonancia con nuestro objetivo, las preguntas de investigación que nos planteamos en este trabajo son:

- 1) ¿podemos diferenciar mediante rasgos lingüísticos el estilo de escritura de GPT-3.5 de un posible estilo general humano en español?
- 2) ¿podemos diferenciar mediante rasgos lingüísticos el estilo de escritura de GPT-4 de un posible estilo general humano en español?

Entendemos por estilo general humano un posible conjunto de rasgos lingüísticos compartidos (si existen) por los autores humanos y que son claramente diferentes de los rasgos lingüísticos compartidos (si existen) por los sistemas de generación automáticos.

La hipótesis de partida es que los modelos GPT-3.5 (concretamente, GPT-3.5-Turbo) y GPT-4 tienen un estilo propio distinto de un estilo general humano y que, entre otros posibles, existen rasgos léxicos, ortográficos y sintácticos que permitirán diferenciarlos. La Tabla 2 recoge los dieciocho rasgos léxicos que hipotetizamos serán diferenciales entre los estilos de escritura humano y los dos robóticos. La Tabla 3 muestra los veintiún rasgos ortográficos potencialmente discriminativos.

³ En análisis de dependencias, palabra principal de la que dependen directa o indirectamente las demás.

Tabla 2: Propuesta de rasgos léxicos diferenciadores de los estilos de escritura de GPT-3.5 y GPT-4 frente al humano

Rasgos léxicos	Valor
Ratio Tipo de Token o <i>TTR (Type-Token Ratio)</i>	palabras únicas/total palabras
Densidad Léxica o <i>LD (Lexical Density)</i>	palabras léxicas únicas/total palabras
proporción de adjetivos	total adjetivos/total palabras
proporción de adjetivos únicos	total adjetivos únicos/total palabras
proporción de adverbios	total adverbios/total palabras
proporción de adverbios únicos	total adverbios únicos/total palabras
proporción de conjunciones	total conjunciones/total palabras
proporción de conjunciones únicas	total conjunciones únicas/total palabras
proporción de sustantivos	total sustantivos/total palabras
proporción de sustantivos únicos	total sustantivos únicos/total palabras
proporción de preposiciones	total preposiciones/total palabras
proporción de preposiciones únicas	total conjunciones únicas/total palabras
proporción de pronombres	total pronombres/total palabras
proporción de pronombres únicos	total pronombres únicos/total palabras
proporción de verbos	total verbos/total palabras
proporción de verbos únicos	total verbos únicos/total palabras
densidad de <i>hápax legómenon</i>	palabras con frecuencia 1/vocabulario único
densidad <i>dís legómenon</i>	palabras con frecuencia 2/vocabulario único

Tabla 3: Propuesta de rasgos ortográficos diferenciadores de los estilos de escritura de GPT-3.5 y GPT-4 frente al humano

Rasgo	Valor
proporción de símbolos ortográficos total	total puntuación/total tokens
proporción de comillas (comillas rectas dobles o simples, comillas españolas, comillas inclinadas)	total comillas/total tokens
proporción de dos puntos	total dos puntos/total tokens
proporción de comas	total comas/total tokens
proporción de puntos suspensivos o la abreviatura ‘etc.’	total ps_etc/total tokens
proporción de corchetes de apertura	total carácter/total tokens
proporción de corchetes de cierre	total carácter/total tokens
proporción de guiones cortos	total carácter/total tokens
proporción de admiraciones de apertura	total carácter/total tokens
proporción de admiraciones de cierre	total carácter/total tokens

(Tabla 3, continúa en la página siguiente)

(Tabla 3, continúa de la página anterior)

Rasgo	Valor
proporción de paréntesis de apertura	total carácter/total tokens
proporción de paréntesis de cierre	total carácter/total tokens
proporción de puntos (punto y seguido o punto con cambio de línea)	total puntos/total tokens
proporción de barras inclinadas	total carácter/total tokens
proporción de interrogaciones de apertura	total carácter/total tokens
proporción de interrogaciones de cierre	total carácter/total tokens
proporción de llaves de apertura	total carácter/total tokens
proporción de llaves de cierre	total carácter/total tokens
proporción de símbolo de porcentaje ‘%’	total carácter/total tokens
proporción de punto y comas	total carácter/total tokens
proporción de otra puntuación (p. ej. guion normal, guion largo, marca de párrafo, signo + o asteriscos)	total otra puntuación/total tokens

Finalmente, los rasgos sintácticos discriminatorios pueden ser:

- 1) La proporción del número de oraciones por el total de palabras en el texto.
- 2) La proporción de oraciones con estructura canónica SVO.

4. EL CORPUS ROBOT-TALK

Para comprobar las hipótesis hemos utilizado una muestra del corpus ROBOT-TALK.⁴ Se trata de un corpus comparable de textos en español (artículos científicos de lingüística, noticias y reseñas de cine) escritos por humanos y por cuatro LLM diferentes. La selección de los tres géneros se ha realizado de forma que estén representados textos conforme diferentes niveles de uso de lenguaje formulaico. Actualmente, consta de un total de 765 textos, la mitad son humanos y la otra mitad son textos comparables generados automáticamente por los LLM ChatGPT-3.5-turbo de OpenAI⁵, ChatGPT-4 de OpenAI⁶, Gemini de Google⁷ y Mixtral-8x7B-Instruct-v0.1 de Mixtral AI⁸. El periodo de recogida de datos abarca desde julio de 2023 hasta abril de 2024. La Tabla 4 muestra el número de textos por autor y por dominio y la Tabla 5 recoge el número de tokens por autor y por dominio.

Desde el punto de vista de la clasificación automática, el corpus ROBOT-TALK es un corpus multiclase respecto a los cinco posibles autores: humano o uno de los cuatro modelos de lenguaje. En este sentido, para la tarea de detección de textos generados automáticamente, el corpus puede utilizarse, bien para la clasificación multiclase (atribución de autoría) con las etiquetas *humano*, *modelo1*, *modelo2*, *modelo3* y *modelo4*, bien para la clasificación binaria (Turing Test) con las etiquetas *humano* y *modelo*.

⁴ El corpus ROBOT-TALK se ha creado en el marco del proyecto de investigación PID2022-140897OB-I00.

⁵ <https://openai.com/index/gpt-3-5-turbo-fine-tuning-and-api-updates/>.

⁶ <https://openai.com/index/gpt-4/>.

⁷ <https://gemini.google.com/app?hl=es-ES>.

⁸ <https://mistral.ai/news/mixtral-of-experts/>.

Tabla 4: Número de textos por autor y dominio del corpus ROBOT-TALK

Dominio de los textos	Autor					N.º de textos por dominio
	Humano	GPT-3.5-Turbo	GPT-4	Gemini	Mixtral	
Artículos científicos	23	23	23	23	23	115
Noticias	65	65	65	65	65	325
Reseñas de cine	65	65	65	65	65	325
Número de textos	153	153	153	153	153	765

Tabla 5: Número de tokens por autor y dominio del Corpus ROBOT-TALK

Dominio de los textos	Autor					N.º de tokens por dominio
	Humano	GPT-3.5-Turbo	GPT-4	Gemini	Mixtral	
Artículos científicos	33.189	21.833	23.508	27.791	33.043	139.364
Noticias	30.003	24.821	30.811	25.084	29.013	139.732
Reseñas de cine	24.667	21.969	24.300	23.928	24.541	119.405
Número de tokens	87.859	68.623	78.619	76.803	86.597	398.501

Hasta donde sabemos, por el momento, no existe otro corpus comparable de textos escritos por hablantes nativos de español, cuya autoría humana haya sido rigurosamente comprobada⁹, y por textos equivalentes y no híbridos (combinaciones de oraciones escritas por humanos y oraciones producidas por modelos generativos), generados por GPT-3.5-Turbo, GPT-4, Gemini y Mixtral-8x7B-Instruct-v0.1 que, además, contemplen los tres dominios (artículos, noticias y reseñas).

5. MÉTODO

Para comprobar si los rasgos lingüísticos definidos en la hipótesis asumen valores que resulten ser característicos de los dos modelos generativos considerados en este estudio, GPT-3.5 y GPT-4, y de un estilo humano común de escritura, hemos utilizado los análisis estadísticos de varianza y covarianza que nos permiten conocer si las medias obtenidas para cada rasgo lingüístico son significativamente diferentes entre los autores.

5.1. Muestra

La muestra utilizada tiene 180 textos del corpus ROBOT-TALK, elegidos al azar y de forma equilibrada respecto a la tipología (20 artículos, 20 noticias y 20 reseñas) y a cada tipo de autor: humano, GPT-3.5 y GPT-4 (Tabla 6).

⁹ Sarvazyan, González, Franco-Salvador, Rangel, Chulvi y Rosso (2023) utilizan textos en español en el corpus AuTexTification, pero (1) no se ha comprobado la autoría humana de los textos etiquetados como humanos y (2) no es un corpus comparable.

Tabla 6: Número de textos por autor y dominio de la muestra utilizada

Dominio de los textos	Autor			N.º de textos por dominio
	Humano	GPT-3.5-Turbo	GPT-4	
Artículos científicos	20	20	20	60
Noticias	20	20	20	60
Reseñas de cine	20	20	20	60
Número de textos	60	60	60	180

5.2. Herramientas

Hemos utilizado la herramienta de análisis textual Sketch Engine (Kilgarriff, Baisa, Bušta, Jakubiček, Kovář, Michelfeit, Rychlý y Suchomel, 2014) para construir el corpus de muestra y obtener automáticamente los valores de cada una de las variables.¹⁰ Los datos se han almacenado en hojas de cálculo. El análisis estadístico se ha realizado con el programa SPSS.

5.3. Hipótesis

H1.1: existen diferencias significativas entre los valores medios de las variables evaluadas en los textos de GPT-3.5 y humanos.

H1.2: existen diferencias significativas entre los valores medios de las variables evaluadas, estudiadas por dominios (artículos, noticias y reseñas), en los textos de GPT-3.5 y humanos.

H2.1: existen diferencias significativas entre los valores medios de las variables evaluadas en los textos de GPT-4 y humanos.

H2.2: existen diferencias significativas entre los valores medios de las variables evaluadas, estudiadas por dominios (artículos, noticias y reseñas), en los textos de GPT-4 y humanos.

5.4. Procedimiento

Hemos realizado cuatro análisis comparando los textos dos a dos:

- 1) GPT-3.5 vs. humano en todos los textos
- 2) GPT-3.5 vs. humano en cada tipo de texto
- 3) GPT-4 vs. humano en todos los textos
- 4) GPT-4 vs. humano en cada tipo de texto

Los análisis realizados han sido:

- 1) en las comparaciones con todos los textos, prueba de t-Student pareada, si la muestra presenta una distribución normal tras un test Shapiro-Wilk y rangos signados de Wilcoxon, si la distribución se aleja de la normalidad.
- 2) en las comparaciones múltiples respecto a cada tipo de texto utilizamos ANOVA y test de Duncan, en caso de normalidad, y ANOVA y test de Kruskal-Wallis, si la muestra se aleja de la normalidad.

¹⁰ <http://www.sketchengine.eu>.

6. RESULTADOS Y DISCUSIÓN

Los resultados detallados de las diferencias obtenidas para los 41 rasgos estudiados entre GPT-3.5 y humanos y entre GPT-4 y humanos pueden encontrarse en la plataforma Open Science Framework.¹¹ Resumimos y discutimos los resultados más sobresalientes en las tres subsecciones siguientes.

6.1. Diferencias en los rasgos léxicos

Sin tener en cuenta las distinciones de dominio al que pertenecen los textos (Figura 1):

- 1) Existe una diferencia moderadamente significativa en el TTR en favor de los humanos, que indica que los textos humanos tienen mayor riqueza léxica que los textos de GPT-3.5 y GPT-4 (+2,3% y +2,8%).
- 2) En relación con la LD podemos observar que no hay diferencias significativas, por lo que el grado de informatividad de los textos es similar entre los textos LLM y los humanos. Curiosamente, GPT-4 utiliza, con una diferencia significativa baja, menos palabras léxicas únicas que los humanos, mientras que esta diferencia no es significativa entre GPT-3.5 y los humanos.
- 3) Las categorías léxicas de adjetivo, adverbio y pronombre, tanto en su frecuencia de uso como en el vocabulario (número de palabras únicas), presentan diferencias muy significativas entre los LLM (GPT-3.5 y GPT-4) y los humanos (Figura 1). Así, los textos humanos presentan un mayor uso y variedad de adverbios y pronombres mientras que, curiosamente, los textos generados por GPT-3.5 y GPT-4 exhiben más adjetivos y adjetivos únicos en comparación con los textos humanos. Además, GPT-3.5 se diferencia de los humanos en el uso de los adverbios únicos y las conjunciones únicas de forma muy significativa. Sin embargo, esto no ocurre en los textos de GPT-4.
- 4) Hay seis rasgos en los que las diferencias entre LLM y humanos no son significativas o son poco significativas: nombres y nombres únicos, preposiciones y preposiciones únicas y verbos y verbos únicos.
- 5) Las diferencias entre GPT-4 y los humanos son ligeramente menos marcadas en algunos rasgos que entre GPT-3.5 y los humanos. En español, y respecto a estos rasgos, parece que el estilo de GPT-4 es algo más cercano al humano que el de GPT-3.5. Esta observación se mantiene también en los tres dominios (Figuras 2, 3 y 4).

Las diferencias observadas entre los rasgos léxicos en las comparaciones entre GPT-3.5 vs. humanos y GPT-4 vs. humanos también se confirman en cada uno de los dominios (Figuras 2, 3 y 4). Esto podría indicar que, además de diferenciarse de los humanos, GPT-3.5 y GPT-4 podrían tener sus propios estilos de escritura.

Analizando los resultados por dominios observamos que el dominio de las reseñas es el más productivo, pues presenta el mayor número de rasgos diferenciales, doce, que permitirían distinguir el estilo de los textos de los LLM de los humanos (Figura 4). Resulta llamativo que, en el dominio más estructurado y formulaico de los artículos lingüísticos, exista también un número importante de rasgos distintivos, once para GPT-3.5 y ocho para GPT-4 (Figura 2). Sin embargo, el dominio de las noticias (Figura 3), con un nivel intermedio de restricciones estilísticas, ha resultado ser el menos adecuado para caracterizar el estilo humano o robótico (GPT-3.5 y GPT-4).

¹¹ https://osf.io/3ahb7/?view_only=076c6b452d68430e812826d91accd9d0

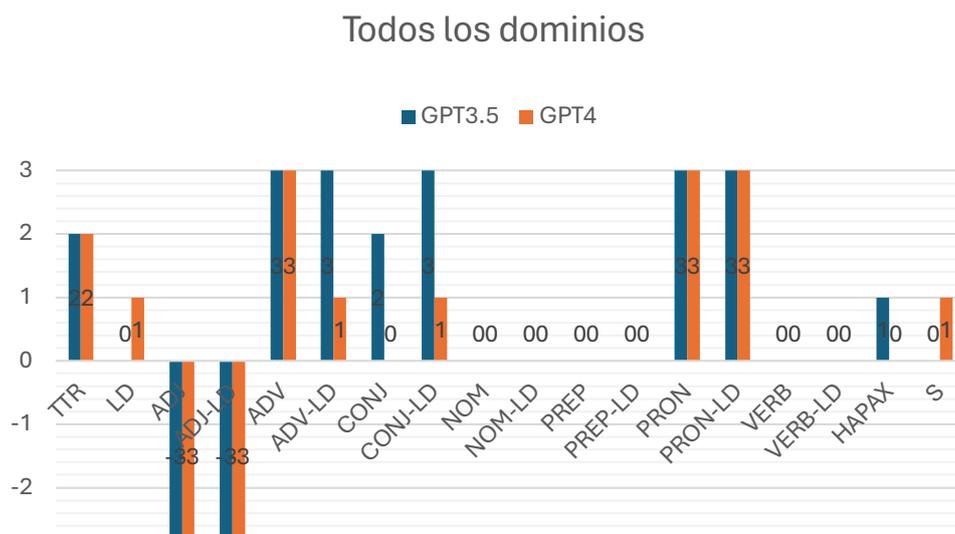


Figura 1: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos léxicos analizados en todos los dominios

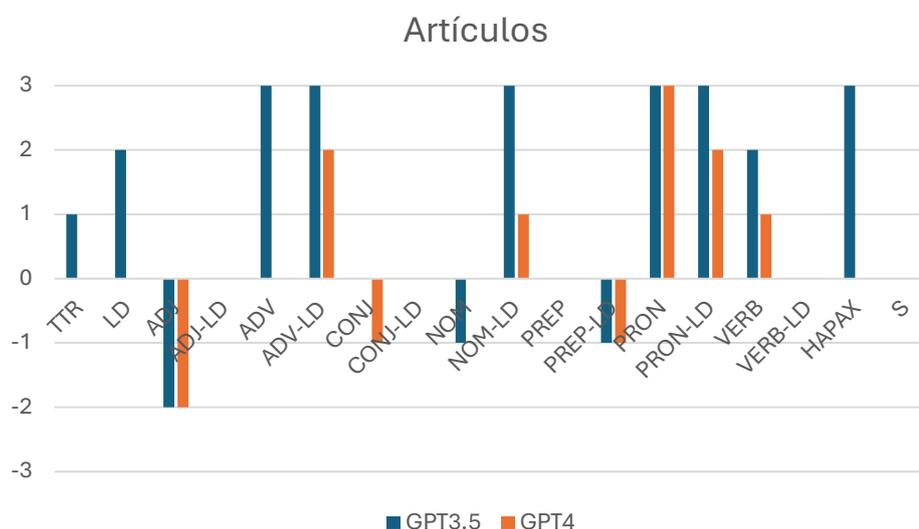


Figura 2: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos léxicos analizados en los textos pertenecientes al dominio de los artículos científicos

Centrándonos en el dominio de los artículos científicos, cabe destacar tres resultados (Figura 2):

- 1) En este dominio, GPT-4 presenta una riqueza de vocabulario semejante a la de los humanos, por los valores nulos que presenta en la diferencia modelo vs. humano en los rasgos TTR y LD, y superior a la de GPT-3.5.
- 2) En los rasgos adjetivos, conjunciones únicas (sólo GPT4), nombres (sólo GPT3.5) y preposiciones únicas los modelos presentan una mayor riqueza de uso o de vocabulario que los humanos.
- 3) Los humanos utilizan significativamente más hápax (palabras únicas) que GPT-3.5; sin embargo, no se han hallado diferencias para este rasgo en la comparación GPT-4 y humanos.

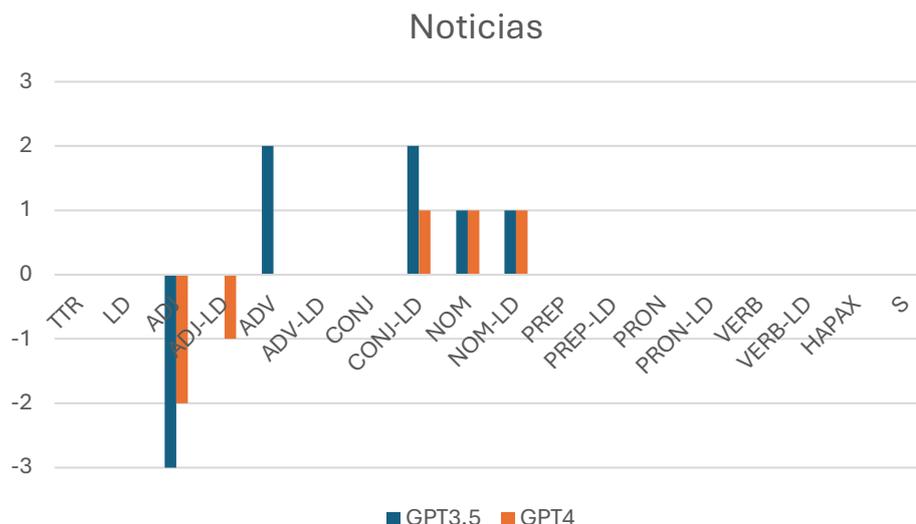


Figura 3: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos léxicos analizados en los textos pertenecientes al dominio de las noticias

En el dominio de noticias, es reseñable que GPT4 es el modelo en el que las diferencias con los humanos tienen menos significancia respecto a GPT3.5 (Figura 3). Podríamos decir que, en este dominio, es en el que existe más parecido entre GPT4 y los humanos. Los resultados son coherentes con los obtenidos en el ámbito de los artículos científicos.



Figura 4: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos léxicos analizados en los textos pertenecientes al dominio de las reseñas de cine

En el dominio de las reseñas, ámbito en el que encontramos más rasgos diferenciales, constatamos que:

- 1) Al contrario de lo que observábamos para los artículos, GPT-3.5 presenta una densidad léxica mayor en comparación con los humanos.

- 2) Los modelos presentan una mayor riqueza en el uso de adjetivos y adjetivos únicos (también ocurre en artículos y noticias), nombres (solo GPT-4) y verbos únicos (solo GPT-3.5) respecto a los textos humanos.
- 3) Existen diferencias entre los rasgos distintivos entre GPT-3.5 vs. humanos y GPT-4 vs. humanos. Esto podría indicar que GPT-3.5 y GPT-4 podrían tener sus propios estilos de escritura diferentes entre sí, además de diferenciarse, en conjunto, del estilo de escritura humano.

6.2. Diferencias en los rasgos de puntuación

Considerando el conjunto de todos los textos, observamos que, en español y en el corpus ROBOT-TALK, los humanos utilizan, en total, más signos de puntuación que GPT-3.5 y GPT-4 —de forma muy significativa en GPT-3.5 y de forma moderada en GPT-4— excepto en la utilización del punto. En este caso, los modelos usan significativamente más puntos que los humanos (Figura 5 y Tabla 7).

Específicamente, encontramos diferencias muy significativas en diez signos de puntuación: comillas, dos puntos, comas, guion corto, paréntesis de apertura y de cierre, punto, barra inclinada, punto y coma y otros signos que incluyen guion normal, guion largo, marca de párrafo, signo + y asteriscos. Esta importante cantidad de rasgos de puntuación diferencial es un resultado relevante de cara a poder identificar los estilos de los modelos frente a los humanos.

También encontramos siete rasgos que parecen no ser relevantes, entre los que se encuentran, de forma inesperada, las admiraciones. Sin embargo, conviene tener en cuenta que, respecto a cuatro de ellos, apertura y cierre de los corchetes y las llaves, no disponemos de datos porque no aparecen en los textos del corpus. Respecto a los otros tres (apertura y cierre de admiraciones y tanto por ciento), no disponemos de datos suficientes para poder extraer conclusiones.

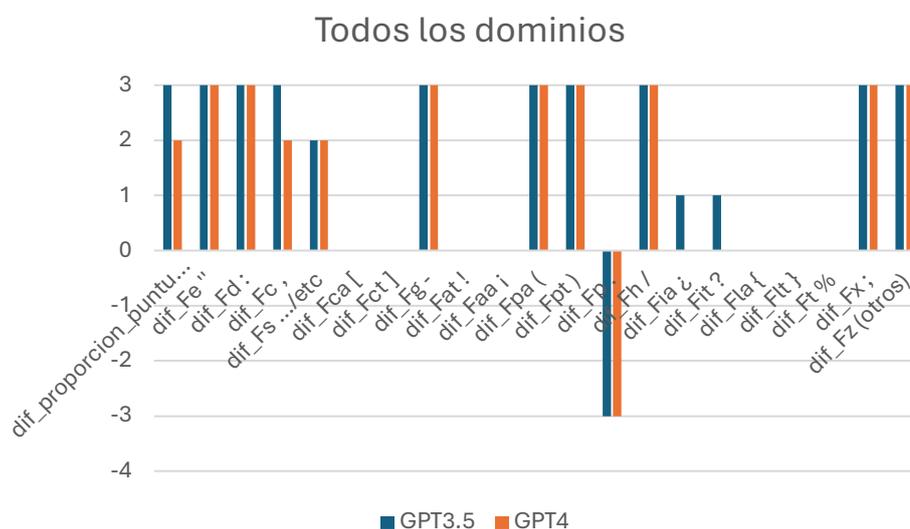


Figura 5: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos de puntuación analizados en todos los dominios

Tabla 7: Nivel de significancia de la diferencia entre GPT-3.5 y GPT-4 respecto a los humanos para los rasgos de puntuación analizados en todos los dominios

	GPT3.5	GPT4
dif_proporcion_puntuacion	3	2
dif_Fe "	3	3
dif_Fd :	3	3
dif_Fc ,	3	2
dif_Fs .../etc	2	2
dif_Fca [0	0
dif_Fct]	0	0
dif_Fg -	3	3
dif_Fat !	0	0
dif_Faa ¡	0	0
dif_Fpa (3	3
dif_Fpt)	3	3
dif_Fp .	-3	-3
dif_Fh /	3	3
dif_Fia ¿	1	0
dif_Fit ?	1	0
dif_Fla {	0	0
dif_Flt }	0	0
dif_Ft %	0	0
dif_Fx ;	3	3
dif_Fz (otros)	3	3

Considerando cada dominio de forma independiente, en los artículos (Figura 6) se corroboran prácticamente todos los resultados hallados a nivel global. Las pequeñas discrepancias deberían estudiarse de forma cualitativa.

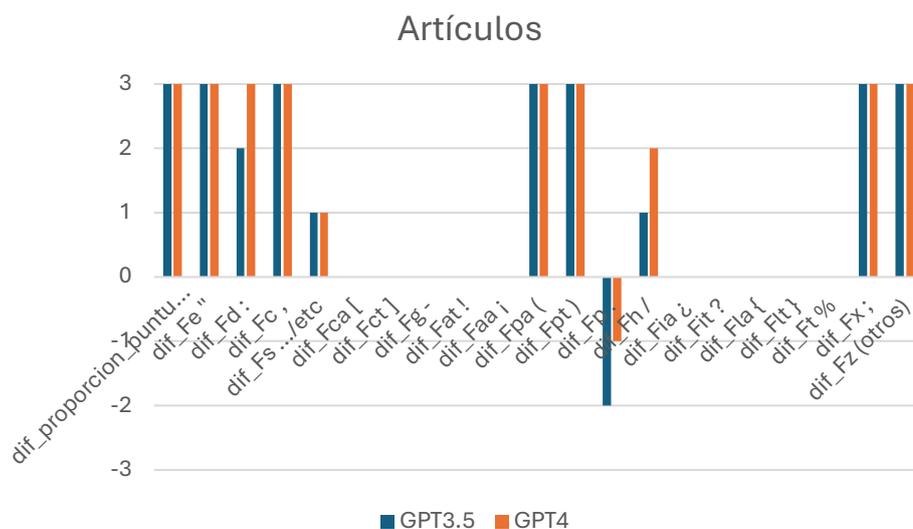


Figura 6: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos de puntuación analizados en todos los textos pertenecientes al dominio de los artículos científicos

En las noticias ocurre, igual que sucede con los rasgos léxicos, que los rasgos diferenciales de puntuación son sensiblemente menores que en los otros dos dominios, corroborando que los estilos de GPT-3.5 y, especialmente el de GPT-4, son más similares al estilo humano. Únicamente el uso de comillas, de paréntesis y de puntos pueden ayudar a distinguirlos para GPT-4 y, además, las comas para GPT-3.5 (Figura 7).

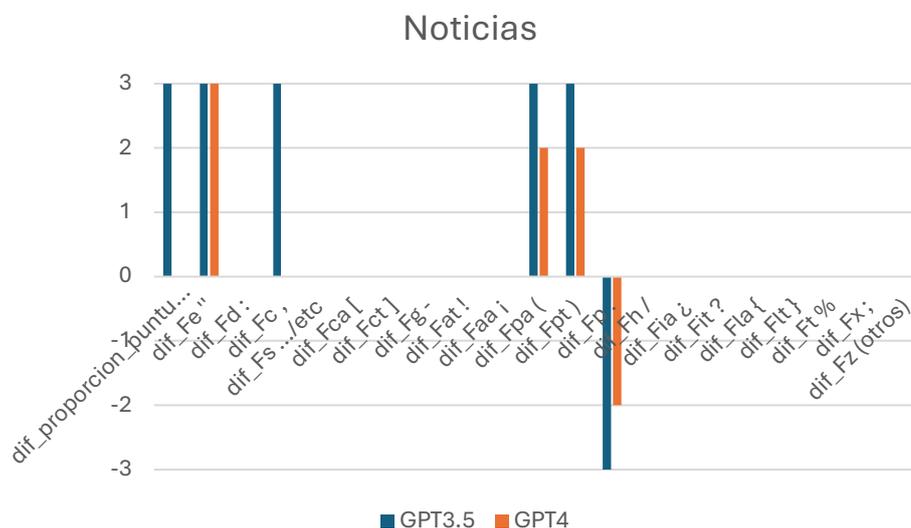


Figura 7: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos de puntuación analizados en los textos pertenecientes al dominio de las noticias

Finalmente, y a diferencia de los rasgos léxicos, las reseñas es el dominio donde los signos de puntuación son menos relevantes para identificar los estilos de GPT-3.5 y GPT-4 frente a los humanos. Se siguen manteniendo como rasgos relevantes el uso de las comillas, los paréntesis y los puntos (Figura 8).



Figura 8: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos de puntuación analizados en los textos pertenecientes al dominio de las reseñas de cine

En definitiva, en el análisis por dominios, es en los artículos donde se hallan las mayores diferencias entre los LLM y los humanos en cuanto al uso de la puntuación. Además, parece que los signos más informativos son las comillas, los paréntesis y los puntos: los humanos utilizan más comillas y paréntesis que GPT-3.5 y que GPT-4 y estos modelos utilizan más puntos que los humanos.

6.3. Diferencias en los rasgos sintácticos

Considerando todos los dominios, se observa que la proporción del número de oraciones por texto es significativamente mayor en los textos generados por GPT-3.5 y GPT-4 que en los textos humanos (Figura 9). El uso del orden canónico SVO es mayor en GPT-3.5 que en los humanos, aunque con una significancia baja y no se observan diferencias entre GPT-4 y los humanos. El análisis por dominios confirma estos resultados, siendo más relevante en los artículos que en las noticias y reseñas. De forma más detallada, el número de oraciones por texto es significativamente mayor en GPT-3.5 que en los humanos en todos los dominios mientras que en GPT4 la significancia es baja en las noticias y reseñas, La diferencia en la proporción del uso del orden canónico SVO tiene una significancia baja en los artículos y reseñas y no se observan diferencias en las noticias. Así pues, el único rasgo sintáctico relevante parece ser la proporción de oraciones en los textos (Figura 10).



Figura 9: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos sintácticos analizados en todos los dominios



Figura 10: Nivel de significancia de la diferencia entre los LLM GPT-3.5 y GPT-4 y los humanos para los rasgos sintácticos analizados por dominios

7. CONCLUSIONES, TRABAJO ACTUAL Y FUTURO

La pregunta de investigación que motiva este trabajo es si es posible verificar que, en español, los modelos GPT-3.5 y GPT-4 presentan un estilo de escritura característico identificable mediante rasgos lingüísticos y, si este estilo es diferente de un posible estilo general humano. Los resultados muestran que en 17 de los 41 rasgos analizados se exhiben diferencias significativas muy altas al comparar textos generados por GPT-3.5 y por GPT-4 con textos humanos.

Concretamente, se trata de los rasgos léxicos de uso de adjetivos y adjetivos únicos que es mayor en los LLM que en los humanos, así como de adverbios y adverbios únicos y de pronombres y pronombres únicos que es mayor en los humanos que en los LLM (aunque en GPT-4 el rasgo de adverbios únicos tiene una significancia más baja). Los rasgos de puntuación, uso de puntuación total, paréntesis, punto, coma, dos puntos, comillas, guion corto, barra inclinada, punto y coma y otros signos, y el rasgo sintáctico del número de oraciones por texto también han resultado ser caracterizadores del estilo de escritura de los LLM estudiados. En este caso, estos rasgos se presentan en mayor cantidad en los textos humanos que en los generados por GPT-3.5 y GPT4 excepto en el caso del punto que es a la inversa: los LLM utilizan más puntos que los humanos, lo que es congruente con el uso de más oraciones por texto.

Aunque el tamaño de la muestra escogida $n=180$ (60 textos por cada *autor* repartidos en 20 textos por dominio) proporciona suficiente confianza en los resultados obtenidos para todos los dominios, hemos encontrado algunos resultados referidos a cada tipo textual (artículos científicos, noticias y reseñas) que no son siempre concluyentes, pero que, por su interés, merecerían comprobarse repitiendo el estudio con una muestra aumentada. Además, es necesario confirmar con una muestra más amplia aquellas diferencias que resultaron ser significativas de forma moderada o baja. Sería interesante también añadir otras variables lingüísticas para continuar verificando las diferencias de estilo entre los LLM y los humanos. En este sentido, este estudio cuantitativo se está completando actualmente con un análisis cualitativo del uso de estos rasgos diferenciadores y con la búsqueda de otros posibles rasgos distintivos.

Los rasgos explorados en este estudio habían sido utilizados en trabajos anteriores para la construcción de clasificadores automáticos que diferencian texto humano de texto automático, fundamentalmente en lengua inglesa. Sin embargo, hasta el momento no se había verificado de forma intrínseca el grado de informatividad de éstos. Tampoco se había llevado a cabo un estudio de estas características en lengua española. Verificar este resultado es relevante para mejorar los métodos de detección de textos generados por modelos, en los que se apliquen soluciones informáticas y también lingüísticas, especialmente de lingüística forense, en los que interesa conocer quién es el autor de textos generados con fines maliciosos. Asimismo, creemos que contribuirá a mejorar la construcción de clasificadores automáticos y permitiría entrenar de forma más eficaz estos sistemas, mediante la eliminación de los rasgos que no pertenecen a los estilos de escritura de los modelos. Esta conclusión, en todo caso, todavía debe corroborarse empíricamente, por ejemplo, con la construcción de los clasificadores.

Otro resultado relevante de este estudio es que parece que GPT-3.5 y GPT-4 podrían tener sus propios estilos de escritura independientes. Esto nos lleva a pensar que podríamos hablar de que los modelos de lenguaje tienen un idiolecto propio, de manera semejante a lo que ocurre con autores humanos. Definir y verificar la existencia de idiolectos de los LLM requiere de la aplicación de modelos y métodos cualitativos y cuantitativos de naturaleza lingüística.

Finalmente, pensamos que las aportaciones de este trabajo constituyen un paso importante en el esfuerzo conjunto para entender mejor la generación automática de textos.

AGRADECIMIENTOS

Esta publicación es parte del proyecto de I+D+i Proyecto ROBOT-TALK PID2022-140897OB-I00 financiado por MCIN/AEI/10.13039/501100011033/ y FEDER/UE.

Agradecemos a Ricardo García Mata del Departamento de Apoyo Investigación (Servicios Informáticos UCM) su asesoramiento técnico para el análisis estadístico de los datos y al Dr. Miguel Jiménez-Bravo sus orientaciones en los experimentos preliminares. Finalmente, agradecemos las sugerencias y comentarios de los revisores anónimos.

REFERENCIAS

Alonso Simón, L., Gonzalo Gimeno, J. A., Fernández-Pampillón Cesteros, A. M.^a, Fernández Trinidad, M. y Escandell Vidal, M.^a V. (2023). Using Linguistic Knowledge for Automated Text Identification. En M. Montes y Gómez et al. (Eds.), *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2023)*. Jaén, España, 26 de septiembre. <https://ceur-ws.org/Vol-3496/autextification-paper17.pdf>

Berber Sardinha, T. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics*, 4(1). <https://doi.org/10.1016/j.acorp.2023.100083>

Cañete, J., Chaperon, G., Fuentes, R., Ho, J-H., Kang, H. y Pérez, J. (2020). Spanish pre-trained BERT model and evaluation data. *arXiv:2308.02976v1*. <https://doi.org/10.48550/arXiv.2308.02976>

Cardenuto, J. P., Yang, J., Padilha, R., Wan, R., Moreira, D., Li, H., Wang, S., Andaló, F., Marcel, S. y Rocha, A. (2023). The Age of Synthetic Realities: Challenges and Opportunities. *APSIPA Transactions on Signal and Information Processing*, 12(1), 1–62. <https://doi.org/10.1561/116.00000138>

Casal, J. E. y Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, 2(3). <https://doi.org/10.1016/j.rmal.2023.100068>

Corizzo, R. y Leal-Arenas, S. (2023). A Deep Fusion Model for Human vs. Machine-Generated Essay Classification. En D. Wang y T. Toyoizumi (Eds.), *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*. Gold Coast, Australia, 18-23 de junio. <https://doi.org/10.1109/IJCNN54540.2023.10191322>

Crothers, E. N., Japkowicz, N. y Viktor, H. L. (2023). Machine-Generated Text: A Comprehensive Survey of Threat Models and Detection Methods. *arXiv:2210.07321*, Oct. 2023. <https://doi.org/10.1109/ACCESS.2023.3294090>

Desaire, H., Chua, A. E., Isom, M., Jarosova, R. y Hua, D. (2023). Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools. *Cell Reports Physical Science*, 4(6). <https://doi.org/10.1016/j.xcrp.2023.101426>

Fernández Vítors, D. (2023). El español: una lengua viva. Informe 2023. En C. Pastor Villalba (dir.), Instituto Cervantes (coord.), *El español en el mundo. Anuario del Instituto Cervantes 2023* (pp. 19-142). Madrid: Instituto Cervantes.

Fröhling, L. y Zubiaga, A. (2021). Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover. *PeerJ Computer Science*, 7, 1–23. <https://doi.org/10.7717/PEERJ-CS.443>

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J. y Wu, Y. (2023). How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation y Detection. *arXiv:2301.07597v1*. <https://doi.org/10.48550/arXiv.2301.07597>

Hadi, M. U., Al-Tashi, O., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M., Akhtar, N., Wu, J. y Mirjalili, S. (2023). *Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects*. TechRxiv. <https://doi.org/10.36227/techrxiv.23589741.v4>

He, Z., Mao, R. y Liu, Y. (2024). Predictive model on detecting ChatGPT responses against human responses. *Applied and Computational Engineering*, 44(1), 18–25. <https://doi.org/10.54254/2755-2721/44/20230078>

Jawahar, G., Abdul-Mageed, M. y Lakshmanan, L. V. S. (2020). Automatic Detection of Machine Generated Text: A Critical Survey. En D. Scott, N. Bel, y C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 2296–2309). Barcelona: International Committee on Computational Linguistics. *arXiv:2011.01314*. <https://doi.org/10.48550/arXiv.2011.01314>

Jurafsky, D. y Martin, J. H. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed. draft)*. Stanford University. Recuperado de <https://web.stanford.edu/~jurafsky/slp3/>

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. y Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>

Liao, W., Liu, Z., Dai, H., Xu, S., Wu, Z., Zhang, Y., Huang, X., Zhu, D., Cai, H., Li, Q., Liu, T. y Li, X. (2023). Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study. *JMIR Medical Education*, 9(1). <https://doi.org/10.2196/48904>

Ma, Y., Liu, J., Yi, F., Cheng, Q., Huang, Y., Lu, W. y Liu, X. (2023). AI vs. Human-Differentiation Analysis of Scientific Content Generation. *arXiv:2301.10416v2*. <https://doi.org/10.48550/arXiv.2301.10416>

Maloyan, N., Nutfullin, B. y Ilyushin, E. (2022). DIALOG-22 RuATD Generated Text Detection. En *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue”* (2022) (pp. 396–401). Moscú: RSUH. *arXiv.2206.08029*. <https://doi.org/10.48550/arXiv.2206.08029>

Mitrović, S., Andreoletti, D. y Ayoub, O. (2023). ChatGPT or Human? Detect and Explain. Explaining Decisions of Machine Learning Model for Detecting Short ChatGPT-generated Text. *arXiv:2301.13852v1*. <https://doi.org/10.48550/arXiv.2301.13852>

Nguyen, T. T., Hatua, A. y Sung, A. H. (2023). How to Detect AI-Generated Texts? En *IEEE 14th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)* (pp. 0464–0471). IEEE; Nueva York. <https://doi.org/10.1109/UEMCON59035.2023.10316132>

Pavlyshenko, B. M. (2022). Methods of Informational Trends Analytics and Fake News Detection on Twitter. *arXiv:2204.04891v1*. <https://doi.org/10.48550/arXiv.2204.04891>

Pizarro, J. (2019). Using n-grams to detect bots on Twitter, notebook for PAN at CLEF 2019. En L. Cappellato, N. Ferro, D. E. Losada, and H. Müller (Eds.), *CLEF 2019 Labs and Workshops, Notebook Papers*. https://ceur-ws.org/Vol-2380/paper_183.pdf

Radford, A., Narasimhan, K., Salimans, T. y Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI Blog*. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

Sarvazyan, A. M., González, J. Á., Franco-Salvador, M., Rangel, F., Chulvi, B. y Rosso, P. (2023). Overview of AuTextification at IberLEF 2023: Detection and Attribution of Machine-Generated Text in Multiple Domains. *Procesamiento del Lenguaje Natural*, 71, 275–288. <https://doi.org/10.26342/2023-71-21>

Savoy, J. (2020). *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*. Neuchatel: Springer. <https://doi.org/10.1007/978-3-030-53360-1>

Shijaku, R. y Canhasi, E. (2023). *ChatGPT Generated Text Detection*. Artículo presentado en The Fifteenth International Conference on Future Computational Technologies and Application. Future Computing 2023. Nice, Saint-Laurent-du-Var, France. 26-30 de junio de 2023. <https://doi.org/10.13140/RG.2.2.21317.52960>

Uchendu, A., Le, T., Shu, K. y Lee, D. (2020). Authorship attribution for neural text generation. En B. Webber et al. (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8384–8395. Online. <https://doi.org/10.18653/v1/2020.emnlp-main.673>

Uchendu, A., Le, T. y Lee, D. (2023). Attribution and Obfuscation of Neural Text Authorship: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 25(1), 1–18. <https://doi.org/10.1145/3606274.3606276>

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. y Polosukhin, I. (2017). Attention Is All You Need. En U. von Luxburg et al. (Eds.), *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, (pp. 6000–6010). Long Beach, California, Estados Unidos. <https://doi.org/10.5555/3295222.3295349>

Yu, P., Chen, J., Feng, X. y Xia, Z. (2024). CHEAT: A Large-scale Dataset for Detecting ChatGPT-writtEn AbsTracts. *arXiv:2304.12008v2*. <https://doi.org/10.48550/arXiv.2304.12008>

Zaitsu, W. y Jin, M. (2023). Distinguishing ChatGPT(-3.5, -4)-generated and human-written papers through Japanese stylometric analysis. *PLoS ONE*, 18(8). <https://doi.org/10.1371/journal.pone.0288453>



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 55-75
Artículo recibido: 04/07/2024
Artículo aceptado: 21/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/636>
DOI: <https://doi.org/10.58859/rael.v23i1.636>

Anglicismos y creación léxica en el discurso de *gamers* españoles: el caso de *Valorant*

Anglicisms and lexical creation in the discourse of Spanish gamers: the case of *Valorant*

ELENA FERNÁNDEZ DE MOLINA ORTÉS
UNIVERSIDAD DE GRANADA

Los videojuegos online han creado, en los últimos años, un espacio lingüístico en el que jugadores profesionales y aficionados están generando un sociolecto nuevo con rasgos propios. En este nuevo contexto comunicativo, el léxico tiene una gran relevancia y, en muchos casos, se puede considerar como una marca de identidad. El objetivo de este trabajo es comprobar las lexías que utilizan estos usuarios y conocer cuál es la influencia del inglés en sus interacciones. Para ello, hemos elaborado un corpus oral a partir del análisis del discurso de cuatro jugadores y comentaristas del videojuego *Valorant*. Se han recopilado más de 900 variantes léxicas y se han clasificado los términos según el tipo de anglicismo. En nuestros resultados hemos comprobado que los anglicismos no adaptados (*kill*, *spot*) tienen una gran presencia en el corpus. También hemos registrado una alta frecuencia de derivados híbridos como *killlear*; (*e*)*spotear* que revelan, sin duda, una gran productividad de los anglicismos en este nuevo contexto comunicativo.

Palabras clave: *pragmática; videojuego; neologismo; anglicismo; léxico; Valorant*

In recent years, online video games have created a linguistic space in which professional and amateur players are generating a new sociolect with their own characteristics. In this new communicative context, the lexicon has a significant relevance, and, in many contexts, it can be considered as a mark of identity. The aim of this work is to check which lexical items these users use and to check the influence of English in their interactions. To this end, we have compiled an oral corpus based on the analysis of the discourse of four players of the video game *Valorant*. We collected more than 900 lexical variants and classified the terms according to the type of Anglicism. In our results we found that non-adapted anglicisms (*kill*, *spot*) have a high presence in the corpus. We have also recorded a high frequency of hybrid words such as *killlear*; (*e*)*spotear*, which undoubtedly reveal a high productivity of Anglicisms in this new communicative context.

Keywords: *pragmatic; videogame; neologism; anglicism; lexicon; Valorant*

1. INTRODUCCIÓN

El estudio de los anglicismos en español tiene una larga tradición. A partir del siglo XIX, y sobre todo desde el siglo XX, palabras y expresiones procedentes de la lengua inglesa comenzaron a introducirse en español desde el inglés británico (Torrebadella-Flix y Nomdedeu-Rull, 2013) aunque, sin duda, la importancia política y económica de EE. UU. y su indudable influencia en espacios culturales como la música, el cine, la moda o la publicidad favorecieron la expansión del inglés en nuestra lengua (Medina-López, 2004; Rodríguez-González, 2023). Posteriormente, con la llegada de internet hace tres décadas, las sociedades se interconectaron y los anglicismos se expandieron con más fuerza, también, en otros ámbitos. Generalmente, los hablantes toman estas nuevas voces procedentes del inglés desde un punto de vista estratégico, según la necesidad que tenga su lengua para hacer referencia a nuevas realidades y, también, estilístico, utilizando una lengua emisora que goza de un prestigio lingüístico mundial en ciertos contextos.

La comunicación *online* y las tendencias que desde las redes sociales y las plataformas de contenido llegan diariamente a los usuarios de internet influyen directamente en la adquisición y en el uso de nuevos términos procedentes del inglés en nuestra época. De hecho, son abundantes los trabajos que demuestran estas nuevas lexías en distintos ámbitos como la prensa y otros medios de comunicación (Balteiro y Campos, 2012; Estornell-Pons, 2012; Rodríguez-Medina, 2021; Tejedor-Martínez, 2021; Vázquez-Amador, 2018, Bernal y Freixa, 2023), la publicidad (van Hooft, 2006; García-Morales et al., 2016), las redes sociales (Rodríguez-Medina, 2022), el cine (Giménez-Folqués, 2022), la música (Gómez-Capuz, 2023) o en contextos de comunicación específicos como el deporte (Rodríguez-González, 2012; Rodríguez-Medina, 2014; Contreras, 2018; Castañón y Rodríguez-González, 2023), la economía (Medina-López, 2021; Hoyos, 2023), la medicina (Roig y Roig, 2023), la política (Crespo-Fernández, 2023) o la informática (Cabanillas et al., 2007; Pano-Alaman, 2007; Luján-García y Bolaños-Medina, 2014).

En este trabajo queremos ampliar el estudio de los anglicismos en español en otro sector en el que el inglés está, indudablemente, muy presente: los videojuegos. Estas plataformas de entretenimiento fueron popularizadas en los años 90 y, a lo largo de los años, esta forma de ocio se ha reinventado significativamente. Aunque en sus inicios la relación entre el videojuego y el jugador era directa e individual, desde finales del siglo XX y, sobre todo, en el siglo XXI, la diversión con los videojuegos es distinta: el acceso a internet ha supuesto un cambio radical en la forma de entretenimiento, porque desde ese momento los jugadores pueden crear partidas con desconocidos de todo el mundo. Esto ha facilitado, sin duda, la creación de una nueva manera de relacionarse jugando (Lafrance, 2003; Revuelta-Domínguez y Bernabé, 2012); asimismo, también se ha observado un nuevo método de intercambio lingüístico entre los jugadores.

En las siguientes líneas, a partir de los conceptos de neologismo y creación léxica, y también partiendo de los trabajos realizados hasta la actualidad sobre el uso de anglicismos en los videojuegos (Morales Ariza, 2015; Fernández de Molina, 2020), presentamos los resultados de un análisis sobre léxico y expresiones específicas de algunos jugadores online en el videojuego *Valorant*. En esta investigación se ampliará el corpus presentado por Fernández de Molina (2020), en donde se analizaron otros videojuegos como *League of Legends (LoL)* (a partir de ahora) y *Fortnite*. Como novedad, estos corpus –el de Fernández de Molina (2020) y el que presentamos en este trabajo– se han obtenido a partir de muestras orales de jugadores españoles que cuentan con bastantes seguidores en sus canales de *streaming*. Esto nos permite analizar el uso espontáneo de los anglicismos, a diferencia de lo que ocurre, por ejemplo, en corpus escritos, creados generalmente mediante la recopilación de las conversaciones en foros en las que los jugadores o los aficionados pueden consultar los términos que deben usar, o, por ejemplo, las expresiones más relacionadas con el juego.

Por lo tanto, con estos nuevos datos podemos comprobar qué préstamos utilizan los jugadores y cuáles son las estrategias más frecuentes para la elección de neologismos en *Valorant* y, de forma contrastiva, también en *LoL* y *Fortnite*. Teniendo en cuenta que, tal y como se ha demostrado en trabajos anteriores (Morales Ariza, 2015; Fernández de Molina, 2020), existe una variación creciente en este ámbito, consideramos que en este análisis la frecuencia de uso de anglicismos habrá aumentado y que, además, algunos términos se usarán con una mayor productividad y funcionalidad que en los estudios previos.

2. LOS ANGLICISMOS EN ESPAÑOL: NEOLOGIZACIÓN Y CREACIÓN LÉXICA

Generalmente, el concepto de anglicismo se refiere a una variante importada, procedente del inglés, que se utiliza en una lengua extranjera (en este caso, el español). Así se define, por ejemplo, en el *Diccionario de la Lengua Española (DLE)* en su edición de 2014: “1. Giro o modo de hablar propio de la lengua inglesa. 2. Vocablo o giro de la lengua inglesa empleado por otra. 3. Empleo de vocablos o giros ingleses en distintos idiomas”. Sin embargo, existen definiciones y propuestas más elaboradas y específicas sobre el término. Pratt (1980: 115) lo define como un “elemento lingüístico, o grupo de los mismos, que se emplea en castellano peninsular contemporáneo y que tiene como étimo inmediato un modelo inglés”; por su parte, López-Morales (1987: 303) utiliza también el origen del término y su forma de inclusión al español: “aquellas [palabras] que proceden de otras lenguas, pero que han entrado al español a través del inglés”. Utilizando como referencia estas definiciones, podríamos decir que un anglicismo es una variante procedente del inglés que se ha implantado en el español. No obstante, el origen de estas unidades puede ser directamente la lengua inglesa, pero también podemos encontrar variantes de otras lenguas que han sido importadas a través del inglés al español como, por ejemplo, “iceberg”, de origen holandés. Asimismo, en esta definición habría que incluir el componente cultural del anglicismo dado que, como todo préstamo, estas unidades proceden de relaciones culturales, económicas y políticas entre los hablantes (Sala, 1988; Stone, 1957). Se necesitaría, por lo tanto, una definición más amplia de anglicismo, como la propuesta por Gottlieb (2005: 163, a partir de Rodríguez-González, 2023: 45), que lo define como “cualquier rasgo individual o sistémico de una lengua adaptado o adoptado del inglés, o inspirado o potenciado por modelos ingleses, empleado en la comunicación interlingüística en una lengua distinta del inglés”¹.

Teniendo en cuenta que el anglicismo se introduce en una lengua como una palabra nueva independientemente de su forma más o menos patente o adaptada en el sistema del español estas nuevas voces se incluyen mediante un proceso neológico y de creación léxica (Díaz-Hormigo y Vega-Moreno, 2018; Fernández de Molina (2020). De hecho, cuando una palabra se exporta al español son necesarias varias etapas de conformación y de integración de esa voz. Siguiendo a autores como Gómez-Capuz (2005), se pueden establecer tres fases: la primera, la creación léxica o su transferencia, en donde la nueva voz llega a la lengua y se utiliza, sobre todo, en ámbitos específicos; la segunda, la asimilación de la palabra y su acomodación, en la que se comienza a generalizar su uso y se extiende por contextos lingüísticos distintos; por último, la tercera y última fase sería la de explotación, en donde la palabra se incluye en el vocabulario de la lengua receptora en cualquier ámbito y es reconocida como propia de la lengua. En la primera fase el anglicismo se encuentra en un proceso de “creación léxica”, dado que esta palabra se exporta a la lengua receptora y se comienza a usar en ámbitos específicos. Posteriormente, tras

¹ La traducción es propia. La referencia original es: “any individual or systemic language feature adapted or adopted from English, or inspired or boosted by English models, used in intralingual communication in a language other than English.”

la asimilación del término, se realiza un proceso de “creación neológica” en donde el anglicismo se incluye en la nueva lengua y los hablantes crean nuevas lexías utilizando, por ejemplo, procedimientos de formación de palabras (Vega-Moreno, 2022).

La transición de una fase a otra se debe, principalmente, a la difusión del anglicismo. En la actualidad, las plataformas de información y de ocio *online* son muy productivas para la propagación de estas nuevas voces y, sin duda, los videojuegos y su expansión mundial son un foco de creaciones léxicas y neológicas. En el apartado 4.5 se realiza una descripción de cómo los jugadores asimilan y usan estos términos y qué funcionalidad tienen los anglicismos en estos contextos comunicativos.

2.1. La creación léxica y neológica en el sociolecto *gamer*

Actualmente, los videojuegos online constituyen un espacio de entretenimiento integrado por jugadores de todas las edades; además, desde finales del siglo XX han generado nuevos perfiles sociales y nuevos tipos de entretenimiento. En este entorno virtual podemos diferenciar, desde un punto de vista general, dos tipos jugadores: el profesional y el aficionado.

En el primer grupo se incluyen los jugadores que se dedican profesionalmente a los videojuegos, bien porque individualmente han tenido un gran impacto en un videojuego concreto y pueden ganar dinero con su actividad, bien porque forman parte de un equipo de eSports (deportes electrónicos) y se dedican a jugar internacionalmente en un contexto profesional. Estos jugadores exponen públicamente sus partidas en canales como *Youtube* o *Twitch* y generalmente cuentan con un gran número de visualizaciones. Asimismo, en este entorno online los jugadores profesionales suelen realizar partidas de forma síncrona con jugadores nacionales e internacionales.

Por su parte, los seguidores generalmente son jugadores, pero también son receptores “pasivos” de los profesionales; reciben informaciones sobre el juego y adquieren nuevas habilidades para desarrollar posteriormente su actividad como jugadores. En este caso, la relación es, sobre todo, asíncrona (excepto en los directos que realizan los profesionales).

En este espacio se produce una interacción entre el jugador profesional y los aficionados. Desde un punto de vista lingüístico, la experiencia del profesional y su comunicación con otros jugadores profesionales internacionales propician la creación de un nuevo estilo lingüístico de los videojuegos. Concretamente, uno de los aspectos en donde más se observa este estilo es en el uso de un léxico específico: los jugadores profesionales crean un sociolecto, porque usan términos y expresiones asociados a ese contexto comunicativo y sus seguidores adoptan y aprenden los rasgos lingüísticos del mismo. De hecho, y como se ha observado en algunos estudios sobre neologismos creados en los videojuegos en español (Morales Ariza, 2015; Fernández de Molina, 2020), durante las partidas es común utilizar el español como lengua de comunicación, pero, en muchos casos, también se utilizan palabras de origen inglés relacionadas con características de personajes o acciones que se realizan en el juego. Estas estrategias lingüísticas, que cada vez son más usadas por los *gamers* y que, además, están asociadas a las generaciones más jóvenes (Balteiro, 2019), están generando un nuevo tipo de comunicación que está a medio camino entre la variedad coloquial y los tecnicismos.

En ese ámbito, los jugadores profesionales y los jugadores aficionados utilizan una variedad koinética compartida por la comunidad en donde, además de usar la lengua nativa —en nuestro caso, el español— se elige una lengua franca: el inglés². Un ejemplo representativo de esta variedad se refleja en la selección léxica de los jugadores. En las últimas investigaciones realizadas sobre este tema (Morales-Ariza, 2015; Fernández de Molina, 2020) se ha observado

² Los jugadores no son bilingües generalmente; sin embargo, en algunas investigaciones los propios usuarios confirman que las interacciones *online* han beneficiado su aprendizaje del inglés. Así lo ha comprobado Salas Quinteros (2020), quien observó que los usuarios que fueron encuestados creían que los videojuegos habían ayudado a generar nuevas habilidades lingüísticas, sobre todo en contextos informales y en la comunicación fluida.

que, efectivamente, durante la interacción entre los jugadores es habitual encontrar préstamos directos no adaptados al español como, por ejemplo, *flash*, *kill*, *gamer*. Asimismo, la influencia del inglés queda patente, también, en variantes adaptadas que se crean mediante un proceso de hibridación entre las dos lenguas (*flashear*, *killlear*), sobre todo a partir de un lexema inglés y un método de formación de palabras en español por derivación, sobre todo verbal, con el sufijo *-ar* o *-ear*: Así lo pudo comprobar, por ejemplo, Fernández de Molina (2020), que recopila términos creados mediante esa combinación: “lutear”, “rushear”, “farmear”, “killlear”, “trolelear”, “banear”. Asimismo, en el trabajo la autora observa una alta productividad de estos derivados híbridos, porque su uso no se realiza exclusivamente con infinitivos; también se pueden encontrar algunas formas no personales (“luteado”, “farmeando”) y formas verbales conjugadas (“pulee”, “tilgueen”).

En los ejemplos de derivación nominal, el trabajo de Fernández de Molina (2020) recopila nuevas palabras con el sufijo *-(e)o* (“luteo”, “farmeo”); Morales-Ariza (2015), por su parte, también incluye el uso del sufijo *-er* (“gamer”, “farmer”, “pusher”) y *-or* (“pusheador”, “pokeador”, “zoneador”) como formas agentivas. Por último, en el análisis de *LoL* y *Fortnite* también se recopilaron algunas formas de derivación por sufijación en adjetivos, como “esnipazo”, para definir que el jugador realiza “un disparo muy bueno”. Además, en ocasiones también se usa un intensificador, en expresiones como “está muy luteado”, “voy mega fedeadado” o “súper tilteado”. En estos casos, la hibridación es doble: el intensificador es español (*muy*, *súper*, *mega*), la base léxica es inglesa (*loot*, *feed*, *tilt*) y la derivación se hace mediante un sufijo de participio (*-ado*).

Los estudios sobre el léxico de los *gamers* españoles, por lo tanto, revelan resultados representativos: en su interacción online utilizan anglicismos con frecuencia, sobre todo préstamos sin modificar. A partir de estas palabras, de hecho, crean nuevas formas híbridas, en donde se observa la gran productividad de los anglicismos como parte de la forma de hablar de estas comunidades.

3. CORPUS Y METODOLOGÍA

Para realizar esta investigación se ha elaborado un corpus oral a partir de las intervenciones de cuatro jugadores del videojuego *Valorant*, un juego de disparos en primera persona (*hero shooter*) en formato de multijugador. Fue presentado oficialmente por la compañía Riot Games (también productora de *League of Legends*) en junio de 2020. *Valorant* se desarrolla en distintos mapas en los que se llevan a cabo las batallas entre los participantes. En el inicio de cada partida, los jugadores eligen un bando: atacantes o defensores. Los primeros tienen que detonar la *spike* (una bomba), y los segundos deben desactivarla, defendiendo el terreno de los atacantes. En cada ronda, el equipo vencedor gana 1 punto, y el primer equipo que gana 13 rondas gana la partida.

Además de la posibilidad de jugar en primera persona y en grupo (características que comparte con *Counter Strike*, por ejemplo), *Valorant* ofrece varios agentes o personajes que tienen distintas habilidades: duelistas (Jett, Raze, Neon, Phoenix, Reyna), iniciadores (Breach, Sova, Skye), controladores (Omen, Brimstone, Viper, Skye) y centinelas (Sage, Chamber, Cypher). Asimismo, para la lucha los personajes tienen un arsenal bastante completo: armas de mano, fusiles y subfusiles, escopetas, rifles, armas pesadas o navajas.

Desde el año de su lanzamiento, *Valorant* ha alcanzado un gran número de usuarios. Según los datos de *Active Player*, en septiembre de 2022 contaba con más de 20 millones de jugadores activos y casi 120 millones de seguidores en *Twitch*. Al igual que en otros juegos, en *Valorant* participan equipos de jugadores profesionales (como los e-Sports) que compiten en convocatorias internacionales.

3.1. Datos del corpus

El corpus oral que se ha utilizado en esta investigación ha sido elaborado a partir de la transcripción y el análisis de 3,40 horas de grabación (204,53 minutos). En total, han sido recopiladas 23653 palabras (*Tabla 1*):

*Tabla 1: Jugadores de Valorant. Descripción del corpus con duración y número de palabras analizadas*³

	Videos (N)	Minutos totales	Palabras totales
Black Valorant	7	53,3	6900
Hitbox	6	54,64	5232
Mixwell	3	50,24	8367
TheGrefg	2	50,24	3154
Totales	18	204,53	23653

Para elaborar el corpus se seleccionaron cuatro jugadores de *Valorant*: Black Valorant, Hitbox, Mixwell y TheGrefg. Son jugadores profesionales con una gran presencia en *Youtube* y *Twitch*, y suelen compartir sus partidas en estas plataformas. Para realizar este trabajo se seleccionaron intervenciones en las que estos jugadores comentan sus propias partidas y las de otros jugadores. En los vídeos se describen, por ejemplo, estrategias del juego y, también, se realizan recomendaciones para que los usuarios aprendan a jugar o para que se especialicen en habilidades de algunos personajes (ver enlaces en el apartado 7). Para crear el corpus se optó por seleccionar este tipo de vídeos porque son los contextos en donde se pueden recopilar más muestras orales: generalmente, cuando los jugadores están en una competición, suelen estar concentrados y no se producen conversaciones abiertas.

Las muestras orales fueron elegidas al azar y se recopilaron durante los meses de julio y agosto de 2022. Generalmente, la duración de los vídeos de todos los integrantes del corpus es similar, pues oscila entre 7 y 12 minutos; no obstante, también se eligió una intervención de Mixwell con una duración de 14 minutos. Aunque es un tiempo un poco más elevado, se analizaron todos los minutos de grabación para no eliminar alguna parte que, posteriormente, pudiera ser relevante.

3.2. Clasificación de los anglicismos para el análisis del corpus

En este trabajo utilizaremos la clasificación de anglicismos propuesta por Pulcini, Furiassi y Rodríguez-González (2012) para poder contrastar este nuevo corpus con trabajos anteriores de Fernández de Molina (2020).

Los autores realizan una distribución de los anglicismos en dos grandes grupos: préstamos directos o préstamos indirectos. En los primeros, la lexía se adapta a la lengua receptora, pero en su representación (escrita y oral) existen evidencias de su origen inglés (*spoiler*, *streamer*). En los préstamos indirectos, sin embargo, la palabra se adapta completamente a la lengua receptora, y su origen pasa inadvertido (*baloncesto*, *implementar*).

Por una parte, en los directos, los autores establecen tres tipos de préstamos: los préstamos (adaptados y no adaptados), los falsos préstamos y las formaciones híbridas. A continuación, se utilizarán ejemplos del corpus para describir cada tipo.

³ En el siguiente enlace se puede acceder a los vídeos utilizados para la creación del corpus: https://docs.google.com/document/d/16RkQy3ke0ZK7nL_t0RFtGsmjxGDap0q-HJ3O413ofww/edit?usp=sharing

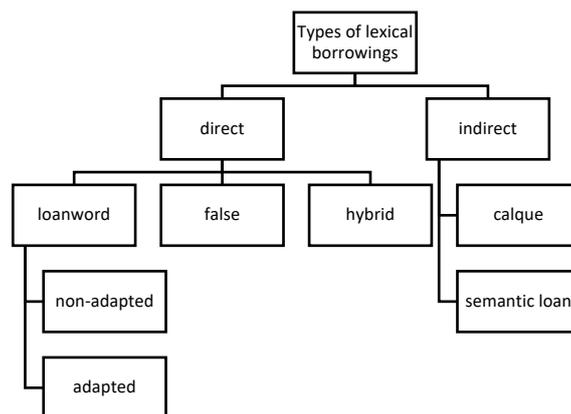


Figura 1: Clasificación de los tipos de préstamos a partir de la propuesta de Pulcini et al. (2012:6).

Comenzamos analizando los préstamos no adaptados, que conservan la forma de la lengua inglesa y no se asimilan a la lengua receptora, en nuestro caso, el español. Así se observa en los ejemplos (1) y (2), con las variantes *flash*, *arrow* o *fake* o en (3), donde TheGrefg no adapta el anglicismo al español, con una pronunciación como ['timĩŋg], sino que usa ['tamiŋ].

- 1) Soak se lleva dos *kills* si no me equivoco porque es que no te lo esperas en esa posición o sea la jugó muy bien el ruso. Vale, vamos a ver aquí en *bind*, lanza *flash* (BlackValorant, 26/07/2022)
- 2) sabes que te ha estado anclando esto, sabe que estás a la derecha, ha tirado la *arrow*, puede hacer un *fake*, puede abusar del tirón y te la juegas (Mixwell, 27/02/2022)
- 3) el portal que ha pillado el Omen es una pedazo de... no puede ser este *timing* (TheGreg, 9/05/2022)

Los préstamos adaptados, sin embargo, conservan algunas características de la lengua nativa, pero se modifican formalmente acomodándose a la lengua receptora.

- 4) Fijaos el movimiento tan inteligente que ha hecho de matar a uno, cambia de spot, y todas las micros han ido en torno a, al *córner*, porque es donde se ha hecho el primer *kill* (Hitboxing, 23/07/2021)
- 5) pero si no recuerdo mal, vais a *flipar* lo que hace Jett (Mixwell, 22/02/2022)
- 6) la verdad es que cómo ha evolucionado el juego desde el principio ¿eh? Está jugando con *kiles*... vaya flecha (Hitboxing, 23/07/2021)
- 7) Never *surrender*, *guys*, ni de coña me creo que hayamos ganado en ataque (TheGrefg, 9/05/2022)

En los ejemplos 4, 5 y 6 se advierten tres variantes adaptadas al español: *córner* (4), *flipar* (5) y *kiles* (6). Otras adaptaciones pueden ser fonéticas, como cuando TheGrefg, en el ejemplo (7), pronuncia [su'reŋder] y no [sə'rendəɪ] (7).

Para finalizar, un préstamo directo híbrido se define generalmente como una combinación de una parte nativa y elementos de la lengua receptora. En nuestro corpus se han observado, sobre todo, nuevas voces creadas a partir de la base inglesa y un método de formación de palabras del español.

- 8) cuando mis dos primeros compañeros mueren, alguien malo intentaría salir a través del humo a intentar *refraguear* a alguien o a intentar pillar a alguien desprevenido (Mixwell, 22/02/2022)

- 9) Dos contra cinco con Faith, vale no parece que haya nadie porque le han (*e*)*spoteado* (Hitboxking, 30/08/2022)
- 10) es muy inteligente porque siempre suele haber alguien ahí *lurkeando* (BlackValorant, 26/07/2022)

Así se observa, por ejemplo, en *refraguear* (8), que se compone de *frag*, el prefijo *re-* y la marca de infinitivo *-ear*; (*e*)*spoteado* (9), que se configura con *spott* y la forma de participio *-ado* con una prótesis vocálica *e-* para adaptarla fonética y gráficamente al español (*e-spott-(e)-ado*)⁴. En (10) encontramos *lurkeando*, un gerundio formado mediante la raíz inglesa *lurk* y la forma de gerundio en español *-ando*.

- 11) encima que es como muy como que le *toxiquean* mucho porque la gente se cree que le están troleando (BlackValorant, 21/04/2022)
- 12) tiene un método muy peculiar dispara con la phantom. Es maravilloso, tío, no *paniquea* (Hitboxking, 21/01/2022)

Por otra parte, en el corpus hemos encontrado también préstamos indirectos. Concretamente, se han observado algunos calcos como *toxiquear* (11) y *paniquear* (12).

- 13) imaginaos ellos intentan entrar. Todo se hace aquí para *plantar* la bomba. Tú abres el one way y empiezas a dispararles mientras están pasando por debajo (Mixwell, 19/06/2020)
- 14) esto es una brutalidad esto lo tienen que *parchear* de alguna manera hay un *lineup* con el cual puedes hacer esto (BlackValorant, 9/05/2022)
- 15) tiene que rotar porque si no se pueden colar por *cocina* por la espalda, porque nadie lo está mirando (Mixwell, 27/02/2022)
- 16) por lo tanto, no hay trampas, la presión de medio al tapar esta zona obliga a tener a alguien mirando *kitchen* (Mixwell, 27/02/2022).

Por último, los préstamos semánticos se producen cuando una palabra de la lengua receptora (el español) se usa con un significado distinto asignado a otra lengua (el inglés). Es lo que hemos observado, por ejemplo, en (13), cuando se dice “plantar una bomba”, procedente de “to plant a bomb”, o en (14), donde se usa *parchear* (14) “poner un parche” (“to patch up”), una variante técnica relacionada con el ámbito de la informática y que ha sido trasladada, también, al contexto de los videojuegos. Asimismo, en este grupo, se han incluido algunas alternancias entre voces del español y el inglés. Uno de los espacios en los que se desarrolla *Valorant* es “la cocina”, un término usual en español. Sin embargo, en el juego original el espacio es “kitchen”, dado su origen inglés. En las intervenciones de jugadores como Mixwell se alterna el uso de la palabra española (16) y la inglesa (17), realizando así una traducción del término inglés a su variante en español.

⁴ Utilizaremos la variante “(e)spotear”, dado que la doble grafía *tt* no es frecuente en español. Siguiendo las recomendaciones de Gutiérrez Ordóñez (2020), representaremos la (e) delante de /s/ cuando el anglicismo comienza con una /s/ líquida, para adaptar el anglicismo a la pronunciación española.

4. RESULTADOS Y ANÁLISIS

4.1. Descripción general del corpus

En el análisis del corpus se han recopilado un total de 982 anglicismos de las intervenciones de cuatro *gamers*; del total de ocurrencias se han analizado y localizado 199 lexías únicas. Por lo tanto, teniendo en cuenta que el total de palabras recopiladas en esta investigación fue de 23653 (ver *Tabla 1*), la frecuencia de anglicismos de este corpus es de un 4,15%. Concretamente, y si analizan las lexías de forma independiente según la clasificación propuesta en este trabajo, en nuestro corpus la mayoría de las variantes léxicas de procedencia inglesa son extranjerismos (861/94%). No obstante, han sido también recopilados préstamos (66/7,2%), calcos (24/2,6%) y préstamos semánticos (31/3,4%).



Gráfico 1: *Préstamos directos en el corpus Valorant (porcentajes a partir del total de anglicismos)*

En cuanto a los préstamos directos, que constituyen una gran parte de nuestro corpus, comprobamos que la mayoría son préstamos no adaptados (706/77,1%) (*lineup, stream, stack, push*) (ver *Gráfico 1*). Esta cifra es interesante, dado que refuerza las conclusiones que se han propuesto en investigaciones anteriores (Morales-Ariza, 2015; Fernández de Molina, 2020): generalmente, los jugadores españoles eligen anglicismos sin modificaciones. El origen de este videojuego y también la proyección internacional que tienen los jugadores son factores que pueden influir en la generalización de voces inglesas. Asimismo, y como se ha observado en otras investigaciones, las variantes híbridas también son comunes (146/15,9%) (*lurkear, bostear, troleear, whifear, whifeada o timear*). En estos ejemplos, los jugadores realizan composiciones constituidas por una base inglesa y un elemento de formación de palabras en español. Por último, aunque con menos frecuencia, en el corpus también han sido recopiladas variantes que, aunque tienen una procedencia inglesa, han sido adaptadas gráfica y/o fonéticamente al español (64/7%) (*radar, máster, rol*). A continuación, se analizarán los tipos de préstamos con una mayor profundidad.

4.2. Préstamos directos (adaptados, no adaptados y derivados híbridos)

Como se indicó en la sección anterior, los préstamos directos son las variantes más usadas por los jugadores de *Valorant*. De hecho, dentro de este gran grupo, los préstamos más frecuentes son no adaptados, es decir, conservan la estructura inglesa en la producción. También podemos encontrar préstamos híbridos y préstamos adaptados, aunque con menos uso.

4.2.1. *Préstamos no adaptados*

La mayoría de los anglicismos de este corpus son préstamos no adaptados (77,1%). Concretamente, en cuanto a las clases de palabras, las lexías pueden ser categorizadas como sustantivos (89,1%) y, esporádicamente, también, como expresiones (3,3%) y adjetivos (2,50%).

Tabla 2: Anglicismos no adaptados con, al menos, 2 ocurrencias

Palabra	Referente /significado	N	Palabra	Referente /significado	N
Jett	Personaje Valorant	44	kitchen	cocina	5
Viper	Personaje Valorant	37	aim	apuntar	5
Flash	deslumbrar	28	fail	error	5
Frenzi	Tipo de pistola	27	Cypher	Personaje Valorant	4
Chamber	Personaje Valorant	26	lobby	zona de espera	4
kill	matar	24	Bind	Mapa Valorant	4
Sage	Personaje Valorant	24	scream	gritar	3
ultimate	Habilidad del juego	20	Omen	Personaje	3
Sova	Personaje Valorant	16	tap	disparo	3
Top	superior	16	dron	dron	3
one way	sentido único	16	stack	Unir jugadores para atacar	3
lineup	alineamiento	15	arrow	flecha	3
clip	clip	14	prefire	Disparo a la pared	3
Orcus	Personaje Valorant	13	immortals	inmortal	3
crosshair placement	Lugar de punto de mira	12	Ascent	Mapa Valorant	3
stream	transmitir	11	Split	Mapa Valorant	2
Site	lugar	11	war	guerra	2
Flick	Jugador de Valorant	9	mode	modo	2
Raze	Personaje Valorant	8	under	debajo	2
Ice	Personaje Valorant	8	shortie	Arma	2
fake	falso	8	Major	Jugador de Valorant	2
retake	recuperar el control	8	container	contenedor	2
ranked	rango alcanzado	8	Ping	latencia de la red	2
smoke	humo	8	teleport	portal	2
tip	consejo	7	flex	flexibilidad en el juego	2
all-in	con todo	7	Striker	Jugador de Valorant	2
spike	Tipo de bomba	6	strafe shooting	disparo de lado	2
operator	Arma	6	game	juego	2
let's go	¡vamos!	5	jump peek	asomarse saltando	2
what?	¿qué?	5	free	libre	2
gamesense	forma de pensar el juego	5	shoulder peek	asomarse "a hurtadillas"	2

En los resultados de la *Tabla 2* podemos comprobar que los anglicismos no adaptados más frecuentes se corresponden con algunos elementos del juego. Por ejemplo, la palabra más usada, *Jett* (44), es un personaje, y también *Sage* (24), *Sova* (16), *Orcus* (13) o *Raze* (8) y *Cypher* (4). Asimismo, los préstamos se usan para denominar las armas, tal y como se observa en *Viper* (37), *Frenzi* (27), *Ultimate* (20) y “ulti”, *Spike* (6), *Arrow* (3) o *Smoke* (8). Los jugadores también se refieren con bastante frecuencia a los mapas y a algunas zonas del juego como *Ice* (8), *Bind* (4) o *Kitchen* (5). Todas estas léxias se usan con su forma y pronunciación inglesa, bien porque no pueden ser traducidas, o bien porque en todas las representaciones del juego (tutoriales, información en la página web, vídeos promocionales etc.) se emplean estas denominaciones.

Sin embargo, existen otras palabras relacionadas con el juego que son utilizadas generalmente como voces no adaptadas y que podrían ser sustituidas por un elemento en español, tal y como veremos en los ejemplos del corpus que se presentarán a continuación.

- 17) te das cuenta de que la *flash* no la lanzas donde crees que va a estar él, sino dónde vas a estar tú (Hitboxking, 26/01/2022)
- 18) pensad en una cosa también que es muy importante y es que los humos de *Viper* si te tira con un *flash de Breach*, por ejemplo, a través de la ventana (...) (Mixwell, 19/06/2020)

La palabra *flash* es una de las más frecuentes. Esta voz también fue recopilada por Fernández de Molina (2020) en el juego *League of Legends* con el mismo sentido: es una habilidad o un instrumento para realizar un destello ante los jugadores contrarios para intentar derrotarlos (así se usa en el ejemplo 17). En *Valorant*, además, se refiere a una habilidad concreta de un personaje, la “flash de Breach” (18). Se trata de un calco procedente del inglés, “Breach flash”, que alude a un tipo de destello muy potente que traspasa paredes y ciega a los adversarios; en este caso, es propio de este personaje, *Breach*.

Ya hemos mencionado, también, la relevancia de la palabra *kill* (ver ejemplos 1 y 4), otra de las léxias que tienen una gran presencia en nuestro corpus (ver tabla 2). Esta variante fue recopilada con anterioridad en el análisis de *Fortnite* y *League of Legends* por Fernández de Molina (2020) y, de hecho, este fue el anglicismo más usado en ambos juegos. La alta frecuencia de *kill* en los corpus es lógica, dado que el objetivo de los tres juegos es matar a adversarios (aunque en distintos escenarios y con distintos objetivos). Parece que, en este caso, como en todos los juegos los usuarios emplean el anglicismo como recurso para hacer referencia a esta acción, esta variante se ha generalizado de forma rápida, y es habitual usar la forma inglesa en estos contextos.

Por otra parte, también podemos organizar otras expresiones por campos léxicos:

- 19) qué locura lo que acaba de Jett, me encanta además el *crosshair placement* que tiene, es que la mecánica de los coreanos es muy buena (Hitboxking, 21/01/2022)
- 20) rompe la flecha mirad la tolerancia que tiene el juego al *strafe shooting* ha disparado la flecha prácticamente moviéndose (Hitboxking, 23/07/2021)
- 21) solo queda la enemiga. Hace el *fake plan* porque si no estaba muerta (BlackValorant, 4/06/2022)
- 22) *Viper* siempre está jugando al *retake*, porque he visto que ha puesto el muro antes del *retake* (Mixwell, 27/02/2022)
- 23) vale tiene un *Sova* en el *site* (...) está apuntando para el *jump peek* (Hitboxking, 26/01/2022)
- 24) esto se llama *shoulder peek*, hago *shoulder peek* en inglés que es como asomar el hombro ¿vale?, en español se hace para ver si el tío te está asomando o no, porque la gente tiene ventaja cuando espera (Mixwell, 15/04/2020)

Así, por ejemplo, en el corpus encontramos variantes que se refieren a la batalla, como *lineup* (14), que es usada con bastante frecuencia. También es interesante el uso de expresiones como *Crosshair placement* (19) o *Strafe shooting* (20) y *Trigger discipline*, que revelan la capacidad léxica de los jugadores: eligen términos complejos en inglés que, además, son específicos del ámbito de la batalla, lo que revela una adaptación lingüística relevante al videojuego. Asimismo, también utilizan variantes para referirse a estrategias como *flick*, *fake* (21), *retake* (22), *jump peek* (23) y *sholder pick* (24).

Asimismo, en la *Tabla 2* se muestran las lexías *top* y *clip*, que fueron incluidas en las ediciones de 2001 y 1993 del Diccionario de la lengua española respectivamente. Asimismo, se han recopilado expresiones como *What*, *What the fuck*, *Let's go*. Estos ejemplos se corresponden con variantes que ya están asimiladas en español, bien porque su inclusión como variantes propias de la lengua receptora se realizó hace tiempo (veinte años, en el caso de las palabras mencionadas anteriormente), o bien porque son expresiones que se usan inconscientemente en contextos comunicativos coloquiales y se generalizan entre los hablantes de la lengua. Finalmente, es necesario hacer referencia a algunos anglicismos no adaptados pero que realmente revelan una tendencia hacia el uso del inglés en estos contextos –ver “préstamos de lujo” en Rodríguez-González (2012, 2005)–. Por ejemplo, Mixwell pronuncia la palabra *inmortal* como [ĩm'mortal] y no [ĩmmor'tal], incluyendo el acento en la segunda sílaba manteniendo, por lo tanto, la variante inglesa.

Por último, se observan sustantivos que son sustituidos por anglicismos como *war*, *mode*, *game*, *free* y que, en realidad, tienen una palabra en español que se puede usar de la misma forma en este tipo de plataformas. La sustitución de estas lexías por préstamos no adaptados supone una transformación y una adaptación total desde el punto de vista léxico a la lengua franca del videojuego, el inglés.

4.2.2. *Préstamos adaptados*

Por otra parte, aunque los préstamos adaptados no tienen una alta frecuencia en el corpus, se han recopilado algunos cambios interesantes. Por una parte, ya veíamos en el apartado metodológico algunos ejemplos de adaptaciones fonéticas como en *surrender* (ver ejemplo 7). Sin embargo, se pueden observar, también, algunos préstamos adaptados ortográficamente y morfológicamente que, de hecho, ya están en el español general como *parchear*, *implementar*, *córner* y *radar*. Estas lexías, por ejemplo, están presentes en el *Diccionario de la Lengua española* desde 1992, en donde es reconocido su origen etimológico (del inglés), pero con adaptaciones a las normas ortográficas y fonéticas del español.

4.2.3. *Derivados híbridos*

En esta sección haremos referencia a las palabras que han sido adaptadas mediante la unión de un elemento del inglés y un método de formación de palabras del español.

Como se puede comprobar en la *Tabla 3*, los préstamos híbridos en *Valorant* se crean mediante procesos de formación verbal, adjetival o nominal en donde se integra una parte de la palabra inglesa, generalmente el lexema, y otra parte de la forma española. Comenzaremos analizando la formación de préstamos híbridos en formas personales y no personales del verbo:

- 25) no puedo encontrar cuatro, tíos... qué bien eso de toda esa ulti para *baitear* y delegar toda la utilidad (BlackValorant, 26/07/2022)
- 26) quién quiere *rolear*, oh my god bro (TheGrefg, 9/05/2022)
- 27) Es el pibe que está encargado de *bustearle* siempre dice que va a picar y *tapear* (BlackValorant, 24/07/2022)
- 28) qué pedazo de *flick*, abajo la pone en la mano porque sabe que la va a *poppear* (Hit-boxking, 21/01/2022)

29) no lo hagas que tiene que *pushearlo*

30) el tema es que, si yo me espero, el veneno se consume, a lo mejor tardan mucho en venir, me *pushean* por la izquierda, y me vende a alguien (Mixwell, 27/02/2022)

Tabla 3: *Palabras híbridas (inglés-español) utilizadas en el corpus*

Variante híbrida	Origen inglés	Traducción al español	Variante híbrida	Origen inglés	Traducción al español
baitear	to bait	Apuntar	holdear	to hold	Mantener
bostear	to boost	Hacer de cebo	lurkear	to lurk	Merodear
fedear	to feed	Alimentar	nerfear	to nerf	Bajar la calidad, reducir características de personajes, armas etc.
callear	to call	Llamar	pickear	to peek	Escoger (un agente al inicio del juego)
cheatear	to cheat	Engañar	pinguear	to ping	Sonar la campana, avisar con un sonido
crouch	to crouch	Agacharse	poppear	popping	Aparecer repentinamente
dashear	to dash	Lanzarse/moverse rápidamente	push	to push	Empujar
defuse	to defuse	Desactivar (la bomba)	ranked	rank	Clasificación
deletear	to delete	Eliminar	rollear	to rol	Elegir un rol
droppear	to drop	Caer, soltar	rush	to rush	Atacar
(e)spotear	to spott	Detectar, vigilar	(e)staquear	to stack	Defensa del personaje con otros jugadores
(e)strafear	to strafe	Desplazarse para evitar bombardeo	shooter	to shoot	Disparar
fail	to fail	Errar	spoilear	to spoil	Estropear, arruinar
flash	to flash	Falsear	tapear	to tap	Explotar
frag, refraguear	frag	Torturar (matar de forma repetida)	tepear	TP (<Teletransport)	Teletransportar
grindear	to grind	Machacar	trolear	to troll	Provocar

En nuestro corpus, este tipo de creaciones léxicas aparecen con una alta frecuencia en infinitivos (23,3%) y se componen de una base inglesa conservada tanto en su forma como en su pronunciación a la que se añade una terminación de infinitivo de la primera conjugación en *-ar/-ear*. Así se puede observar en *bait-ear* (25), *rol-ear* (26), *boost-ear* (27), *pop-ear* (28) o *push-ear* (29). En estos ejemplos, para la formación de palabras se incluye una *-e-* epentética que sirve para unir las dos partes de la palabra (ver también *Tabla 3*).

- 31) porque, aunque me mate Viper, Reyna *tradea* a Viper y los tres que quedan están en *As*⁵ (Mixwell, 27/02/2022)
- 32) dentro del túnel le tiras el muro, le dices que *defuse*, ¿vale? Tu compañero *defusa* y te quedas mirando (Mixwell, 15/04/2022)
- 33) muy bien, este ice es bonito porque además *timea* muy bien (Hitboxking, 30/08/2020)
- 34) ¿Os acordáis cuando *stremeaba*, que me iba a 10 fps? (TheGrefg, 9/05/2022)

Asimismo, es interesante observar que el anglicismo híbrido se utiliza, cada vez más, en el contexto comunicativo de los videojuegos. Por ejemplo, en el enunciado (31) observamos un anglicismo híbrido, *tradea*, que procede del término inglés *to trade*, “intercambiar”. Con este uso Mixwell describe que, en esa jugada, Viper quiere matar a su personaje “haciéndole daño”; pero su personaje “le devuelve a Viper ese daño” y lo intercambia. Como se puede comprobar en este enunciado, el jugador incluye cuatro neologismos de una forma natural: tres préstamos directos (*Viper*, *Reyna* y *As*), y también un verbo en forma personal (*tradea*) creado mediante una derivación híbrida para hacer referencia a una acción entre los personajes.

Estas tendencias pueden ser comprobadas también en las variantes *defuse* y *defusa* (32), que Mixwell usa para la tercera persona del singular. Otro ejemplo que, además, se constituye como préstamo híbrido tanto gramatical como fonético es la variante “*timea*” (33). Hitboxking dice [tai'mea] y no [ti'mea], es decir, utiliza la formación de palabras española conjugando la tercera persona del singular (del híbrido *timear*), pero conserva la pronunciación inglesa. Por último, en (34) comprobamos la creación de un nuevo verbo mediante una composición híbrida en pretérito, (*e*)*stremeaba*, una formación que no es tan habitual en los videojuegos, dado que en este caso no se trata de una descripción del videojuego en sí mismo, pero interesante desde el punto de vista estructural.

Por otra parte, aunque con una frecuencia menor, también se han recopilado participios (15,1%) y gerundios (9,6%) híbridos inglés-español en el corpus.

- 35) Vale, bien (*e*)*spoteado* con la E, además con la E y tira con el *ulti*... yo prácticamente (*e*)*spoteaba* todo el *site* (Hitboxking, 26/07/2022)
- 36) El chaval está marcado, ya que el chamber está *nerfeado*. (TheGrefg, 9/05/2022)
- 37) ahí tenía que haberlo *holdeado* muy bien porque realmente quería haber jugado un poquito más safe, un dos contra uno (Hitboxking, 10/07/2022)
- 38) Se piensan que es él, como la jugada de G2 solo *baiteando* con el agente correcto han hecho pensar te queda el otro *site* (BlackValorant, 26/07/2022)
- 39) de verdad, que *faileando* hasta en los puntos del *Valorant* (TheGrefg, 9/05/2022)
- 40) Lo han hecho en dos rondas seguidas. Les pido que jueguen más lento para que el muro se vaya a caer y que no estén *estaqueándola* (Mixwell, 27/02/2022)

Los participios se utilizan para describir habilidades y estrategias de los jugadores. Por ejemplo, (*e*)*spoteado* (35) se emplea para referirse a un apoyo de un jugador a su compañero. En (36) se recoge la expresión “estar nerfeado” para comentar que Chamber ha perdido una habilidad. Por otra parte, en (37) Hitboxking utiliza “haberlo holdeado” para decir que un personaje debería haber realizado un ataque sigiloso para hacer una lucha más segura.

Asimismo, en el corpus también hemos recogido algunos gerundios que expresan acciones de los jugadores, por ejemplo, *baiteando* (38), con el sentido apuntar; *faileando*, un gerundio que indica una estrategia en la batalla (ver *Tabla 3*) o (*e*)*staqueándola* (40), donde Mixwell explica que algunos jugadores apoyan a otro personaje para que no pierdan la partida.

⁵ “As” procede de la expresión “Aim down Sight”, que se refiere a la acción de apuntar la mirada hacia abajo. Los jugadores utilizan esta expresión cuando un equipo mata a todos sus enemigos.

4.2.4. Análisis del uso y la productividad de nuevas palabras en el sociolecto *gamer*

En las secciones anteriores se ha podido corroborar que, en contextos comunicativos como los videojuegos online, el uso de neologismos es bastante frecuente, funcional y diverso. De hecho, el 22,9% de los datos de nuestro corpus procede de la combinación o adaptación de variantes inglesas al español. Estas lexías pueden ser consideradas neologismos porque cumplen con los criterios de neologicidad (Lavale-Ortiz, 2019, 2020; Díaz-Hormigo y Vega-Moreno, 2018; Vega-Moreno, 2022; ver también Guilbert, 1975; Rey, 1976; Cabré, 1993): son palabras de reciente aparición, no han sido integradas aún en obras lexicográficas y, en la mayoría de los casos, desde un punto de vista psicológico, los hablantes las toman como novedosas. No obstante, es importante señalar que la novedad de un neologismo puede ser arbitraria en algunos casos. Algunas palabras muy usadas en los videojuegos (podemos pensar, por ejemplo, en las más frecuentes en nuestro corpus, como *kill*, *lineup*, *aim* o *nerfear* y *fraguear*) pueden ser novedosas para hablantes de español que no estén relacionados con ese ámbito; sin embargo, los jugadores o los aficionados de videojuegos online las tienen totalmente integradas en su vocabulario (Lavale-Ortiz, 2020; Cabré, 2022). Esto ocurre porque, además de la novedad, para que un neologismo se pueda usar debe existir un conocimiento compartido entre grupos que, partiendo de actividades y características comunes (ámbito profesional, generación, grupo social, etc.), se relacionan con otros usando conceptos y términos nuevos.

De hecho, en los trabajos que se han realizado sobre la escritura de jugadores en foros de internet se ha afirmado que, generalmente, estos usuarios utilizan una jerga específica con expresiones compartidas que les permiten comunicarse en estos contextos. Como explica Balteiro (2019) a partir de Ensslin (2012), los jugadores crean un léxico específico llamado “*techspeak*”. Estos grupos “necesitan aprender un código o un sociolecto especializado dentro del grupo que se usa en el lenguaje cara a cara con sus compañeros de juego”.

4.3. Palabras funcionales en corpus orales de jugadores

Como se indicó en el apartado 2 de este trabajo, y siguiendo a Vega-Moreno (2022) a partir de Gómez-Capuz (2004) y Castillo-Carbacho (2006), el proceso de adaptación de los neologismos a una lengua receptora se produce en tres etapas: la transferencia del préstamo, la asimilación, y la explotación del uso. Las lexías que hemos descrito hasta ahora en el trabajo son usadas generalmente en un ámbito específico, los videojuegos. Sin embargo, la productividad léxica de algunos neologismos indica que muchas palabras están ya en fase de explotación. ¿Cómo podemos comprobar qué léxico es el más utilizado y cuál tiene una mayor funcionalidad en los videojuegos? Una opción sería analizar el corpus que ha sido utilizado en este estudio (ver también *Tabla 3*) y comparar las palabras con mayor variabilidad léxica con las de los corpus orales de Fernández de Molina (2020), dado que ambos trabajos comparten el mismo método y el mismo tipo de análisis, aunque con distintos videojuegos.

Para realizar esta comparación se analizarán 14 variables léxicas que tienen una gran funcionalidad en nuestro corpus actual dado que, a partir de un derivado híbrido, se han usado en expresiones con, al menos, 2 categorías gramaticales diferentes. Además, tenemos en cuenta que, de estas 14 variables, 6 de ellas han aparecido en los corpus de *LoL* y *Fortnite* del año 2020⁶, lo que revela, sin duda, un uso cada vez más generalizado de estas lexías en los videojuegos.

⁶ En la *Tabla 4* hemos marcado las variables que los jugadores usan en los tres corpus con un asterisco (*).

Tabla 4: Variantes léxicas de palabras propias de videojuegos

Lema	Variantes
<i>aim</i>	“tiene crisp aim”, “tiene mucho aim”, “los mejores aimer”, “estamos súper bien de aim”, “no es aimer”
<i>baitear*</i>	baitear (inf.), baiteado (part.), baiteo (sust.)
<i>dashear</i>	dashear (inf.), dash (sust.), “se ha tirado un dash” (sust.exp.)
<i>drop*</i>	dropear (inf.), drop (sust.)
<i>(e)spotear</i>	(e)spotear (inf.), (e)spotear (verb), (e)spotear (verb.) “lo han (e)spotear” (part.), (e)spot (sust.), “cambiarse de (e)spot” (sust. exp.)
<i>fake</i>	fakear (inf.), “es el fake” (sust. exp.), “te puede hacer un fake” (sust. exp.), “es un muro fake” (adj. exp.), fakeplan (sust. adj.), faketp (adj. exp.)
<i>flash*</i>	flashear (inf.), flasheado (part.), flasheas (verb), te flashea (verb), “tirar una flash” (sust. sing.), “las flashes (sust. pl.)”, “tener flash” (sust. exp.)
<i>flex</i>	flexear (inf.), “un poquito de flex” (adj.), “es flex” (adj.)
<i>head</i>	headhunter (sust.), headshot (sust. action), head of team
<i>hold</i>	holdear (inf.), holdeado (part.)
<i>lurkear*</i>	lurkear (inf.), lurkeando (ger.), lurkeo (sust.),
<i>nerfear</i>	nerfear (inf.), nerfeado (part.)
<i>pushear*</i>	pushear (inf.), pushearla (inf. + pron.), pusheando (ger.), pushea (verb), este/ese push (sust.)
<i>trolear*</i>	troleo (inf.), troleando (ger.), “así se trolea” (verb), “enorme troleada (adj.)”, “es una troleada de locos (sust.)”, troleo (sust.)

Según los datos de la *Tabla 4*, las palabras que tienen más variabilidad son *spot*, *fake*, *flash* y *trol*. Con estas bases se generan verbos conjugados (“te trolea”, “te flashea”, “(e)spotear”); los jugadores también usan formas no verbales, como participios (*flasheado*, *(e)spotear*) y gerundios (*troleando*). Asimismo, en el corpus hemos encontrado sustantivos y adjetivos que se incluyen en expresiones propias del español: “te puede hacer un *fake*”, “tener *flash*”, “es una *troleada* de locos”, “es un muro *fake*” o “es *flex*”.

En otras variables observamos palabras con gran funcionalidad expresiva; por ejemplo, *aim* se usa en expresiones como “tiene *crisp aim*”, “tiene mucho *aim*”. También se han recopilado ejemplos como “los mejores *aimer*”, “no es *aimer*”, en donde se habla de una característica de los personajes usando la variante inglesa y el sufijo *-er*, muy productivo en este tipo de creaciones léxicas (Morales-Ariza, 2015). Con esta base léxica se crea, también, una estructura calificativa (“estamos súper bien de *aim*”).

Por otra parte, se ha observado que los jugadores realizan algunas composiciones interesantes con la palabra “head”. Están relacionadas con el principal objetivo de *Valorant* (derrotar al contrincante) y generalmente se crean por el método de composición: el primer elemento es “head” y se le añade una característica del personaje (*headhunter*, *head of team*) o una acción (*headshot*), por ejemplo.

A partir de las ocurrencias que se han mostrado en la tabla 5, y mediante la comparación del corpus de *Valorant* con los trabajos anteriores de *LoL* y *Fortnite* (Fernández de Molina, 2020), podemos establecer en qué fase se encuentra cada neologismo. Teniendo en cuenta nuestros datos, las palabras *holdear* o *nerfear* aún se sitúan en un proceso de asimilación, porque los jugadores no han realizado todavía derivaciones complejas. Sin embargo, *baitear* o *lurkear* tienen más funcionalidad: estas palabras aparecen con tres tipos de categorías y, además, los jugadores utilizan estas variantes en los tres juegos (*LoL*, *Fortnite*, *Valorant*). Por último, se

podría afirmar que las palabras recopiladas que están en proceso de explotación de uso son los préstamos directos no adaptados *aim*, *fake*, *flash* y *head*, y también los híbridos (*e*)*spotear*, *pushear*, *troleear*. Los jugadores usan estas lexías con una gran funcionalidad, porque no solo las eligen como variantes aisladas, sino que las integran en sus enunciados recategorizándolas y creando expresiones nuevas. Además, son variantes que tienen una gran presencia en los tres videojuegos, lo que corrobora nuestras explicaciones.

5. CONCLUSIONES

En esta investigación hemos recopilado y analizado los neologismos que los jugadores españoles de videojuegos online usan en *Valorant*. Asimismo, hemos observado que, frecuentemente, los jugadores comparten un léxico en este ámbito comunicativo. Mediante la comparación del uso de los anglicismos de *Valorant* con otros estudios sobre *LoL* y *Fortnite* (Morales-Ariza, 2015; Fernández de Molina, 2020) hemos encontrado similitudes en la elección de anglicismos y, también, en estrategias de formación de préstamos híbridos.

El análisis del corpus nos ha permitido observar algunas tendencias lingüísticas en los jugadores online. Por una parte, comprobamos que, efectivamente, los jugadores profesionales seleccionados emplean términos en inglés y préstamos por derivación híbrida de forma frecuente. Los anglicismos se utilizan como variantes léxicas propias del sociolecto *gamer*, que se usan generalmente para representar conceptos que, en su mayoría, tienen equivalentes en español. Sin embargo, el uso del inglés en los videojuegos revela una pertenencia del hablante a ese ámbito, y refuerza su identidad lingüística en ese espacio comunicativo y social. De la misma forma, y siguiendo estas tendencias, prevemos que los jugadores aficionados, siguiendo a sus ídolos, utilizarán un léxico similar y, seguramente, con cada vez más frecuencia, también utilizarán variantes creadas por composición. De esta forma se producirá una identificación del jugador novel con ese espacio lingüístico.

No obstante, en nuestro trabajo no hemos visto un aumento significativo de la frecuencia de anglicismos con respecto a los corpus anteriores. En el análisis de la muestra de *Valorant* se han observado nuevas expresiones y hemos corroborado que la mayoría de los préstamos son directos no adaptados. Sin embargo, el uso de derivados híbridos es menos frecuente. Por lo tanto, aunque en esta investigación se preveía un aumento de este tipo de variantes, parece que en los dos últimos años el uso de anglicismos en los videojuegos es similar al observado en trabajos anteriores.

Por último, hemos podido comprobar que, efectivamente, algunos anglicismos son muy funcionales en los videojuegos. Algunas lexías aparecen en los tres videojuegos, y eso refleja una extensión de los términos. Es posible advertir que el uso que hacen los jugadores de lexías y expresiones en este contexto comunicativo se realiza inconscientemente, dado que ya han sido asimiladas previamente en su discurso. No obstante, no todas las palabras recopiladas en nuestro corpus tienen una gran funcionalidad; la mayoría están en una fase de asimilación y explotación en este contexto. Aun así, no se puede obviar la generalización de nuevas variantes en estos contextos comunicativos, ni la creación de un sociolecto específico en este ámbito.

Aunque desde nuestro punto de vista este análisis contribuye al conocimiento de la nueva realidad léxica en un contexto tan nuevo como los videojuegos online, aún se necesitan más trabajos de investigación en este ámbito. De hecho, un corpus más amplio sobre *Valorant*, *LoL* o *Fortnite* en el que se analicen sesiones de partidas en tiempo real o las intervenciones de jugadores y comentaristas en torneos nacionales e internacionales permitirían examinar con mucha más precisión la generalización de lexías y expresiones de nueva creación. Asimismo, debemos tener en cuenta que este es un ámbito nuevo que está en constante actualización, por lo que las

investigaciones y las recopilaciones de datos en este ámbito deben ser constantes y progresivas. Con ello, sin duda, se podrán realizar estudios contrastivos a lo largo del tiempo que permitan conocer qué cambios se están produciendo en el léxico de los videojuegos.

REFERENCIAS

Balteiro, I. (2019). Lexical and Morphological Devices in Gamer Language in *Fora*. En Astrid Ensslin y Isabel Baiteiro (Eds.), *Approaches to videogame discourse: lexis, interaction, textuality* (pp. 39–57). Alicante: Universidad de Alicante. <https://doi.org/10.5040/9781501338489.0008>

Balteiro, I. y Campos, M. Á. (2012). False anglicisms in the Spanish language of fashion and beauty. *Ibérica*, 24, 233–260.

Bernal, E. y Freixa, J. (2023). Anglicismos en los medios de comunicación. En F. Rodríguez-González (Ed.), *Anglicismos en el español contemporáneo* (pp. 45–64). Peter Lang Verlag. <https://doi.org/10.3726/b20002>

Cabanillas, I. de la C., Tejedor-Martínez, C., Díez Prados, M. y Cerdá Redondo, E. (2007). English loanwords in Spanish computer language. *English for Specific Purposes*, 26(1), 52–78. <https://doi.org/10.1016/j.esp.2005.06.002>.

Cabré, M.T. (1993). La terminología. Teoría, metodología, aplicaciones. Barcelona: Editorial Antártida/Empúries.

Cabré, M. T. (2022). Los neologismos no nos sirven únicamente para expresarnos y para denominar, nos sirven también para distinguirnos de los demás. *CLINA Revista Interdisciplinaria de Traducción Interpretación y Comunicación Intercultural*, 7(1), 95–102. <https://doi.org/10.14201/clina20217195102>.

Castañón-Rodríguez, J. y Rodríguez-González, F. (2023). Anglicismos en el deporte. En F. Rodríguez-González (Ed.), *Anglicismos en el español contemporáneo* (pp. 275–293). Peter Lang. <https://doi.org/10.3726/b20002>

Castillo-Carbacho, M. A. (2006). El préstamo lingüístico en la actualidad. Los anglicismos. *Liceus: E-Excellence*.

Contreras, L. (2018). Los anglicismos en el lenguaje deportivo chileno. *Boletín De Filología*, 7, 177–341.

Crespo-Fernández, E. (2023). Anglicismos en la política. En F. Rodríguez-González (Ed.), *Anglicismos en el español contemporáneo* (pp. 187–207). Peter Lang Verlag. <https://doi.org/10.3726/b20002>

Díaz-Hormigo, M. T. y Vega-Moreno, É. (2018). Algunas de las aplicabilidades actuales de las investigaciones en neología y sobre los neologismos. *Pragmalinguística*, 26, 54–68. <https://doi.org/10.25267/Pragmalinguistica.2018.i26.03>

Ensslin, A. (2012). *The Language of Gaming*. Nueva York: Palgrave Macmillan. <https://doi.org/10.1007/978-0-230-35708-2>

Estornell-Pons, M. (2012). Préstamos del inglés en revistas femeninas: entre la necesidad denominativa y la estrategia pragmática. *Pragmalingüística*, 20, 61–91. <https://doi.org/10.25267/Pragmalinguistica.2012.i20.03>

Fernández de Molina-Ortés, E. (2020). Usos lingüísticos en los videojuegos. Creaciones léxicas y variantes en los jugadores online. En E. Waluch (ed.), *Las lenguas ibéricas en la traducción y la interpretación* (pp. 75-96). Biblioteka Iberykska.

García-Morales, M. G., Luján-García, C., González-Cruz, I. y Rodríguez-Medina, M. J. (2016). *La presencia del inglés en la publicidad televisiva española (2013-2015)*. Madrid: Síntesis.

Giménez Folqués, D. 2022. Anglicismos en el cine. En F. Rodríguez-González (ed.). *Anglicismos en el español contemporáneo. Una visión panorámica* (255-274). Peter Lang Verlag.

Gómez-Capuz, J. (2004). *Préstamos del español: lengua y sociedad*. Madrid: Arco/Libros.

Gómez-Capuz, J. (2005). *La inmigración léxica*. Madrid: Arco/Libros.

Gómez-Capuz, J. (2023). Anglicismos en el lenguaje de la música rock. En F. Rodríguez González. *Anglicismos en el español contemporáneo: una visión panorámica. Anglicismos en el español contemporáneo: Una visión panorámica*. Peter Lang.

Gottlieb, H. (2005). “Chapter 11. Anglicisms and Translation”. En Anderman, G. y Bristol, M. *Multilingual Matters*, 2005, pp. 161-184. <https://doi.org/10.21832/9781853597893-014>

Guilbert, L. (1975). *La créativité lexicale*. Paris: Librairie Larousse,

Gutiérrez-Ordóñez, S. (2020). Híbridos ortográficos. En Real Academia Española (Ed.), *Crónica de la lengua española* (pp. 303–345). Barcelona: Editorial Planeta.

Hoyos, J. C. (2023). Anglicismos en la economía. En F. Rodríguez-González (Ed.), *Anglicismos en el español contemporáneo* (pp. 167–186). Peter Lang Verlag. <https://doi.org/10.3726/b20002>

Lafrance, J. P. (2003). El juego interactivo: el primer medio de masas de la era electrónica. *Quaderns Del CAC*, 15, 59–68.

Lavale-Ortiz, R. M. (2019). Bases para la fundamentación teórica de la neología y el neologismo: la memoria, la atención y la categorización. *Círculo de Lingüística Aplicada a La Comunicación*, 80, 201–226. <https://doi.org/10.5209/clac.66608>

Lavale-Ortiz, R. M. (2020). El sentimiento de novedad en la identificación de neologismos: configuración de corpus y metodología desde una dimensión cognitiva. En *Cognitivismo y neología* (pp. 35–56). Iberoamericana Vervuert. <https://doi.org/10.31819/9783968690247-003>

López-Morales, H. (1987). Anglicismos léxicos en el habla culta de San Juan de Puerto Rico. *LEA: Lingüística Española Actual*, 9(2), 285–304.

Luján-García, C. I. y Bolaños-Medina, A. (2014). Disponibilidad léxica y anglicismos informáticos en los centros de interés: internet, software y hardware. *Odisea*, 15, 101–126.

Medina-López, J. (2004). *El anglicismo en el español actual*. Madrid: Arco/Libros.

Medina-López, J. (2021). El inglés y el español ante la globalización: una muestra de anglicismos en el ámbito de la economía. En C. Luján-García (Ed.), *Anglicismos en los nuevos medios de comunicación Tendencias actuales* (pp. 13–44). Granada: Comares.

Morales-Ariza, L. (2015). La terminología “gamer” en el contexto del videojuego multijugador en línea. *Revista Electrónica Del Lenguaje*, 2.

Pano-Alaman, A. (2007). Anglicismos en el lenguaje de la informática en español. El misterioso mundo del tecnicismo a través de foros y glosarios en línea. *Quaderni Del CeSLiC*, 3–18.

Pratt, C. (1980). *El anglicismo en el español peninsular contemporáneo*. Madrid: Gredos.

Pulcini, V., Furiassi, C. y Rodríguez-González, F. (2012). The lexical influence of English on European languages. In *Anglicization of European Lexis* (pp. 1–24). John Benjamins. <https://doi.org/10.1075/z.174.03pul>

Reuelta-Domínguez, F., y Bernabé, A. (2012). El videojuego en red social: un nuevo modelo de comunicación, *Tejuelo*, 6, 157–176.

Rey, A. (1976). Le néologisme: un pseudoconcept? *Cahiers de Lexicologie*, 45, 3-19.

Rodríguez-González, F. (2005). Calcos y traducciones del español actual. In P. Fuertes Olivera (Ed.), *Lengua y sociedad: investigaciones recientes en lingüística aplicada* (pp. 177–191). Universidad de Valladolid.

Rodríguez-González, F. (2012). Anglicismos en el mundo del deporte variación lingüística y sociolingüística. *Boletín de La Real Academia Española*, 92(306), 317–341.

Rodríguez-González, F. (2023). Estudios sobre el anglicismo en el español actual. Perspectivas lingüísticas. Lausana: Peter Lang. <https://doi.org/10.3726/b20397>

Rodríguez-Medina, M. J. (2014). Anglicismos en el léxico de las actividades deportivas de los gimnasios españoles. *Lexis*, 38(2), 401-427. <https://doi.org/10.18800/lexis.201402.006>

Rodríguez-Medina, M. J. (2021). Este jugador está en el top ten de los cracks: anglicismos en la prensa deportiva en las redes. En C. Luján-García (Ed.), *Anglicismos en los nuevos medios de comunicación Tendencias actuales* (pp. 121–140). Granada: Comares.

Rodríguez-Media, M.J. (2022). Anglicismos en las redes sociales. En F. Rodríguez González (ed.). *Anglicismos en el español contemporáneo. Una visión panorámica* (p. 335-376). Peter Lang Verlag.

Roig-Marín, A. y Roig-Marín, N. (2023). Anglicismos en medicina. En F. Rodríguez-González (Ed.), *Anglicismos en el español contemporáneo* (pp. 139–165). Peter Lang Verlag. <https://doi.org/10.3726/b20002>

Sala, M. (1988). *El problema de las lenguas en contacto*. Universidad Nacional Autónoma de México.

Salas-Quinteros, G. (2020). Variedad de la lengua en los jugadores de videojuegos. *Psicogamer*, 1–10.

Stone, H. (1957). Los anglicismos en España y su papel en la lengua oral. *Revista de Filología Española*, 41(1/4), 141–160. <https://doi.org/10.3989/rfe.1957.v41.i1/4.1046>

Tejedor-Martínez, C. (2021). El “look” de las “celebrities” o la imagen de los famosos. La influencia del inglés analizada a través de las secciones sobre gente y estilo en la prensa. En C. Luján-García (Ed.), *Anglicismos en los nuevos medios de comunicación Tendencias actuales* (pp. 45–72). Granada: Comares.

Torreadella-Flix, X. y Nomdedeu-Rull, A. (2013). Foot-ball, futbol, balompié... Los inicios de la adaptación del vocabulario deportivo de origen anglosajón. *RICYDE. Revista Internacional de Ciencias del Deporte*, IX(31), 5-22. <https://doi.org/10.5232/ricyde2013.03101>

van Hooft, A. P. J. V. (2006). El valor añadido de los anglicismos en la publicidad española. El impacto y la valoración social de las voces inglesas en los anuncios de la revista Elle. En Calvi, M.V.; Chierichetti, L. (ed.), *Nuevas tendencias en el discurso de especialidad Linguistics Insights. Studies in Language and Communication*, pp. 219-243

Vázquez-Amador, M. (2018). Los anglicismos de la moda en la prensa rosa española. *Onomázein Revista de Lingüística Filología y Traducción*, 40, 49–55. <https://doi.org/10.7764/onomazein.40.03>

Vega-Moreno, É. (2022). *Caracterización lingüística de los procedimientos de creación léxica. Iberoamericana Vervuert*. <https://doi.org/10.31819/9783968693408>



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 76-100
Artículo recibido: 15/07/2024
Artículo aceptado: 22/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/640>
DOI: <https://doi.org/10.58859/rael.v23i1.640>

Una primera aproximación descriptiva a la morfosintaxis de los hablantes de catalán y español con síndrome de Williams

A preliminary descriptive approach to the morphosyntax of Catalan and Spanish speakers with Williams syndrome

ELGA CREMADES
LLUÍS BARCELÓ-COBLIJN
UNIVERSITAT DE LES ILLES BALEARS

Este trabajo pretende determinar si existen diferencias en la producción de errores morfosintácticos y en la complejidad de las estructuras oracionales entre los hablantes con síndrome de Williams (SW) y los hablantes con desarrollo típico (DT). El SW, un trastorno del neurodesarrollo producido por la deleción de material genético en el cromosoma 7, ha suscitado debate en relación con el lenguaje, pues algunos autores sugieren que es un aspecto preservado y otros defienden que sí está afectado. El trabajo constituye, pues, una aportación más en este debate. Para ello, se han analizado 30 muestras de habla espontánea mediante el análisis de errores. Los resultados muestran que, en general, los hablantes con SW producen más errores morfosintácticos que los hablantes con DT y que, además, estos errores suelen ser de tipo distinto. Todo ello sugiere que la capacidad sintáctica de las personas con síndrome de Williams difiere, al menos en parte, de la de personas con DT.

Palabras clave: *análisis de errores, catalán, español, síndrome de Williams, sintaxis*

This paper aims to determine if there are differences in the production of morphosyntactic errors and in the complexity of sentence structures between speakers with Williams syndrome (WS) and speakers with typical development (TD). WS, a neurodevelopmental disorder caused by the deletion of genetic material on chromosome 7, has sparked debate regarding language, as some authors suggest that it is a preserved aspect while others argue that it is indeed affected. This paper thus constitutes another contribution to this debate. To this end, 30 spontaneous speech samples were analyzed through error analysis. The results show that, in general, speakers with WS produce more morphosyntactic errors than speakers with TD, and moreover, these errors tend to be of a different type. All this suggests that the syntactic ability of individuals with Williams syndrome differs, at least in part, from that of individuals with TD.

Keywords: *error analysis, Catalan, Spanish, Williams syndrome, syntax*

Este artículo es posible gracias al proyecto PID2021-128404NA-I00 financiado por MCIN/AEI /10.13039/501100011033

Citar como: Cremades, E. y Barceló-Cobljin, L. (2024). Una primera aproximación descriptiva a la morfosintaxis de los hablantes de catalán y español con síndrome de Williams. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 76-100. <https://doi.org/10.58859/rael.v23i1.640>

1. INTRODUCCIÓN

Está ampliamente aceptado que ciertos trastornos pueden contribuir a confirmar o rechazar las bases de la teoría sintáctica (Penke, 2015), por lo que estudiar las particularidades sintácticas del discurso de las personas con síndrome del neurodesarrollo (por ejemplo, el síndrome de Down o el síndrome de Williams) parece fundamental para conseguir un mejor conocimiento de los procesos de adquisición de la sintaxis en cualquier contexto.

En concreto, este trabajo se centrará en las particularidades sintácticas del lenguaje de las personas afectadas por el síndrome de Williams, también conocido como síndrome de Williams-Beuren, que, como explica Brock, es “a fascinating condition that can provide important insights into the processes of language and cognitive development, particularly when seen in the broader context of other developmental disorders”¹ (2007: 16).

1.1. *Caracterización fenotípica y psicolingüística de los hablantes con síndrome de Williams*

El síndrome de Williams (en adelante, SW) es un trastorno genético descrito por primera vez en 1961 por el cardiólogo neozelandés John C.P. Williams y el pediatra alemán Alexander J. Beuren (Brock, 2007) que afecta principalmente al desarrollo físico y cognitivo de una persona. El SW es causado por una delección submicroscópica de material genético (se estima que entre 19 y 25 genes) en la región 7q11.23 del cromosoma 7 (Ewart, Morris, Atkinson, Jin, Sternes, Spallone, Stock, Leppert y Keating, 1993a; Ewart, Morris, Ensing, Loker, Moore, Leppert y Keating, 1993b; Ewart, Jin, Atkinson, Morris y Keating, 1994) y, aunque inicialmente se estimaba que aparecía en 1 de cada 20.000 nacimientos (Karmiloff-Smith, Grant, Berthoud, Davies, Howlin y Udwin, 1997; Karmiloff-Smith, Tyler, Voice, Sims, Udwin, Howlin y Davies, 1998; Schubert, 2009), actualmente se estima que afecta a 1 de cada 7.500 nacimientos (Brock, 2007).

Dicha delección incluye el gen ELN, que codifica la elastina y que suele vincularse a cuestiones como las anomalías cardíacas asociadas con el SW, aunque no es este gen el que tiene efectos sobre el desarrollo cognitivo (Brock, 2007). En general, las personas con SW presentan unas características físicas entre las que se encuentran la afectación de varios órganos (por ejemplo, el corazón o los riñones), la hipercalcemia (una cantidad excesiva de calcio en sangre), unos rasgos faciales específicos (nariz respingada, labios gruesos, dientes pequeños y espaciados y sonrisa amplia), hiperacusia (hipersensibilidad auditiva, aunque no aparece en todos los individuos) o un volumen cerebral generalmente reducido (Niego y Benítez-Burraco, 2019). En este sentido, varios estudios demuestran reducciones específicas en la regiones parietales (Eckert, Hu, Eliez, Bellugi, Galaburda, Korenberg, Mills y Reiss, 2005), y el cuerpo calloso (Schmitt, Watts, Eliez, Warsofsky, Bellugi y Reiss, 2001), una reducción de entre el 15 y el 21% de la materia blanca y una reducción de entre el 6 y el 8% de la materia gris, significativa en todos los lóbulos (Niego y Benítez-Burraco, 2019). Al mismo tiempo, hay un aumento de células gliales en el núcleo caudado (Hanson, Lew, Hrvoj-Mihic, Groeniger, Halgren, Bellugi y Semedeferi, 2018) y una regionalización diferente del cerebro (Niego y Benítez-Burraco, 2019). Asimismo, el SW puede estar asociado con otros problemas médicos, como cardiopatías congénitas, problemas de crecimiento, problemas de visión y alteraciones del sistema endocrino.

La percepción social y las habilidades cognitivas relacionadas con la comprensión de las acciones sociales (que incluyen actos de comunicación verbal y no verbal) también muestran un perfil atípico, que difiere tanto del perfil de los niños con desarrollo típico como del perfil de los niños con, por ejemplo, trastornos del espectro autista (Hepburn, Fidler, Hahn y Philofsky, 2011). En este sentido, las personas con SW suelen ser socialmente afectuosas, sobre todo las

¹ “Una condición fascinante que puede proporcionar conocimientos importantes sobre los procesos de desarrollo del lenguaje y la cognición, en particular cuando se observa en el contexto global de otros trastornos del desarrollo” (la traducción es nuestra).

más jóvenes –si bien pueden mostrar alteraciones en la adaptación a nuevas situaciones. Así, como explica Brock (2007), las personas con SW se suelen caracterizar como extremadamente amistosas o *hipersociables*, aunque sus resultados en actividades de razonamiento social no son mejores que los precedibles por su edad mental (Brock, 2007). En palabras de Niego y Benítez-Burraco, “although in WS prosocial aspects of social functioning are not usually impaired, difficulties with the social-cognitive aspects of social functioning are frequently observed, impacting negatively on communication and cognition”² (Niego y Benítez-Burraco, 2019: 16).

El desarrollo cognitivo en las personas con SW, pues, parece ser atípico, con afectación en las habilidades visuoespaciales (Mervis Robinson, Bertrand, Morris, Klein-Tasman y Armstrong, 2000) y en las habilidades numéricas (Brock, 2007). Además, de acuerdo con Howlin, Davies y Udwin (1998), el coeficiente intelectual (CI) verbal en adultos con SW se sitúa, de media, en 64,50, mientras que el CI de rendimiento, en 60,80.

A pesar de estas atipicidades cognitivas, hay autores que han proporcionado evidencias para mostrar que la capacidad lingüística de las personas afectadas por el SW parece estar intacta. De hecho, esto se ha utilizado, en el pasado, como prueba de una estructura de la mente modular (Bellugi, 1994; Pinker, 2011), aunque otros expertos en desarrollo atípico sugieren que la perspectiva de desarrollo ontogenético, especialmente en casos como el SW, no parece apoyar la teoría de la modularidad de la mente (Stojanovik, Perkins y Howard, 2004; D’Souza y Karmiloff-Smith, 2011).

Así, en primer lugar, hay estudios que sugieren que los hablantes con SW tienen una buena capacidad sintáctica. Concretamente, Bellugi, Lichtenberger, Jones, Lai y St. George afirman que “the general cognitive impairment seen in adolescents and adults with Williams syndrome stands in stark contrast to their relative strength in language, their facility and ease in using sentences with complex syntax”³ (Bellugi et al., 2000:12). Para estos autores, las atipicidades tienen que ver con el procesamiento léxico (por ejemplo, Bellugi Marks, Bihle y Sabo, 1988; Clahsen y Almazan, 1998). Otros estudios muestran algunas alteraciones en el desarrollo del lenguaje de los individuos con SW. Autores como Perovic y Wexler (2007), por ejemplo, postulan que el proceso de adquisición del lenguaje está demorado, pero no está fundamentalmente alterado, es decir, que las características *únicas* del lenguaje en los hablantes con SW son consecuencias indirectas de alteraciones cognitivas no lingüísticas, como la hipersociabilidad o las alteraciones visuoespaciales. Finalmente, para otros autores, los errores morfológicos y sintácticos no son infrecuentes (Karmiloff-Smith et al., 1998; Benítez-Burraco, Garayzábal y Cuetos, 2016). Es lo que afirman Niego y Benítez-Burraco cuando comentan que “syntax is not spared in WS, but delayed or perhaps impaired across diverse domains (subcategory constraints, passives, Wh-questions, agreement, etc.)”⁴ (Niego y Benítez-Burraco, 2019: 5). Desde esta perspectiva, las capacidades lingüísticas de los hablantes con SW no se deberían formular en términos de *preservación* o *módulos intactos*, sino como el producto de restricciones alteradas en el desarrollo lingüístico y cognitivo (Karmiloff-Smith et al., 1998). Es decir, se considera que los hablantes con SW adquieren y procesan el lenguaje de un modo distinto al de los hablantes típicos como resultado de las características del desarrollo cognitivo general de estos individuos y no porque la mente sea modular.

² “Aunque en SW los aspectos de funcionamiento prosocial no suelen estar alterados, se observan dificultades en aspectos sociocognitivos del funcionamiento social, lo que impacta negativamente en la comunicación y la cognición” (la traducción es nuestra).

³ “El deterioro cognitivo general observado en adolescentes y adultos con síndrome de Williams contrasta drásticamente con su relativa fortaleza en el lenguaje, su facilidad y comodidad en utilizar frases con sintaxis compleja” (la traducción es nuestra).

⁴ “La sintaxis no está preservada en SW, sino retrasada o a lo mejor deteriorada en varios dominios (restricciones de subcategorización, pasivas, preguntas-*cu*, concordancia, etc.)” (la traducción es nuestra).

No todos los niveles del lenguaje parecen afectados del mismo modo en los hablantes con SW. Por ejemplo, la revisión bibliográfica de Brock (2007) muestra diferentes patrones en fonología, léxico y semántica, pragmática y gramática (entendida como morfosintaxis). En cuanto a la fonética y la fonología, parece que los hablantes con SW tienen una producción del habla relativamente buena, así como una memoria fonológica a corto plazo mejor que de la de individuos con el síndrome de Down (pero no que la de los hablantes típicos). Ciertamente, algunos estudios apuntan al hecho de que los hablantes con SW manifiestan problemas para segmentar el habla o, por lo menos, ciertas cadenas de sonidos, pero no hay, por el momento, suficientes pruebas que permitan afirmar que las habilidades fonológicas de los hablantes con SW sean atípicas.

En lo que respecta al nivel lexicosemántico, el panorama es más confuso que en el caso de la fonética y la fonología. Por un lado, estudios como el de Clahsen y Almazan afirman que “the excessive use of the regular -s plural in WS results from an impairment of the lexical system and/or its access mechanisms”⁵ (Clahsen y Almazan, 1998: 729). En la misma línea, Ring y Clahsen (2005) señalan que la gramática no está afectada en hablantes con SW, mientras que el léxico sí lo está. En palabras de los autores:

[T]he two core modules of language are dissociated in WS such that the computational (rule-based) system for language is not affected, at least not beyond a general developmental delay, while lexical representations and/or their access procedures are impaired.⁶ (Ring y Clahsen, 2005:480)

Brock, sin embargo, afirma que “overall, the available evidence suggests that young children with Williams syndrome have vocabularies that are no better or worse than one would expect based on their nonverbal abilities”⁷ (Brock, 2007: 9) y que, en la adolescencia, el vocabulario receptivo es, sin duda, una fortaleza en los hablantes con SW. Varios autores (Clahsen y Almazan, 1998; Ring y Clahsen, 2005) ponen de manifiesto un uso poco habitual de algunos elementos léxicos en el habla cotidiana, lo que, según Brock (2007) se podría deber a cuestiones extralingüísticas como las herramientas para llamar la atención o los mecanismos de mediación de las interacciones sociales. Ahora bien, parece que, en sus discursos espontáneos, estos hablantes tienden a producir más colocaciones léxicas o unidades fraseológicas que frases complejas o enunciados largos, lo que indicaría más habilidad léxica que sintáctica (Sederias, Krakovitch, Stojanovik y Zimmerer, 2024).

En el nivel morfosintáctico (que se desarrolla con más detalle en el apartado 1.2), parece que los resultados son todavía más confusos que en el caso del nivel lexicosemántico, pues, por un lado, hay autores (Clahsen y Almazan, 1998; Stavrakaki, 2004; Ring y Clahsen, 2005) para los que hay muy pocas pruebas de que la sintaxis o la morfología de los hablantes con SW tenga un desarrollo atípico. Por ejemplo, como señala Brock (2007), los estudios con niños sugieren una relación normal entre la complejidad de las frases y el repertorio léxico. Asimismo, Llull, Barceló-Coblijn y Cremades (2024) demuestran que no parece que los hablantes adultos de español con SW cometan significativamente más errores en la asignación y la concordancia de género y número que los hablantes con desarrollo típico. Sin embargo, varios autores (entre los que destacan Karmiloff-Smith et al., 1997, o Benítez-Burraco et al., 2016) muestran, en

⁵ “El uso excesivo de la -s del plural regular en SW resulta de un deterioro del Sistema léxico o sus mecanismos de acceso” (la traducción es nuestra).

⁶ “Los dos módulos centrales del lenguaje están disociados en el SW, de manera que el sistema computacional (basado en reglas) del lenguaje no se ve afectado, al menos no más allá de un retraso general en el desarrollo, mientras que las representaciones léxicas y/o sus procedimientos de acceso están deteriorados” (la traducción es nuestra).

⁷ “En general, la evidencia disponible sugiere que los niños pequeños con síndrome de Williams tienen vocabularios que no son ni mejores ni peores de lo que cabría esperar según sus habilidades no verbales” (la traducción es nuestra).

contraste con otros estudios, que los hablantes con SW sí presentan errores relativamente atípicos en el orden de la oración, la concordancia de género, la flexión verbal o la comprensión de determinadas estructuras gramaticales complejas. A su vez, estudios recientes de Cremades y Barceló-Coblíjn (2023), y de Barceló-Coblíjn y Cremades (2023, 2024) sugieren que, en los análisis de redes sintácticas complejas, la sintaxis de los hablantes con SW es significativamente distinta de la de los hablantes con desarrollo típico.

En cuanto a la pragmática, Díez-Itza, Viejo y Fernández-Urquija (2022), por un lado, y Moraleda y López Resa (2024), por el otro, indican que los hablantes con SW experimentan dificultades significativas a lo largo de su desarrollo. Así, se ha encontrado que, aunque a menudo estos individuos se caracterizan como hipersociables, presentan un retraso en la aparición de habilidades de atención conjunta (como coordinar la atención entre una persona y un objeto o evento de interés mutuo), tienen dificultades para inferir la intención comunicativa detrás de los gestos de señalar y el contacto visual, cometen más transgresiones de las máximas de Grice que los hablantes típicos (como disfluencias repetitivas y más proporción de lenguaje no contingente en las conversaciones), y a veces no comprenden adecuadamente el lenguaje figurativo (Díez-Itza et al., 2022). Asimismo, Brock (2007) encuentra, en los hablantes con SW, dificultades para entender los requerimientos informativos de sus interlocutores, así como en la inferencia de estados mentales, tanto en las tareas narrativas como en otras situaciones sociales.

La cuestión de cómo se desarrolla el lenguaje en los hablantes con síndrome de Williams, pues, sigue abierta, y es precisamente esto lo que pretende abordar el trabajo que presentamos. Después de describir el SW y el marco teórico en el que se basa el análisis, con una descripción más detallada del desarrollo de la sintaxis en los hablantes típicos y atípicos, se mostrarán los objetivos del estudio y la metodología utilizada. A continuación, se presentarán los resultados, que demuestran que se puede afirmar que en la producción de habla espontánea los hablantes con SW producen más errores sintácticos que los hablantes típicos, y que su sintaxis es, cuando menos, particular y, por lo tanto, debe ser objeto de un análisis más profundo.

1.2. El desarrollo de la sintaxis en hablantes con SW

La capacidad de utilizar las reglas gramaticales para formar oraciones y estructurar el lenguaje de manera adecuada forma parte de la cognición humana. Los estudios sobre la adquisición del lenguaje han demostrado que el desarrollo de la sintaxis comienza desde los primeros años de vida, cuando los niños empiezan a producir sus primeras palabras y frases, y, a medida que crecen, su habilidad para producir oraciones se va volviendo más compleja y sofisticada. El desarrollo de esta habilidad pasa por diferentes etapas, que, si bien puede variar entre los niños (pues está influenciada por factores como el entorno lingüístico, las interacciones sociales y la exposición al lenguaje), en general se observa una progresión sistemática y predecible a medida que los niños adquieren habilidades lingüísticas más complejas. Tradicionalmente, se ha dividido el desarrollo de la sintaxis de diferentes maneras según la edad y la producción de palabras, y la consiguiente evolución de la capacidad sintáctica. McNeill (1970), por ejemplo, propuso el uso inicial de holofrases a los 12-18 meses, y después consideró que hay dos fases más: a los 18-24 meses aparecen combinaciones de 2 palabras (normalmente estructuras sin verbo) y a los 24-36 meses aparecen cadenas de más de 2 elementos léxicos. Más adelante, Brown (1973) dividió el proceso de adquisición del lenguaje en cinco fases, y la sintaxis se desarrollaría a partir de la fase III, donde aparece la adquisición de reglas y marcadores sintácticos. Otros trabajos posteriores se han centrado en el tipo de estructuras sintácticas y lo han contabilizado con métodos como el Índice de Sintaxis Productiva (Scarborough, 1990). Asimismo, mientras Tomasello (1992) siempre ha defendido un crecimiento lineal y progresivo, Ninio (2006) explica que hay datos que muestran que el desarrollo del uso de los verbos sólo puede describirse mediante una ley de potencias, es decir, un progreso no lineal.

Los datos de los estudios sobre hablantes con SW son desconcertantes, sin embargo, para más de un marco teórico, ya que no parecen encajar con muchos de los supuestos de las diferentes teorías sobre la adquisición del lenguaje. Para empezar, los niños con SW no suelen hablar hasta los 3-4 años, precisamente cuando los niños con un desarrollo típico han alcanzado la madurez en muchos aspectos del lenguaje. A partir de aquí, los múltiples estudios que han proliferado en cuanto a la relación de la sintaxis y SW conducen, a menudo, a conclusiones opuestas.

Por un lado, como reporta Brock (2007), varios estudios (Singer-Harris, Bellugi, Bates, Jones y Rossen, 1997; Vicari, Caselli, Gagliardi, Tonucci y Volterra, 2002; Zukowski, 2004, o Mervis y Becerra, 2007) indicarían un desarrollo típico de la sintaxis o, por lo menos, parecido en cuanto a la complejidad (que parece ser siempre superior a la de hablantes con otros trastornos como el síndrome de Down). Por ejemplo, Mervis, Morris, Bertrand y Robinson (1999) demuestran, utilizando el Índice de Sintaxis Productiva, que los hablantes con SW presentan una complejidad oracional típica de acuerdo con la longitud media de sus enunciados, y Vicari et al. (2002,) muestran que los niños con SW producen oraciones de una complejidad similar a la de los niños con desarrollo típico.

Esta perspectiva es la que apoyan, también, estudios como los de Clahsen y Almazan (1998), para quienes las dificultades de los hablantes con SW (como ya hemos mencionado) tienen que ver más con el acceso a ciertos tipos de información léxica que con el acceso a la gramática (que consideran intacta). Del mismo modo, Ring y Clahsen (2005) consideran que los hablantes con SW muestran una buena actuación en tareas relacionadas con el acceso al sistema computacional de la gramática (es decir, a las operaciones gobernadas por reglas), mientras que manifiestan problemas para las tareas que dependen del acceso a la información del lexicón.

Sin embargo, son numerosos los estudios que cuestionan estos puntos de vista y que proporcionan evidencia que demuestra un procesamiento lingüístico atípico, especialmente en morfosintaxis, por parte de la población con SW. Entre ellos, destacan los de Karmiloff-Smith et al. (1998), Karmiloff-Smith, Thomas, Annaz, Humphreys, Ewing, Brace, van Duuren, Pike, Grice y Campbell (2004), Mervis et al. (2000), Grant, Valian y Karmiloff-Smith (2002), Volterra, Caselli, Capirci, Tonucci y Vicari (2003), Stojanovic et al. (2004), Joffe y Varlokosta (2007), Thomas, Dockrell, Messer, Parmigiani, Ansari y Karmiloff-Smith (2007), Benítez-Burraco et al. (2016), Perovic y Wexler (2007) o Díez-Itza et al. (2019), entre muchos otros. Así, por ejemplo, Grant et al. (2002) demuestran que los individuos con síndrome de Williams desarrollan peor que los hablantes típicos de 6 años la tarea de repetir oraciones relativas complejas, y Volterra et al. (2003) explican que la sintaxis no parece tener un desarrollo típico en los hablantes con SW. Del mismo modo, Joffe y Varlokosta afirman que, en los individuos con SW, existe “evidence for significant language impairment, and more specifically, impaired syntactic processing”⁸(Joffe y Varlokosta, 2007: 718).

Hay que tener en cuenta, además, que, en la mayoría de los estudios sobre SW, la capacidad sintáctica se mide solo indirectamente, a través de informes de los tutores (que son los que confirman si el niño produce o dice determinadas expresiones), pruebas neuropsicológicas que miden el vocabulario receptivo (a partir de ahí se extrapola la capacidad sintáctica), la comprensión de determinadas frases (el niño las escucha y señala un dibujo concreto entre 4 posibles), etc. Como señalan Brock et al. (2007), el vocabulario de los hablantes con SW parece ser un punto fuerte de su perfil lingüístico, pero, en lo que respecta a la sintaxis, los datos no son concluyentes, en gran parte debido a la metodología utilizada para medir la capacidad sintáctica del hablante, que rara vez sigue las pretensiones de la lingüística teórica.

Algunas excepciones a este panorama son, por ejemplo, los trabajos de Joffe y Varlokosta (2007), Perovic y Wexler (2007), Perovic, Modyanova y Wexler (2013) o Benítez-Burraco et al. (2016). En primer lugar, Joffe y Varlokosta estudian las oraciones interrogativas de acuerdo con

⁸ “Evidencia de un deterioro significativo del lenguaje y, más específicamente, un deterioro en el procesamiento de la sintaxis” (la traducción es nuestra).

la función sintáctica desarrollada por el interrogativo, y llegan a la conclusión de que “the individuals with WS (...) performed similarly to individuals with DS”⁹ (Joffe y Varlokosta, 2007: 717). Por otro lado, Perovic y Wexler (2007) estudian los pronombres personales (reflexivos y no reflexivos) y la elevación de sujeto, y observan que los hablantes con SW adquieren los pronombres significativamente más tarde que los hablantes típicos y que, además, parece que la elevación puede incluso que no se llegue a dominar. En otro trabajo, Perovic et al. (2013) detectan, desde la perspectiva de la gramática generativa, diferencias entre hablantes típicos y hablantes con SW en la ligadura y en el comando-c. Asimismo, Benítez-Burraco et al. (2016) estudiaron un grupo de hablantes de español con SW y su capacidad para procesar referencias sintácticas cruzadas en oraciones con estructuras de control como *El policía amenazó al ladrón con detenerlo*. Los resultados indican que los participantes con SW tienen problemas para comprender las referencias del sujeto lógico PRO del verbo *detener* y la referencia del clítico *lo* en este tipo de oraciones: *El policía_i amenazó al ladrón_k con PRO_i detenerlo_k*. Es decir, los participantes con SW entienden quiénes son los actores que intervienen, pero tienen dificultades en entender quién es el sujeto y quién el objeto.

Otra forma de analizar la competencia sintáctica de un hablante es su capacidad narrativa. Reilly, Klima y Bellugi (1990) analizaron las capacidades narrativas de niños con SW y niños con síndrome de Down (en adelante, SD) y demostraron que los primeros producen textos narrativos orales más ricos en vocabulario y en número de conectores textuales; sin embargo, la prosodia afectiva no se corresponde con la del grupo control neurotípico. La conclusión es que los hablantes con SW pueden llevar a cabo una narración bien estructurada (presentación, nudo y desenlace) y bien organizada desde el punto de vista gramatical. En otro trabajo más reciente, Reilly, Losh, Bellugi y Wilfleck (2004) compararon los resultados en tareas narrativas entre cuatro grupos: SW, Trastorno Específico del Lenguaje (TEL), niños con lesión cerebral focal y niños con desarrollo típico. Los participantes completaron una tarea de narración sin imágenes y, curiosamente, los resultados alinean a los participantes con SW con aquellos con TEL debido a los tipos de oraciones y los errores morfológicos que produjeron. Lorusso, Burigo, Tavano, Milani, Martelli, Borgatti y Molteni (2017) también recogieron muestras de narraciones en italiano, lo que les permitió mostrar que los hablantes con SD utilizan más palabras abstractas que los hablantes con SD. Díez-Itza et al. (2019) demuestran que los hablantes de español con SW recuerdan los acontecimientos que tuvieron lugar en las películas y los pueden explicar de una forma coherente.

En resumen, los estudios iniciales sobre hablantes con SW concluían que su gramática permanecía intacta y remarcaban un buen desarrollo en estructuras complejas como las pasivas, las negaciones o las oraciones condicionales (Bellugi, 1994; Clahsen y Almazan, 1998). Sin embargo, trabajos más recientes han señalado que, en realidad, sí está afectada la gramática, y que las observaciones de los autores como Ring y Clahsen (2005) se debían a que se comparaba a hablantes con SW con hablantes con SD. Estos últimos trabajos señalan que, aunque los hablantes con SD cometan errores morfológicos y sintácticos con menos frecuencia que los hablantes con otros trastornos genéticos como el SD, “their type and distribution remain atypical (...) which points towards differential trajectories of language development”¹⁰ (Díez-Itza et al. 2019: 219). De hecho, como hemos dicho, Joffe y Varlokosta (2007) muestran que, en cuanto a la producción de interrogativas, los resultados de los hablantes con SW son parecidos a los de los hablantes con SD, y en ambos casos están lejos de los resultados obtenidos por los hablantes típicos. Además, Perovic y Wexler (2007) demostraron que algunos aspectos de la gramática compleja que maduran tarde en poblaciones típicas no se desarrollan en los hablantes con SW, entre los que destacan ciertos tipos de pasivas o algunos aspectos de la rección y el ligamiento.

⁹ “Los individuos con SW (...) actuaron de forma parecida a los individuos con SD” (la traducción es nuestra).

¹⁰ “el tipo y la distribución se mantienen atípicos (...) lo que sugiere trayectorias distintas en el desarrollo del lenguaje” (la traducción es nuestra).

Al mismo tiempo, Grant et al. (2002) ponen de manifiesto la necesidad de investigar con más profundidad la capacidad sintáctica de los hablantes con SW, algo a lo que pretende contribuir este trabajo mediante el análisis de la producción espontánea de hablantes adultos con este síndrome.

2. MARCO TEÓRICO: EL ANÁLISIS DE ERRORES

El estudio se fundamenta en el análisis de errores, que surgió a partir del trabajo *The significance of learner's errors*, en el que Corder (1967) proponía la necesidad de analizar sistemáticamente los errores de los estudiantes de lenguas segundas, extranjeras o adicionales con el objetivo de determinar qué es lo que el estudiante ha interiorizado y qué no de la lengua objeto de aprendizaje. El análisis de errores es, pues, una corriente de investigación que nació como una rama de la lingüística aplicada y que busca estudiar y analizar los errores de los aprendices de lenguas “para descubrir sus causas y conocer las estrategias que utilizan los alumnos en el proceso de aprendizaje” (Martín Peris, 2008). Así, mientras que en enfoques como el análisis contrastivo parten de la comparación entre la primera lengua del aprendiz y la lengua meta (recordemos que esta corriente se aplica inicialmente a los procesos de adquisición y aprendizaje de lenguas adicionales), el análisis de errores se basa en las producciones reales de los aprendices y debe seguir los pasos siguientes: (1) identificación de los errores en su contexto; (2) clasificación y descripción de los errores; (3) explicación del origen de los errores, buscando los mecanismos o estrategias psicolingüísticas y las fuentes de cada error, y (4) evaluación de la gravedad del error y búsqueda de un posible tratamiento.

Según Santos Gargallo (1993), esta metodología representa una contribución significativa para la disciplina de la lingüística aplicada, en la medida que cambia el concepto de *error* que se tenía hasta el momento: así, en lugar de algo negativo que hay que corregir, en este tipo de análisis el error se entiende como un signo positivo de que se está produciendo un proceso de aprendizaje o de adquisición, muestra en qué punto se encuentra el proceso de aprendizaje y qué estrategias se están llevando a cabo, ayuda a establecer una jerarquía de dificultades (por lo tanto, ayuda a determinar las prioridades en la enseñanza o la actuación y a crear materiales útiles y significativos) y ofrece herramientas para evaluar al hablante en cada punto de su evolución.

En definitiva, se pasó a defender la necesidad del error a lo largo del aprendizaje de las lenguas porque puede contribuir a determinar las etapas por las que pasa el hablante, además de a crear materiales. El error es, en definitiva, una parte integrante del proceso de evolución que indica, por un lado, el estado de conocimiento lingüístico del aprendiz y, por el otro, cómo se aprende la lengua meta. Corder (1967) ya afirmaba que no hay nada negativo en el término *error* y, del mismo modo, Gràcia y Mayans consideraban que el error es “*un pas necessari en el camí cap al domini de l’L2*” (Gràcia y Mayans, 2009: 59).¹¹ En este sentido, Vázquez (1999) explica que analizar errores sirve para tener un conocimiento más profundo sobre el proceso de adquisición y aprendizaje y, por lo tanto, para enseñar mejor; para obtener información lingüística sobre problemas gramaticales todavía no generalizados, y para tomar decisiones en cuanto a los criterios sobre qué, cuándo y cómo corregir.

Alexopoulou (2005), siguiendo a Corder (1967) y Norrish (1983) advierte que hay que distinguir entre *error*, *lapsus* y *falta*. Así, considera que un error es una desviación sistemática que no puede ser corregida por el hablante, una falta es una desviación inconsistente y eventual que puede ser corregida y un lapsus es una desviación debida a factores extralingüísticos, como problemas de memoria, estados físicos (por ejemplo, cansancio) o estados psicológicos. Sin embargo, la propia autora reconoce que “es verdad que la distinción entre ‘error’ y ‘falta’, en

¹¹ “Un paso necesario en el camino hacia el dominio de la L2” (la traducción es nuestra).

la práctica, muchas veces es difícil de comprobar” (Alexopoulou, 2005: 48) y, en este sentido, en nuestro trabajo, basado en una muestra de habla espontánea cuyo contexto no hemos podido controlar, hemos considerado objeto de análisis todas las desviaciones de la norma.

Ligado con esto, como explica Rodríguez García (2020), la definición misma del concepto de la *norma* a partir de la cual se producen las desviaciones también puede resultar conflictiva, ya que tanto podemos tomar en consideración la norma prescriptiva (es decir, el sistema de reglas que recogen la ortografía, la gramática, la adecuación funcional de los mensajes lingüísticos y que sirve como base para la variedad estándar) como la norma descriptiva (que se refiere al uso común y corriente de la lengua por parte de una comunidad concreta de hablantes).

En este trabajo hemos tenido en cuenta solo la norma descriptiva. En este contexto, como explica Rodríguez García (2020), hay que tomar en consideración la variación lingüística, que puede depender del lugar geográfico, del momento histórico y del entorno social en el que el hablante ha crecido o ha adquirido la lengua, de modo que se ha considerado *norma* cualquier elemento que fuera adecuado al contexto comunicativo (a menudo, informal) en cualquier variedad geográfica de la lengua, siempre que el hablante fuera coherente. Por ejemplo, en el caso del catalán, se han dado por buenos tanto el uso del artículo definido derivado de *IPSU IPSA* (*es, sa*) como el derivado de *ILLU, ILLA* (*el, la*), o las distintas terminaciones de la primera persona del presente de indicativo (*cant, canto, cante, canti*).

Una vez determinado el concepto de *error*, sin embargo, hay que esclarecer cómo se clasifican los errores. Como explican Santos Gargallo (1993) y Vázquez (2010) a lo largo de la historia se han seguido varios criterios: (1) descriptivo, (2) pedagógico, (3) etiológico, (4) comunicativo, y (5) gramatical. En primer lugar, el criterio descriptivo, de acuerdo con los primeros estudios del generativismo, clasifica los errores según la forma en que la estructura superficial de los enunciados se ve alterada. Puede haber omisión, adición (doble marca, regularización o adición simple), falsa elección, o falsa colocación. En segundo lugar, el criterio pedagógico se basa en la distinción entre competencia y actuación y distingue los errores transitorios, permanentes, colectivos, individuales, fosilizables, fosilizados y ambiguos. En tercer lugar, el criterio etiológico distingue el origen del error, que puede ser extralingüístico (un error inducido o de método, uno causado por la personalidad, por el contexto de aprendizaje, por la clase social, por la corrección o por la memoria), lingüístico (interlingüístico, si se debe a la primera lengua o a otras lenguas del aprendiz, o intralingüístico) o bien psicológico (error por analogía, hipergeneralización, neutralización o asociación cruzada). En cuarto lugar, el criterio comunicativo se basa en su efecto sobre la comunicación desde la perspectiva del oyente. Algunos autores distinguen entre errores locales (si afectan sólo a elementos individuales de la oración) y globales (si afectan a la construcción del discurso), y otros distinguen entre ambigüedad, irritación y estigmatización. Finalmente, el criterio gramatical consiste en la clasificación basada en los niveles de análisis de la lengua y sus categorías.

A nuestro parecer, este último criterio es el que permite una presentación más metódica de los errores a partir de los sistemas de la lengua (y, por lo tanto, es el que hemos utilizado en este trabajo). Es por este motivo que, como ya proponían Abukahema et al. (2008), Dickinson y Ledbetter (2012), Rodríguez García (2020) y Cremades (2021, 2022), para este trabajo los errores se clasificaron según fueran morfológicos (como la forma verbal, forma pronominal, género, número) o sintácticos (como la estructura argumental, el uso de pronombres, el uso de preposiciones, la concordancia de género y número, el uso de artículos, el orden de palabras, el uso de tiempo verbal y modo y la selección de categorías). Se tuvieron en cuenta las frecuencias absolutas y relativas de errores, con una atención especial a los fenómenos más controvertidos en cuanto a la caracterización de la morfología y la sintaxis de los hablantes con SW (v. 4.2 para más información sobre la clasificación de los errores). Entre ellos se encuentran el reconocimiento y formación de género, número y concordancia (Benítez-Burraco et al., 2016); la producción y la comprensión de verbos regulares e irregulares (Clahsen y Almazan, 1998, Ring y Clahsen, 2005), la producción de oraciones relativas (Grant et al., 2002); la interpretación de

pronombres reflexivos y anafóricos (Benítez-Burraco et al., 2016), los principios del ligamiento (Perovic y Wexler, 2007) y las pasivas (Karmiloff-Smith et al., 1998; Benítez-Burraco et al., 2016, entre otros).

3. OBJETIVOS

El presente trabajo pretende dar respuesta a la pregunta de si existen diferencias estadísticamente significativas entre los hablantes con SW y las poblaciones típicas en ciertos aspectos de la sintaxis, es decir, si la gramática de los hablantes con SW está marcada, al menos hasta cierto punto, en la producción de la sintaxis. Por lo tanto, la pregunta de investigación de este trabajo es la siguiente: ¿los hablantes con SW producen espontáneamente una sintaxis marcada y con más errores respecto a los hablantes con desarrollo típico?

Esta pregunta de investigación implica dos objetivos secundarios: por un lado, descubrir si los hablantes con SW producen, en su discurso espontáneo, más errores morfosintácticos que los hablantes con desarrollo típico. Por otro lado, observar si, desde el punto de vista del tipo de oración que se produce, o de algunos rasgos que se podrían situar en la interfaz sintaxis-pragmática, las producciones de los hablantes con SW son marcadas con respecto a las de los hablantes con desarrollo típico. Como explica Turell (2011: 79), el uso de formas “inesperadas” en lugar de las más comunes en un contexto determinado puede ayudar a identificar el estilo de los individuos y, por tanto, facilitar la caracterización y descripción de su sintaxis.

3.1. METODOLOGÍA

3.2. Configuración del corpus

Para configurar el corpus, el *Laboratori d'Investigació en Complexitat i de Lingüística Experimental* (LICLE) colabora con el proyecto Àncora de la Universitat Pompeu Fabra.¹² El objetivo principal de este proyecto es la recopilación de datos lingüísticos de hablantes con SW. Entre otras cosas, se están recopilando corpus de habla espontánea, de donde se extrajo el corpus utilizado para el presente estudio. El proyecto “Conocimiento semántico en el síndrome de Williams” recibió la aprobación del Comité de Ética de la Investigación con medicamentos del Parc de Salut MAR. Una parte de este proyecto se concentra en la recopilación de muestras de habla espontánea y el análisis de errores gramaticales. Cada fragmento tenía una duración de entre 4 y 6 minutos y se transcribió ortográficamente en todos los casos. Los participantes son un total de 30, con un grupo control de 15 personas lingüísticamente adultas (media de edad 23,20, DE \pm 15,93) y un grupo de 15 personas con SW (media de edad 23,25 años, DE \pm 13,55). De cada grupo, 4 informantes tenían el catalán como L1 y 11 informantes, el español. Los participantes fueron informados del procedimiento y dieron su consentimiento para la realización de la prueba.

3.3. Análisis: clasificación de los errores y análisis estadístico

Como ya se ha comentado, en este trabajo se optó por emplear el criterio gramatical, ya que es el que permite categorizar de un modo más metódico el tipo de error (Cremades, 2021, 2022). Como el artículo se centró en el análisis morfosintáctico, en la clasificación de los errores se tuvieron en cuenta, por un lado, los errores morfológicos y, por el otro, los errores sintácticos.

¹² Asimismo, se recurrirá a datos descriptivos para dar detalles sobre diferencias que no resultaron estadísticamente significativas. Se puede encontrar más información sobre el proyecto Àncora en la siguiente web: <https://clarasoberats.wixsite.com/investigawilliams>.

En cuanto a los errores morfológicos, es decir, referidos a cuestiones relacionadas con la forma de las palabras o de los morfemas, se analizaron los errores en el marcaje de género (es decir, las marcas de femenino y masculino propiamente dichas), el de número (es decir, las marcas de singular y de plural), la forma de los pronombres, la forma de los artículos, etc., pero no cuestiones vinculadas al uso de estas categorías.

En cuanto a los errores sintácticos, se analizaron los posibles errores relacionados con el ámbito verbal (aspecto, modo o tiempo, por ejemplo), el ámbito nominal (uso de los artículos y de los pronombres), los usos de los verbos atributivos (*ser, estar*), el uso de las preposiciones, el orden de las oraciones, las conjunciones, el uso de ciertos determinantes, errores en la selección categorial, el régimen verbal, el régimen adjetival, etc.

También se tuvieron en consideración los tipos de oración producidos, y se distinguió entre las oraciones simples y las subordinadas (relativas, completivas y adverbiales). Las etiquetas de los tipos de errores son las que presentamos en la *Tabla 1*.

Tabla 1: Etiquetas de clasificación de los errores

Etiqueta	Significado
[SIN_TV]	sintaxis, tiempo verbal
[SIN_Pron]	sintaxis, pronombres
[SIN_Prep]	sintaxis, preposiciones
[SIN_Pos]	sintaxis, posesivos
[SIN_Pers]	sintaxis, persona
[SIN_Orden]	sintaxis, orden de constituyentes
[SIN_O]	sintaxis, construcción oracional
[SIN_Num]	sintaxis, número
[SIN_Neg]	sintaxis, negación
[SIN_MV]	sintaxis, modo verbal
[SIN_Gen]	sintaxis, género
[SIN_Dem]	sintaxis, demostrativos
[SIN_Conj] Polisíndeton	sintaxis, polisíndeton
[SIN_Conj] altra	sintaxis, otras conjunciones
[SIN_Art]	sintaxis, artículos
[SIN_Adv]	sintaxis, adverbios
[SIN_RegV]	sintaxis, régimen verbal
Morf_PL	Morfología, forma del plural

Para el análisis estadístico se recurrió a la prueba de Mann-Whitney, una prueba no paramétrica, puesto que el tamaño reducido de la muestra no permitía garantizar las suposiciones de normalidad ni homogeneidad de varianzas; la prueba aplicada permitió, por lo tanto, más fiabilidad que las pruebas paramétricas). Asimismo, se utilizan datos descriptivos para dar detalles sobre diferencias que no resultaron estadísticamente significativas.

4. RESULTADOS

4.1. Análisis de errores

En general, la primera observación que se puede hacer en relación con la producción de errores es que, globalmente, los hablantes con SW producen más errores que los del grupo control, tanto en términos absolutos como en términos relativos (es decir, en porcentajes de formas erróneas según el número total de palabras del discurso). Es lo que se puede comprobar en la *Figura 1*, que contiene el número total de errores de los hablantes con SW (izquierda) en comparación con los hablantes con desarrollo típico, que se ha utilizado como grupo control (derecha). Mientras que en un caso los errores pueden llegar casi al 8% de las producciones totales de los hablantes, en el otro, apenas superan el 3%.

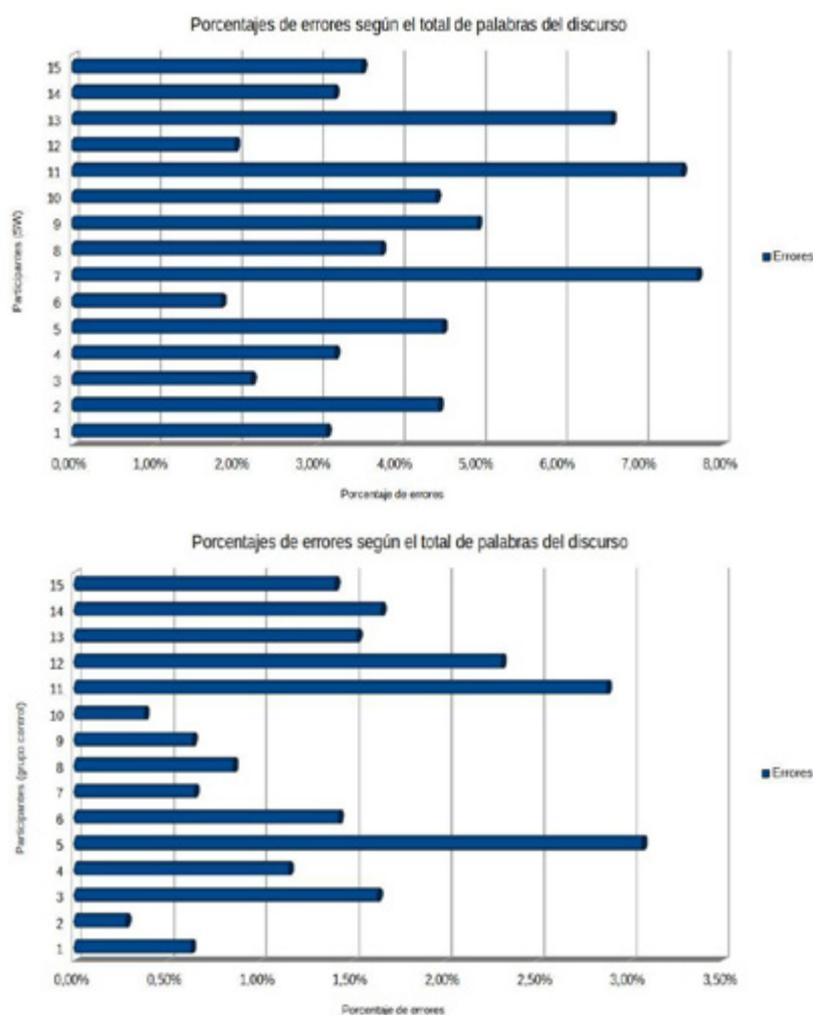


Figura 1: Producción total de errores (SW y grupo control)

Las diferencias que se observan en los gráficos son significativas. Así, tal como se ve en la *Tabla 2*, los resultados de la prueba de Mann-Whitney muestran que el número de errores sintácticos en el habla espontánea de las personas con SW ($n = 15$) es significativamente superior al de los del grupo control ($n = 15$), en total ($p = < 0,001$), con un tamaño del efecto grande ($g = -0,916$).

Tabla 2: Errores sintácticos y morfológicos. Prueba U de Mann-Whitney. Tamaño del efecto calculado con la g de Hedges (intervalo de confianza 95%)

Tipo de error	W	p	g
Tiempo verbal	40,500	< ,001	-0,640
Pronombre	83,500	0,198	-0,258
Preposición	96,500	0,277	-0,142
Posesivo	NaN		
Persona	NaN		
Orden de constituyentes	41,000	< ,001	-0,636
Construcción oracional	118,500	0,811	0,053
Concordancia de número	125,500	0,456	0,116
Negación	NaN		
Modo verbal	NaN		
Concordancia de género	104,000	0,524	-0,076
Demostrativo	NaN		
Polisíndeton con la conjunción copulativa	80,500	0,061	-0,284
Otras conjunciones	103,500	0,674	-0,080
Artículo	79,000	0,160	-0,200
Adverbio	NaN		
Régimen verbal	NaN		
Forma del plural	NaN		
Total	9,500	< ,001	-0,915

Cuando se analiza de manera desglosada el conjunto de errores, se observa que hay categorías que sobresalen. Una subcategoría de error sintáctico que ha resultado significativa es la de tiempo verbal (TV) ($p = < 0,001$) pero con un tamaño de efecto moderado ($g = -0,640$); otra es la referente al orden sintáctico ($< 0,001$) con un tamaño de efecto moderado ($g = -0,636$). El resto de subcategorías no han alcanzado el suficiente número de errores para indicar que sea algo significativo.

Sin embargo, se pueden constatar algunas diferencias. Así, como se puede comprobar en la *Figura 2*, en términos absolutos, el número de errores relacionados con la morfología del plural, el género, el número, los artículos, la negación y los pronombres nunca llega, en los hablantes del grupo control, a los cuatro errores en términos absolutos. En cambio, entre los hablantes con SW puede haber hasta 30 errores en total (dentro de este grupo).

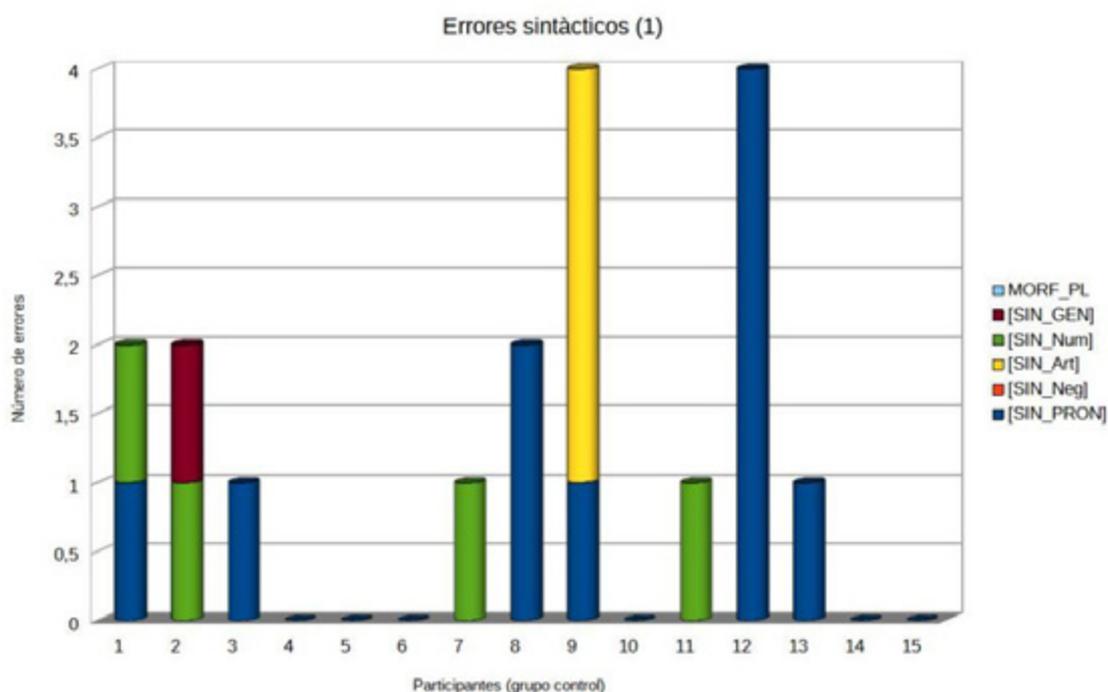
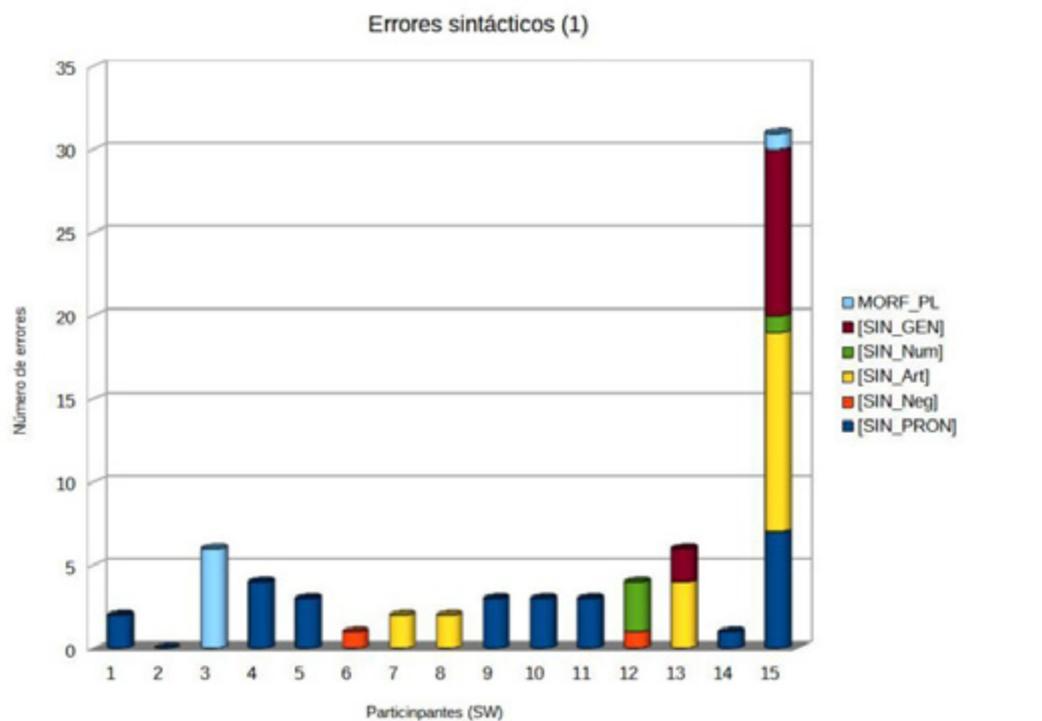


Figura 2: Producción de errores sintácticos: plural (morfología), género, número, artículo, negación, pronombre

De un modo parecido al anterior, en la *Figura 3* se constata que, mientras que todos los hablantes con SW han producido algún error vinculado a los adverbios, las conjunciones, los posesivos o las preposiciones (no las de régimen verbal, sino las semánticamente explícitas), los hablantes del grupo control solo producen, en algunos casos, errores en las conjunciones.

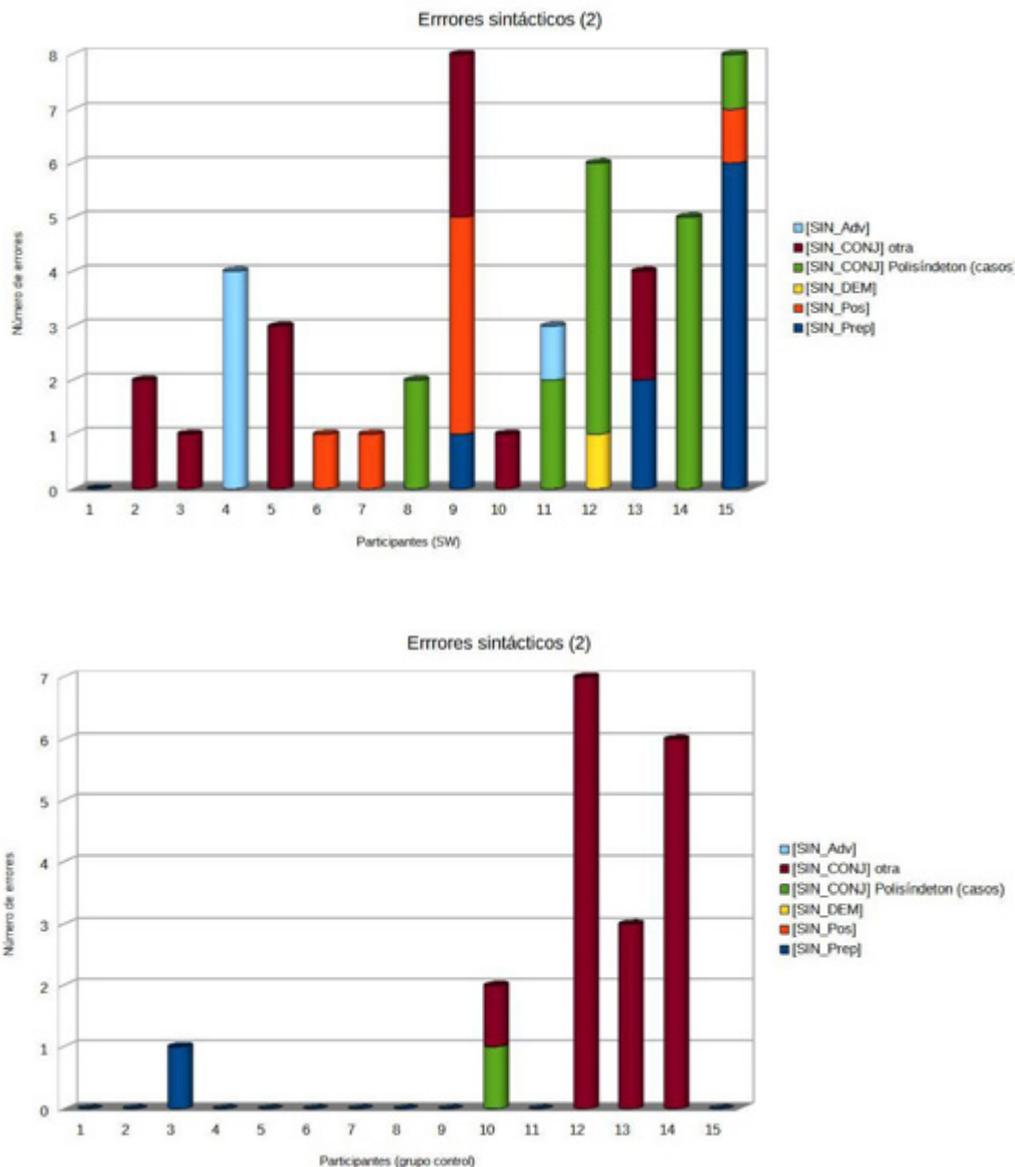


Figura 3: Producción de errores sintácticos: adverbios, conjunciones (polisíndeton), conjunciones (otras), demostrativos, posesivos, preposiciones

Finalmente, en la Figura 4 se observa que, mientras que los hablantes con SW producen, en casi todos los casos, errores en los tiempos verbales y en la construcción de la oración, este tipo de error no es frecuente entre los hablantes del grupo control.

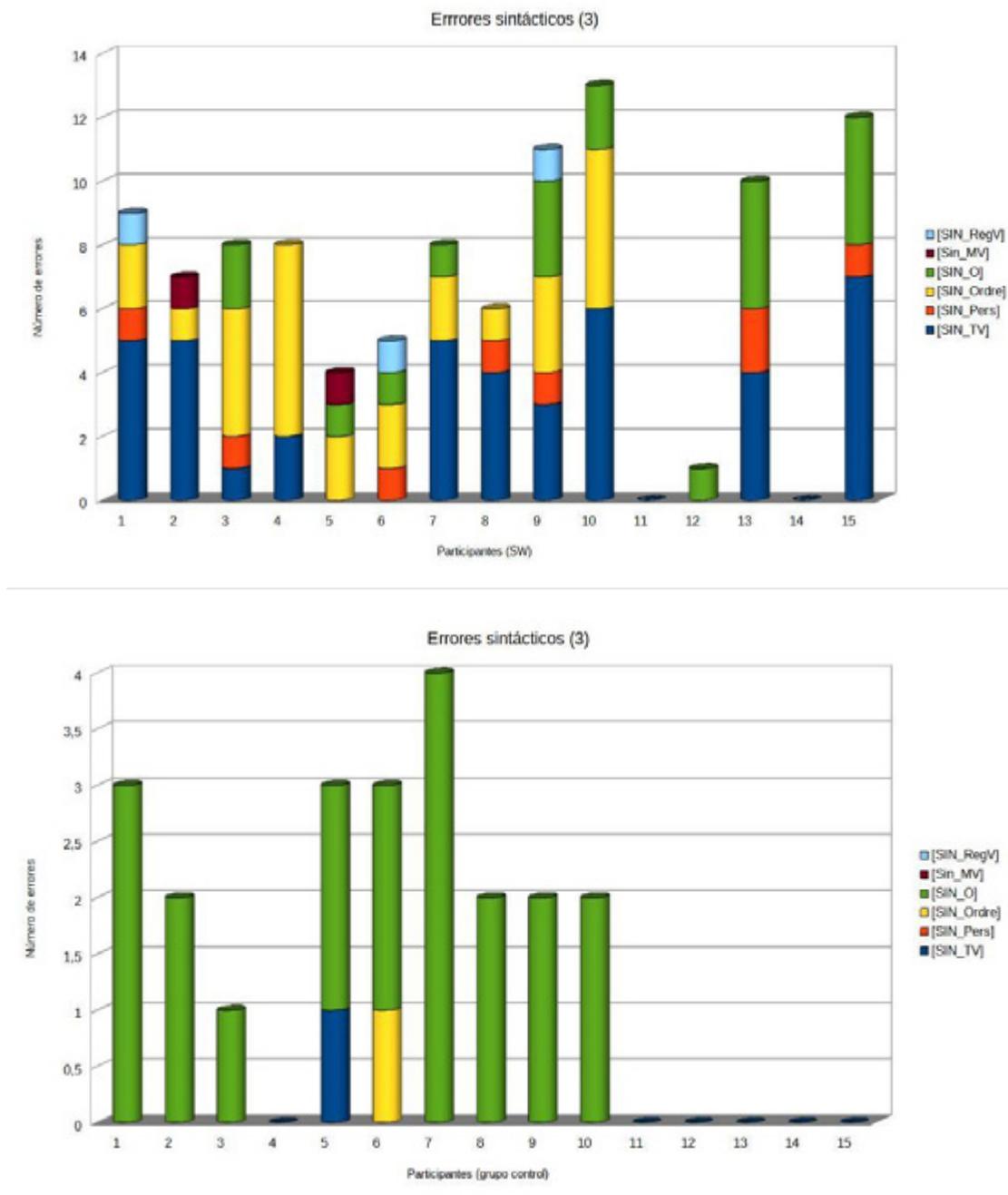


Figura 4: Producción de errores sintácticos: régimen verbal, modo verbal, construcción de la oración, orden de constituyentes

4.2. Producción de estructuras subordinadas

Una cuestión ampliamente debatida en la relación entre la sintaxis y el SW es la complejidad oracional. Recordemos que Brock (2007) explicaba que, aparentemente, los hablantes con SW producen oraciones con grados de complejidad similares a los de los hablantes con desarrollo típico, algo que ha sido puesto en cuestión por varios autores (recientemente, por Barceló-Coblign y Cremades, 2024).

Estas diferencias de complejidad se pueden observar en la *Figura 5*, en la que se constata que, entre los hablantes del grupo control, la producción total de oraciones subordinadas llega a 60 en el mismo periodo de 5 minutos de discurso espontáneo, mientras que, entre los hablantes con SW, dicha producción es considerablemente menor (especialmente en cuanto a producción de relativas). Además, a menudo las subordinadas adverbiales que producen son respuestas a preguntas, encabezadas por la conjunción *porque*, de modo que parece que, en cuanto a la complejidad oracional, sí existen diferencias en los adultos con SW y los hablantes del grupo control.

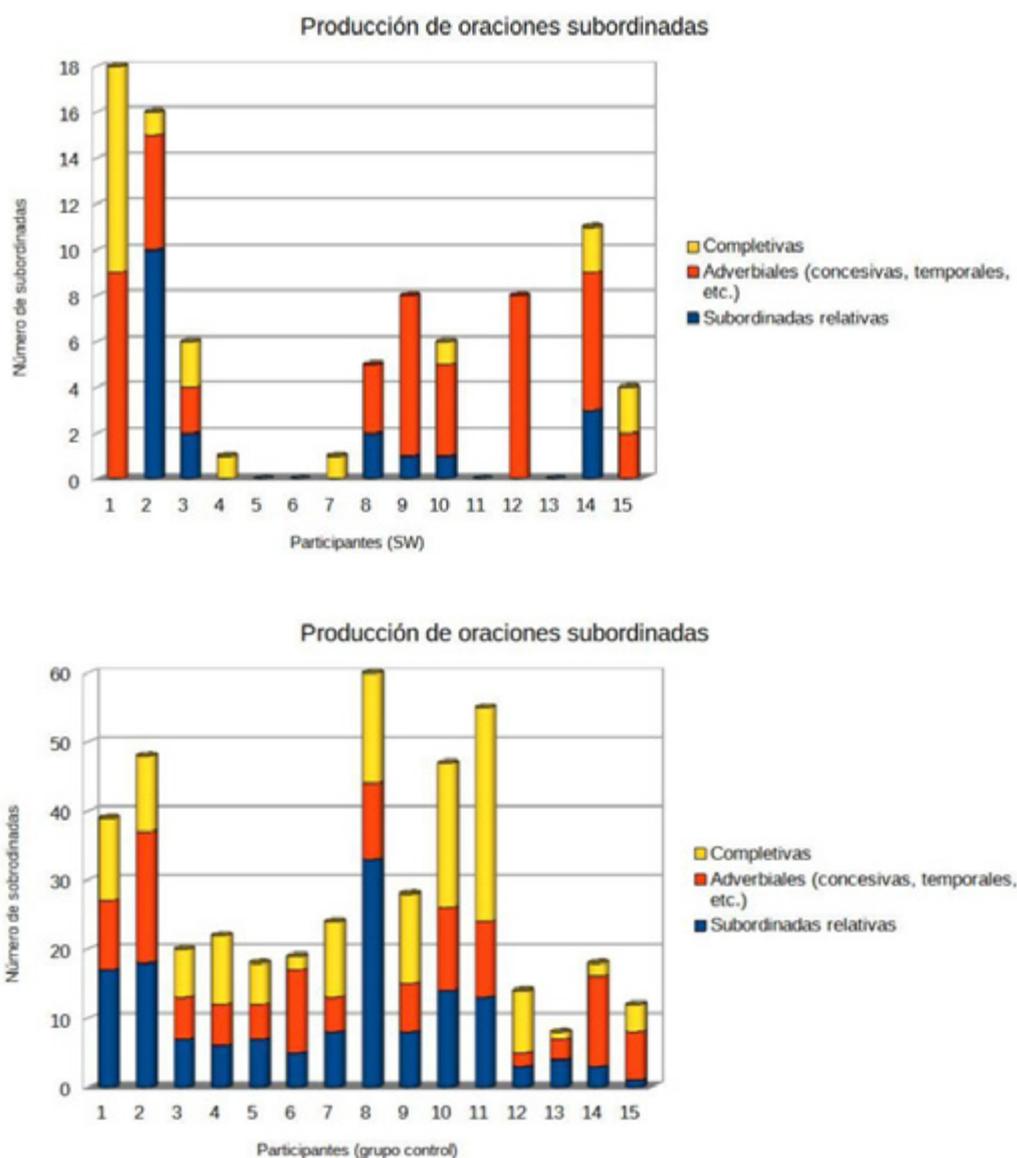


Figura 5: Producción de oraciones subordinadas: completivas, adverbiales y relativas

De hecho, en la *Tabla 3* podemos constatar que estas diferencias son significativas tanto en el número total de oraciones subordinadas como en cada uno de los tres tipos de subordinadas. Así, se puede afirmar que el grupo control produce más estructuras subordinadas que los participantes con SW. Aunque significativo, el efecto es más atenuado en la subcategoría de subordinadas adverbiales ($p = 0,001$; $g = 0,684$).

Tabla 3: Producción de oraciones subordinadas. Prueba U de Mann-Whitney. Tamaño del efecto calculado con la g de Hedges (intervalo de confianza 95%)

Tipo de subordinada	W	p	g
Relativa	185,000	< ,001	0,867
Adverbial (concesivas, temporales, etc.)	162,000	0,001	0,684
Completiva	181,500	< ,001	0,871
Total	190,000	< ,001	0,938

5. DISCUSIÓN

En este trabajo se han examinado 30 producciones espontáneas de hablantes adultos de catalán y español (15 con SW y 15 con desarrollo típico) con el objetivo principal de determinar si, por un lado, los hablantes con SW cometen más errores morfosintácticos en la producción de discursos espontáneos y si, por el otro, su sintaxis, en cuanto a la complejidad de las estructuras producidas, es de algún modo marcada en el sentido de Turell (2011) –es decir, entendiéndolo por *marcades* el uso de formas inesperadas o atípicas en lugar de las más frecuentes en un contexto determinado. Esto, de algún modo, ya había sido señalado por Reilly et al. (2004) o Thomas et al. (2007), para quienes el lenguaje expresivo era estilísticamente diferente, con vocabulario atípico, frases estereotipadas, lenguaje excesivamente familiar y un uso abundante de mecanismos de evaluación social, si bien estas observaciones no eran unánimes.

Los resultados permiten constatar que la sintaxis de los hablantes con SW es, como se había presupuesto, marcada de distintos modos. Para empezar, producen significativamente más errores morfosintácticos que los hablantes típicos en segmentos de duración similar. Estos resultados se encuentran en la línea de Díez-Itza, Martínez, Fernández-Urquiza y Antón (2017), que encuentran una media de errores morfológicos significativamente más alta entre los hablantes con SW que entre los hablantes con DT. De hecho, señalan “clear-cut deficits in the production of aspects of morphosyntax”¹³ (Díez-Itza et al., 2017: 321). Además, los tipos de error son distintos. En este sentido, en los hablantes del grupo control (es decir, con DT) predominan errores que se pueden asociar al habla espontánea, como los errores de régimen verbal o de concordancia de género y número, algo explicable por la naturaleza informal y fluida de la conversación espontánea. En cambio, entre los hablantes con SW predominan los errores en los tiempos verbales, en el orden oracional, en el uso de determinadas conjunciones, en los pronombres y en los artículos. Estos errores son, en algunos casos, significativos, como demuestran los análisis estadísticos, lo que se ve en los tiempos verbales y el orden oracional. Esto podría sugerir dificultades subyacentes en la manipulación de estructuras jerárquicas del lenguaje, así como limitaciones en la memoria de trabajo o en el control ejecutivo que afecta la producción sintáctica.

En este sentido, pues, los resultados encajarían con la idea manifestada por Brock, de que hay estudios que demuestran que “children with Williams syndrome produce grammatical errors that are extremely rare in typical development”¹⁴ (Brock, 2007: 13), y que esta producción atípica se mantendría hasta la edad adulta. Es, asimismo, lo que aseñalan Díez-Itza et al. cuando afirman que, más allá de las diferencias cuantitativas en la frecuencia de errores, “the

¹³ “Déficits claros en la producción de algunos aspectos morfosintácticos” (la traducción es nuestra).

¹⁴ “Los niños con SW producen errores gramaticales que son extremadamente raros en el desarrollo típico” (la traducción es nuestra).

patterns observed in the WS individuals did not respond to a typical path of development”¹⁵ (Díez-Itza et al., 2017: 324), y que no se trata tan solo de un retraso en el desarrollo, sino de trayectorias distintas de las de los individuos con desarrollo típico. Sin embargo, puesto que el corpus analizado no contenía suficientes ocurrencias de algunos de los tipos de error, habrá que investigar más a fondo otras categorías.

Por otro lado, la complejidad de la estructura oracional también parece distinta entre los hablantes con SW y los del grupo control. Así, los hablantes con SW producen significativamente menos oraciones subordinadas que los del grupo control, especialmente en relación con las relativas, pero también con las completivas y las causales. Estos resultados se podrían relacionar con los movimientos provocados por los pronombres relativos (las palabras *cu-*) que, como demostraban Joffe y Varlokosta (2007) para las interrogativas, presentaban serias dificultades para los hablantes con SW. Asimismo, las observaciones de este estudio encajan con las de Sederias et al., quienes consideran que los hablantes con SW son más dependientes de las representaciones holísticas (es decir, de las representaciones estadísticas) que de las reglas gramaticales abstractas, por lo que caracterizan el lenguaje de las personas con SW como un lenguaje que contiene “short and syntactically simple utterances with a tendency to overuse familiar (strongly collocated) word combinations”¹⁶ (Sederias et al., 2024: 9).

Cabe destacar la alta variabilidad que existe entre los hablantes con SW, algo que también ha sido señalado por Romero-Rivas, Rodríguez-Cuadrado, Sabater, Rodríguez Gómez, Hidalgo de la Guía, Moreno y Garayzábal, que ponen de manifiesto que “people with WS present very diverse cognitive and linguistic profile”¹⁷ (Romero-Rivas et al., 2023: 12). Esto dificulta, en algunos casos, establecer generalizaciones y que requeriría profundizar en la investigación. Sin embargo, parece que, a pesar de la variabilidad individual, los hablantes con SW producen más errores que los hablantes del grupo control y suelen construir menos oraciones subordinadas.

6. CONCLUSIONES

Los resultados del presente estudio sugieren que la capacidad sintáctica de las personas con SW es diferente de la de personas con DT, puesto que presentan diferencias significativas tanto en la precisión morfosintáctica como en la complejidad sintáctica. En el ámbito de la morfosintaxis, los hablantes con SW cometen más errores, entre los que destacan fallos en los tiempos verbales, el orden oracional y el uso de conjunciones, pronombres y artículos. A diferencia de los hablantes con DT, cuyos errores se asocian a la espontaneidad del discurso (como la concordancia de género y número o el régimen verbal), los errores en los hablantes con SW parecen reflejar dificultades subyacentes en el manejo de estructuras jerárquicas y en la memoria de trabajo, lo que coincide con investigaciones previas que documentan déficits morfosintácticos claros en este grupo (Benítez-Burraco et al., 2016; Díez-Itza et al., 2017).

En cuanto a la complejidad sintáctica, los hablantes con SW tienden a producir menos subordinadas, especialmente relativas, completivas y causales, en comparación con los hablantes del grupo control. Este patrón puede estar relacionado con dificultades en el manejo de movimientos sintácticos complejos, como los asociados con pronombres relativos. Asimismo, se observa una preferencia por estructuras más simples y frases estereotipadas, lo que sugiere una dependencia mayor de representaciones estadísticas (es decir, combinaciones familiares de palabras) en lugar de reglas gramaticales abstractas (Sederias et al., 2024). Esto refuerza la

¹⁵ “Los patrones observados en los individuos con SW no responde a la trayectoria de desarrollo típico” (la traducción es nuestra).

¹⁶ “Enunciados cortos y simples con una tendencia a abusar de las combinaciones de palabras familiares (colocaciones fuertes)” (la traducción es nuestra).

¹⁷ “Las personas con SW presenta un perfil cognitivo y lingüístico muy diverso” (la traducción es nuestra).

caracterización del lenguaje de las personas con SW como estilísticamente diferente y menos complejo. Ello sugiere que no solo el procesamiento sino la producción, ambos vinculados a la memoria de trabajo, se habrían desarrollado de manera diferente en el caso de los fenotipos afectados por el SW.

Es destacable la alta variabilidad entre los hablantes con SW, lo que refleja una diversidad significativa en sus perfiles lingüísticos y cognitivos. Aunque esta variabilidad dificulta la generalización, los hallazgos permiten afirmar que, en términos generales, los hablantes con SW producen más errores morfosintácticos y construyen estructuras menos complejas que los hablantes con DT. Próximos trabajos sobre habla espontánea, sin embargo, podrían ampliar y consolidar tales sugerencias.

REFERENCIAS

- Abuhakema, G., Faraj, R., Feldman, A. y Fitzpatrick, E. (2008). Annotating an Arabic learner corpus for error. A N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis y D. Tapias (Eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* (pp.1347-1350). European Language Resources Association. http://www.lrec-conf.org/proceedings/lrec2008/pdf/343_paper.pdf
- Alexopoulou, A. (2005). *Análisis de errores en la interlengua de aprendientes griegos de español*. Ediciones Clásicas.
- Barceló-Coblijn, L. y Cremades, E. (2023, 6 de julio). A first look into the capability of Catalan and Spanish-speaking individuals with Williams syndrome for syntactically combining words (comunicación oral). *ICPLA'23* (Salzburg, 4-7 de julio de 2023).
- Barceló-Coblijn, L. y Cremades, E. (2024, 20 de junio). Williams Syndrome through the lens of syntax: analyzing spontaneous speech patterns. *VII Congreso Internacional de Lingüística Clínica* (Salamanca, 19-21 de junio de 2024).
- Bellugi, U., Marks, S., Bihrlé, A., Sabo, H. (1988). Dissociation between language and cognitive functions in Williams syndrome. En D. Bishop y K. Mogford (Eds.), *Language development in Exceptional Circumstances* (pp. 177-189). Psychology Press.
- Bellugi, U. (1994). Williams syndrome: an unusual neuropsychological profile. En S. Brosnan y J. Grafman (Eds.), *Atypical Cognitive Deficits in Developmental Disorders: Implications for Brain Function* (pp. 23-56). Lawrence Erlbaum Associates.
- Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z. y St. George, M. (2000). The neurocognitive profile of Williams syndrome: a complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, 12:S, 7-29. <https://doi.org/10.1162/089892900561959>
- Benítez-Burraco, A., Garayzábal, E. y Cueto, F. (2016). Syntax in Spanish-speaking children with Williams syndrome. *Journal of Communication Disorders* 60, 51-61. <https://doi.org/10.1016/j.jcomdis.2016.03.001>
- Brock, J. (2007). Language abilities in Williams syndrome: a critical review. *Developmental Psychopathology*, 19(1), 97-127. <https://doi.org/10.1017/S095457940707006X>
- Brown, R. (1973). *A first language: the early stages*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674732469>

Clahsen, H. y Almazan, M. (1998). Syntax and morphology in Williams syndrome. *Cognition*, 68(3), 167-198. [https://doi.org/10.1016/S0010-0277\(98\)00049-3](https://doi.org/10.1016/S0010-0277(98)00049-3)

Corder, S.P. (1967). The significance of learners' errors. *IRAL*, 5, 161-170. <https://doi.org/10.1515/iral.1967.5.1-4.161>

Cremades, E. (2021). Anàlisi d'errors en l'expressió escrita del català com a llengua addicional: contrast entre l'alumnat serbi i l'anglòfon. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education*, 4(2), 21-34. <https://doi.org/10.5565/rev/clil.65>

Cremades, E. (2022). Expressió escrita i català com a llengua addicional: anàlisi d'errors en parlants de llengües eslaves. *Reseracle*, 3, 50-69. <https://doi.org/10.31009/reseracle.2022.03>

Cremades, E. y Barceló-Coblijn, L. (2023, 28 de abril). Grammar in speakers of Catalan and Spanish with Williams syndrome: a first descriptive approach (comunicación oral). *40º Congreso Internacional de la Asociación Española de Lingüística Aplicada* (Mérida, 26-28 de abril de 2023).

D'Souza, D., Karmiloff-Smith, A. (2011). When modularization fails to occur: a developmental perspective. *Cognitive Neuropsychology*, 28(3-4), 276-87. <https://doi.org/10.1080/02643294.2011.614939>

Dickinson, M. y Ledbetter, S. (2012). Annotating errors in a Hungarian learner corpus. En A.N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk y S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 1659–1664). European Language Resources Association.

Díez-Itza, E., Martínez, V., Fernández-Urquiza, M., & Antón, A. (2017). Morphological Profile of Williams Syndrome: Typical or Atypical? In *Language Development and Disorders in Spanish-Speaking Children* (pp. 311-327). Cham: Springer. https://doi.org/10.1007/978-3-319-53646-0_15

Díez-Itza, E., Miranda, M., Pérez, V. y Martínez, V. (2019). Profiles of grammatical morphology in Spanish-speaking adolescents with Williams Syndrome and Down Syndrome. En E. Aguilar-Mediavilla, L. Buil-Legaz, R. López-Penadés, V. A. Sánchez-Azanza y D. Adrocer-Roig. *Atypical Language Development in Romance Languages* (pp. 219-234). John Benjamins Publishing Company. <https://doi.org/10.1075/z.223.13die>

Díez-Itza, E., Viejo, A. y Fernández-Urquiza, M. (2022). Pragmatic Profiles of Adults with Fragile X Syndrome and Williams Syndrome. *Brain Sci.*, 2022, 12(3), 385. <https://doi.org/10.3390/brainsci12030385>

Eckert, M. A., Hu, D., Eliez, S., Bellugi, U., Galaburda, A., Korenberg, J., Mills, D. y Reiss, A. L. (2005). Evidence for superior parietal impairment in Williams Syndrome. *Neurology*, 64, 152-153. <https://doi.org/10.1212/01.WNL.0000148598.63153.8A>

Ewart, A. K., Morris, C. A., Atkinson, D., Jin, W., Sternes, K., Spallone, P., Stock, A. D., Leppert, M. y Keating, M. T. (1993a). Hemizygoty at the elastin locus in a developmental disorder, Williams syndrome. *Nature genetics*, 5(1), 11–16. <https://doi.org/10.1038/ng0993-11>

- Ewart, A. K., Morris, C. A., Ensing, G. J., Loker, J., Moore, C., Leppert, M. y Keating, M. (1993b). A human vascular disorder, supravalvular aortic stenosis, maps to chromosome 7. *Proceedings of the National Academy of Sciences of the United States of America*, 90(8), 3226–3230. <https://doi.org/10.1073/pnas.90.8.3226>
- Ewart, A. K., Jin, W., Atkinson, D., Morris, C. A. y Keating, M. T. (1994). Supravalvular aortic stenosis associated with a deletion disrupting the elastin gene. *The Journal of clinical investigation*, 93(3), 1071–1077. <https://doi.org/10.1172/JCI117057>
- Gràcia, L. y Mayans, P. (2009). *Llengua i immigració. La influència de la primera llengua en l'adquisició del català com a segona llengua*. Eumo.
- Grant, J.; Valian, V. y Karmiloff-Smith, A. (2002). Study of relative clauses in Williams syndrome. *Journal of Child Language*, 29, 403-416. <https://doi.org/10.1017/S030500090200510X>.
- Hanson, K. L., Lew, C. H., Hrvoj-Mihic, B., Groeniger, K. M., Halgren, E., Bellugi, U. y Semendeferi, K. (2018). Increased glia density in the caudate nucleus in Williams syndrome: Implications for frontostriatal dysfunction in autism. *Developmental neurobiology*, 78(5), 531–545. <https://doi.org/10.1002/dneu.22554>
- Hepburn, S. L., Fidler, D., Hahn, L. y Philofsky, A. (2011). Social-perceptual and social-cognitive skills in young children with Williams syndrome: evidence for discontinuity. *International Review of Research in Developmental Disabilities* 40, 181-210. <https://doi.org/10.1016/B978-0-12-374478-4.00007-1>
- Howlin, P., Davies, M. y Udwin, O. (1998). Cognitive functioning in adults with Williams syndrome. *Journal of Child Psychology and Psychiatry*, 39, 183-189. <https://doi.org/10.1111/1469-7610.00312>
- Joffe, V. y Varlokosta, S. (2007). Patterns of syntactic development in children with Williams syndrome and Down's syndrome: evidence from passives and wh- questions. *Clinical Linguistics & Phonetics*, 21(9), 705-727. <https://doi.org/10.1080/02699200701541375>
- Karmiloff-Smith, A., Grant, J., Berthoud, I., Davies, M., Howlin, P. y Udwin, O. (1997). Language and Williams syndrome: how intact is intact? *Child Development*, 68, 246-262. <https://doi.org/10.1111/j.1467-8624.1997.tb01938.x>
- Karmiloff-Smith, A., Tyler, L., Voice, K., Sims, K., Udwin, O., Howlin, P., Davies, M. (1998). Linguistic dissociations in Williams syndrome: evaluating receptive syntax in on-line and off-line tasks, *Neuropsychologia*, 36(4), 343-351. [https://doi.org/10.1016/S0028-3932\(97\)00133-4](https://doi.org/10.1016/S0028-3932(97)00133-4)
- Karmiloff-Smith, A., Thomas, M. S. C., Annaz, D., Humphreys, K., Ewing, S., Brace, N., van Duuren, M., Pike, G., Grice, S. y Campbell, R. (2004). Exploring the Williams syndrome face-processing debate: the importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 45(7): 1258-1274. <https://doi.org/10.1111/J.1469-7610.2004.00322.X>
- Llull, A., Barceló-Coblijn, L. y Cremades, E. (2024, 17 de abril). Williams syndrome and morphosyntax: the case for gender and number agreement (comunicación oral). *41º Congreso Internacional de la Asociación Española de Lingüística Aplicada* (València, 17-19 de abril de 2024).

- Lorusso, M. L., Burigo, M., Tavano, A., Milani, A., Martelli, S., Borgatti, R. y Molteni, M. (2017). Learning and using abstract words: evidence from clinical populations. *BioMed research international*, 2017, 8627569. <https://doi.org/10.1155/2017/8627569>
- Martín Peris, E. (dir.) (2008). *Diccionario de términos clave de ELE* [en línea]. Sociedad General Española de Librería, Centro Virtual Cervantes.
- McNeill, D. (1970). Language before symbols: very early children's grammar. *Interchange*, 1(3), 127-133. <https://doi.org/10.1007/BF02214683>
- Mervis, C. B., Morris, C. A., Bertrand, J. y Robinson, B. F. (1999). Williams syndrome: findings from an integrated program of research. En H. Tager-Flusberg (Ed.). *Neurodevelopmental disorders* (pp. 65-110). The MIT Press. <https://doi.org/10.7551/mitpress/4945.003.0009>
- Mervis, C. B., Robinson, B. F., Bertrand, J., Morris, C. A., Klein-Tasman, B. P. y Armstrong, S. C. (2000). The Williams syndrome cognitive profile. *Brain and Cognition*, 44, 604-628. <https://doi.org/10.2147/PRBM.S63474>
- Mervis, C. B. y Berra, A. M. (2007). Language and communicative development in Williams syndrome. *Mental Retardation Development and Disabilities Research Reviews*, 13, 3-15. <https://doi.org/10.1002/mrdd.20140>
- Moraleda, E. y López Resa, P. (2024). Pragmatic skills in people with Williams syndrome: the perception of families. *Orphanet journal of rare diseases*, 19(1), 95. <https://doi.org/10.1186/s13023-024-03016-0>
- Niego, A. y Benítez-Burraco, A. (2019). Williams Syndrome, human self-domestication, and language evolution. *Frontiers in Psychology*, 10, 521. <https://doi.org/10.3389/fpsyg.2019.00521>
- Ninio, A. (2006). *Language and the learning curve: A new theory of syntactic development*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199299829.001.0001>
- Norrish, J. (1983). *Language learners and their errors*. The Macmillan Press.
- Penke, M. (2015). Syntax and language disorders. En T. Kiss, A. Alexiadou (Eds.). *Syntax: theory and analysis. An international handbook* (pp.1833-1874). Walter de Gruyter. <https://doi.org/10.1515/9783110363685-013>
- Perovic, A. y Wexler, K. (2007). Complex grammar in Williams syndrome. *Clinical Linguistics & Phonetics*, 21(9), 729-745. <https://doi.org/10.1080/02699200701541409>
- Perovic, A., Modyanova, N. y Wekler, N. (2013). Comprehension of reflexive and personal pronouns in children with autism. A syntactic or pragmatic deficit? *Applied psycholinguistics*, 34(4), 813-835. <https://doi.org/10.1017/S0142716412000033>
- Pinker, S. (2011). *Words and rules*. Harper Perennial.
- Reilly, J., Klima, E. y Bellugi, U. (1990). Once more with feeling: affect and language in atypical populations. *Development and Psychopathology*, 2(4), 367-391. <https://doi.org/10.1017/S0954579400005782>

- Reilly, J., Losh, M., Bellugi, U. y Wulfeck, B. (2004). "Frog, where are you?" Narratives in children with specific language impairment, early focal brain injury, and Williams syndrome. *Brain and language*, 88(2), 229-247. [https://doi.org/10.1016/S0093-934X\(03\)00101-9](https://doi.org/10.1016/S0093-934X(03)00101-9)
- Ring, M. y Clahsen, H. (2005). Distinct patterns of language impairment in Down's syndrome and Williams syndrome: the case of syntactic chains. *Journal of Neurolinguistics*, 18, 479-501. <https://doi.org/10.1016/j.jneuroling.2005.06.002>
- Rodríguez García, C. (2020). Anàlisi d'errors en català com a llengua estrangera en l'expressió oral d'estudiants universitaris txecs. *Estudis Romànics*, 42, 83-99. <https://raco.cat/index.php/Estudis/article/view/370594>
- Romero-Rivas, C., Rodríguez-Cuadrado, S., Sabater, L., Rodríguez Gómez, P., Hidalgo e la Guía, I., Moreno, E. M., Garayzábal, E. (2023). Beyond the conservative hypothesis: a meta-analysis of lexical-semantic processing in Williams syndrome. *Language and Cognition* 15(3), 1-25. <https://doi.org/10.1017/langcog.2023.15>
- Santos Gargallo, I. (1993). *Análisis contrastivo, análisis de errores e interlengua en el marco de la lingüística contrastiva*. Síntesis.
- Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguist*, 11, 1-22. <https://doi.org/10.1017/S0142716400008262>
- Schmitt, J. E., Watts, J., Eliez, S., Warsofsky, I. S., Bellugi, U. y Reiss, A. L. (2001). Corpus callosum morphology of Williams syndrome: relation to genetics and behavior. *Developmental Medicine and Child Neurology*, 43, 155-159. <https://doi.org/10.1017/s0012162201000305>
- Schubert, C. (2009). The genomic basis of the Williams-Beuren syndrome. *Cellular and molecular life sciences: CMLS*, 66(7), 1178-1197. <https://doi.org/10.1007/s00018-008-8401-y>
- Sederias, I., Krakovitch, A., Stojanovik, V. y Zimmerer, V. C. (2024). Overuse of familiar phrases by individuals with Williams syndrome masks differences in language processing. *Journal of Child Language*, 1-15. <https://doi.org/10.1017/S0305000924000436>
- Singer Harris, N. G., Bellugi, U., Bates, E., Jones, W. y Rossen, M. (1997). Contrasting profiles of language development in children with Williams and Down syndromes. *Developmental Neuropsychology*, 13, 345-370. <https://doi.org/10.1080/87565649709540683>
- Stavrakaki, S. (2004). Wh-questions in Greek children with Williams syndrome: a comparison with SLI and normal development. In S. Bartke y J. Sigmüller (Eds.). *Williams Syndrome across Languages* (pp. 295-318). John Benjamins. <https://doi.org/10.1075/lald.36.17sta>
- Stojanovik, V., Perkins, M. y Howard, S. (2004). Williams syndrome and specific language impairment do not support claims for developmental double dissociations and innate modularity. *Journal of Neurolinguistics*, 17(6), 403-424. <https://psycnet.apa.org/doi/10.1016/j.jneuroling.2004.01.002>
- Thomas, M. S. C., Dockrell, J. E., Messer, D., Parmigiani, C., Ansari, D. y Karmiloff-Smith, A. (2007). Speeded naming, frequency and the development of the lexicon in Williams syndrome. *Language and Cognitive Processes*, 21(6), 721-759. <https://doi.org/10.1080/01690960500258528>
- Tomasello, M. (1992). The social bases of language acquisition. *Social development*, 1(1), 67-87. <https://doi.org/10.1111/j.1467-9507.1992.tb00135.x>

Turell, M. (2011). La tasca del lingüista detectiu en casos de detecció de plagi i determinació d'autoria en textos escrits. *Llengua, Societat i Comunicació*, 9, 69-85.

Vázquez, G. (1999). *¿Errores? ¡Sin falta!* Edelsa

Vázquez, G. (2010). Hacia una valoración positiva del concepto de error. *Monográficos MarcoELE*, 11, 164-177.

Vicari, S., Caselli, M. C., Gagliardi, C., Tonucci, F. y Volterra V. (2002) Language acquisition in special populations: a comparison between Down and Williams syndromes. *Neuropsychologia* 40, 2461–2470. [https://doi.org/10.1016/S0028-3932\(02\)00083-0](https://doi.org/10.1016/S0028-3932(02)00083-0)

Volterra, V., Caselli, M. C., Capirci, O., Tonucci, F. y Vicari, S. (2003). Early linguistic abilities of Italian children with Williams Syndrome. *Developmental Neuropsychology*, 23(1-2): 33-58. <https://doi.org/10.1080/87565641.2003.9651886>

Zukowski, A. (2004) Investigating knowledge of complex syntax: insights from experimental studies of Williams syndrome. *Developmental language disorders: From phenotypes to etiologies*, 99–119.



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 101-118
Artículo recibido: 30/07/2024
Artículo aceptado: 25/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/649>
DOI: <https://doi.org/10.58859/rael.v23i1.649>

Investigating recontextualisation processes in scientific digital practices: the SciDis database

Investigando procesos de recontextualización en prácticas científicas digitales: la base de datos SciDis

DANIEL PASCUAL
ANA EUGENIA SANCHO-ORTIZ
UNIVERSIDAD DE ZARAGOZA

Scientific communication is exploiting innovative dynamics of research dissemination and knowledge transfer through digital platforms which allow for the wide circulation of disciplinary information for diversified audiences. Against this backdrop, recontextualisation is a key phenomenon, enhanced by medium-dependent affordances, the combination of semiotic modes and users' discursive choices. In this paper, we present and justify the SciDis Database, collected to analyse this complex phenomenon in various digital scientific dissemination practices, including X/Twitter and Instagram posts, research digests, feature articles, *Ask an Expert* sites, *The Conversation* platform, explainers, and lay summaries. We conceptualise each of them, considering two parameters to frame them: whether they represent web-hosted or social media practices, and whether they entail author-generated or writer-mediated knowledge. Exploitations and applications of the SciDis Database are emphasised in connection with digital discourse analysis and digital pragmatics, multimodally-oriented approaches, knowledge communication frameworks, and ethnographically-informed studies.

Keywords: *digital discursive practices; scientific dissemination; recontextualisation processes; websites; social media.*

La comunicación científica está maximizando nuevas dinámicas de difusión y transferencia de conocimiento mediante plataformas digitales que permiten la amplia circulación de información especializada para audiencias diversificadas. Para ello, la recontextualización es un fenómeno clave, potenciada por las prestaciones del medio digital, la combinación de modos semióticos y la singularidad discursiva de los usuarios. En este artículo, presentamos y justificamos la base de datos SciDis, que representa este complejo fenómeno en diversas prácticas digitales de divulgación científica, tales como *posts* en X/Twitter e Instagram, *research digests*, *feature articles*, sitios web de *Ask an Expert*, la plataforma *The Conversation*, *explainers* y *lay summaries*. Conceptualizamos cada práctica estableciendo dos parámetros: si aparecen alojadas en webs o redes sociales, y si recontextualizan

Citar como: Pascual, D. y Sancho-Ortiz, A.E. (2024). Investigating recontextualisation processes in scientific digital practices: the SciDis database. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 101-118. <https://doi.org/10.58859/rael.v23i1.649>

conocimiento generado o mediado por el autor del texto. Se enfatizan las aplicaciones de la base de datos SciDis para el análisis del discurso digital y la pragmática digital, la multimodalidad, el marco de comunicación del conocimiento y los estudios etnográficos.

Palabras clave: *prácticas discursivas digitales; divulgación científica; procesos de recontextualización; sitios web; redes sociales.*

1. DISSEMINATION AND RECONTEXTUALISATION IN SCIENTIFIC DIGITAL PRACTICES

The innovative ways of making science nowadays are paralleled in innovative ways of communicating science. Scholarly systems and professional communities growingly endorse values such as collaboration, interdisciplinarity, innovation and excellence. This is part of a loop in which scholars and professionals are imbued in demand from institutional policies and social needs, which they try to meet through their work. Science, as a very complex endeavour, has undergone significant changes in the way it is conceptualised. We have clearly moved beyond traditional laboratory work and experiments to a much broader notion that also encompasses processes of knowledge creation, knowledge transfer, participatory culture and movements towards openness. This is translated at present into scientists' efforts to attain a global, fast and effective communication of science, as well.

Accordingly, dissemination processes must be situated within the conceptual model currently predominating in science communication: the Participation Model. This model stresses the urge for knowledge to be understood as a social asset that is democratically enacted through the participation of expert and non-expert social agents (Bucchi & Trench, 2016). In such a context, scientific production is inevitably conceived as initially triggered by social needs and eventually restored to society (Engberg, 2023). Making an impact on society can only be achieved if knowledge is communicated, for which several cumulative phases can be discerned, following Bartling and Friesike (2014). These include, first, the professionalisation of knowledge creation, which is ensured by the value of research; later, attempts to transfer knowledge, which was linked to the emergence of the journal system; and, finally, these days, an unprecedented scientific revolution motivated by the possibilities enabled by the Internet.

Recent research on discourse analysis and communication studies has placed the focus precisely on the role of the digital medium in how science is communicated in different disciplinary fields and through different practices. Some of these approaches have looked into the emergence and reconfiguration of digital genres (Luzón & Pérez-Llantada, 2019), the construction of credibility in knowledge transmission (Bondi & Cacchiani, 2021), the principles of knowledge recontextualisation from philosophical and interdisciplinary perspectives (Kramer & Gottschling, 2021), the recent proliferation of parascientific practices (Mur-Dueñas & Lorés, 2022), and the construction of identity and visibility in research dissemination (Plo-Alstrué & Corona, 2023), to name just a few.

Drawing from many of these assets, it is our main intention to address how scientific dissemination is manifested in scholarly and professional digital practices. In particular, we investigate which specific processes of recontextualisation take place in scientific dissemination to make knowledge accessible for diversified audiences. In understanding these processes, we also aim to uncover how the notions of identity, visibility, self-branding and credibility are foregrounded discursively. As Lorés and Diani contend (2021), digitalised discourse aiming at knowledge dissemination and resulting from novel professional and scholarly practices is shaping and reshaping aspects of language, genre and identity.

“Recontextualisation” serves as a core concept which enables us to look into the relationship between scientific dissemination practices and the particular enactment of experts’ identity and visibility in the environment where these practices are hosted. Recontextualisation implies the dynamic transference and transformation of a text (Linell, 1998; 2001) in the communication between addressers and addressees, operationalised at various levels (cognitive, pragmatic, discursive, linguistic). Put differently, recontextualisation is regarded as “a process by which some part of discourse is extracted from one communicative context and conveyed into another” (Bondi et al., 2015: 2). Recontextualisation often emerges from experts’ interest in catering for knowledge building processes, such as expansion, extension and enhancement (Engberg & Maier, 2015), which may help bridge the potential knowledge asymmetries between them and their audiences (e.g., Bondi et al., 2015; Gotti, 2014; Maier & Engberg 2021).

Consequently, recontextualising scientific information stands as a process whereby users’ knowledge gaps are attested and addressed to mediate comprehension of specialised discourse (Mur-Dueñas, 2024). Research analyses have traditionally linked recontextualisation processes to the rising of practices of “popularisation” (e.g., Calsamiglia & van Dijk, 2004; Gotti, 2014, Valeiras-Jurado & Ruiz-Madrid, 2023), but now a plethora of more complex scientific dissemination practices needs to be researched as a consequence of the sophistication of science production and scientific communication.

These novel and developing practices that foster scientific dissemination bring with them other interrelated discursive phenomena, such as “interdiscursivity” (Fairclough, 1995) and “hybridity” (Bhatia, 2004), which hint at the appropriation of conventions across different discourses. Linell (1998) already pointed at this particularity, signalling that recontextualisation involves a blending of voices in professional discourse. This means that scientific discourse can borrow discursual characteristics from other disciplinary areas, such as corporate discourse, educational discourse, medical discourse, as well as from non-academic, informal discourses.

Moreover, the role of multimodality should be integrated in the discursive exploration of such recontextualisation processes. The “resemiotisation” (Iedema, 2003) of texts brings along a reshaping of such texts from one mode to another through semiotic meaning-making resources. Thus, from this multimodal perspective, another layer should be added to the study of recontextualisation, which is conceived as “moving *meaning material* from one context with its social organisation of participants and its modal ensembles to another, with its different social organisation and modal ensembles” (Bezemer & Kress, 2008: 184). This translates into the adjustment of specialised knowledge to different audiences, both expert and non-expert, by exploiting and combining diverse modes, such as the visual and the aural modes.

In the particular scenario of scientific dissemination carried out by experts, recontextualisation is defined as a multidimensional process by which scientific knowledge is extracted from its original expert source and reshaped discursively, so that it may fit in other (non-) expert textual environments. When this occurs online, the resultant shape is determined by the digital affordances available in the platform where the scientific dissemination practice takes place. The implication is that recontextualisation through digital discursive practices requires the transformation of specialised, expert knowledge beyond linguistic reformulation, with the aim of making it accessible, explanatory and accepted by multiple audiences.

The rest of the paper is organised as follows. In Section 2, we introduce the SciDis (Scientific Dissemination) Database, first, considering a number of variables for the compilation of scientific digital practices and, then, by defining the features of the digital practices included so far. In Section 3, we highlight the applications of the SciDis Database to various strands in Applied Linguistics and Communication Studies. In Section 4, we summarise the main ideas and provide some final remarks.

2. THE SciDis DATABASE

The SciDis Database is a dynamic collection of digital texts that involves updated communicative practices related to scientific dissemination and knowledge transfer. These digital practices respond to today's global urge of transforming science into a public good and transferring scientific knowledge to the wide public. Consequently, the SciDis Database comprises complex discursive phenomena instantiating current scientific dissemination, in general, and the recontextualisation of knowledge, in particular. Departing from their "primary output" (Puschmann, 2015), scientists, researchers and professionals growingly publish their investigations in other channels and media that require new registers, multimodal formats and targeted audiences. To meet these demands, the leverage of digital affordances is essential in scientists' recontextualisation practices to discursively facilitate disciplinary expert knowledge to other expert and non-expert diversified audiences.

Stemming from the rationale of the SciDis Database, in the next subsections we discuss the methodological decisions taken for its compilation and provide an overview of the digital discursive practices it features.

2.1. Methodological decisions for the compilation of the SciDis Database

Salient digital practices harnessed for knowledge communication and scientific dissemination are collected in what we have coined as the SciDis Database. Several criteria were followed in the process of creating the database to ensure methodological consistency and prospective analytical rigour.

A first decision to frame the scope of meaningful scientific practices was to focus on Sustainable Development Goals (SDGs) as a thematic criterion capturing the updated, medium-term interests and concerns of our society. A further step to narrow down our research involved the selection of particular SDGs around which scientific dissemination practices are flourishing. Under the overriding umbrella of Goal no. 17, which emphasises the transversal need to revitalise the global partnership for sustainable development and strengthen its implementation, we chose four specific SDGs. Taking the rationales given to these goals by the United Nations (<https://sdgs.un.org/goals>), our selected goals include Goal no. 3, which refers to good health and well-being and aims to "ensure healthy lives and promote well-being for all at all ages"; Goal no. 7, which is concerned with efficient and clean energy and intends to "ensure access to affordable, reliable, sustainable and modern energy for all"; Goal no. 12, which aspires to "ensure sustainable consumption and production patterns" by promoting responsible behaviours; and Goal no. 13, which is related to environmental action and seeks to "take urgent action to combat climate change and its impacts".

Our contention is that these scientific issues are of utmost relevance in today's professional activities, political agendas and social structures. They trigger an enormous interest on the part of the civil society, whose demands for knowledge consumption need to be met through broad dissemination. *Figure 1* visually represents the consideration of SDGs for our research:

To move from this theoretical layer, we concurred that these SDGs needed to be reflected in specific macro-disciplines and particular sub-topics with high social relevance, which is precisely what drives the interest to produce and disseminate specialised knowledge about them for the wide public. Thus, the SciDis Database comprises practices from three main disciplines: 1) health, 2) economy and 3) natural sciences. Two fine-grained sub-topics tightly linked to the SDGs presented above were pointed out for each of the three disciplines: 1a) physical activity and nutrition and 1b) mental health; 2a) sustainability and 2b) circular economy; and 3a) climate change and 3b) energy efficiency. Although categorical differences between these topics may be sometimes hard to find, this procedure helped us in identifying well-delimited digital discursive practices according to a classification of topics related to scientific dissemination.



Figure 1: SDGs structuring the SciDis Database (United Nations, n.d.)

From these criteria, our next step was to focus on the notions of “dissemination” and “recontextualisation”, which we have deeply discussed in Section 1. The possible knowledge asymmetries among expert and non-expert audiences steer the dissemination efforts and recontextualisation processes in multiple contexts maximised in the digital medium. Subsequently, our iterative exploration of scientific dissemination practices online disclosed some relevant trends. These practices primarily include feature articles, research digests, *Ask an Expert* practices, personal and institutional X/Twitter accounts, lay summaries, explainers, and institutional blogs. Figure 2 is the result of compiling the SciDis Database and displays a map of the digital discursive practices related to scientific dissemination that we have found out as salient:

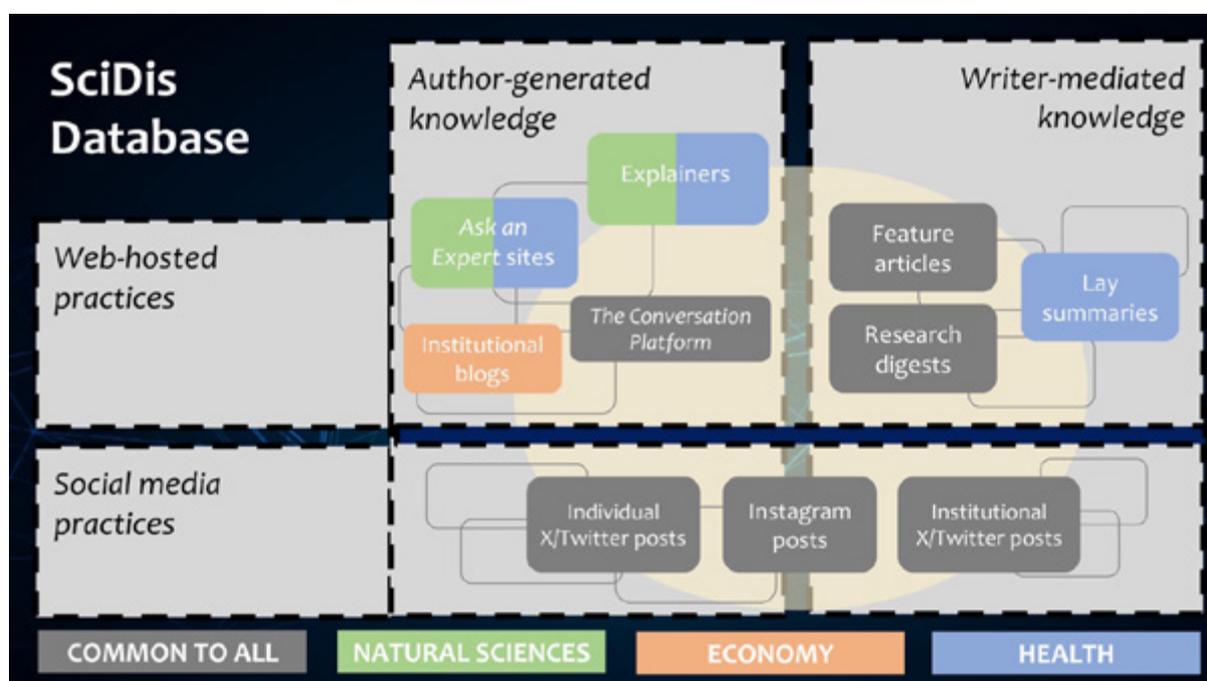


Figure 2: Digital discursive practices and disciplines of scientific knowledge included in the SciDis Database

The selection and classification of these digital scientific practices was structured around two variables which have been determined as relevant after an iterative shaping of the database based on collaborative exploration and observation of selected data. The first variable concerns the typology of the practices based on the platform where they are hosted. This enables us to establish a dichotomy between websites and social media, as well as to capture the complementary realities of both digital practices in knowledge transfer and scientific dissemination. Our starting hypothesis is that, whereas web-hosted practices may fulfil roles of informativity and visibility, social media accounts may help boost interactivity and networking.

As such, web-hosted practices in the SciDis Database are defined as dissemination practices aimed at the construction and circulation of scientific knowledge, and rhetorically and visually organised in a “website hierarchy” (Djonov, 2007), by which webpages are generated as layers of scientific information through which users can navigate. In turn, social media practices are defined as dissemination practices which take place in platforms with unique affordances whose driving force is the promotion of user interaction -in this case, social interaction around scientific knowledge. These include X/Twitter, Instagram, YouTube, TikTok and ResearchGate, among others, and their use has been coined as Social Media for Research Dissemination Purposes (Pascual & Mur-Dueñas, 2022; Mur-Dueñas & Pascual, 2023) and explored by these authors by focusing on X/Twitter and the landscape of international research project communication.

The second variable touches upon the agency of the texts, that is, who is involved in creating the content that shapes the digital discursive practices under scrutiny in our SciDis Database. Two scenarios are contemplated to make it more accurate. A first scenario lies in practices in which the individual user “holding” scientific knowledge produces and distributes the information and data. The source of knowledge and the user disseminating that knowledge coincide and, hence, these are referred to as author-generated practices. A second scenario is represented by practices in which the author of the text partially or completely differs from the source of knowledge and/or the individual user generating that knowledge. This implies cases in which other professionals in communicating scientific knowledge intervene in the creation and publication of content, as well as in which users, based on other authors and/or sources of knowledge, produce their own scientific content. These are considered writer-mediated practices.

Bearing these criteria in mind, we downloaded textual instantiations of the scientific dissemination practices selected for the SciDis Database. Through web scraping techniques and manual compilation, documents containing these texts were stored as one-page documents in PDF format. This decision guaranteed that we captured real-time documents preserving their particular layout at one point, which is helpful for our future analyses and to observe how the scientific dissemination practices compiled evolve. We also transformed the PDF files into Word documents for practical reasons. We proposed a coding system for the identification of digital scientific dissemination practices and, within each of these, for the different sources from which texts are compiled. The addition of metadata was helpful in the organisation of texts within the database, retrieving information about the disciplines and sub-topics featuring the texts, the access route to the original sources, the authorship of the texts, the dates of publication, and further comments on specificities of the digital practice in question.

In our research studies, we resort to a number of software tools that assist us in retrieving numerical, empirical and contextual data. For our qualitative analyses, we rely on NVivo 12 and 14, which is a CAQDAS with which we code our texts through nodes and carry out pragmatic, discursive and multimodal analyses. For our quantitative analyses, we employ corpus-assisted software tools such as SketchEngine, AntConc and Wordsmith Tools that allow us to look into the frequency, distribution and patterning of lexico-grammatical elements.

At this stage of our project, we have already compiled a core corpus that contains 30 texts per each of the digital discursive practices related to scientific dissemination contemplated in the SciDis Database. These are further subdivided into 10 texts per each of the three macro-disciplines, and 5 texts from a different source. This ensured that we counted on two different salient sources of scientific dissemination practices per each of the macro-disciplines. Specific information that is periodically updated can be found at the X website (to be included after revision).

In Subsection 2.2 we conceptualise in further detail the scientific dissemination practices included in the SciDis Database.

2.2. *An overview of digital scientific practices*

A wide variety of practices has been compiled in the SciDis Database to track trends and shifts in current online scientific dissemination, as shown in *Figure 2*. Conceptualisations and analyses of these practices may help understand how new knowledge is communicated and recontextualised to adapt to diversified audiences. Although our focus is on three particular macro-disciplines in which knowledge is recurrently disseminated and recontextualised, we acknowledge that the role of these practices may also be salient in other disciplinary fields. Thus, we broadly define the most representative digital practices within the SciDis Database, which include X/Twitter and Instagram posts, as representatives of social media platforms, as well as research digests, feature articles, *Ask an Expert* websites, knowledge-sharing platforms like *The Conversation* platform, explainers and lay summaries.

As social media practices, X/Twitter and Instagram offer platforms in which both author-generated and writer-mediated practices find room in the pursuit of scientific communication. This means that individual accounts of experts as well as institutional accounts from scientific organisations and research communities managed by a collective writer coexist within the same space. X/Twitter was chosen due to its still prominent position within scholarly contexts and research transfer (Pascual & Mur-Dueñas, 2022; Mur-Dueñas & Pascual, 2023), with the aim of looking into further uses related to knowledge dissemination and recontextualisation processes. In turn, Instagram was added to the SciDis Database in later stages of its compilation and revisitation, since it is gradually gaining ground as another digital reality that should be deemed as significant to disseminate science and distribute specialised knowledge to heterogeneous readers.

“Institutional X/Twitter posts” are published in official profiles managed by institutions, organisations, or entities such as universities, government agencies, corporations, non-profits, or other structured groups. Their accounts serve as “a means for the global transfer of information and the sharing of academic publications” (Sancho-Ortiz, 2024: 135). Through these X/Twitter post, institutions can communicate with a broader audience that includes the public, stakeholders, customers, students, and community members. The content shared in institutional X/Twitter posts typically reflect the organisation’s mission, values, and objectives in a semi-formal register. It may include announcements, news updates, event promotions, research findings, policy changes, and community engagement. Behind the publication of these posts, there is often a specific communication strategy designed to build and maintain relationships with their followers, disseminate important information, answer non-expert questions, promote their brand, engage in broader discussions related to their area of focus, and gather feedback. Overall, institutional X/Twitter accounts may be managed by a team of social media specialists, public relations professionals, or communication experts, who ensure, alongside scientists and researchers, that the posts are consistent with the organisation’s image and messaging guidelines.

“Individual X/Twitter and Instagram posts” are comprised in official profiles managed by individual users who hold academic and professional credentials which ensure the scientific validity of the content provided. They are authored by experts from the academic world such as researchers, university professors and lecturers, as well as experts from outside academia who work as professionals in the private sectors. The posts featured in these individual accounts address a diverse array of topics, ranging from strictly professional to purely personal content, and often mixing both. Professional content refers to topics related to academic and professional practices, which include publishing research articles, participating in conferences, receiving academic prizes, delivering specialised courses, and explaining specialised terminology and concepts, among others. Personal content alludes to aspects of the user’s interests and life, such as their personal status, hobbies and events related to family matters. The combination of professional and personal content in these posts responds to experts’ intention to establish a close relationship with their followers. Individual X/Twitter and Instagram posts often endorse audience engagement strategies intended for expert users to assert their position as professionals, disseminate knowledge and promote their personal brand.

A “research digest” is “a condensed version of a previously published research article” (Lorés, 2024a: 180). In other words, it is a concise summary or collection of summaries that distils and presents key insights from one scientific study or research article on a specific topic. The aim of a research digest is to provide a quick, accessible overview of the latest findings, trends, or developments in a field, allowing readers to stay informed without having to delve into lengthy or highly technical documents. Research digests are commonly used in academic and professional settings to help readers grasp the essence of a body of research, identify emerging patterns, and keep up with new advancements. They typically use clear language and organise information in a way that is easy to understand and navigate. Research digests incorporated into the SciDis Database include sources of information such as [The British Psychological Society](#), [Science Daily](#), and [European Commission](#).

A “feature article” is a comprehensive and in-depth piece of writing typically found in magazines, newspapers, or online publications. Unlike news articles that focus on current events or breaking news, feature articles explore a specific disciplinary topic, issue, person, or event in more detail and from various perspectives. They often use a narrative approach, incorporating storytelling techniques, character development, and detailed descriptions to engage readers. Feature articles may cover a wide range of topics beyond the ones selected for the SciDis Database, such as culture, travel, history, health and social issues. They are designed to be informative and entertaining, offering a deeper understanding of the subject matter. Writers of feature articles often conduct extensive research and interviews to present a well-rounded and compelling narrative. The tone can vary from formal and investigative to personal and reflective, depending on the publication and target audience (Mur-Dueñas, 2024). Feature articles in our SciDis Database have been retrieved from online portals such as [Smithsonian Magazine](#), [Popular Science](#) and [Nature](#). In comparing the rationale of research digests and feature articles, whereas research digests transfer the results of a single research paper, the starting point for a feature article is a topic on which mediators elaborate from various points, bringing multiple expert voices into the text.

“Ask an Expert websites” are online asynchronous platforms where users can ask questions and receive personalised answers from knowledgeable specialists or experts within a particular disciplinary field in a user-friendly, accessible manner (Pascual, 2025). These websites facilitate the exchange of peer or expert support on a wide range of topics, which comprises not only health, economy and natural sciences, but also technology, legal issues, education and psychology. They provide users seeking guidance or clarification with reliable and accurate information, and boost opportunities for affective and cognitive engagement (Ocepek & Westbrook, 2015). Experts who contribute to these websites usually have recognised credentials, qualifications, or substantial experience in their respective fields. They are not only character-

ised by the quality and accuracy of the information provided, but also for their communicative and interactive value (Pounds, 2018). Responses by experts often include practical advice, explanations, or recommendations, and can serve as a valuable resource for individuals seeking formal insights without the need to look into academic research. Examples of sources of *Ask an Expert* websites collected in the SciDis Database comprise [MIT Climate Portal](#), [GoAskAlice](#), [PsychHelp](#), [IDONTMIND](#), [EatFresh](#), and [Child Mind Institute](#).

The Conversation is an independent, non-profit digital space that publishes news and analysis articles written by academic experts and researchers. Prior to their publication, these texts undergo an editing process supported by professionals working in this platform. The content on *The Conversation* aims to provide accessible and reliable information to the public, helping readers understand complex issues from a scholarly perspective. The platform covers a wide range of topics, including science, health, education, environment, politics, technology, and culture. The articles on *The Conversation* are typically based on rigorous academic research and are written in a way that is clear and engaging for a general audience. The platform seeks to bridge the gap between academia and the public, allowing researchers to communicate their insights and expertise directly to a broader audience (Herrando-Rodrigo, 2022). It also aims to promote informed public discourse by providing high-quality, fact-based content that is free from commercial bias. In addition to articles, *The Conversation* often offers commentary, expert opinion, and analysis on current events, making it a valuable resource for anyone interested in gaining a deeper understanding of various topics from credible academic sources. The platform is also open to republishing and sharing content to promote wider dissemination of knowledge. A similar digital space to distribute knowledge to non-academic audiences is [Science Media Centre](#), which has also been considered in subsequent revisions of the SciDis Database.

“Explainers” are informative pieces of content designed to simplify complex concepts, processes, or events for a general audience, providing background information and boosting public outreach (Zou & Hyland, 2024). They aim to break down technical information into clear, understandable terms, for which they can take the formats of written articles, videos, infographics, or interactive multimedia presentations. They are frequently used in journalism, education, marketing, and science communication to help lay people quickly grasp specialised topics without requiring deep prior knowledge. They serve to inform and educate audiences about a wide range of subjects, from news events and scientific breakthroughs to technology trends and public policy issues. Key characteristics of explainers include clarity, as they are written in simple language, avoiding jargon, and focusing on essential points; logical structure, with information organised using headings, bullet points and step-by-step explanations; and engagement, incorporating visuals, animations, or storytelling elements to maintain interest and improve understanding. They are a valuable tool for disseminating information and promoting understanding in an accessible and engaging way, evidencing recontextualisation processes in which technical information stems from other academic sources (Herrando-Rodrigo, 2023). Instantiations of explainers have been retrieved from digital platforms like [The Guardian](#), [Vox](#), [MIT Climate Portal](#), and [Our World in Data](#).

A “lay summary” is a brief and simplified explanation of complex research findings, concepts, or technical details, designed to be easily understood by a general audience who lack expertise in that field. Lay summaries are often used in academic and scientific contexts accompanying traditional, publicly available abstracts in online publications and push the transparency and visibility of a piece of research (Kuehne & Olden, 2015). By avoiding technical jargon and using straightforward language to convey the key points and significance of disciplinary topics, information is made accessible and comprehensible not only to the public and non-specialist professionals, but also to collaborators, policymakers and stakeholders. In connection with the latter, lay summaries are also employed to communicate research findings back to participants in trials and surveys as well as to pitch proposals of investigation to non-expert funding boards

(Nunn & Pinfield, 2014). The production of lay summaries is tightly connected with research in the fields of health and biomedicine, so these texts have been collected from a variety of specialised journals hosted in the online repository *PubMed*.

In future research steps, it is our aim to continue to expand the SciDis Database with other digital practices that may be flourishing as part of experts' ecosystem of digital scientific practices. These could include institutional science blogs, pedagogical science websites, online popularising magazines, crowdfunding websites, TikTok as a Social Medium for Research Dissemination Purposes, science-themed podcasts run by experts to communicate new knowledge, educational YouTube channels devoted to the dissemination of science, and citizen science projects, which directly invite the public to participate in scientific research.

As follows in Section 3, we single out the analytical approaches and the practical applications that our research can unfold stemming from the exploration of the SciDis Database.

3. APPLICATIONS OF THE SCIDIS DATABASE FOR THE ANALYSIS OF DIGITAL RECONTEXTUALISATION

3.1. Research strands for the exploitation of the SciDis Database

As a dynamic compilatory collection of scientific digital texts to identify new discursive trends in online environments, the SciDis Database can be taken as the object for analyses comprising a myriad of discursive perspectives and frameworks. These may include digital discourse analysis, corpus studies, pragmatics, genre studies, metadiscourse or multimodally-driven approaches, among others.

From the conventional perspective of Genre Studies (e.g., Bhatia, 2004; Swales, 2004), the analysis of the texts conforming the database can serve to distinguish between fossilised, hybrid, and newly arising rhetorical structures in digital disseminating discourses. This would facilitate the characterisation of new rhetorical patterns and moves (Swales, 2004), and the study of concepts traditionally applied to academic texts, such as generic hybridisation and genre colonies (Bhatia, 2004, 2006). These generic connections can be further explored thanks to the distinction made in the database upon the typology of the practices and the different set of technological affordances available in the sites and platforms where they are hosted. Specifically, insightful remarks could be drawn as regards intertextuality and hypertextuality as means for knowledge recontextualisation, particularly focusing on the alternation of “reading and navigation modes” (Askehave & Nielsen, 2005) and on the role of “interactivity” (Adami, 2015) in digital media.

A more thorough look into the SciDis Database can also involve investigating the discursive differences inherent in the role of the expert as either knowledge-mediator or knowledge-generator. The database could be approached from metadiscursive perspectives, shedding light on concepts such as authorial stance (Hyland, 2005; Mur-Dueñas, 2011) and interpersonality (Lorés-Sanz, Mur-Dueñas & Lafuente-Millán, 2010), and refining conventional frameworks and categories in light of the digital medium. Moreover, pragmatic perspectives could be undertaken from various standpoints. Departing from speech-act, relevance-theoretic and politeness theories, our analyses may help uncover new patterns in the evolution of pragmatics online (Scott, 2022). Particularly, contributions to the field can be made by looking into data-driven analyses of pragmatic strategies that are salient in situated communicative contexts (Pascual, 2023). Pragmatic studies may likewise prove helpful in making connections with experts' construal of their identity and credibility (Sancho-Ortiz, 2025).

The SciDis Database could also be approached from a multimodal analytical perspective, as it is made of texts displaying multimodally complex discursive phenomena. Analyses could explore aspects like the combinations of semiotic resources and the degree of modal density

used to generate and transfer new scientific meanings making them comprehensible to multiple audiences. Thus, a multimodal approach to processes of “knowledge extension and enhancement” (Engberg & Maier, 2015) would contribute to the exploration of issues particular to digital communication such as “context collapse” and the need to differentiate between the intended and the imagined audience (Marwick & Boyd, 2011).

The study of these phenomena could also be enriched with the implementation of corpus-assisted methods such as collocation, frequency, concordance, and keyword analysis, as these would unveil recurrent linguistic-discursive patterns in the recontextualisation of expert knowledge. Therefore, quantitative findings in these corpus-based analyses could be complementary to qualitative data obtained through pragmatic, generic and multimodal approaches to eventually identify expert identity types for each of the variables included in the database (i.e., platform typology, that is, websites or social media, and agency of the text, that is, expert or mediator roles).

Furthermore, from the perspective of Knowledge Communication approaches (e.g., Engberg, 2023; Engberg, Fage-Butler & Kastberg, 2023; Engberg & Maier, 2015; Kastberg, 2010, 2019), the analysis of the SciDis Database could intend to “investigat[e] the intentional communication of specialised knowledge in professional settings performed as part of the practice of being an expert” (Engberg, 2023: 152). Departing from the understanding of knowledge as a social phenomenon, the database could be exploited to explore emerging communicative dynamics that concern both expert-to-expert and expert-to-non-expert knowledge communication. This includes the exploration of how knowledge is discursively recontextualised when catering for the various knowledge processing needs of an audience with diverse levels of expertise. Eventually, the study of the database from this perspective could cast some light on general knowledge processing differences between specialised and lay audiences, and on the blurring limits between knowledge-related notions such as dissemination and popularisation.

The analyses of the SciDis database presented above could be complemented with ethnographically-oriented methods involving both expert and non-expert participants. This would include, for example, the distribution of surveys and questionnaires among specialised expert groups to gain insights into their reliance and involvement in digitally-mediated scholarly and scientific practices to reach varied audiences. Similarly, semi-structured interviews and focus groups with both experts and non-experts could be conducted to assess their perceptions on the relevance of transferring knowledge through digital platforms and gather information on newly emerging disseminating practices that could be added to the SciDis Database. These ethnographic methods overall facilitate the adoption of a combined emic-etic perspective (Lillis, 2008) which demonstrate the consistency of the SciDis Database as a contribution to current discussions on the democratisation of science, particularly within the framework of the Participation Model as the currently governing trend in scientific production and communication.

3.2. Research approximations to the analysis of the SciDis Database

A major motivation driving the compilation of the SciDis Database is to foster its exploitation and reproduction as a means to explore practices of knowledge recontextualisation and scientific dissemination in digital environments. Bearing this in mind, some questions are proposed to apply the frameworks explained in Section 3.1 according to two variables. Text-oriented questions tackle processes of recontextualisation affecting the textual nature, structure and layout characterising the instantiations of the digital practices in the SciDis Database. Discourse-oriented questions deal with a more fine-grained level in which pragmatic and lexico-grammatical aspects of recontextualised texts are researched. They revolve around notions such as intention, terminology, evaluation, and engagement.

Table 1: Guiding questions for the exploration of knowledge recontextualisation in digital scientific dissemination practices in the SciDis Database

Selection and analysis of digital scientific dissemination practices and recontextualisation processes		
Text-oriented	<ol style="list-style-type: none"> 1. Is there any specific textual relationship between the recontextualised digital text and the original source text? <hr/> 2. Are there explicit intertextual connections in the recontextualised digital text? <hr/> 3. Are there any interdiscursive patterns in the recontextualised digital text? <hr/> 4. Are there differences in the rhetorical structures of the recontextualised digital text and the source text? <hr/> 5. Does the multimodal layout of the source text change to accommodate the recontextualised digital text? <hr/> 6. Do the recontextualised digital text and the source text present different degrees of modal density? 	
Discourse-oriented	<ol style="list-style-type: none"> 7. Does the authorial voice in the recontextualised digital text influence how knowledge is distributed? <hr/> 8. Are the author's intentions made self-evident in the processes of recontextualisation? <hr/> 9. Does evaluation play a significant role in how knowledge is conveyed in the recontextualised digital text? <hr/> 10. Are pragmatic strategies deployed to fulfil novel communicative goals in the recontextualised digital text? <hr/> 11. Is specialised terminology explained to make disciplinary knowledge accessible in the recontextualised digital text? <hr/> 12. Does the recontextualised digital text show differences in the engagement cues employed to involve the readership? 	

These questions have already guided pieces of research which adopt some of the theoretical perspectives outlined in Section 3.1 to explore the digital practices of science dissemination and knowledge recontextualisation comprised in the SciDis Database. Some of the research conducted so far includes analyses on knowledge comprehensibility in web-hosted practices of author-generated knowledge (namely, *The Conversation*) and writer-mediated knowledge (feature articles and research digests), approached through the study of verbal and non-verbal explanatory strategies like elaboration, explicitation, exemplification, enumeration, comparison/analogy, spatial organisation and visual representations (Mur-Deñás 2024). Similarly, studies have also concentrated on specific dissemination practices, for instance, online research digests, exploring the introduction of (meta)discoursal features in varied semiotic and hyper-textual formats to attain comprehensibility, credibility and engagement (Lorés 2023), the use of interactive metadiscourse markers as resources for comprehensibility (Lorés 2024a), and the exploitation of expert-oriented and audience-oriented pragmatic strategies to achieve a balance between authority enactment and knowledge accessibility in the recontextualisation process (Lorés 2024b).

These pieces of text-oriented and discourse-oriented research underpin the applications of the SciDis Database to work alongside those sectors of society interested in and benefited by scientific advancements. Future applications are already being considered from a pedagogical standpoint, looking into the potential exploitation of disseminating practices such as pedagog-

ical science websites as means for the teaching and learning of English as a Foreign Language (EFL) and into the replication of effective practices analysed for English for Specific Purposes (ESP). This pedagogical approach to the database would effectively enhance its alignment with the promotion of the Sustainable Development Goals given the current emphasis placed within current pedagogical trends on the development of students' socio-civic competence and sense of citizenship as key 21st century skills (Binkley et al., 2012).

Lastly, as applications of our work, the fact that the SciDis Database covers a diverse array of disseminating practices motivates us to share it as a public resource repository in the near future. We believe that this database constitutes a comprehensive illustration of up-to-date effective knowledge disseminating practices that might serve as a replicable model for other experts to contribute to the promotion of open science, and hence, its democratisation. Thus, we aim to draw on the results from our multi-layered analyses to design materials, such as video pills, infographics and deliverables. Additionally, we intend to offer training workshops for experts in which we can summarise and transfer our main findings on dissemination practices and they can learn how to integrate them when communicating online.

4. CONCLUSIONS

The SciDis Database stands as a contribution to the understanding of how scientific dissemination and knowledge recontextualisation operate nowadays in digital environments. This paper has provided a description of the database, including the methodological criteria established for its compilation as an encompassing, yet dynamic, representation of science dissemination practices online. Thematically speaking, the database collects digital disseminating practices from three scientific disciplines which recurrently trigger social interest and are openly subject to discussion: health, economy, and natural sciences. The focus on these disciplines demonstrates our intention to ponder issues that arise high social concern from a perspective that is aligned with Sustainable Development Goals (SDGs) proposed by the United Nations.

In this article we have also explained the two variables around which the SciDis Database is structured. First, we have taken a look at the typology of the practices based on the site in which they are hosted, resulting in a mutually-exclusive dichotomy between web-hosted and social media practices. Second, we have deemed the role of the expert as determinant in the transfer process, either as a generator or a mediator of knowledge. Author-generated knowledge has been argued to refer to cases in which experts are the authors of the final products in which recontextualisation processes can be observed, while writer-mediated knowledge has been claimed to involve an external agent other than the author in the circulation of scientific knowledge. The two variables and the practices included in the SciDis Database are believed to account for the different communicative scenarios afforded by the most popular digital platforms currently used by experts.

Moreover, an overview of the scientific dissemination practices conforming the database according to the aforementioned variables has been given. We have also highlighted other emerging practices we would like to explore in future research to keep on feeding the database, as it is our contention that it should be regarded as a living ecosystem in permanent development. Some examples are pedagogical science websites, science-themed podcasts, TikTok as a Social Medium for Research Dissemination Purposes, and online popularising magazines. Thus, we conceive the SciDis Database as a faithful representation of scientific dissemination digital practices to be updated as disseminating trends keep developing further.

Future applications for the database have also been outlined, with a focus on its potential to enrich the research carried out from discourse-oriented perspectives and frameworks. Specifically, our emphasis has been on the utility of the database to address, in an integrated manner, newly emerging discursive phenomena, such as genre hybridisation and intertextuality, authori-

al stance and online pragmatics, context collapse and expert identity, and intersemiosis, among others. Similarly, we have stressed the functionality of the database to explore the development of science dissemination dynamics from non-discursive perspectives, as are the framework of Knowledge Communication and ethnographic methods. In order to illustrate how some of these theoretical and analytical tenets can be specifically approached, we have offered a list of guiding questions to drive prospective analyses. It is our hope that these analytical cues may help other researchers face the study of recontextualisation processes within and outside the field of scientific dissemination.

The variability of the potential uses of the SciDis Database demonstrates its functionality to approach scientific dissemination and knowledge transfer as dynamic endeavours. Overall, we aim to contribute to the open discussion and circulation of scientific knowledge by compiling and analysing a database which comprises examples of current and trendy disseminating practices. Our ultimate goal is to make this database and its future analytical results accessible to other expert and non-expert communities through our website (<https://intergedi.unizar.es/>). Scientists, scholars and experts from diverse disciplinary fields may benefit from the findings of our research, and may raise awareness of the digital practices pinpointed in this paper in relation to specific analyses. On the whole, the SciDis Database may assist interested professionals in endorsing informed digital scientific dissemination practices, as well as users in consuming scientific content in more critical ways.

ACKNOWLEDGEMENTS

We are tremendously grateful to Dr Lorés and Dr Mur-Dueñas for their valuable feedback in structuring this paper and improving its quality through their accurate comments. This piece of research is a contribution to the InterGedi research group (<https://intergedi.unizar.es/>). Grant PID2021-122303NB-100, funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”, as well as by Gobierno de Aragón (H16_23), Spain.

REFERENCES

- Adami, E. (2015). What's in a click? A social semiotic framework for the multimodal analysis of website interactivity. *Visual Communication*, 14(2), 133-153. <https://doi.org/10.1177/1470357214565583>
- Askehave, I. & Nielsen, A. E. (2005). Digital genres: A challenge to traditional genre theory. *Information Technology & People*, 18(2), 120-141. <https://doi.org/10.1108/09593840510601504>
- Bartling, S. & Friesike, S. (2014). Towards another scientific revolution. In S. Bartling & S. Friesike (Eds.), *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing* (pp. 3-15). Cham: Springer. https://doi.org/10.1007/978-3-319-00026-8_1
- Bezemer, J. & Kress, G. (2008). Writing in multimodal texts: A social semiotic account of designs for learning. *Written Communication*, 25(2), 166-195. <https://doi.org/10.1177/0741088307313177>
- Bhatia, V. K. (2004). *Worlds of Written Discourse: A Genre-Based View*. New York: Continuum.

Binkley, M., Ertad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M. & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 17-66). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-2324-5_2

Bondi, M., Cacchiani, S. & Mazzi, D. (Eds.). (2015). *Discourse in and through the Media: Recontextualizing and Reconceptualizing Expert Discourse*. Newcastle upon Tyne: Cambridge Scholars.

Bondi, M. & Cacchiani, S. (2021). Knowledge communication and knowledge dissemination in a digital world. *Journal of Pragmatics*, 186, 117-123. <https://doi.org/10.1016/j.pragma.2021.10.003>

Bucchi, M. & Trench, B. (Eds.). (2016). *Handbook of Public Communication of Science and Technology*. London: Routledge. <https://doi.org/10.4324/9780203483794>

Calsamiglia, H., & van Dijk T. A. (2004). Popularization discourse and knowledge about the genome. *Discourse & Society*, 15(4), 369-389. <https://doi.org/10.1177/0957926504043705>

Djonov, E. (2007). Website hierarchy and the interaction between content organization, webpage and navigation design: A systemic functional hypermedia discourse analysis perspective. *Information Design Journal*, 15(2), 144-162. <https://doi.org/10.1075/idj.15.2.07djo>

Engberg, J. & Maier, C. D. (2015). Exploring the hypermodal communication of academic knowledge beyond generic structures. In M. Bondi, S. Cacchiani & D. Mazzi (Eds.), *Discourse in and through the media: Recontextualizing and Reconceptualizing Expert Discourse* (pp. 46-65). Newcastle upon Tyne: Cambridge Scholars.

Engberg, J. (2023). Between infotainment and citizen science: Degrees of intended non-expert participation through knowledge communication. In R. Plo-Alastrué & I. Corona (Eds.), *Digital Scientific Communication: Identity and Visibility in Research Dissemination* (pp. 149-170). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-38207-9_7

Engberg, J., Fage-Butler, A. & Kastberg, P. (Eds.). (2023). *Perspectives on Knowledge Communication: Concepts and Settings*. New York: Routledge. <https://doi.org/10.4324/9781003285120>

Fairclough, N. (1995). *Critical Discourse Analysis: The Critical Study of Language*. New York: Longman.

Gotti, M. (2014). Reformulation and recontextualization in popularization discourse. *Ibérica*, 27, 15-34.

Herrando-Rodrigo, I. (2022). “Our editors work with academics and researchers to bring you stories no one else in the media can”: Credibility building in The Conversation.com. Paper presented at CMC2022: 9th Conference on Computer-Mediated Communication (CMC) and Social Media Corpora. Santiago de Compostela, Galicia, Spain. September, 25–29. <https://doi.org/10.5281/zenodo.8064594>

Herrando-Rodrigo, I. (2023). Exploring the genre and discourse practices of web-text explainers. Paper presented at AELFE-LSPPC International Conference 2023. Zaragoza, Zaragoza, Spain. June, 28–30. <https://doi.org/10.5281/zenodo.8087846>

- Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- Iedema, R. (2003) Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual Communication*, 2, 29-57. <http://dx.doi.org/10.1177/1470357203002001751>
- Kastberg, P. (2010). Knowledge communication: Formative ideas and research impetus. *Programmatic Perspectives*, 2(1), 59-71.
- Kastberg, P. (2019). *Knowledge Communication: Contours of a Research Agenda*. Berlin: Frank & Timme.
- Kramer, O. & Gottschling, M. (2021). *Recontextualized Knowledge: Rhetoric – Situation – Science Communication*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110676310>
- Kuehne, L. M. & Olden, J. D. (2015). Lay summaries needed to enhance science communication. *Proceedings of the National Academy of Sciences*, 112(12), 3585-3586. <https://doi.org/10.1073/pnas.1500882112>
- Lillis, T. (2008). Ethnography as method, methodology, and “deep theorizing”: Closing the gap between text and context in academic writing research. *Written Communication*, 25(3), 353-388. <https://doi.org/10.1177/0741088308319229>
- Linell, P. (1998). Discourse across boundaries: On recontextualizations and the blending of voices in professional discourse. *Text & Talk*, 18(2), 143-157. <https://doi.org/10.1515/text.1.1998.18.2.143>
- Linell, P. (2001). *Approaching Dialogue: Talk, Interaction and Context in Dialogical Perspectives*. Amsterdam: John Benjamins.
- Lorés-Sanz, R., Mur-Dueñas, P. & Lafuente-Millán, E. (Eds.). (2010). *Constructing Interpersonality: Multiple Perspectives on Written Academic Genres*. Newcastle upon Tyne: Cambridge Scholars.
- Lorés, R. (2023). Dual voices, hybrid identities: The recontextualization of research in digital dissemination scientific discourse. *Círculo de Lingüística Aplicada a la Comunicación*, 93, 69-84. <https://doi.org/10.5209/clac.85566>
- Lorés, R. (2024a). Digesting psychology: Metadiscourse as a recontextualising tool in the digital communication of disciplinary research. *Chinese Journal of Applied Linguistics*, 47(2), 178-195. <https://doi.org/10.1515/CJAL-2024-0202>
- Lorés, R. (2024b). Mediating expert knowledge: The use of pragmatic strategies in digital research digests. *Journal of Pragmatics* 232, 26-40. <https://doi.org/10.1016/j.pragma.2024.08.007>
- Lorés, R. & Diani, G. (2021). Disseminating knowledge: The effects of digitalised academic discourse on language, genre and identity. *European Journal of English Studies*, 25(3), 249-258. <https://doi.org/10.1080/13825577.2021.1988262>
- Luzón, M. J. & Pérez-Llantada, C. (Eds.). (2019). *Science Communication on the Internet: Old Genres Meet New Genres*. <https://doi.org/10.1075/pbns.308>

- Maier, C. D. & Engberg, J. (2021). Harvard Business Review's reframing of digital communication: From professional expertise to practical guidance. *Journal of Pragmatics*, 176, 186-197. <https://doi.org/10.1016/j.pragma.2021.02.005>
- Marwick, A. E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114-133. <https://doi.org/10.1177/1461444810365313>
- Mur-Dueñas, P. (2011). An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. *Journal of Pragmatics*, 43, 3068-3079. <https://doi.org/10.1016/j.pragma.2011.05.002>
- Mur-Dueñas, P. (2024). Digital dissemination practices: An analysis of explanatory strategies in the process of recontextualising specialised knowledge. *Discourse & Interaction*, 17(1), 94-114. <https://doi.org/10.5817/DI2024-1-94>
- Mur-Dueñas, P. & Lorés, R. (Eds.). (2022). *Scientific and Parascientific Practices*. Basel: MDPI. <https://doi.org/10.3390/books978-3-0365-3822-8>
- Mur-Dueñas, P. & Pascual, D. (2023). Metadiscoursal realisation of pragmatic strategies @ ResearchProject Twitter accounts. *Atlantis: Journal of the Spanish Association of English Studies*, 45(2), 119-150. <https://doi.org/10.28914/Atlantis-2023-45.2.06>
- Nunn, E. & Pinfield, S. (2014). Lay summaries of open access journal articles: Engaging with the general public on medical research. *Learned Publishing*, 27(3), 173-184. <https://doi.org/10.1087/20140303>
- Ocepek, M. G. & Westbrook, L. (2015). Question, answer, compare: A cross-category comparison of answers on question and answer websites. *New Review of Hypermedia and Multimedia*, 21(3-4), 212-226. <https://doi.org/10.1080/13614568.2015.1036134>
- Pascual, D. (2023). Scientific dissemination and professional practices through digital media: The study of pragmatic strategies in the communication of international research projects. Unpublished Doctoral Dissertation. Universidad de Zaragoza.
- Pascual, D. (2025). Dialogic markers in *Ask an Expert* webpages on environmental discourse. *Language & Dialogue*, in press. <https://doi.org/10.1075/ld.00191.pas>
- Pascual, D. & Mur-Dueñas, P. (2022). Dialogic interaction with diversified audiences in Twitter for Research Dissemination Purposes. *Círculo de Lingüística Aplicada a la Comunicación*, 90, 61-79. <https://doi.org/10.5209/clac.81307>
- Plo-Alastrué, R. & Corona, I. (Eds.). (2023). *Digital Scientific Communication: Identity and Visibility in Research Dissemination*. Cham: Palgrave Macmillan. <https://doi.org/10.1007/978-3-031-38207-9>
- Pounds, G. (2018). Patient-centred communication in ask-the-expert healthcare websites. *Applied Linguistics*, 39(2), 117-134. <https://doi.org/10.1093/applin/amv073>
- Puschmann, C. (2015). A digital mob in the ivory tower? Context collapse in scholarly communication online. In M. Bondi, S. Cacchiani & D. Mazzi (Eds.), *Discourse in and through the Media: Recontextualizing and Reconceptualizing Expert Discourse* (pp. 22-45). Newcastle upon Tyne: Cambridge Scholars.

Sancho-Ortiz, A. E. (2024). Analysing the functionality of Twitter for science dissemination in EFL teaching and learning. *PROFILE: Issues in Teachers' Professional Development*, 26(1), 133-147. <https://doi.org/10.15446/profile.v26n1.105685>

Sancho-Ortiz, A.E. (2025). Environmental discourse on social media: Exploring engagement in X/Twitter for environmental purposes. *Language & Dialogue*, in press. <https://doi.org/10.1075/ld.00189.san>

Scott, K. (2022). *Pragmatics Online*. New York: Routledge. <https://doi.org/10.4324/b22750>

Swales, J. (2004). *Research Genres: Exploration and Application*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524827>

United Nations. (n.d.). Sustainable Development Goals. Retrieved 9 May, 2024 from <https://sdgs.un.org/goals>

Valeiras-Jurado, J. & Ruiz-Madrid, N. (2023). The influence of discipline, medium and target audience in multimodal recontextualization practices: The case of popular science online videos. *Ibérica*, 46, 237-269. <https://doi.org/10.17398/2340-2784.46.237>

Zou, H. J. & Hyland, K. 2024. "Let's start with the basics of the virus": Engaging the public in two forms of explainers. *Journal of English for Academic Purposes*, 68, 101353. <https://doi.org/10.1016/j.jeap.2024.101353>



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 119-135
Artículo recibido: 31/07/2024
Artículo aceptado: 21/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/650>
DOI: <https://doi.org/10.58859/rael.v23i1.650>

Expresión e interacción orales y autoevaluación: estudio descriptivo y de intervención con hablantes avanzados de español como lengua adicional

Speaking and Self-assessment: a Descriptive and Intervention Study with Advanced Speakers of Spanish as an Additional Language

MARTA F. NOGUEROLES LÓPEZ
UNIVERSIDAD DE ALCALÁ

El presente artículo recoge dos estudios que, conjuntamente, abordan el papel de la autoevaluación en la expresión e interacción orales en español como lengua adicional. El primero de ellos es un estudio de intervención centrado en el desarrollo de la autoevaluación sobre la competencia oral de los participantes (diez hablantes de español como lengua adicional de nivel avanzado); mientras que el segundo constituye un estudio descriptivo focalizado en la expresión e interacción orales a través de la técnica de recuerdo estimulado. Los objetivos que nos planteamos son, en primer lugar, analizar los efectos de la intervención implementada sobre la capacidad de expresión e interacción orales en español de los participantes; en segundo lugar, examinar la autoevaluación de los participantes en relación con su capacidad de expresión e interacción orales en español y, por último, conocer su valoración respecto a los protocolos de autoevaluación implementados. Los resultados ponen de manifiesto el impacto positivo de la autoevaluación para la expresión e interacción orales en lengua adicional. Además, el estudio arroja luz sobre las fortalezas y debilidades de los participantes como hablantes de español y da cuenta de la buena acogida del protocolo por parte de los estudiantes.

Palabras clave: *Autoevaluación; expresión e interacción orales; intervención; técnica de recuerdo estimulado.*

The present article embraces two studies which, together, address the role of self-assessment in speaking in Spanish as an additional language. The first is an intervention study based on the development of self-assessment on the oral competence of the participants (ten advanced speakers of Spanish as an additional language); meanwhile the second is a descriptive study of the participants' speaking skills through the technique of stimulated recall. Our objectives were, first, to analyze the effects of the implemented intervention on the participants' speaking in Spanish; second, to learn about the participants' self-assessment of their speaking skills in Spanish; and finally, to examine the participants' evaluation of the implemented self-assessment protocols. The results show the positive impact of self-assessment on the participants' speaking skills. In addition, the study sheds light on the speakers' strengths and weaknesses and shows that the protocol was well received by the students.

Key words: *Self-assessment; oral expression and interaction; intervention; stimulated recall technique.*

Citar como: Nogueroles López, M.F. (2024). Expresión e interacción orales y autoevaluación: estudio descriptivo y de intervención con hablantes avanzados de español como lengua adicional. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 119-135. <https://doi.org/10.58859/rael.v23i1.650>

1. INTRODUCCIÓN

La interacción oral se caracteriza por la presencia y la participación simultánea de los interlocutores, así como por la relación interpersonal que estos activan, construyen y negocian (Calsamiglia y Tusón, 2008). La interacción oral facilita el desarrollo de la lengua adicional (Keck, Iberri-Shea, Tracy-Ventura y Wa-Mbaleka, 2006; Mackey y Goo, 2007) y, además, constituye una actividad de la lengua fundamental en la comunicación.

Algunos estudios previos (Keck et al., 2006; Mackey y Goo, 2007; Lyster y Saito, 2010) indican que la interacción juega un papel facilitador en el desarrollo de la lengua adicional (a partir de ahora, LA) en tanto en cuanto posibilita a los aprendientes la producción y la recepción de información entrante comprensible y de retroalimentación interactiva.

En el presente estudio, los resultados de un cuestionario de análisis de necesidades, administrado entre los estudiantes de un curso universitario de lengua española de nivel avanzado, pusieron de manifiesto que la expresión e interacción orales eran las actividades comunicativas de la lengua en la que los participantes decían tener menor competencia.

Con base en esta necesidad verbalizada por los propios estudiantes de mejorar su expresión e interacción orales, se decidió, en primer lugar, implementar una intervención longitudinal basada en la promoción de la autoevaluación y medir sus efectos sobre la expresión e interacción orales y, en segundo lugar, realizar un estudio descriptivo focalizado de autoinforme retrospectivo a través de la técnica de recuerdo estimulado centrado en la autovaloración por parte de los participantes de su capacidad de hablar e interactuar en español.

La intervención didáctica implementada se ha centrado en el desarrollo de la autoevaluación, entendida como el acto de monitorizar los propios procesos y productos para hacer ajustes y adaptaciones que profundicen el aprendizaje y mejoren el desempeño (Andrade, 2019). El uso de la autoevaluación favorece la participación activa por parte de los estudiantes en reflexiones y discusiones sobre su desempeño en el aprendizaje, y, con ello, promueve la autonomía del estudiantado. Algunos estudios previos señalan que la autoevaluación puede ayudar a los estudiantes a: examinar su desempeño de aprendizaje de acuerdo con criterios impuestos o desarrollados conjuntamente por aprendientes y docentes (Chen, 2008); desarrollar habilidades de pensamiento crítico y lograr un aprendizaje profundo (Cassidy, 2006); continuar con su aprendizaje con una mayor comprensión de cómo pueden hacerlo (Boud, 1995); aumentar la seguridad y la autosuficiencia con respecto a su desempeño (Boud, 1995); asumir plena responsabilidad sobre sus decisiones y acciones (Boud, 1995); tomar conciencia sobre su éxito en el aprendizaje (Gardner, 2000); estar más motivados y ser más autónomos en su aprendizaje (Gardner, 2000); desarrollar su autorregulación (Zimmerman, 2002).

Por otro lado, en el estudio descriptivo recogido en estas páginas, hemos puesto en práctica la técnica de recuerdo estimulado. Dicha técnica constituye un método de informe verbal en el que se les pide a los participantes que verbalicen las acciones físicas y mentales que experimentan a medida que se ven o escuchan a sí mismos realizando una tarea comunicativa que sirve como estímulo para evocar sus recuerdos sobre la misma (Mackey y Gass, 2005). A través de esta técnica, la experiencia o acontecimiento de aprendizaje se graba y, en un momento posterior y casi siempre también en otro lugar, se muestra a los protagonistas con el fin de estimular el recuerdo sobre lo que pasaba por sus mentes en el momento en que estaban implicados en dicha experiencia (Cohen y Macaro, 2010). La idea subyacente a esta técnica es que la visualización del evento estimula el recuerdo hasta el punto de que los participantes puedan recuperar y verbalizar lo que estaba ocurriendo en sus mentes durante el evento (Gass y Mackey, 2000). La recuperación puede ser consecutiva (inmediatamente después del momento en el que se hace uso de la lengua), retrasada (menos de una semana después del momento) o no reciente (más de una semana desde el instante de uso). Si bien la calidad del recuerdo inevitablemente sufre con

el lapso del tiempo, según Ericsson (2002), el recuerdo estimulado es la técnica menos reactiva de todas las técnicas introspectivas, ya que los procesos de pensamiento específicos no se ven afectados por el procedimiento de la técnica.

A partir del estudio de intervención y del estudio descriptivo, en el presente trabajo nos planteamos, en primer lugar, analizar los efectos que tiene la intervención implementada sobre la capacidad de los participantes para hablar e interactuar en español; en segundo lugar, examinar cómo se autoevalúan los informantes como hablantes de español y, por último, conocer la valoración de los distintos procedimientos de autoevaluación por parte de los participantes.

2. ESTADO DE LA CUESTIÓN

2.1. *Efectos de la autoevaluación sobre el aprendizaje y uso de la lengua*

En el ámbito del aprendizaje de segundas lenguas en general, Butler y Lee (2010) llevaron a cabo un estudio de intervención dirigido a examinar la efectividad de la autoevaluación por parte de estudiantes coreanos de inglés como LA durante un semestre. En este estudio, se les administró una serie de actividades de autoevaluación, diseñadas específicamente para cada unidad del plan de estudios, así como una actividad de autoevaluación general que se introdujo con fines sumativos. Los resultados cuantitativos ponen de manifiesto algunos efectos positivos de la autoevaluación en el rendimiento en inglés de los estudiantes, así como en su confianza en el aprendizaje de la lengua meta. Los datos también apuntan que los estudiantes mejoraron su capacidad para autoevaluar su desempeño con el tiempo.

Por otro lado, Winke, Zhang y Pierce (2023), con base en la creencia de que la autoevaluación podía ser válida para medir la evolución de la competencia oral de estudiantes universitarios de diferentes niveles curriculares, exploraron la relación entre la autoevaluación y las calificaciones de las pruebas de competencia de 807 estudiantes de español como lengua extranjera que realizaron un examen asistido por ordenador. Con ello, descubrieron que era más probable que un estudiante pasara a subsecciones de autoevaluación más altas si tenía una calificación mayor en la Entrevista de la Competencia Oral a través del ordenador (ACTFL – computadora). Por último, los autores consideran que las autoevaluaciones adaptativas por ordenador son apropiadas para medir el dominio de la L2, especialmente porque son rentables, tienen un sentido intuitivo y promueven el aprendizaje de la L2.

2.2. *Autoevaluación de procesos y estrategias de competencia oral*

De Saint Léger (2009) investigó la evolución de la percepción de los estudiantes sobre sí mismos como hablantes de una segunda lengua (L2) durante un período de 12 semanas. Los participantes autoevaluaron sus habilidades para hablar y su nivel de participación en tareas orales en francés en tres momentos distintos del curso y, en consecuencia, establecieron metas de aprendizaje. La autopercepción evolucionó de manera positiva con el tiempo, particularmente en relación con la fluidez, el vocabulario y la confianza general para hablar en la L2. Además, el establecimiento de metas individuales alentó al estudiantado a asumir una mayor responsabilidad hacia su propio aprendizaje, aunque esta mayor conciencia no siempre se tradujo en acciones concretas para mejorar el aprendizaje.

En un trabajo conjunto, De Saint Léger y Storch (2009) investigaron las percepciones de los participantes sobre su competencia oral, sobre sus contribuciones a las actividades orales de la clase (debates de toda la clase y en grupos pequeños), así como sobre sus actitudes hacia estas actividades, y cómo tales percepciones y actitudes influyeron en su disposición para comunicarse en francés L2. Los datos se recogieron, fundamentalmente, mediante cuestionarios

de autoevaluación. Tales cuestionarios animaban a los estudiantes, por un lado, a reflexionar sobre su entorno de aprendizaje inmediato en varios momentos del semestre y, por otro, a autoevaluar sus habilidades para hablar. Este estudio concluyó que, en general, a medida que los aprendientes ganaban confianza en sí mismos, aumentaba su disposición para comunicarse en francés en clase.

2.3. *Valoración de la intervención*

Si bien la investigación en el aprendizaje de lenguas adicionales se ha centrado principalmente en la orientación evaluativa de la autoevaluación, los investigadores están prestando cada vez más atención a la evaluación formativa o evaluación para el aprendizaje (Butler, 2016), en la que la evaluación se considera parte de la instrucción, por lo que está integrada en todos los aspectos de la enseñanza y el aprendizaje (Black, Harrison, Lee, Marshall y Wiliam, 2003). En términos generales, las investigaciones realizadas con estudiantes universitarios indican que estos comprendieron el objetivo de la autoevaluación (Ratminingsih, Marhaeni y Vigayanti, 2018) y, en su mayoría, la consideraron valiosa para guiar tanto la evaluación como la revisión de su trabajo (Micán y Medina, 2017). Además, los estudios indican que la autoevaluación ayudó a los participantes a entender cómo responsabilizarse de su propio aprendizaje (Bourke, 2014), así como a desarrollar su pensamiento crítico (Siow, 2015). A través de la autoevaluación, también lograron aplicar nuevas habilidades adquiridas (Murakami, Valvona y Broudy, 2012) y promovieron su aprendizaje autorregulado, ya que este les facilitó el establecimiento de objetivos, la planificación, el autocontrol y la reflexión (Wang, 2017).

De entre los estudios que se han centrado en analizar la percepción de los estudiantes con respecto a la autoevaluación, Hanrahan e Isaacs (2001) realizaron un análisis de las opiniones de 233 estudiantes que acababan de experimentar la autoevaluación y la evaluación por pares como parte de sus asignaturas. Los resultados muestran que los estudiantes sintieron que, a pesar de que no estaban capacitados para evaluarse a sí mismos, se beneficiaron de la intervención. El análisis de los beneficios y los aspectos problemáticos de la autoevaluación reveló ocho dimensiones generales: Dificil, Mejor comprensión de la calificación, Incómodo, Productivo, Problemas de aplicación, Leer el trabajo de los demás, Desarrollar la empatía, Motivación. Dado su carácter cualitativo, el estudio de Hanrahan e Isaacs (2001) ofrece una imagen más detallada de los beneficios e inconvenientes que los estudiantes percibieron en cuanto a la autoevaluación.

3. ESTUDIO

3.1. *Participantes y contexto*

Los participantes son diez estudiantes (seis mujeres y cuatro hombres) del último curso de comunicación oral de la Licenciatura de Lengua Española y Culturas Hispánicas de la Universidad de Copenhague. Todos los estudiantes matriculados en dicho curso participaron en el estudio, pues, tal y como indican Black et al. (2003), este fue incluido como parte del programa del curso. La asignación de los participantes, por tanto, es la propia de una clase intacta en la que todo el grupo participó en la actividad propuesta.

La lengua de comunicación habitual entre los participantes es el danés, pues es la lengua materna (L1) del 80% de ellos. Una estudiante tiene el alemán como lengua materna y otra es bilingüe en español y alemán. De este modo, todos los estudiantes tienen una lengua germánica como L1 y, además, todos poseen competencia avanzada en inglés. Muchos de ellos, asimismo, hablan otras lenguas extranjeras, por lo que se trata de un grupo altamente plurilingüe. Aunque

existen diferencias en cuanto al nivel de competencia en español de los estudiantes, todos ellos poseen un nivel avanzado-superior de español (B2-C1, según el Consejo de Europa, 2002), pues todos eran recién graduados del *Bachelor i spansk og latinamerikansk sprog og kultur* de la Universidad de Copenhague, que se corresponde con un grado o primer ciclo de estudios universitarios, de tres años de duración. En la Universidad de Copenhague, el *Bachelor i spansk og latinamerikansk sprog og kultur* se puede complementar con el programa de *Kandidat i spansk og latinamerikansk sprog og kultur*, que es el programa que estaban cursando los participantes del presente estudio y que constituye un segundo ciclo de grado o licenciatura.

Al inicio del curso de licenciatura, se administró un cuestionario de análisis de necesidades con el fin de averiguar, entre otros aspectos, la percepción de dificultad de los estudiantes con respecto a las distintas actividades comunicativas y conocimientos de la lengua. El análisis de las respuestas a dicho cuestionario, tal y como muestra la *Figura 1*, pone de manifiesto que la expresión e interacción orales eran la actividad comunicativa de la lengua en la que los participantes decían tener menos competencia.

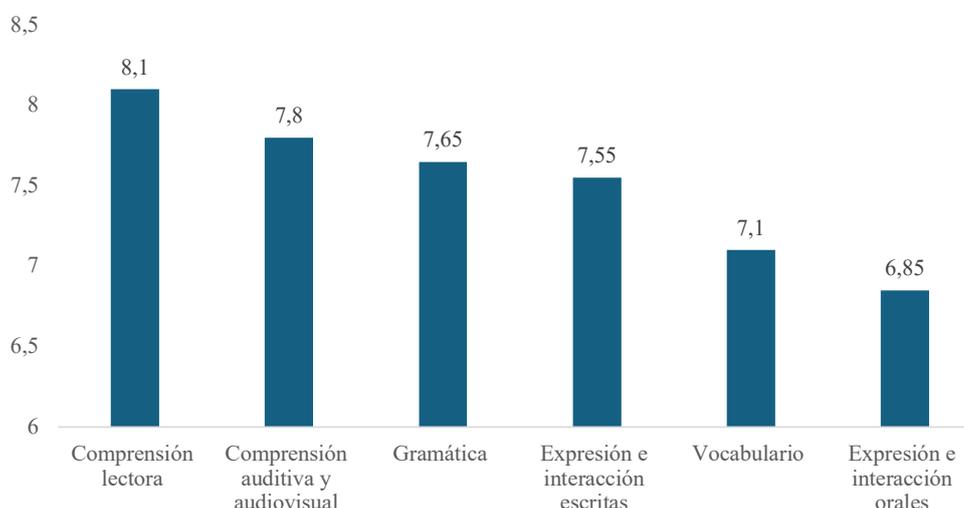


Figura 1: Análisis de necesidades: promedio de resultados de la percepción del nivel/competencia (1-10) por actividad de la lengua

El presente estudio se llevó a cabo en un curso de producción e interacción orales de nivel avanzado, que se desarrolló a lo largo de un cuatrimestre, de 14 semanas de duración, durante dos horas semanales de clase.

3.2. *Objetivos*

En este trabajo, partimos, por un lado, de los resultados del análisis de necesidades que identifica la expresión e interacción orales como las actividades comunicativas de la lengua en la que los participantes perciben tener una menor competencia y, por otro lado, de la hipótesis de que la toma de consciencia por parte de los estudiantes sobre sus fortalezas y debilidades como hablantes e interlocutores de español como LA revertirá positivamente en su competencia comunicativa oral en lengua meta. El objetivo general es explorar la autoevaluación de los participantes sobre su capacidad de expresión e interacción orales en español como LA. Para cumplir dicho objetivo, hemos llevado a cabo, en primer lugar, un estudio de intervención basado en la promoción de la autoevaluación aplicada a la expresión e interacción orales en español

y, en segundo lugar, un estudio descriptivo focalizado para conocer las fortalezas y debilidades en materia de competencia oral por parte de los hablantes de español como LA. A partir del objetivo expuesto, formulamos las siguientes preguntas de investigación:

- P1. ¿Qué efectos tiene la instrucción implementada sobre la expresión e interacción orales en español como LA de los participantes?
- P2. ¿Cómo se autoevalúan los participantes como hablantes en español?, ¿qué fortalezas y debilidades identifican?, ¿qué propuestas de mejora proponen?
- P3. ¿Cómo valoran los participantes los distintos procedimientos de autoevaluación?

3.3. *Diseño y procedimiento*

La primera pregunta responde a un diseño preexperimental de corte longitudinal, con un grupo experimental, en el que las variables dependientes están representadas por la calificación en la prueba inicial y en la prueba final de expresión e interacción orales; mientras que la variable independiente viene determinada por la instrucción en expresión e interacción orales basada en la autoevaluación. Las dos últimas preguntas responden a un diseño descriptivo focalizado. Los instrumentos de recogida de datos que se han utilizado son los siguientes: dos pruebas de expresión e interacción orales, una inicial y otra final; transcripciones de las entrevistas; cuestionario de valoración del estudio descriptivo; diario de la profesora. Los resultados, tanto cuantitativos como cualitativos, derivados de los mencionados instrumentos nos ayudan a dar respuesta a las preguntas de investigación planteadas.

3.3.1. *Estudio de intervención*

El presente artículo recoge un estudio de intervención que se enmarca en la tradición de investigación en el aula conocida como investigación-acción, en la que los profesores investigan ciertos aspectos de su propia práctica docente con el fin de cambiar y mejorar lo que está aconteciendo en clase (Nunan y Bailey, 2009).

La intervención o instrucción didáctica se implementó durante las 14 semanas de duración del cuatrimestre universitario y consistió en integrar la autoevaluación en el proceso educativo de manera que formara parte de la metodología regular del curso, tanto en las clases como en el trabajo fuera de ellas.

El curso estaba articulado en torno a diez temas de actualidad relacionados con las culturas hispanas, uno por sesión. Algunos ejemplos de los temas sobre los que se organizó el curso eran el proceso de paz en Colombia, la situación de los hispanos en los Estados Unidos o la des/igualdad de género en los países de habla hispana. Dichos temas fueron seleccionados por la investigadora a partir de las propuestas de los estudiantes en función del interés generado y la susceptibilidad para argumentar al respecto. Puesto que los estudiantes confesaron no tener predilección por ningún tema en especial, la asignación temática se hizo mediante un sorteo. La metodología de trabajo para cada sesión consistía en la presentación oral del tema por parte del estudiante correspondiente y la posterior discusión/conversación sobre diferentes aspectos del mismo, primero, en parejas y, después, en grupos de tres o cuatro. A esta discusión en dobles parejas, le seguía una puesta en común en clase de todos los aspectos comentados. La duración de cada sesión semanal era de dos horas.

Cuando un estudiante adoptaba el papel de presentador, además de ser evaluado por sus compañeros de clase, asumía la tarea de reflexionar y autoevaluar su propia práctica mediante la compleción (la misma tarde de la presentación) de la hoja de autoevaluación (véase el Apéndice 1). Dicho documento está compuesto por diez preguntas abiertas que se centran en la valoración de los siguientes aspectos: estado anímico y emocional en el momento de la presentación; familiaridad y control del tema; preparación y práctica; propia actuación con

respecto a la estructura y claridad de las ideas presentadas, pronunciación, entonación y fluidez, corrección lingüística y vocabulario; puntos fuertes; puntos débiles; aspectos a mejorar y nivel de satisfacción. Por último, los estudiantes deben valorar su presentación asignándose una nota del 1 al 10 (siendo 10 la nota más alta), así como reflexionar sobre los aspectos en los que, en general, deberían concentrarse para mejorar su expresión e interacción orales en español. Una vez completado el documento de autoevaluación, el estudiante, en su papel de presentador, se lo hacía llegar a la profesora por correo electrónico. Por su parte, la docente valoraba la actuación tanto individual del presentador/a como grupal de todo el estudiantado en la discusión mediante la anotación y explicación de valoraciones y comentarios. En el caso de los comentarios referentes a las presentaciones orales, la docente se los hacía llegar al presentador correspondiente vía correo electrónico después de la clase. Al respecto, cabe considerar que la retroalimentación es mucho más efectiva si se proporciona inmediatamente después o cerca del momento de la actuación a la que se refiere, pues, en ese momento, el estudiante es plenamente consciente de su práctica y, en palabras de Dörnyei (2001), tiene una consciencia en curso de su propio progreso. Por lo que se refiere a la discusión de los estudiantes, la profesora expone y explica los comentarios y valoraciones a toda la clase inmediatamente después del final de la discusión. Es importante señalar que la docente no solamente comentaba los errores o aspectos susceptibles de mejora, sino que también enfatizaba los recursos lingüísticos y las estrategias que habían sido utilizados de manera adecuada con el fin de hacerlos conscientes de sus buenas prácticas y de su progreso a lo largo del curso.

Antes y después de la instrucción, los estudiantes llevaron a cabo una prueba de expresión e interacción orales que consistía en exponer oralmente sobre uno de los temas abordados en la asignatura en la que se enmarca el presente estudio, así como mantener una conversación, de índole argumentativa, sobre el tema. Para seleccionar el tema, cada estudiante debía escoger, al azar, dos bolas que contenían un número que se correspondía con uno de los temas y, de entre los dos temas obtenidos al azar, debía seleccionar uno de ellos. Antes de iniciar la exposición oral, el estudiante contaba con cuatro minutos de preparación en los que contaba con una hoja en blanco, por si quería tomar notas, y un diccionario para consultar dudas terminológicas en caso necesario. La duración de cada prueba oral fue de 15 minutos. El mismo tipo de prueba oral se realizó al inicio y al final de la intervención didáctica con el fin de registrar la competencia oral previa y posterior a la instrucción implementada. Ambas pruebas se realizaron fuera de clase en sesiones individuales con cada participante.

3.3.2. *Estudio descriptivo*

Tal y como se ha adelantado y con el fin de dar respuesta a los objetivos planteados, el presente trabajo incluye un estudio descriptivo focalizado que implica el uso de la técnica de recuerdo estimulado dirigido a que los estudiantes reflexionaran sobre sus fortalezas y debilidades como hablantes en español como LA. A través de la técnica de recuerdo estimulado, los estudiantes, en parejas, se grabaron manteniendo una conversación informal en español, transcribieron dicha conversación y acudieron a una entrevista con la profesora en la que visualizaron conjuntamente el vídeo de la conversación y lo comentaron.

El estudio descriptivo se implementó a lo largo de la quinta y sexta semana del curso, cuando los estudiantes ya estaban familiarizados con la reflexión sobre la propia actuación en actividades comunicativas orales en español. El recuerdo estimulado fue la técnica utilizada en la fase descriptiva. A través de esta técnica de informe verbal, los estudiantes, en parejas, se grabaron manteniendo una conversación informal en español, transcribieron dicha conversación y acudieron a una entrevista con la profesora en la que verbalizaron sus procesos de pensamiento durante la conversación, cuyo recuerdo fue estimulado a partir de la visualización del vídeo. Con base en la relevancia del diseño metodológico en los estudios que emplean la técnica de

recuerdo estimulado (Mackey y Gass, 2005: 78), el presente estudio presta especial atención a la metodología implementada. En este sentido, el procedimiento del estudio descriptivo consta de cinco pasos:

1) *Clara explicación del procedimiento por parte de la profesora:*

Tabla 1: Muestra de la parte introductoria del documento de trabajo de los participantes

OBJETIVO: Reflexionar sobre la propia competencia en producción e interacción orales en español (identificar estrategias, fortalezas y debilidades) con el fin de sugerir propuestas de mejora con el fin último de mejorar como hablantes de español.

PROCEDIMIENTO:

- Grabación de una conversación
 - Transcripción de la conversación
 - Visualización pausada y análisis de la conversación. Podéis detener la reproducción en cualquier momento para comentar dificultades, estrategias, sentimientos, etc.
 - Autoevaluación de la propia actuación e identificación de aspectos susceptibles de mejora
 - Propuestas de mejora
 - Evaluación de la actividad
-

2) *Desarrollo y grabación de la conversación:*

Los estudiantes se agrupan en parejas fuera del aula y llevan a cabo una conversación de, aproximadamente, diez minutos de duración, que graban en vídeo para su posterior visualización. La conversación, de temática libre, debía llevarse a cabo no más de cinco días antes de la fecha de entrevista acordada.

3) *Transcripción de la conversación por los propios estudiantes y envío a la profesora:*

Cada pareja de estudiantes escucha la grabación de su propia conversación y la transcribe. Esta acción favorece la toma de consciencia por parte de los participantes de sus fortalezas y debilidades en cuanto a la expresión e interacción orales.

Tabla 2: Muestra de transcripción de la conversación entre dos estudiantes

INVESTIGADORA: Del 1 al 10, ¿qué nota te darías en esta actividad?, ¿por qué?

ESTUDIANTE G: Yo creo que 4, porque estoy pensando demasiado y, no sé, tengo que trabajar mucho o practicar, no tener estas pausas muy largas, pero, por otro lado, mmmm... uso las frases como “distinto punto de vista”, “a mi juicio”.

INVESTIGADORA: Dices que piensas mucho, ¿en qué piensas?

ESTUDIANTE G: En la gramática, en el vocabulario, en el vocabulario en este contexto, muchas cosas.

INVESTIGADORA: ¿Y tú, ESTUDIANTE D?

ESTUDIANTE D: No sé, 6, quizás, porque veo que me he mejorado desde comparado con el verano, desde junio, por ejemplo, antes de ir a Andalucía. Cometo menos errores que había esperado y hablo con más confianza que había esperado. Pero había demasiadas pausas.

INVESTIGADORA: ¿Con qué dificultades te encontraste?

ESTUDIANTE G: Sí, y cuando quiero expresar una frase, por ejemplo, y no tengo la palabra, por ejemplo, es más fácil no decir nada. Y me gusta tu estrategia de expresarlo en otra manera, sin la palabra.

4) *Técnica de recuerdo estimulado durante entrevistas:*

El recuerdo estimulado se desarrolló durante una entrevista con la investigadora en su despacho, un máximo de 5 días después del momento en que los estudiantes habían mantenido la conversación para evitar, como advierten Cohen y Scott (1996), el deterioro de la memoria en la actividad de autoinforme. Puesto que las conversaciones se mantuvieron en parejas, los participantes acudieron a la entrevista también en parejas, por lo que se llevaron a cabo un total de cinco entrevistas de un promedio de 75 minutos de duración cada una.

Durante la entrevista, la profesora e investigadora explicó los objetivos de la entrevista, así como el procedimiento que se iba a seguir en el transcurso de esta. La docente se cercioró de que lo habían entendido y, en el único caso que se planteó una duda, la resolvió. Con el fin de romper el hielo y de reducir la ansiedad de los estudiantes, la profesora formuló algunas preguntas personales, seguidas de cuestiones más generales sobre la trayectoria de cada estudiante en el aprendizaje del español. Trascurrida esta fase inicial, los dos estudiantes junto con la investigadora visualizaron el vídeo de la conversación de forma pausada, de manera que tanto la profesora como los estudiantes podían detener la reproducción en cualquier momento para verbalizar oralmente cuestiones referentes a los procesos de pensamiento que estaban experimentando en el momento de la conversación. Como ejemplo ilustrativo, observamos cómo la estudiante G reflexiona sobre su fluidez en la expresión oral: “Estoy pensando mucho, mucho, creo que es mi problema, que estoy pensando demasiado antes de simplemente hablar.”, o cómo la estudiante A razona sobre su riqueza léxica y sobre una posible falta de confianza sobre su propia competencia: “Aunque estaba diciendo ANTIGUO, *empieza a decir VIEJO, quizás por la influencia de F que decía esto.”. En otro ejemplo, vemos al estudiante C reflexionando sobre sus fortalezas: “Pues, mis fuertes ..., me siento bastante cómodo hablando español.”

Una vez visualizada y comentada toda la conversación, los estudiantes, guiados por la investigadora, autoevaluaban su propia competencia en producción e interacción orales (en materia de fortalezas, debilidades y propuestas de mejora), así como el protocolo en sí mismo.

Tabla 3: Muestra literal de valoración del protocolo

ESTUDIANTE A: A mí me sorprendía que muchas veces si no recuerdo la palabra intento llegar a la palabra de otro modo o explicar entonces tú dices Ah EXPERIENCIA.
INVESTIGADORA: ¿No eras consciente de eso?
ESTUDIANTE A: No tanto, pero lo hice varias veces en el vídeo.
INVESTIGADORA: Muy bien. Entonces, ¿crees que te ha hecho más consciente de cómo hablas?
ESTUDIANTE A: Sí.

4. RESULTADOS

4.1. *Efectos de la instrucción sobre la competencia en expresión e interacción orales*

Para dar respuesta a la pregunta 1, mediante la herramienta Jamovi (2020), llevamos a cabo una prueba T de muestras relacionadas con el fin de averiguar si existen diferencias significativas entre la nota conseguida en la prueba de expresión e interacción orales inicial y la obtenida en la final.

Tabla 4: *Resultados prueba T de muestras relacionadas. Calificaciones iniciales y finales*

Paired Samples T-Test			statistic	df	p
Pretest	Posttest	Student's t	-19,0	8,00	<0,001

La Tabla 4 pone de manifiesto que sí existen diferencias significativas entre las notas iniciales y finales.

4.2. Resultados de la autoevaluación

Con el fin de dar respuesta a la segunda pregunta de investigación, analizamos y categorizamos la transcripción de las entrevistas mantenidas entre los participantes y la profesora. De acuerdo con el objetivo del presente trabajo, los datos procedentes de la autoevaluación se analizaron en función de si los participantes los consideraban como dificultades, como fortalezas o como propuestas de mejora. En cada una de estas tres grandes categorías, y tras la lectura repetida de las transcripciones, se optó por clasificar los resultados según los cuatro aspectos más influyentes en la comunicación oral que apunta el Instituto Cervantes en la *Guía del examen DELE CI* (Instituto Cervantes, 2014): coherencia, fluidez y pronunciación, corrección lingüística y alcance léxico. A estas cuatro categorías se les sumaron otras tres que se corresponden con los tres grandes grupos de estrategias (a saber, metacognitivas, cognitivas y socioafectivas), que identifican O'Malley y Chamot (1990) en su conocida clasificación.

El análisis descriptivo focalizado de las transcripciones de las entrevistas desvela que el número de dificultades verbalizadas (total: 56) supera el número de fortalezas (total: 39) y de propuestas de mejora (total: 12). No obstante, considerando la tendencia natural de priorizar la valoración negativa en comparación con la positiva (Pérez Morales, 2007), consideramos que el número de fortalezas verbalizado es muy positivo.

Si analizamos las categorías de forma mixta, comprobamos que, por lo que respecta al componente fónico, las dificultades verbalizadas son la falta de fluidez (6), las dificultades en la pronunciación (4), la articulación poco marcada o murmullo (2), la influencia de la L1 en la entonación (1) y la excesiva velocidad del habla. De estos datos se desprende la necesidad de los participantes de mejorar la fluidez y la pronunciación, aspectos no siempre desarrollados en las clases regulares de español como LA. Por lo que se refiere a las fortalezas fónicas, la subcategoría con mayor puntuación es, paradójicamente, la fluidez (4), seguida de la imitación del acento del interlocutor (3) y la pronunciación (2). En este caso, al comparar las fortalezas con las dificultades en cuanto a fluidez y pronunciación, encontramos que los alumnos que no tienen problemas de fluidez consideran dicha fluidez como su fortaleza, lo cual pone de manifiesto las diferencias individuales en un mismo grupo, incluso cuando el número de estudiantes es muy reducido. Con respecto a las propuestas de mejora en el plano fónico, solo encontramos una, la del participante E, que considera la reducción de la velocidad del habla como su primordial propuesta de mejora. En el caso específico de este participante, la actividad de autoevaluación fue absolutamente reveladora en tanto en cuanto, al transcribirse y entrevistarse con la profesora, se percató de que hablaba en exceso y con muchísima rapidez, según él, por encontrarse nervioso, lo cual dificultaba enormemente la comprensión por parte de los demás compañeros.

En relación con la coherencia, el 60% de los participantes asegura tener problemas de interacción, especialmente porque están más centrados en la formulación y la codificación de lo que van a decir a continuación que en la interpretación de las palabras de su compañero/a. Esta es una de las dificultades más notables de la interacción oral, ya que, como recuerdan Goh y Vandergrift (2012), la interacción oral exige que el oyente responda de forma inmediata a las palabras del hablante, por lo que este debe preparar y formular una respuesta al mismo tiempo

que presta atención y procesa el discurso del interlocutor, lo cual supone una carga cognitiva considerable. Asimismo, los participantes también aseguran tener dificultades para tomar y ceder los turnos conversacionales (2), cumplir con el principio de cortesía (1), terminar las frases cuando se desconoce el léxico (2), así como para no desviarse del hilo de la conversación e irse por las ramas (1). En cuanto a las fortalezas relacionadas con esta categoría de coherencia, solo aparece la cualidad de mantener la conversación viva a pesar de la falta de vocabulario. Por lo que se refiere a las propuestas de mejora, dos participantes se proponen usar más conectores, especialmente focalizadores del discurso con el fin de organizar su discurso oral.

La corrección lingüística a la hora de expresarse oralmente es otro aspecto en el que casi la mitad de los estudiantes (4) asegura tener dificultades y, por ello, propone prestar más atención a la gramática mientras habla, aunque sea a costa de perder cierta fluidez. Por otro lado, la otra mitad de los estudiantes considera que la corrección gramatical es uno de sus puntos fuertes. Estos datos ponen de manifiesto la adecuación de la autoevaluación para reflexionar sobre las propias dificultades y proponer soluciones o propuestas de mejora adecuadas en base a ellas. La información extraída de estos resultados, además, es extremadamente valiosa para los docentes, ya que dan a conocer la realidad individual de cada estudiante.

Por lo que se refiere al léxico, la mitad de los participantes asegura carecer del vocabulario necesario para poder expresar sus ideas. Uno de los estudiantes añade que experimenta cierta interferencia léxica entre las lenguas que habla, de manera que, cuando en el proceso de codificación, está intentando encontrar la palabra correcta en español, en muchas ocasiones, le viene a la cabeza la palabra en otra de sus lenguas extranjeras aprendidas más tarde, ya que, como asegura, el inglés, al ser una lengua estudiada y usada durante muchos años, se encuentra en un casillero aparte en su mente, mientras que el español comparte casillero con otras dos lenguas que ha aprendido de forma tardía. Según De Houwer (2018), esta diferencia en el momento de exposición y de adquisición a otra lengua marca la diferencia en la trayectoria de aprendizaje temprano de la lengua de los niños pequeños. De hecho, la edad o momento de adquisición no solo puede afectar la competencia en esa lengua adquirida, sino también el vínculo emocional que se establece con ella. Ante estas dificultades léxicas, el 20% de los estudiantes se propone leer y escribir para enriquecer su vocabulario, mientras que el 10% se plantea empezar a emplear prefabricados léxico-gramaticales con el fin de descargar la memoria de trabajo y agilizar el proceso de formulación lingüística.

Además de los aspectos lingüístico-comunicativos, los participantes también informaron sobre cuestiones estratégicas, que pasamos a comentar siguiendo la clasificación tripartita de O'Malley y Chamot (1990), que distingue entre estrategias metacognitivas, cognitivas y socio-afectivas. Por lo que se refiere a la metacognición, el 30% de los estudiantes asegura no haber planificado ni organizado su discurso lo suficiente, mientras que el 10% afirma exactamente lo opuesto, es decir, haber planificado en exceso, de manera que su producción carecía, en cierta manera, de la espontaneidad esperada. En cuanto a las fortalezas metacognitivas, el 20% apunta la autocorrección, en tanto en cuanto se corrige cuando percibe que ha cometido un error; mientras que el 10% se considera consciente de su progreso. De entre las propuestas de mejora de índole metacognitivo, destaca la de una mayor planificación del discurso (10%).

Por lo que respecta a las acciones cognitivas, el 30% de los participantes asegura utilizar coletillas o rellenos, del tipo “Mmmm”, con el fin de ganar tiempo para poder pensar en su respuesta. En otras palabras, los participantes emplean rellenos para contar con más tiempo para conceptualizar y formular lo que quieren decir. El 20% de los participantes asegura emplear la estrategia cognitiva de traducción. En cuanto a la valoración de esta última estrategia por parte de los usuarios, la mitad considera que la excesiva traducción ralentiza en exceso su proceso de codificación y de comprensión; mientras que la otra mitad opina que la traducción agiliza la formulación en la expresión, así como la construcción de significado en la comprensión. Otro 20% hace uso de la inferencia y valora esta estrategia como beneficiosa para la comprensión

auditiva, así como para el parafraseo dirigido a minimizar los efectos de las limitaciones léxicas. Finalmente, el 10% afirma hacer uso del lenguaje corporal, fundamentalmente gestos y movimientos de las manos, para facilitar la comprensión de su interlocutor.

En cuanto a lo afectivo, el 10% admite haberse medido la temperatura emocional y haberse sentido excesivamente nervioso/a y, por ello, se propone intentar reducir la ansiedad en el futuro. El 20%, por otra parte, se siente muy seguro/a de sí mismo, lo cual lo considera como una fortaleza. Por lo que se refiere a lo social, el 30% apunta haber copiado los errores de los compañeros; por ejemplo, un estudiante expresa la palabra “invertir” correctamente y, al escuchar que su compañero se refiere a esa acción como “invertar”, el primero, por la falta de confianza en sí mismo, termina por utilizar “invertar” de forma equivocada. Las fortalezas sociales más informadas son la búsqueda de oportunidades de práctica oral (50%) y la cooperación entre compañeros (40%).

4.3. *Valoración del protocolo*

Los datos cuantitativos extraídos de la parte del cuestionario de evaluación final apuntan que los estudiantes valoran la efectividad de la intervención basada en la autoevaluación con un promedio de 8.14 sobre 10. Todos los estudiantes coinciden con Cameron (2001) en que la autoevaluación promueve la toma de consciencia sobre aspectos tales como el propio proceso de aprendizaje, la propia competencia, la habilidad de producir mensajes orales en español y la actuación en la presentación oral realizada en clase. Una estudiante afirma que, aunque ella siempre acostumbra a autoevaluarse después de una presentación, en este caso, hacerlo por escrito le ha permitido analizar más detalladamente los distintos aspectos de su actuación y, por ello, le ha resultado muy útil. Otra estudiante añade que ha aprendido mucho al autoevaluarse. Como aspectos negativos, una estudiante destaca la cantidad de tiempo invertido en este proceso; mientras que otro señala que, dado que él se autoevalúa mentalmente con frecuencia, en su caso, la redacción de dicha reflexión resultaba innecesaria.

Con respecto al estudio descriptivo, fundamentado en la aplicación de la técnica de recuerdo estimulado, el promedio de valoración de los estudiantes es de 9,35 sobre 10, lo cual demuestra la amplia satisfacción de los estudiantes con esta técnica. De entre los múltiples comentarios positivos de los alumnos referidos a esta actividad, destacan su utilidad, pertinencia, validez como métodos de reflexión e, incluso, una participante la califica de “imprescindible”. De todo el protocolo, los estudiantes destacan como especialmente útiles la fase de visionado y transcripción de las conversaciones, así como las entrevistas con la investigadora. La fase que menos les agradó, *a priori*, fue la grabación de la conversación, pues les resultaba extraño saber que estaban siendo grabados.

5. DISCUSIÓN

Con respecto a los efectos de la instrucción sobre la competencia en expresión e interacción orales, nuestros resultados (véase la *Tabla 4*) ponen de manifiesto que sí existen diferencias significativas entre las notas iniciales y finales. De estos datos, podemos deducir que la intervención implementada ha sido beneficiosa para la competencia oral de los participantes. No obstante, cabe considerar que la práctica oral propuesta sobre la que los alumnos han reflexionado (y que no puede ser desligada de la autoevaluación) también puede haber repercutido favorablemente en la competencia oral de los participantes. Nuestros datos, aunque más cuantitativos, están en consonancia con los de De Saint Léger (2009), quien halló efectos positivos de la autoevaluación sobre la fluidez percibida, el vocabulario, la confianza y el sentido de responsabilidad hacia

su propio aprendizaje. De manera similar, de De Saint Léger y Storch (2009) encontraron que la autoevaluación tenía una influencia positiva en la disposición de los estudiantes adultos para comunicarse en la lengua meta.

Por otro lado, del análisis cualitativo se desprende que el estudio descriptivo implementado favorece la evaluación y la toma de consciencia no solo de las dificultades y problemas, sino también de las fortalezas y de las propuestas de mejora; de manera que parece desarrollar la estrategia metacognitiva de evaluación en todas sus variantes. Cuando los estudiantes valoran sus actuaciones, normalmente de manera informal, suelen centrarse fundamentalmente en las dificultades y problemas que les han surgido. El hecho de que la instrucción aquí presentada haya propiciado la valoración de fortalezas es especialmente beneficioso en tanto en cuanto les ofrece una imagen mucho más precisa de su estado, les hace conscientes de su progreso, lo cual repercute positivamente en su motivación, así como en su autoestima, ya que repara también en los aspectos positivos. Por otro lado, la instrucción incide no solo sobre aspectos lingüísticos y comunicativos, como son el alcance léxico o la corrección gramatical, sino también sobre cuestiones metacognitivas, cognitivas y socioafectivas.

Por último, los resultados de los cuestionarios de evaluación, tanto del estudio de intervención como del estudio descriptivo, ponen de manifiesto la estupenda valoración de las técnicas e instrumentos propuestos. Por lo que acontece a la valoración de la intervención, nuestros resultados están en concordancia con los de De Saint Léger (2009), quien, tras analizar la valoración de la instrucción por parte de los participantes, concluyó que todos los estudiantes entrevistados percibieron el proceso general de autoevaluación como una oportunidad útil para reflexionar sobre sí mismos y para tomar consciencia de sus patrones de aprendizaje, comportamiento, y logros lingüísticos. Con relación al estudio descriptivo, los resultados informados por los estudiantes a partir de la técnica de recuerdo estimulado coinciden con los datos resultantes del análisis de las conversaciones y las transcripciones de estas por parte de la investigadora. Tales coincidencias ponen de manifiesto la validez de la técnica como herramienta de autoevaluación.

A la luz de estos resultados, nos atrevemos a concluir, igual que lo hacen Hanrahan e Isaacs (2001), que los participantes en el presente estudio se beneficiaron de las técnicas de autoevaluación presentadas.

6. CONCLUSIONES

A pesar del reducido número de participantes, el presente estudio nos ha permitido diseñar, implementar y medir los efectos en la expresión e interacción orales de una intervención didáctica basada en la autoevaluación dirigida a un grupo de estudiantes avanzados de español como lengua adicional.

Aunque el número de participantes no nos permite generalizar, los resultados sugieren que, a través de la autoevaluación, los estudiantes toman consciencia sobre sus fortalezas y debilidades como hablantes y, en general, como aprendientes de español como LA. En este sentido, la implementación de actividades de autoevaluación puede ser muy beneficiosa en cursos con estudiantes con diferentes niveles o necesidades, para que cada uno se conciencie de su situación y pueda sacar el máximo provecho al curso en función de sus fortalezas y debilidades individuales.

En este orden de cosas, los resultados también parecen apuntar que la técnica de recuerdo estimulado puede ser especialmente provechosa en cursos universitarios para que, a la vez que los estudiantes reflexionan sobre su propia competencia, se familiaricen con técnicas de investigación en lingüística aplicada.

REFERENCIAS

- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Front. Educ.*, 27. <https://doi.org/10.3389/feduc.2019.00087>
- Black, P., Harrison, C., Lee, C., Marshall, B. y Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Nueva York: Open University Press.
- Boud, D. (1995). *Enhancing learning through self assessment*. Londres: Routledge Falmer.
- Bourke, R. (2014). Self-assessment in professional programmes within tertiary institutions. *Teach. High. Educ.*, 19, 908-918. <https://doi.org/10.1080/13562517.2014.934353>
- Butler, Y. G. y Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5-31. <https://doi.org/10.1177/0265532209346370>
- Butler, Y. G. (2016). Self-Assessment of and for Young Learners' Foreign Language Learning. En M. Nikolov (Ed.), *Assessing Young Learners of English: Global and Local Perspectives* (pp. 291-315). Springer International. https://doi.org/10.1007/978-3-319-22422-0_12
- Calsamiglia, B. H. y Tusón, A. (2008). *Las cosas del decir: manual de análisis del discurso*. Barcelona: Ariel.
- Cameron, L. (2001). *Teaching Languages to Young Learners*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511733109>
- Cassidy, S. (2006). Learning style and student self-assessment skill. *Education & Training*, 48(2/3), 170-177. <https://doi.org/10.1108/00400910610651791>
- Chen, Y.M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2), 235-262. <https://doi.org/10.1177/1362168807086293>
- Cohen, A. D. y Macaro, E. (2010). Research Methods in Second Language Acquisition. En E. Macaro (Ed.), *Continuum Companion to Second Language Acquisition* (pp. 107-133). Londres: Continuum. <https://doi.org/10.5040/9781474212397.ch-003>
- Cohen, A. D. y Scott, K. (1996). A synthesis of approaches to assessing language learning strategies. En Oxford, R. L. (Ed.), *Language Learning Strategies Around the World: Cross-cultural perspectives* (pp. 89-106). Honolulu, Hawaii: University of Hawai'i Press. <https://doi.org/10.1515/9780824897376-008>
- Consejo de Europa (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Anaya.
- De Houwer, A. (2018). Input, context and early child bilingualism: Implications for clinical practice. En A. Bar-On & D. Ravid (Eds.), *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp. 599-616). Berlín: Walter de Gruyter. <https://doi.org/10.1515/9781614514909-030>
- De Saint Léger, D. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals*, 42, 158-178. <https://doi.org/10.1111/j.1944-9720.2009.01013.x>

De Saint Léger, D., y Storch, N. (2009). Learners' perceptions and attitudes: Implications for willingness to communicate in an L2 classroom. *System*, 37, 269-285. <https://doi.org/10.1016/j.system.2009.01.001>

Dörnyei, Z. (2001). *Motivational Strategies in the Language Classroom*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667343>

Ericsson, K. A. (2002). Towards a procedure for eliciting verbal expression of non-verbal experience without reactivity: Interpreting the verbal overshadowing effect within the theoretical framework for protocol analysis. *Applied Cognitive Psychology*, 16, 98-987. <https://doi.org/10.1002/acp.925>

Gardner, D. (2000). Self-assessment for autonomous language learners. *Links and Letters*, 7, 49-60.

Gass, S.M. y Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Routledge. <https://doi.org/10.4324/9781410606006>

Goh, C. C. M. y Vandergrift, L. (2012). *Teaching and Learning Second Language Listening. Metacognition in Action*. Nueva York: Routledge. <https://doi.org/10.4324/9780203843376>

Hanrahan, S.J. y Isaacs, G. (2001). Assessing Self- and Peer-assessment: The students' views. *Higher Education Research & Development*, 20(1), 53-70. <https://doi.org/10.1080/07294360123776>

Instituto Cervantes. (2014). *Guía del examen DELE C1*. Recuperado de https://examen.es.cervantes.es/sites/default/files/guia_examen_dele_c1.pdf

Keck, C., Iberri-Shea, G., Tracy-Ventura, N. y Wa-Mbaleka, S. (2006). Investigating the empirical link between task-based interaction and acquisition: a quantitative meta-analysis. En John M. Norris y Lourdes Ortega (Eds.), *Synthesizing Research on Language Learning and Teaching* (pp. 91-131). Hawaii: John Benjamins Publishing Company. <https://doi.org/10.1075/llt.13.08kec>

Lyster, R. y Saito, K. (2010). Oral feedback in classroom SLA: A Meta-Analysis. *Studies in Second Language Acquisition*, 32(2), 265-302. <https://doi.org/10.1017/S0272263109990520>

Mackey, A. y Gass, S. M. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum Associates Publishers.

Mackey, A. y Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. En A. Mackey (Ed.), *Conversational interaction in second language acquisition: a series of empirical studies* (pp. 407-453). Oxford University Press.

Micán, D. A. y Medina, C. L. (2017). Boosting vocabulary learning through self-assessment in an English language teaching context. *Assess. Evalu. High. Educ.*, 42, 398-414. <https://doi.org/10.1080/02602938.2015.1118433>

Murakami, C., Valvona, C. y Broudy, D. (2012). Turning apathy into activeness in oral communication classes: regular self- and peer-assessment in a TBLT programme. *System*, 40, 407-420. <https://doi.org/10.1016/j.system.2012.07.003>

Nunan, D. y Bailey, K. M. (2009). *Exploring Second Language Classroom Research. A Comprehensive Guide*. Canadá: Heinle.

O'Malley, J. M. y Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524490>

Pérez Morales, J. I. 2007. *La evaluación como instrumento de mejora de la calidad del aprendizaje. Propuesta de intervención psicopedagógica para el aprendizaje del idioma inglés*. Tesis doctoral, Universitat de Girona. Recuperado de <https://www.tdx.cat/bitstream/handle/10803/8004/tjipm.pdf>

Ratminingsih, N. M., Marhaeni, A. A. I. N. y Vigayanti, L. P. D. (2018). Self-assessment: the effect on students' independence and writing competence. *Int. J. Instruc.* 11, 277-290. <https://doi.org/10.12973/iji.2018.11320a>

Siow, L. F. (2015). Students' perceptions on self- and peer-assessment in enhancing learning experience. *Malaysian Online J. Educ. Sci.*, 3, 21-35.

The jamovi project. (2020). *jamovi*. (Version 1.2) [Software informático]. Recuperado de <https://www.jamovi.org>.

Wang, W. (2017). Using rubrics in student self-assessment: student perceptions in the English as a foreign language writing context. *Assess. Evalu. High. Educ.*, 42, 1280–1292. <https://doi.org/10.1080/02602938.2016.1261993>

Winke, P., Zhang, X. y Pierce, S. J. (2023). A closer look at a marginalized test method: Self-assessment as a measure of speaking proficiency. *Studies in Second Language Acquisition*, 45(2), 416-441. <https://doi.org/10.1017/S0272263122000079>

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2

APÉNDICES

APÉNDICE 1. EJEMPLO LITERAL DE HOJA DE AUTOEVALUACIÓN DE LA PRESENTACIÓN ORAL COMPLETADA POR UN ESTUDIANTE

Una vez realizada tu presentación oral en clase (esa misma tarde, si es posible), reflexiona sobre tu actuación y responde las siguientes preguntas. No dudes en escribir tanto como necesites.

1) **¿Cómo te has sentido durante la presentación?**

Primero me sentí un poco nerviosa, pero poco a poco me sentía más segura en mí misma.

2) **¿Te sientes cómodo/a con el tema presentado?, ¿hasta qué punto controlas el tema?**

Me sentí muy cómoda con el tema. Lo he estudiado mucho para prepararme para la presentación.

3) **¿Consideras que has preparado y practicado la presentación de manera concienzuda?**

He preparado desde el jueves la semana pasada. Me gustaría haber practicado la presentación en voz alta con alguien, pero me faltaba tiempo para eso.

- 4) **Valora tu actuación con respecto a los siguientes aspectos:**
- Estructura y claridad de las ideas presentadas:**
Pienso que la presentación estaba estructurada de buena manera.
 - Pronunciación:**
A veces sentí que algunas palabras me desafiaron, pero he intentado pronunciarlas bien.
 - Entonación y fluidez:**
Ya sé que me falta un poco la fluidez. Me parecía que tenía toda la presentación en mi cabeza pero las palabras adecuadas se me escaparon.
 - Corrección lingüística (gramática):**
Cometí errores e intenté corregirlos cuando lo descubrí. Cuando tengo que responder o explicar algo rápidamente, tengo problemas con el tiempo y el modo de los verbos.
 - Vocabulario (uso y variedad):**
A veces, cuando quería decir algo sobre el tema, no recordé o no supe la expresión adecuada. Otras veces usé unas expresiones básicas o coloquiales sin pensar en la forma que debiera usar.
- 5) **¿Cuáles dirías que han sido los puntos fuertes de tu presentación?**
He estudiado y preparado el tema profundamente, tanto para poder explicarme y responder a las preguntas como para no atascarme por inseguridad.
- 6) **¿Cuáles dirías que han sido los puntos débiles de tu presentación?**
La fluidez. El léxico – especialmente cuando improvise. A veces la gramática.
- 7) **Si pudieras repetir la presentación, ¿qué aspectos cambiarías?**
Me gustaría o haber discutido el tema con alguien en español antes de presentar o haber practicado la presentación en voz alta.
- 8) **¿Hasta qué punto estás satisfecho/a de tu presentación?**
Estoy bastante satisfecha con mi presentación. Siento que he logrado comunicar el tema de modo adecuado para que los demás lo entiendan, y eso es lo más importante para mí. Por supuesto la gramática y la fluidez son aspectos importantes también, pero la sustancia me parece lo principal.
- 9) **Si tuvieras que ponerte una nota del 1 al 10, ¿qué nota te pondrías?**
No estoy segura. Quizás 8 o 9 por el contenido y 5 o 6 por mi lengua – por consiguiente 7 en total.
- 10) **En general, ¿qué aspectos consideras que debes trabajar para mejorar tu expresión oral en español?**
Debo practicar mi lengua oral con alguien que me corrija cuando cometo errores. También debo pasar un semestre en España para lograr más fluidez.



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 136-150
Artículo recibido: 03/08/2024
Artículo aceptado: 24/10/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/654>
DOI: <https://doi.org/10.58859/rael.v23i1.654>

Etimologías del *caló* en el *Diccionario de la Lengua Española* de la Real Academia Española: comentario y análisis de algunos ejemplos

***Caló* etymologies in the *Spanish Language Dictionary* of the Royal Spanish Academy: commentary and analysis of some examples**

ALBA MACÍAS COUSO
UNIVERSIDAD DE CÁDIZ

En este trabajo partimos de la hipótesis de que la falta de rigor presente en la lexicografía *caló* – llevada a cabo principalmente por aficionados y no lingüistas – se refleja en los diccionarios generales de la lengua. Por lo tanto, nos proponemos analizar las etimologías de los gitanismos incluidos en la versión en línea del *Diccionario de la Lengua Española (DLE)* de la Real Academia Española, analizando la correspondencia de estas etimologías con las informaciones consignadas en distintas obras de lexicografía gitano-español. Para ello, extraeremos la etimología de aquellas palabras del *DLE* que incluyan *caló* como étimo para, posteriormente, localizarlas en una selección de seis diccionarios gitanos. Tras esto, seleccionaremos una muestra de ejemplos, tomando aquellos más relevantes, para llevar a cabo un análisis en base a las informaciones proporcionadas por diferentes autores con la finalidad de detectar problemas etimológicos e intentar clarificar si realmente estamos ante vocablos de origen gitano.

Palabras clave: *lexicografía; etimología; caló; Diccionario de la Lengua Española; lexicografía gitano-español.*

Our main hypothesis is that the lack of rigour existing in *caló* lexicography – mostly developed by amateurs and non-linguists - is shown in the general dictionaries of the language. Therefore, we intend to analyse the gypsy etymologies included in the online version of the *Diccionario de la Lengua Española (DLE)* of the Real Academia Española, analysing the connection of these etymologies with the information included in different works of gypsy-Spanish lexicography. To achieve this, we will extract the etymology of those words in the *DLE* which include *caló* as an etymon and then locate them in a selection of six Gypsy dictionaries. After this, we will select a number of examples, choosing the most relevant ones, to make an analysis based on the information provided by different authors in order to detect etymological problems and try to clarify if they are indeed words of gypsy origin.

Keywords: *lexicography; etymology; caló; Spanish Language Dictionary; gypsy-Spanish lexicography.*

Esta publicación es parte del proyecto de I+D+i ayuda PID2022-139201OB-I00, financiado/a por MCIN/AEI/10.13039/501100011033

Citar como: Macías Couso, A. (2024). Etimologías del *caló* en el *Diccionario de la Lengua Española* de la Real Academia Española: comentario y análisis de algunos ejemplos. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 136-150. <https://doi.org/10.58859/rael.v23i1.654>

1. INTRODUCCIÓN

Una considerable cantidad de léxico *caló* se ha introducido en el español debido al contacto de esta variedad del romaní con el español coloquial, popular y jergal. Tanto es así que, actualmente, en los diccionarios generales de la lengua se consignan gitanismos y, en algunos casos, estas entradas de diccionario incluyen la etimología que le hayan atribuido, lo que debería contribuir a la comprensión de su origen.

Acudir a los diccionarios gitano-español podría ser una de las formas de conocer el origen etimológico de estos vocablos. Sin embargo, la lexicografía del *caló* presenta numerosas deficiencias ya que estamos ante obras que se han realizado principalmente por aficionados, no lingüistas ni filólogos. Debemos tener en cuenta que algunos de los primeros glosarios o vocabularios publicados sobre el *caló* no son diccionarios, sino que se trata de apartados correspondientes a estudios de sociología criminal, como el de Salillas (1896). A esto, además, hay que sumarle la confusión histórica a la que *caló*, como término, ha sido sometida, ya que desde su aparición se utilizó como sinónimo de *germania*, lo que derivó a que posteriormente se utilizase también para designar al *lenguaje de la delincuencia* o *lenguaje de los maleantes* – denominación utilizada en el vocabulario publicado por la Guardia Civil (1949, 1950) – algo que todavía sucede, por ejemplo, en la tradición mexicana. Toda esta amalgama de incoherencias, confusiones y falta de rigor en su estudio han desembocado en que la lexicografía *caló* no resulte una fuente especialmente fiable si lo que se pretende es abordar el estudio del léxico gitano.

La información etimológica de los diccionarios puede resultar útil para conocer el origen de ciertas palabras. Sin embargo, podemos deducir que, a causa de esta falta de rigor en la lexicografía *caló*, surgen deficiencias en las etimologías documentadas en las obras generales. Precisamente, esta es la hipótesis de la que partimos para la realización de este trabajo. Nuestra labor consistirá, por tanto, en seleccionar los gitanismos pertenecientes al diccionario general más utilizado en el mundo hispanohablante, como es el *DLE* de la Real Academia Española, en su versión en línea, y así estudiar las etimologías del *caló* documentadas en él. Tomaremos como muestra un conjunto de obras lexicográficas de *caló*-español, en los que pretendemos localizar estas etimologías y estudiar sus relaciones, similitudes o diferencias con respecto a lo que se documenta en el *DLE*.

2. BREVE PANORAMA DE LA LEXICOGRAFÍA CALÓ¹

El estudio del *caló* ha quedado limitado a reductos muy concretos dentro de la lingüística. Desde que aparecieron los primeros trabajos sobre el tema hasta la actualidad, el *caló* se ha abordado desde un punto de vista lexicográfico. La cantidad de glosarios, diccionarios y recopilaciones de vocabulario de gitanismos es considerable y, sin embargo, también destaca la falta de rigor y de veracidad a la hora de elaborar estas obras lexicográficas (*cf.* Buzek, 2011)².

La mayor parte de los estudios al respecto coinciden en situar la primera obra de referencia en el diccionario de Borrow (1843)³, donde el autor recoge el que podría considerarse el primer repertorio de gitanismos con equivalencias en inglés, su etimología, las clases de palabra y el género en el caso de sustantivos y adjetivos, añadiendo, además, un cancionero y poemario

¹ Somos conscientes de que la producción lexicográfica del *caló* es mucho más extensa. En este trabajo nos limitaremos a comentar aquellas obras a las que hemos tenido acceso, tarea difícil teniendo en cuenta que la mayoría de obras se encuentran descatalogadas, tratándose, prácticamente, de documentación histórica.

² Ivo Buzek realiza un estudio extenso y exhaustivo sobre el tema en su libro *Historia crítica de la Lexicografía gitano-española*, una obra en la que propone una revisión de numerosos diccionarios y glosarios de *caló*. En este apartado nos limitaremos a comentar cronológicamente los diccionarios a los que hemos logrado tener acceso.

³ La primera edición de este diccionario se publicó en 1841. Sin embargo, para nuestro trabajo tenemos disponible una segunda edición publicada en 1843.

gitano junto con un estudio antropológico sobre la cultura y los modos de vida de esta comunidad. Tras esta publicación, a mediados y finales del siglo XIX, le siguen algunos autores del ámbito hispano como Jiménez (1846), Campuzano (1848) y Quindalé (1870). Tanto Jiménez (1846: 7-43) como Campuzano (1848: III-XXIX) realizan extensos prólogos previos al glosario en los que hablan acerca de la comunidad gitana de forma que roza lo peyorativo, tratándolos como una comunidad cuyo modo de vida se difumina con la delincuencia. Ninguno de ellos incluye en el diccionario información etimológica, ciñéndose exclusivamente a la traducción al español de los vocablos, y tan solo Jiménez añade la categoría. Quindalé (1870: 2-48), por su parte, incluye un apartado que dedica al origen del pueblo gitano, su llegada a la península, sus actividades, costumbres y vinculación con la delincuencia y, por último, la persecución que sufrió esta etnia a nivel europeo. Añade también un epítome de gramática gitana (Quindalé, 1870: 49-76) algo que hasta la fecha no había sido contemplado en el resto de obras lexicográficas sobre el *caló*. En cuanto a las entradas de diccionario, estas incluyen, junto a la traducción al español, información sobre la categoría y el género.

Siguiendo con el siglo XX, se publican las obras de Besses (2000) y Rebolledo (2006)⁴. El diccionario de Besses no es exclusivo del *caló*, sino que se trata de un repertorio de argot español en el que el autor documenta multitud de gitanismos. Esto se debe a la inclusión de vocablos del *caló* en el lenguaje coloquial por la cercanía de las comunidades gitanas a los estratos sociales bajos. El hecho de que Besses utilice el término argot para referirse a lo que sería un argot común (aquel vocabulario que, teniendo su origen en comunidades marginales, se traspaasa al lenguaje popular) deja entrever la más que frecuente confusión entre argot y lenguaje coloquial. De hecho, en el subtítulo de esta obra de Besses reza lo siguiente “o lenguaje jergal, gitano, delincuente profesional y popular”, por lo que vemos que la confusión entre todos estos conceptos es más que palpable.⁵ La información que incluye en las entradas se centra en el ámbito de uso. Rebolledo (2006: 5-6) realiza un pequeño prólogo de dos páginas en el que trata el origen del *caló*, rechazando la idea de que se trata de un lenguaje de delincuentes. Por otra parte, las entradas del diccionario se componen de su traducción al español y de la categoría de palabra.

Respecto al siglo XXI, Mójica (2017) publica en abierto un diccionario donde incluye un prólogo en el que critica la inclusión de vocabulario de germanía y del lenguaje de la delincuencia dentro de la lexicografía *caló* (Mójica, 2017: 5-6) y dedica numerosas páginas a describir la gramática del *caló* plagiando el epítome de Quindalé (1870). En cuanto al diccionario, simplemente incluye la palabra en *caló* y su equivalencia en español, sin añadir ningún otro tipo de información.

Sin embargo, el léxico *caló* ha trascendido más allá de su uso por parte de la comunidad gitana, introduciéndose en otras variantes lingüísticas, especialmente en los lenguajes marginales. Algunos ejemplos los vemos en Salillas (1896), que incluye en su trabajo sobre el delincuente español un glosario en el que aparecen gitanismos. También tenemos el caso del *caló mexicano*, definido como lenguaje de delincuentes, y donde se enmarcan los trabajos de Roumagnac (1904), Serrano García (1935), Saavedra (1941) y Chabat (1956), todos dedicados al estudio del lenguaje de la delincuencia en México. También merece la pena tener en cuenta una de las variantes lingüísticas más populares dentro del habla argentina, gracias a su expansión popular por medio del tango, como es el *lunfardo*⁶, reflejado en el diccionario de Gobello y Payet (1959).

⁴ Disponemos de una segunda edición del diccionario de Besses, publicada en el año 2000, y de una edición facsimilar del diccionario de Rebolledo, publicada en el año 2006.

⁵ Para más información terminológica sobre argot puede consultarse un trabajo que hemos realizado con anterioridad donde clarificamos las relaciones conceptuales entre los conceptos de argot y jerga (Macías Couso, 2021).

⁶ Autores como Casas Gómez (2003: 567) van más allá y consideran el *lunfardo* y el *caló* como *dialectos sociales*, que, a pesar de que presentan características propias de la variación diafásica, están fuertemente marcados por el uso de grupos sociales diferenciados, situándose dentro de la variación diastrático-diafásica.

Las investigaciones sobre el *caló* han sido más prolíficas que la propia lexicografía gitano-española. Queremos destacar algunos trabajos previos dentro de este ámbito. Realizando un repaso diacrónico, hablamos en primer lugar el trabajo de Wagner (1941), que trata algunas palabras del *caló* en Andalucía, España y Latinoamérica. Como obra y autor clave debemos resaltar los trabajos de Clavería (1949a, 1949b, 1951, 1953a, 1953b), en los que encontramos una serie de reflexiones que permiten entender las relaciones del *caló* con la lengua española, con el argot de la delincuencia y con el lenguaje flamenco. La misma línea sigue el capítulo de Suárez Solís (1969) en el que se trata el *caló* y sus relaciones con la germanía histórica. Por otra parte, podemos destacar la labor de Casas Gómez (1986) en su trabajo sobre el léxico gitano despectivo para designar a la ‘prostituta’ y su presencia dentro de distintas obras lexicográficas. Indudable resulta también la importancia de las aportaciones de Roperó Núñez (1978, 1991, 1992, 1999 y 2005) al estudio del léxico *caló* dentro de los diccionarios generales de la lengua, como el *DLE*, pero también dentro del mundo flamenco. En último lugar, y más reciente, destaca la tesis doctoral realizada por Mesa Navarro (2013), en la que analiza lexicográfica y etimológicamente el diccionario de Quindalé.

Debemos destacar la labor del grupo de la República Checa, concretamente de la Universidad de Masaryk, liderado por Ivo Buzek, cuya producción sobre estudios de lexicografía *caló* se extiende entre los años 2009-2016, entre los que consideramos especialmente relevantes dos libros publicados: *La imagen del gitano en la lexicografía española* (2010) e *Historia crítica de la lexicografía gitano-española* (2011). Concretamente, la publicación de 2011 resulta muy útil como referencia para aquellos investigadores que se estén iniciando en el estudio del *caló* y quieran conocer cuál es el estado – no tan actual ya que han pasado varios años desde la publicación de este libro – de la lexicografía *caló*.

3. METODOLOGÍA

Nuestra metodología comienza por la confección de una base de datos cronológica de gitanismos. Como referencia, tomamos el estudio de Gamella (2011, 2012, 2015) sobre disponibilidad léxica, con la finalidad de obtener equivalencias en español a raíz de las muestras que aportan sus informantes. Así, disponemos de un punto de partida con el que comenzar la búsqueda en los diccionarios gitanos a partir de palabras del español. El primer resultado que obtenemos consta de una base de datos con más de 400 gitanismos, divididos en los 16 centros de interés que utiliza Gamella (2015): 1) alimentos; 2) cuerpo-acciones; 3) ropa; 4) animales; 5) tiempo-clima; 6) religión; 7) utensilios; 8) dinero-negocios; 9) familia-tratamiento; 10) delincuencia-autoridad-violencia; 11) sentimientos-estados; 12) profesiones; 13) adjetivos-adverbios-interjecciones; 14) pronombres; 15) acciones verbales; 16) lugares.

Durante la confección de este documento comenzamos a observar muestras de inestabilidad formal y de contenido en los gitanismos. Una muestra de esto lo observamos, por ejemplo, en ‘hablar’, para la que encontramos *penar*, *chamullar*, *chamucar* o *chamuyar*, rasgo característico de las variedades lingüísticas de transmisión oral. Nos preguntamos entonces si este fenómeno se ha trasladado a las etimologías del *caló* presente en los diccionarios generales de la lengua y decidimos trabajar con el diccionario de la RAE en su edición en línea. En primer lugar, realizamos una búsqueda de los gitanismos registrados en la edición online del *DLE*. Para ello, utilizamos el buscador de palabras IEDRA (<https://iedra.es/>), que entre sus opciones de búsqueda permite filtrar por la etimología que aparezca en la entrada.

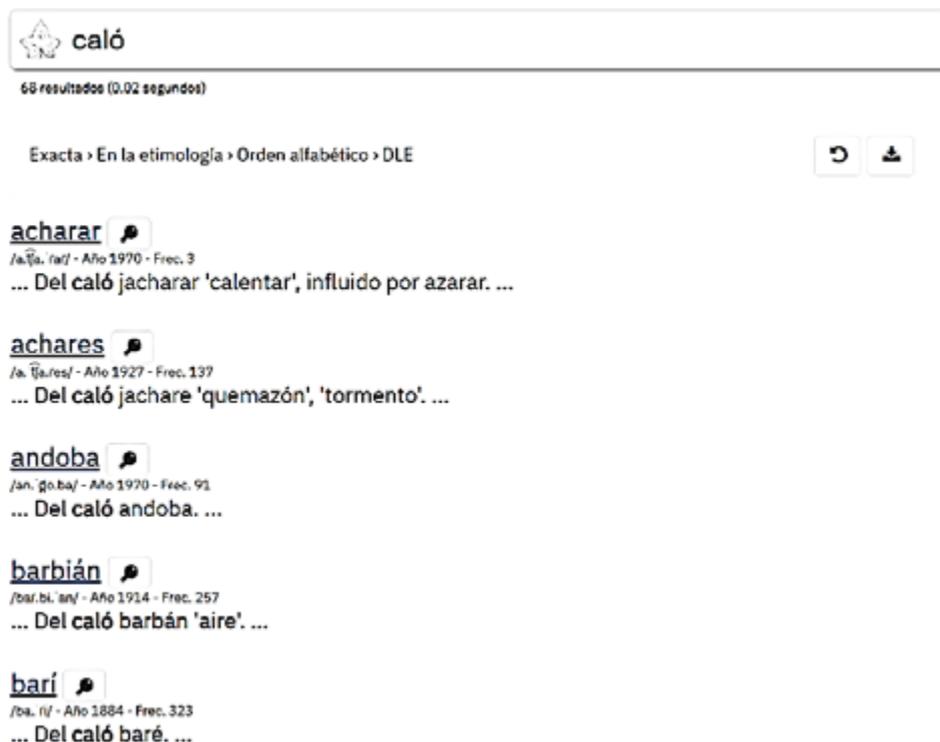


Figura 1: *Ejemplo de búsqueda en IEDRA.*

Como vemos en la Figura 1, al introducir en el buscador ‘caló’, IEDRA nos devuelve todas aquellas palabras incluidas en la versión en línea del diccionario cuya etimología contiene este término. Obtenemos entonces una lista de 67 gitanismos, disponible para descargar en formato de texto plano.

Procedemos entonces a localizar estos gitanismos en las obras lexicográficas a las que tenemos acceso. Para su selección, hemos tomado como referencia la obra de Buzek (2011) en la que realiza un análisis crítico de un repertorio de diccionarios gitanos. De este repertorio, hemos conseguido acceder a cuatro, además de un glosario que no aparece en el análisis de Buzek.

En primer lugar, tenemos a Borrow (1843), que podría representar una fuente fiable ya que algunos expertos afirman que el glosario es resultado de “la recogida de datos realizada por Borrow entre los gitanos españoles durante su estancia en España” (Buzek, 2011: 121). Se trata de uno de los primeros diccionarios de *caló* y ha servido como referencia para la elaboración de obras posteriores. En segundo lugar, figura el diccionario de Jiménez (1846). Los apuntes que hace Buzek (2011: 140-145) sobre esta obra indican que se realiza de forma descuidada y que se incluyen distintas nomenclaturas no normalizadas con la finalidad de engrosar el diccionario. Tampoco se le atribuye mucha fiabilidad, ya que no incluye notas gramaticales o de uso y, además, se indica que, probablemente, Jiménez no fue el creador, sino que su función fue más parecida a la de editor o compilador de materiales que no eran de su propiedad.

Siguiendo nuestro eje cronológico, aparece en tercer lugar la obra de Campuzano (1848). Este diccionario ofrece una ingente cantidad de palabras con las características propias de un diccionario monolingüe, esto es, lo que ofrece no son equivalencias en español, sino definiciones (Buzek, 2011: 152). En este caso sí aporta marcas gramaticales. Además, su consulta por parte del usuario se hace dificultosa debido a la multitud de remisiones que hace a distintos vocablos dentro de la obra. Quindalé (1870) es el siguiente autor al que hemos tenido acceso. La principal particularidad de este diccionario es que por primera vez ofrece un epígrafe de gramática del *caló*. El autor se esfuerza por no incluir vocablos provenientes de la germanía,

en un intento de realizar una obra llena de rigurosidad y científicidad. Sin embargo, tal y como afirma Buzek (2011: 170) encontramos “voces probablemente inventadas por él mismo según las pautas de formación de palabras incluidas en el Epítome”.

En penúltimo lugar encontramos el diccionario de Rebolledo (2006), el primer diccionario bidireccional *caló*-español español-*caló*. Según Buzek, a pesar de tratarse de la obra más amplia publicada hasta la fecha, no es la más fiable, ya que encontramos errores como la falta de homogeneidad en los acentos gráficos, la invención de palabras, la recopilación de vocablos procedentes de fuentes de todo tipo y la falta de unificación o lematización (Buzek, 2011: 181-182). En último lugar, incluimos a Mójica (2017). Este diccionario es el más reciente al que hemos tenido acceso y al publicarse posteriormente a su libro, Buzek no lo incluye en su análisis. El autor asegura que en sus páginas representan un estudio riguroso sobre léxico *caló*, realizando una crítica a la falta de rigurosidad científica de la mayoría de obras de lexicografía *caló* publicadas. En el prólogo afirma que ha eliminado el léxico de la germanía (Mójica Legarre, 2017: 5) y presenta “las palabras bajo su forma científica más genuina”. Incluye un apartado dedicado a la gramática del *caló* que, de hecho, es una copia exacta del epítome de Quindalé (1870: 50-76). Y, de nuevo, estamos ante una obra bidireccional basada en equivalencias. No incluye marcas de ningún tipo y, si lo comparamos con el resto de obras seleccionadas para este trabajo, el aumento de vocablos incluidos es considerable. Encontramos los mismos problemas a los que Buzek hace referencia en su análisis: invenciones, falta de lematización, falta de homogeneidad en la escritura y remisiones a múltiples palabras para un mismo significado. El glosario de Mójica se encuentra publicado en acceso abierto por el mismo autor. El acceso al resto de obras ha sido posible gracias al repositorio lingüístico perteneciente al Instituto Universitario de Investigación en Lingüística Aplicada, en el que se encuentran tanto en formato digital como en formato físico.

De esta manera, se ha obtenido una nueva base de datos en la que incluimos los siguientes campos: 1) gitanismo recogido en el *DLE*; 2) etimología; 3) acepciones; 4) ocurrencias en los distintos diccionarios y 5) significado.

GITANISMO	ETIMOLOGÍA	ACEPCIÓN 1	ACEPCIÓN 2	ACEPCIÓN 3	BORROW: 1843	JIMÉNEZ: 1846	CAMPUZANO: 1848
acharar	jacharar ‘calentar’	Avergonzar, azarar, sobresaltar	Disgustar, enojar, desazonar	Dar achares	Jacharár ‘quemar’	Jacharar ‘calentar’	Jacharar ‘calentar, quemar’
achares	jachare ‘quemazón’, ‘tormento’	Celos, penas o tormentos amorosos			Jachári ‘incendio’	Jachare ‘quemazón’	Jachare ‘quemazón acción y efecto de quemar o quemarse’
andoba	andoba	Persona cualquier que no se nombra			Andoba ‘este’, pronombre determinado	X	Andoba ‘andolaya’
barbián	barbán ‘aire’	Desenvuelto, gallardo, arriscado			Barban ‘viento, aire’	Barban ‘viento’	Barbal ‘viento, aire’
barí	baré	Excelente			Baro ‘grande’	Baró, baré ‘grande’	Baré ‘barial’

Figura 2: Base de datos. Fuente: elaboración propia.

De esta última base de datos, resultado de la búsqueda manual en cada uno de los diccionarios, se han obtenido las informaciones que analizamos en este trabajo.

4. RESULTADOS Y ANÁLISIS

Como primer resultado, producto de la búsqueda por etimologías que contienen ‘caló’ en IEDRA, obtenemos que, en la versión en línea del *Diccionario de la Lengua Española*, se registran un total de 67 gitanismos que presentamos a continuación por orden alfabético, tal y como se muestran en nuestra base de datos: *acharar, achares, andoba, barbián, barí, bato, bulo, burel, calé, caló, camelar, cané, canguelo, canguis, cate, chai, chalar, chamullar, chaval, chavea, chavó, chingar, chipé, chipén, chorar, chori, choro, chulé, chungo, churumbel, clisos, currar, diñar, diquelar, ducas, endiñar, estache, fetén, gaché, gachí, gachó, gilí, guripa, jai, jamar, jindama, jiñar, jonjabar, lacha, mangar, mangue, manús, menda, molar, mollate, mui, muquir, naja, pañí, paripé, parné, pinrel, pira, postín, pureta, terne, trajelar*.

Como avanzamos en la introducción, uno de nuestros objetivos es comprobar la presencia de las etimologías de estos gitanismos del *DLE* en relación con la muestra de lexicografía *caló* que hemos seleccionado. En primer lugar, podemos afirmar que, de estos 67 gitanismos, 38 no presentan variación respecto a la etimología que muestra el *DLE*: *andoba, bato, barbián, burel, calé, caló, camelar, canguelo, chalar, chai, chamullar, chungo, chipé, chipén, diñar, ducas, diquelar, estache, gaché, gachí, gachó, gilí, jamar, jindama, jiñar, jonjabar, lacha, mangar, mangue, menda, molar, mollate, mui, pañí, parné, pinré, postín* y *terne*. Es cierto que encontramos algunas variaciones de forma, como por ejemplo en *bato* ‘padre’ que Quindalé (1870: 11) recoge bajo las formas *bato* y *batú*; en *diquelar* ‘ver’ que en ocasiones se documenta con la forma *dicar* (Borrow, 1843: 44; Jiménez, 1846: 88; Quindalé, 1870: 28) o en *gaché*, definido en el *DLE* como ‘entre los gitanos, andaluz’, que aparece también bajo la forma *gachó* (Borrow, 1843: 54; Campuzano, 1848: 88, Quindalé, 1870: 41) para el masculino. Autores como Wagner (1962) explican que, en el *caló* primitivo, la terminación *-o* se corresponde con el masculino singular, la terminación *-í* se utiliza para el femenino singular y la terminación *-é* para el masculino singular o para el plural.

Sin embargo, los 29 gitanismos restantes sí presentan algunas variaciones etimológicas importantes respecto al léxico encontrado en los diccionarios. De este segundo grupo, por motivos de espacio, comentaremos tan solo aquellos ejemplos que consideramos relevantes.

4.1. *Acharar, achares*

Estos dos vocablos se encuentran íntimamente relacionados tanto en forma como en contenido. *Acharar* se documenta en el *DLE* con tres acepciones: 1) avergonzar, azarar, sobresaltar; 2) disgustar, enojar, desazonar y 3) dar achares. La etimología proporcionada corresponde con *jacharar* ‘calentar’ (RAE, 2024). Esta forma aparece documentada en los diccionarios de Borrow (1843: 45), Jiménez (1846: 43), Campuzano (1848: 73), Quindalé (1870: 66) y Mójica (2017: 71), justo con el significado de ‘calentar, quemar, escaldar, abrasar’. Rebolledo (2006: 143), en cambio, documenta *acharar* con el significado de ‘mortificar, atormentar’, más cercano a lo que se presenta en las acepciones del *DLE*, además de añadir la forma *coriar*.

Achares se recoge en el *DLE* con una acepción: “celos, penas o tormentos amorosos”. La etimología que aparece es la de *jachare* ‘quemazón, tormento’ (RAE, 2024). En los diccionarios gitanos encontramos lo siguiente: Borrow (1843: 60) documenta la forma *jachári* como ‘incendio’; Jiménez (1846: 76), Campuzano (1848: 98), Quindalé (1870: 38), Rebolledo (2006: 192) y Mójica (2017: 85) incluyen la forma *jachare* como ‘quemazón’ y ‘acción y efecto de quemar o quemarse’. Además, Quindalé (1870: 38) y Rebolledo (2006: 192) documentan *jachare* como ‘tormento’, que también consigna *achare* como ‘mortificación’ (Rebolledo 2006: 176) y, por último, Mójica (2017: 46) documenta la forma *achare* para ‘tormento’ y el plural *achares* para ‘celos’.

Este significado metafórico que ha tomado *jachare/achare* de un primer significado literal relacionado con la acción y efecto de quemar, nos parece sumamente interesante, ya que se trata de una metáfora visual de los celos como algo que puede “quemar por dentro”.

4.2. *Bulo*

La palabra *bulo* aparece en el *DLE* como gitanismo con la acepción de “noticia falsa propagada con algún fin” (RAE, 2024). La etimología aportada se corresponde con *bul* ‘porquería’. Sin embargo, en ninguno de los diccionarios hemos encontrado esta forma con ese significado: Borrow (1843: 24) documenta *bul* para ‘ano, orificio’; Campuzano (1848: 40) incluye la forma *bul* para ‘orificio, abertura’; en Quindalé (1870: 46) se encuentran *bulipen*, *buló* y *buró* para ‘embuste, engañifa, mentira’ y en Jiménez (1846: 46), Rebolledo (2006: 46) y Mójica (2017: 167) vemos, en cambio, la forma *bulo* para ‘embuste, mentira, engañifa, trola”, correspondiéndose con lo recogido en el *DLE*.

La hipótesis de Corominas (1987: 111) es que se trata de una voz jergal adaptada del *caló* *bul* ‘porquería, excremento’ –recordemos que no hemos encontrado esta acepción en los diccionarios gitanos–, influenciado por *bola* ‘mentira’. Sin embargo, aquí planteamos que el cambio de forma de *bul* a *bulo* puede deberse a la adopción de rasgos morfológicos del español en el léxico *caló*. En este caso, quizá, esa terminación en *-o* pueda corresponderse con el sufijo masculino del español. Por otra parte, también podemos observar un cambio de contenido de origen metafórico, ya que de *bul* ‘orificio, ano’ se pasa a *bulo* ‘engaño’. Teniendo en cuenta que por el ano sale “porquería”, esto podría servir como explicación del proceso de transformación de este vocablo. En cualquier caso, estamos ante otra manifestación expresa de la confusión entre *caló* con el lenguaje popular, ya que históricamente los límites entre el *caló* y otros conceptos afines no están claramente definidos.

4.3. *Cané*

Esta palabra aparece recogida en el *DLE* con la acepción de ‘juego de azar parecido al monte’ y, como etimología, se aporta “del *caló* *cané*” (RAE, 2024), algo que también sucede en Corominas (1987: 125) que indica que es “de origen gitano”. Borrow no documenta esta forma y el resto de los diccionarios incluyen *cané* pero con el significado de ‘oído’ (Jiménez, 1846: 46); Campuzano, 1848: 47; Quindalé, 1870: 18; Rebolledo, 2006: 180; Mójica (2017: 62). No encontramos cuál puede ser la relación entre la etimología que presenta el *DLE* y esta palabra documentada. Tenemos algunas anotaciones interesantes sobre *cané* en Mohedano Gallardo (2018: 57): “En *caló*, el *cané* era un juego de azar parecido al monte y al *baceta*”. Según ciertas fuentes, era una variante del *monte* y procedía del *sacanete*. Si buscamos en el *DLE* esta palabra, nos devuelve una acepción que indica que designa un “juego de naipes y de envite, en que se juntan y mezclan hasta seis barajas” (RAE, 2024) y, además, añade la etimología *lansquenete* “soldado de infantería alemana” (RAE, 2024) que proviene del francés *lansquenet* y este, a su vez, del alemán *landsknecht* ‘mercenario’, según el Diccionario de la Real Academia.

En base a estas informaciones no deberíamos atrevernos a afirmar que la etimología de *cané* es *caló* y que se trate de un gitanismo, más teniendo en cuenta que este no aparece registrado con este sentido u otro parecido en ninguno de los diccionarios gitanos que hemos consultado.

4.4. *Chaval, chavó, churumbel*

Estas tres palabras se encuentran relacionadas tanto por las etimologías que el *DLE* les otorga como por su definición. Comenzando con las definiciones, para *chaval* tenemos ‘niño o joven’; para *chavó* se remite a ‘chaval’ y para *churumbel* se da la acepción ‘niño, muchacho’ (RAE, 2024). Como vemos, parece ser que las tres palabras se utilizan para designar realidades muy parecidas, prácticamente iguales.

En cuanto a las etimologías, para *chaval* incluyen *chavale* pl. de *chavó* ‘muchacho’ y, a su vez, la etimología de *chavó* también es *chavó*. En cambio, para *churumbel* solo dan la información “origen *caló*”, sin proporcionar más datos. Ahora bien, si acudimos a los diccionarios gitanos, encontramos la siguiente información: Borrow (1843: 34) incluye *chabal* para ‘hijo’; *chaboro* y *chabo* para ‘niño, muchacho’ y *churumbel* o derivados no aparecen documentados en su diccionario; Jiménez (1846: 55) incluye *chaval* para ‘hijo’ y no documenta *chavó* ni *churumbel*; Campuzano (1848: 27) documenta *bedoro* para ‘muchacho’, sin coincidencias con *chaval*, *chavó*, *churumbel* o derivados; Quindalé (1870: 21) registra *chabal* para ‘joven, mozo, hijo’; *chabó* para ‘niño, muchacho’ y *chaboró* para ‘niño’. No encontramos nada relacionado con *churumbel*; Rebolledo (2006: 177) documenta *chaval* y *bedoró* para ‘mozo’; *chaval* y *lacrizuelo* para ‘mozuelo’ y es el único en el que aparece *churumbeles* para ‘hijos’ y finalmente, Mójica (2017: 64) incluye *chaval* para ‘mozuelo, joven’, lo más cercano a lo que aparece en el *DLE*.

Descartamos la posibilidad de que *churumbel* sea un gitanismo basándonos, además de en estas informaciones, en la muy acertada reflexión de Roperó Núñez (1991: 18), que hace referencia a las obras de Wagner (1941), Corominas (1954) y Cuevas (1974) para demostrar que *churumbel* se trata de un andalucismo cuya etimología se remonta a *churumbela*, derivado del latín *calamellus* ‘caña’ que, por alusiones al pene, se convierte en *churumbel* para designar a ‘niño’.

No ocurre lo mismo con *chaval* o *chavó*, para los que, además de tener sus correspondencias en los diccionarios gitanos consultados, también existe un estudio realizado por Wagner (1962) el que se discute el posible origen árabe de la palabra *chaval*, algo que queda descartado por el autor: “La palabra *chavó* ‘niño, muchacho’ es de pura cepa gitana y está difundida en todos los dialectos gitanos europeos desde Turquía a través de Europa hasta Inglaterra y España” (Wagner, 1962: 20)”. El autor propone diferentes formas derivadas: *chavó* para el masculino singular; *chavé* para el masculino singular, originariamente utilizado para el masculino plural; *chavea* ‘chiquillo’ como gitanismo utilizando principalmente en Andalucía, *chai* – vocablo que veremos más adelante – como contracción de *chaví*, el femenino de *chavó*; *chavóro* y *chaboró*, formas que ya vimos en los diccionarios y, por último, *chaval*, forma derivada de *chavãle*, cuyo origen sitúa en el gitano alemán *tsãvale* (Wagner, 1962: 22). Informaciones similares aporta Corominas (1987: 193) en la entrada de *chaval* “del gitano *chavale*, vocativo masculino del plural de *chavó* ‘hijo, muchacho’”.

4.5. *Chingar*

Aparece en el *DLE* con tres acepciones: 1) importunar, molestar a alguien; 2) practicar el coito y 3) beber con frecuencia vino y licores. La etimología que aporta el diccionario es čingarár ‘pelear’ (RAE, 2024). En los diccionarios de *caló* consultados encontramos lo siguiente sobre este vocablo: Borrow (1843: 39) documenta *chingarar* para ‘pelear’; Jiménez no lo incluye en su glosario; Campuzano (1848: 61) y Quindalé (1870: 23) documentan *chicarelar* para ‘pelear’; Rebolledo (2006: 137, 170, 184, 195) consigna *chingar* para ‘cohabitar’ y, para concluir, Mójica (2017: 63) anota el sustantivo *chicarela* para ‘pelea’, *chingarar* para ‘reñir’.

El *DLE* no nos aclara a qué acepción se corresponde esta etimología cuyo significado es ‘pelear’. Pensamos que la acepción que podría corresponderse con esto es la primera ‘importunar, molestar a alguien’, ya que el resto de acepciones parecen estar demasiado alejadas del sig-

nificado original, a no ser que se haya producido un giro metafórico en ‘pelear’ con ‘practicar el coito’. Corominas (1987: 195) también plantea esta duda al indicar que se trata de vocablo de origen jergal y que, a pesar de que proceda del *caló chingarar*, también existen vocablos que comienzan en *-ching* cuyas raíces son aborígenes.

4.6. *Clisos*

Clisos cuenta con una sola acepción y se define como ‘ojos’. La etimología que se presenta es la siguiente “del *caló clisos*, este quizá del dialect. *clisarse* ‘mirar fijamente’, y este de *eclipsarse*” (RAE, 2024). El recorrido etimológico que realiza esta palabra, según el *DLE*, responde a las interacciones entre el *caló* y el español, por lo que el primero se ve influido por el vocabulario del segundo (de *eclipsarse*, *clisarse* y de este, *clisos*). Corominas (1987: 155) da información etimológica sobre *clisar* que se corresponde con esto: “verbo dialectal *clisarse* ‘quedarse mirando algo fijamente, embobarse’ de *eclipsar*. Sin embargo, en ninguno de los diccionarios de *caló* consultados aparece *clisarse*. Sí encontramos, en cambio, lo siguiente: Borrow (1843: 34) documenta *aquia* para ‘ojos’. Esta podría ser la primera forma documentada de lo que luego conoceríamos como *acais*, variante, en nuestra opinión, más extendida, principalmente, gracias a la difusión de palabras del *caló* en el cante flamenco que, sin embargo, no aparece en el *DLE*. Este *aquias* también aparece documentado en Campuzano (1848: 10) –que apunta en la entrada a *sacais*– y Mójica (2017: 77), que también incluye *fanales*. Por otra parte, Jiménez (1846: 68) incluye *sacais* y *glises* para ‘ojos’; Quindalé (1870: 66) documenta *sacais* para ‘ojos’ y, por último, Rebolledo (2006: 113) incluye *aquí* para ‘ojos’ y también *clisos*.

Lo más cercano a *clisarse* y *eclipsarse* que hemos encontrado es la forma *eclisar/eclisarse*: Quindalé (1870: 30) documenta *eclisar* con las acepciones de ‘ojetear, agujerear, herir los ojos’; Rebolledo (2006: 113) incluye *eclisar* para ‘agujerear’ y Mójica (2017: 74) consigna *eclisar* para ‘encantar, maravillar”. Si bien es cierto que todos estos significados se enmarcan dentro de lo relacionado con la vista, no terminamos de comprobar la etimología que plantea el *DLE* para *clisos*.

4.7. *Endiñar*

En el *DLE* figuran dos acepciones para *endiñar*: 1) dar o asestar un golpe; 2) hacer que alguien acepte o cargue con algo, generalmente molesto, inoportuno o perjudicial. La etimología que se aporta sobre esta palabra es, simplemente, “voz *caló*”, sin dar más información (RAE, 2024). La forma *endiñar* no aparece documentada en ninguno de los diccionarios gitanos consultados. Sin embargo, sí aparece *diñar* en todos ellos, con el significado de ‘dar algo a alguien’ (Borrow, 1843: 45; Campuzano, 1848: 73; Quindalé, 1870: 66) y Jiménez y Rebolledo también incluyen *diñelar* (Jiménez 1846: 43; Rebolledo, 2006: 143). Resulta interesante que la formación de *endiñar* parece responder a la unión del prefijo *en-*, definido en el *DLE* como ‘dentro de’ o ‘sobre’, seguido de *diñar*. Este prefijo parece dotar a *diñar* de las connotaciones negativas que se desprenden de las acepciones, ya que lo dotan de cierta agresividad e imposición.

4.8. *Jai*

Esta palabra aparece recogida en el *DLE* con la acepción de “mujer joven y atractiva”, y la etimología que documentan es *jai*, sin aportar significado (RAE, 2024). Esta forma no aparece documentada en ninguno de los diccionarios gitanos que hemos consultado. Sin embargo, queremos aportar aquí una serie de informaciones que podrían contribuir a arrojar algo de luz sobre la etimología de esta palabra. En concreto, planteamos la hipótesis de que el verdadero origen de *jai* podría encontrarse en una forma alterada del *caló chai*, como indican Wagner (1962: 22) y Casas Gómez (1986: 229). Este *chai* también aparece en el *DLE* con dos acepciones: 1) niña

o joven; 2) prostituta y la etimología que aparece es *chai*, sin añadir significado (RAE, 2024). Si acudimos a los diccionarios de *caló* encontramos lo siguiente: Borrow (1843: 35) documenta *chai* como ‘niños, muchachos’; no aparece en Jiménez ni en Campuzano; Quindalé (1870: 22) y Rebolledo (2006: 176) incluyen *chai* para ‘niña, mocita’ y Mójica (2017: 63) consigna *chai* para los equivalentes ‘niños, muchachos, niña, mocita, ramera’.

Existe la posibilidad de que *jai* tan solo sea una alteración fonética de *chai*, ya que la primera forma no aparece en ningún diccionario, mientras que *chai*, además de guardar cierto parecido fonético, comparte ciertos rasgos de significado: ‘mujer joven y atractiva’ con ‘mocita’, ‘prostituta’, ‘muchachos’, ‘mocita’.

Para concluir este apartado, queremos comentar algunas deducciones a las que hemos llegado mediante el análisis de estas etimologías. No cabe duda de que acudir a los étimos resulta útil para el investigador si lo que se pretende es conocer la evolución de un vocabulario concreto. En el caso que aquí nos ocupa, el de los gitanismos, podemos afirmar que la muestra analizada representa, por una parte, la evolución experimentada por el léxico *caló* al mezclarse con el español: es lo que sucede en el caso de *acharar*, *achares* que han sufrido una transformación formal (partiendo de *jacharar*, *jachare*) pero también semántica, al producirse ese giro de significado de lo literal (quemazón, tormento) a lo metafórico (celos, penas o tormentos), algo similar a lo que ocurre con *chingar*. En el plano formal también encontramos modificaciones para vocablos que designan realidades similares, como sucede en *chaval*, *chavó* y las variantes formales que hemos documentado. La palabra *endiñar*, por último, también representa la transformación del léxico *caló* al verse influenciado por el español, en este caso en el plano morfológico mediante un mecanismo de prefijación, pasando de *diñar* a *endiñar* y ampliando, además, el valor connotativo de esta nueva unidad léxica.

Sin embargo, además de la evolución de los gitanismos, existen evidencias de que la lexicografía general no documenta con la exactitud necesaria la información etimológica, al menos en el caso del *caló*. Esta afirmación ha quedado reflejada en los casos de *bulo*, *churumbel*, *clisos* y *jai*, donde la etimología: 1) no refleja la esperable evolución del léxico *caló* si atendemos a lo que encontramos en diccionarios gitano-español, que, inevitablemente, son parte de este problema y 2) la etimología es prácticamente oscura e inexistente.

5. CONCLUSIONES

A la hora de construir la hipótesis sobre la que basamos este trabajo, planteamos la idea de que la lexicografía *caló* o, mejor dicho, la forma en la que se han ido creando diccionarios gitano-español, presenta bastantes deficiencias en lo que a rigor científico y sistematicidad se refiere, al haberse realizado por aficionados y para aficionados. Esto se traslada, inevitablemente, a las informaciones sobre el *caló* que recogen los diccionarios generales de la lengua, sobre todo en el caso de que hayan acudido a diccionarios de *caló* para obtener esas informaciones.

Tras realizar la investigación, hemos corroborado la hipótesis de la que partíamos y es que, como hemos visto en la muestra de ejemplos analizados, vemos casos en los que a) la etimología parece seguir una secuencia lógica en la evolución del *caló* en español; b) casos en los que existen variaciones de forma y de contenido y, por último, c) casos en los que la etimología resulta sorprendentemente oscura.⁷

Por lo tanto, podemos concluir que el *caló* ha sido y continúa siendo una variedad lingüística prácticamente desconocida incluso para los más experimentados, ya que, a pesar de que actualmente para cualquier lingüista o filólogo el *caló* se entiende como una variedad del romaní que hablan los gitanos en España, en la práctica, y centrándonos ahora en el nivel léxico,

⁷ Un claro ejemplo de *pogadolecto* lo encontraríamos en *endiñar*, que hemos analizado en el apartado correspondiente y cuya formación parece responder a un prefijo en- del español + *diñar* del *caló*.

no vemos con claridad qué se corresponde con el *caló* y qué corresponde a otra modalidad lingüística, como en el caso *churumbel* y la modalidad andaluza. De hecho, según algunos autores, el léxico *caló* ha perdido cualquier signo de identidad. Teniendo en cuenta esto, la definición de *caló* más actual y con la que nos mostramos más de acuerdo desde nuestra perspectiva es la siguiente:

El *caló* no es exactamente un dialecto del romanó, sino que es un *pogadolecto*. Este neologismo ha tenido que ser ideado por los lingüistas romaníes para definir una categoría de fenómenos lingüísticos que se dan en múltiples lugares del mundo y en muy variados contextos históricos y sociológicos. Un *pogadolecto* consiste en un habla cuya estructura gramatical está tomada de una lengua A, en la cual se inserta un léxico procedente de una lengua B [...] En el caso del *caló* de España, este *pogadolecto* ha tomado la estructura gramatical del español y sobre ella ha ido insertando el vocabulario romanó. (Jiménez González, 2009: 152)

Las implicaciones son claras y es, precisamente, lo que venimos demostrando en esta investigación: si lo que pretendemos es estudiar el léxico *caló*, los diccionarios gitano-español no representan una fuente verdaderamente fiable y todo apunta a que sucede lo mismo con los diccionarios generales de la lengua, en el caso de que para documentar la etimología se hayan basado en la lexicografía *caló*. Lamentablemente, la fiabilidad que se espera encontrar en los diccionarios, no queda reflejada en el caso de los gitanismos y para solventar las múltiples confusiones que esto provoca, es necesario remontarse a estudios sobre el léxico del *caló* que se alejen de la producción lexicográfica con fines lucrativos, como hemos visto en los trabajos de Wagner (1962) y Corominas (1987). La falta de estudios recientes sobre el *caló* se suma a esta problemática y es que, exceptuando a los trabajos de Buzek (2010, 2011 y 2016), la mayoría de fuentes de consulta fiables datan de años atrás.

Pensamos que la solución a las lagunas que deja la lexicografía podría encontrarse en la sociolingüística, es decir, acudir a las propias comunidades gitanas para conocer cuál es el estado del *caló* a día de hoy, algo que también presenta dificultades (*cf.*: Krinková, 2014). Todo parece indicar que en la actualidad sería complicado encontrar un “caló puro” que no haya estado sujeto a interferencias por el contacto con el español y con distintas modalidades de este, como los distintos argots de la delincuencia o lenguajes marginales. Pensamos que valdría la pena realizar más investigaciones como las de disponibilidad léxica de Gamella, Fernández, Niego y Adiego (2011, 2012, 2015), combinadas con otros estudios lingüísticos que combinen aspectos relacionados con los usos del léxico *caló* por parte de la comunidad o, quizá, el grado de conocimiento que la propia comunidad gitana pueda tener sobre el *caló*, ya que se sospecha que esta variedad ha quedado tan desatendida que incluso los romaníes la desconocen. Esta idea la expresa muy bien Jiménez al exponer sus consideraciones sobre el estado del *caló* dentro de la comunidad gitana (Jiménez González, 2009: 154).

AGRADECIMIENTOS

Este trabajo no hubiera sido posible sin los materiales conseguidos gracias a los proyectos de investigación llevados a cabo dentro del Instituto de Lingüística Aplicada (ILA) relacionados con la lingüística digital y la documentación: *Lingüística y Humanidades Digitales: base de datos relacional de documentación lingüística (PY18-FR-2511)*; *Lingüística y Nuevas Tecnologías de la información: la creación de un repositorio electrónico de documentación lingüística (FEDER-UCA18-107788)* y el proyecto que se encuentra actualmente en desarrollo *Aplicaciones de la lingüística digital al ámbito de la terminología: la creación de un léxico relacional bilingüe de los usos terminológicos de la semántica léxica (Terlexweb) (PID2022-139201OB-I00)*.

REFERENCIAS

- Besses, L. (2000). *Diccionario de argot español* (2ª ed.). Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Borrow, G. (1843). *The Zincali; or, an account of the gypsies of Spain, with an original collection of their songs and poetry, and a copious dictionary of their language*. London: John Murray.
- Buzek, I. (2010). *La imagen del gitano en la lexicografía española*. Brno: Masarykova Univerzita.
- Buzek, I. (2011). *Historia crítica de la lexicografía gitano-española*. Brno: Masarykova Univerzita.
- Buzek, I. (2016). ¿Qué quiere decir caló? Matices de un glotónimo. *Estudios de Lingüística del Español*, 37, 263-283. <https://doi.org/10.36950/elies.2016.37.8670>
- Campuzano, R. (1848). *Orijen, usos y costumbres de los jitanos, y diccionario de su dialecto con las voces equivalentes del castellano y sus definiciones*. Madrid: M. R. y Fonseca.
- Casas Gómez, M. (1986). Gitanismos designativos de la prostituta en el español moderno. *Romanistisches Jahrbuch*, 37, 225-239. <https://doi.org/10.1515/9783110244939.225>
- Casas Gómez, M. (2003). Hacia una tipología de la variación. En Moreno Fernández, F., Samper Padilla, J. A., Vaquero Ramírez, M. de, Gutiérrez Araus, M. L., Hernández Alonso, C., Gimeno Menéndez, F. (Coords.), *Lengua, variación y contexto*. Vol. 2 (pp. 559-574). Madrid: Arco Libros.
- Chabat, C. G. (1956). *Diccionario del caló. El lenguaje del hampa en México*. Guadalajara, Jalisco, México: Biblioteca Nacional de México.
- Clavería, C. (1949a). Miscelánea gitano-española II. Manús ‘hombre’. *Nueva Revista de Filología Hispánica*, 3(2), 158-160. <https://doi.org/10.24201/nrfh.v3i2.138>
- Clavería, C. (1949b). Miscelánea gitano-española III. Menda y mangué en el sistema pronominal español. *Nueva Revista de Filología Hispánica*, 3(3), 267-274. <https://doi.org/10.24201/nrfh.v3i3.3190>
- Clavería, C. (1951). *Estudios sobre los gitanismos del español*. Madrid: CSIC.
- Clavería, C. (1953a). Nuevas notas sobre los gitanismos del español. *Boletín de la Real Academia Española*, 33(138), 73-94.
- Clavería, C. (1953b). Terne. *Nueva Revista de Filología Hispánica*, 7(1-2), 127-133. <https://doi.org/10.24201/nrfh.v7i1/2.303>
- Corominas, J. (1954). *Diccionario crítico etimológico de la lengua castellana*. Madrid: Gredos.
- Corominas, J. (1987). *Breve diccionario etimológico de la lengua castellana*. Madrid: Gredos.
- Cuevas, C. (1974). *Arturo Reyes, su vida y su obra. Un enfoque humano del andalucismo literario*. Málaga: Publicaciones de la Caja de Ahorros Provincial.

Gamella, J., Fernández, C. Nieto, M. y Adiego, I. (2011). La agonía de una lengua. Lo que queda del *caló* en el habla de los gitanos. Parte I. Métodos, fuentes y resultados generales. *Gazeta de Antropología*, 27(2), 1-32. <http://dx.doi.org/10.30827/Digibug.19109>

Gamella, J., Fernández, C. Nieto, M. y Adiego, I. (2012). La agonía de una lengua. Lo que queda del *caló* en el habla de los gitanos. Parte II. Voces y campos semánticos más conocidos. *Gazeta de Antropología*, 28(1), 1-32. <http://dx.doi.org/10.30827/Digibug.20215>

Gamella, J., Fernández, C. y Adiego, I. (2015). Un vocabulario selecto del *caló* con datos sobre su conocimiento actual por una muestra de hablantes gitanos. *Revista de Lexicografía. Monográfico sobre la lexicografía caló*, 6, 129-132.

Gobello, J. y Payet, L. (1959). *Breve diccionario lunfardo*. Buenos Aires: Editorial A. Peña Lillo.

Guardia Civil. (1949). *Vocabulario del 'caló' de los maleantes. Primer volumen*. Madrid: Taller Escuela de Artes Gráficas de Huérfanos de la Guardia Civil.

Guardia Civil. (1950). *Vocabulario del 'caló' de los maleantes. Segundo volumen*. Madrid: Taller Escuela de Artes Gráficas de Huérfanos de la Guardia Civil.

Jiménez, A. (1846). *Vocabulario del dialecto gitano, con cerca de 3000 palabras y una relación exacta del carácter, procedencia, usos, costumbres, modo de vivir de esta jente en la mayor parte de las provincias de España, celebridad en las fiestas, nombres y apellidos más usuales, fisonomía y cuantos antecedentes se puedan tener de ellos, con varios rezos, cuentos, fábulas, versos, brindis, parte de la doctrina cristiana y ordenanza militar*. Sevilla: Imprenta del Conciliador.

Jiménez González, N. (2009). ¿El romanó, el *caló*, el romanó-kaló o el gitañol? Cincuenta y tres notas sociolingüísticas en torno a los gitanos españoles. *Anales de Historia Contemporánea*, 25, 149-161.

Krinková, Z. (2014). *Acerca de la problemática de la investigación sociolingüística de un léxico marginal: los gitanismos presentes en las variantes del español no estándar*. [Trabajo académico, Univerzita Karlova].

Macías Couso, A. (2021). Relaciones conceptuales entre las nociones variacionistas de argot y jerga mediante la terminología. *Revista Estudiantil Alma Mater*, 2, 112-124.

Mesa Navarro, R. M. (2013). *Estudio etimológico y lexicográfica del Diccionario Gitano de Francisco Quindalé. Tomos I y II*. Tesis doctoral. Sevilla: Universidad de Sevilla.

Mohedano Gallardo, M. (2018). Los juegos de naipes: resistencias cotidianas en Rute (Córdoba) en la segunda mitad del siglo XVII. *Revista de Estudios de Ciencias Sociales y Humanidades*, 39, 49-64.

Mójica Legarre, J. M. (2017). *Vocabulario Caló-Español Español-Caló*. Recuperado de: https://www.academia.edu/11649711/Vocabulario_Caló_Español_Español_Caló

Quindalé, F. (1870). *El gitanismo. Historia, costumbres y dialecto de los gitanos. Con un epítome de gramática gitana, primer estudio filológico publicado hasta el día, y un diccionario caló-castellano, que contiene, además de los significados, muchas frases ilustrativas de la acepción propia de las palabras dudosas* (2ª ed.). Madrid: Librería de Victoriano Suárez.

RAE. (2024). *Diccionario de la Lengua Española* (23ª ed.). [versión 23.7 en línea].

Rebolledo, T. (2006). *Diccionario gitano-español y español-gitano* (edición facsimilar de la de 1909). Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.

Ropero Núñez, M. (1978). *El léxico caló en el lenguaje del cante flamenco*. Sevilla: Universidad de Sevilla.

Ropero Núñez, M. (1991). *El léxico caló en el lenguaje del cante flamenco* (2º ed.). Sevilla: Universidad de Sevilla. <https://doi.org/10.12795/9788447220830>

Ropero Núñez, M. (1992). Un aspecto de lexicología histórica marginado: los préstamos del caló. En M. Ariza Viguera, R. Cano-Aguilar, J. M. Mendoza Abreu y A. Narbona Jiménez (Eds.), *Actas del II Congreso Internacional de Historia de la Lengua Española Vol. 1* (pp. 1305-1314). Sevilla: Pabellón de España.

Ropero Núñez, M. (1999). Los préstamos del caló en el DRAE. En P. Gómez Manzano, P. Carbonero Cano y M. Casado Velarde (Coords.), *Lengua y Discurso: estudios dedicados al profesor Vidal Lamiquiz*, pp. 843-852. Madrid: Arco Libros.

Ropero Núñez, M. (2005). Tratamiento lexicográfico y sociolingüístico de los gitanismos en el diccionario, en M. Casas Gómez y G. Fernández Smith (Coords.), *IX Jornadas de Lingüística*, pp. 45-82. Cádiz: Universidad de Cádiz.

Roumagnac, C. (1904). *Los criminales de México: ensayo de psicología criminal*. México D.F.: Tipografía El Fénix.

Saavedra, A. M. (1941). El “Caló” de la Delincuencia y la Expresión Sexual. *Anuario de la Sociedad Folklórica de México*, 2, 23-38.

Salillas, R. (1896). *El delincuente español. El lenguaje. Estudio filológico, psicológico y sociológico, con dos vocabularios jergales*. Madrid: Librería de Victoriano Suárez.

Serrano García, P. (1935). *Delincuentes profesionales contra la propiedad: su clasificación, procedimiento para apoderarse de lo ajeno que usualmente emplean los profesionales del robo, el hurto, la estafa, la falsificación, el chantaje, etc. Estados peligrosos o próximos al delito de delincuentes patológicos. Con un vocabulario del “caló”*. Madrid: Imprenta Justo López.

Suárez Solís, S. (1969). *El léxico de Camilo José Cela*. Madrid: Alfaguara.

Wagner, M. L. (1941). Sobre algunas palabras gitano-españolas y otras jergales. *Revista de Filología Española*, 25, 168-181.

Wagner, M. L. (1962). El abolengo gitano-indio de ‘chavó’ y su familia. *Revista de Filología Española*, 45, 305-310. <https://doi.org/10.3989/rfe.1962.v45.i1/4.932>



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 151-168
Artículo recibido: 14/08/2024
Artículo aceptado: 21/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/665>
DOI: <https://doi.org/10.58859/rael.v23i1.665>

La diversidad del español en *Instagram*. Entre entretenimiento y posibilidades didácticas

Spanish diversity on Instagram. Between entertainment and educational possibilities

LISA GOZZI

SILVIA BETTI

UNIVERSITÀ DI BOLOGNA

En una era de continuo progreso tecnológico, la variación lingüística del español se ha extendido también a las redes sociales. Este fenómeno se ha convertido en tema de publicaciones compartidas por diversos usuarios, ganando notoriedad en el ámbito virtual. En esta contribución, se examinan contenidos audiovisuales de *Instagram* para observar cómo se aborda la riqueza del español en la plataforma. El análisis se centra en cuatro categorías de publicaciones, siendo estas la variación léxica, los choques culturales entre España y Colombia/Argentina, el orgullo nacional canario y el llanito. Asimismo, se observa la postura adoptada por los autores, sea esta humorística o didáctica, ante el tema tratado. Finalmente, se consideran las posibilidades educativas brindadas por el uso potencial de la plataforma en los contextos de español como lengua extranjera. En definitiva, el estudio analiza cómo *Instagram* refleja la realidad lingüística del español y el sistema de creencias asociado a esta.

Palabras clave: *variedades; léxico; cultura; didáctica de ELE; Instagram*

During a time of continuous technological progress, Spanish linguistic variety has extended to social networks as well. This phenomenon has become the focus of posts shared by various users and has gained notoriety in virtual environments. This contribution analyses audiovisual content on Instagram in order to observe how Spanish richness is treated on the social network. The analysis focuses on four categories of posts, namely lexical variety, cultural shocks between Spain and Colombia/Argentina, Canarian national pride and the case of *llanito*. It also inspects the attitude, be it humoristic or didactic, embraced in the contents towards the presented varieties. Lastly, it explores the didactic possibilities created by the potential use of social media in the contexts of Spanish as foreign language. Finally, the work wants to observe how Instagram reflects the linguistic reality of the Spanish language and the belief system associated with it.

Keywords: *varieties; vocabulary; culture; SFL teaching; Instagram*

1. INTRODUCCIÓN

En el contexto actual, caracterizado por una constante transformación y evolución, los entornos cotidianos se ven atravesados por cambios vertiginosos. Entre los elementos que desempeñan un papel crucial en este proceso continuo de mutación se encuentran los productos tecnológicos, particularmente las redes sociales. Estas han revolucionado la forma en que comunicamos a diario, ofreciendo posibilidades de interacción instantánea y proporcionando un alcance masivo para los mensajes publicados. Además, permiten establecer relaciones con otros usuarios de las plataformas, generando comunidades virtuales de considerable magnitud (Mancera Rueda y Pano Alamán, 2013).

Como usuarios de las redes sociales, y en particular de *Instagram*, este estudio se enfoca en el análisis del potencial de la plataforma digital como medio de intercambio cultural, superando las barreras geográficas y facilitando el contacto de los usuarios con culturas distantes a través del contenido compartido en dichas redes. El objetivo de esta investigación exploratoria es examinar la representación de temáticas relacionadas con la variación diatópica y cultural del español. En particular, se propone destacar la propuesta de publicaciones que resaltan la riqueza lingüística y las diferencias derivadas de la vasta extensión territorial de la lengua española, al tiempo que se analiza la actitud de los creadores de contenidos en la presentación de estas características lingüísticas y culturales.

Desde un enfoque metodológico, se abordan cuatro ejes específicos en torno a la variación lingüística: (1) la diversidad léxica a nivel regional, (2) los conflictos culturales que surgen a partir de estas variantes, (3) el orgullo identitario de los canarios respecto a su modalidad de habla y cultura, y, finalmente, (4) la presentación del llanito como una variante lingüística característica de Gibraltar. Para llevar a cabo el análisis, se han examinado 80 publicaciones seleccionadas a través de *hashtags* específicos y por la popularidad alcanzada, sin que se haya podido determinar la periodicidad de su publicación, como se detallará en secciones posteriores.

2. LA RESILIENCIA Y DINAMISMO DEL ESPAÑOL EN EL CONTEXTO GLOBAL

De acuerdo con el último informe del Instituto Cervantes (2024), el español es una lengua de considerable extensión demográfica, con más de 600 millones de hablantes, incluyendo a usuarios nativos, personas con competencias limitadas y estudiantes de español como lengua extranjera (en adelante, ELE). Además, por su amplia difusión geográfica, el español se distingue por ser lengua oficial en 21 países, distribuidos en cinco continentes, y uno de los idiomas oficiales en importantes organismos internacionales como la ONU (Instituto Cervantes, 2024).

En este contexto geolingüístico, se observa una notable variación diatópica de la lengua, que se materializa en diversas modalidades regionales con características específicas. Según Moreno Fernández (2017; 2020; 2021), las variantes geolectales se agrupan en ocho zonas distintas, basándose en los rasgos distintivos de cada una. Tres de estas áreas se localizan en España, correspondientes a las variedades castellana, andaluza y canaria. Las cinco restantes se encuentran en América Latina, identificadas como las variedades caribeña, mexicano-centroamericana, andina, rioplatense y chilena (Moreno Fernández (2017; 2020; 2021). Cada una de estas áreas muestra características lingüísticas distintivas en los distintos niveles del idioma, incluyendo los aspectos fonéticos, morfosintácticos, léxicos y pragmáticos. Estos elementos lingüísticos contribuyen a conformar la identidad hispana de los hablantes en cada región.

Además de estas características, es fundamental destacar que, debido al contacto con otras lenguas, han surgido variedades mixtas que combinan elementos del español con los de la lengua de contacto. Dentro del ámbito hispano, sobresalen dos modalidades de contacto originadas por la convivencia con el inglés: el *spanGLISH* y el llanito.

El término *spanglish* evoca la particular forma de comunicación que caracteriza a las comunidades de origen hispano radicadas en los Estados Unidos. Derivado de la combinación de *span-* (de *Spanish*) y *-glisch* (de *English*), el *spanglish* puede considerarse, en cierta medida, como el fruto del encuentro —o, en algunos casos, del choque— entre dos universos, dos sensibilidades, dos culturas y dos lenguas: la hispana y la anglosajona. Esta práctica representa una forma de comunicación particularmente interesante entre los latinos bilingües, quienes la utilizan como una alternativa a la comunicación monolingüe. El *spanglish* no solo puede convertirse en un medio de comunicación efectivo en contextos específicos, sino que también es visto por quienes lo practican como un signo de una nueva identidad *in-between*. Es posible que simbolice una identidad en proceso de negociación para la integración de los latinos en los Estados Unidos, ya que este fenómeno lingüístico implica mucho más que el simple dominio de dos idiomas; es una “forma de vida” que expresa la riqueza y diversidad de culturas entrelazadas (Betti, 2024).

El llanito, por otro lado, caracteriza la realidad lingüística de Gibraltar: de hecho, a pesar de localizarse en la Península Ibérica a nivel geográfico, políticamente forma parte del Reino Unido, siendo el inglés, por consiguiente, su lengua oficial. En este contexto, el contacto histórico entre el inglés, el español y otras lenguas ha dado lugar a una modalidad híbrida única conocida como llanito (Moreno Fernández, 2020).

Es evidente, por lo tanto, que la lengua española varía notablemente en su dominio territorial y, en el mundo globalmente interconectado actual, esta diversidad ha pasado a formar parte también de los recursos digitales y de las redes sociales.

3. LAS REDES SOCIALES: EL CASO DE *INSTAGRAM*

Hoy en día, como afirma Yus (2017: 55), “Internet está imbricado e hibridado en la vida cotidiana de millones de personas que ya no se conectan a la Red, sino que viven con total naturalidad dentro de ella”. Los avances tecnológicos han creado nuevos medios de comunicación e interacción que han revolucionado profundamente la forma humana de establecer relaciones comunicativas. Entre los nuevos medios de comunicación destacan las redes sociales, las que representan una herramienta tecnológica de significativa importancia. Estas constituyen un recurso en línea a través del que se pueden establecer contextos comunicativos de diversa tipología que permiten una comunicación inmediata y personalizada, que deja la posibilidad de crear un diálogo constante y colaborativo (Mancera Rueda y Pano Alamán, 2013).

En el caso de *Instagram*, los usuarios interactúan principalmente compartiendo imágenes y/o vídeos en la plataforma. Las formas de interacción en la plataforma han evolucionado significativamente desde su nacimiento hasta hoy y pueden ser muy variadas: en concreto, los usuarios pueden utilizar *posts*, es decir, imágenes y vídeos colgados en el propio perfil, e historias, que permanecen visibles durante 24 horas después de su publicación, pero pueden emplear también *hashtags*, comentarios, y *likes* o “me gusta”, así como pueden dialogar a través de grupos o mensajes privados (Figueroa-Benítez, González-Quñones y Machin-Mastromatteo, 2021).

Entre los usuarios de la plataforma, se encuentra tanto a los jóvenes y adultos, que la usan como herramienta privada, como a profesionales y políticos, empresas, científicos académicos e instituciones, llegando a dar espacio también a la figura de los *influencers*. En 2021, *Instagram* era la sexta red social más utilizada en el mundo y contaba con cerca de 1000 millones de usuarios activos mensuales (Figueroa-Benítez et al., 2021).

En tal entorno digital, un acercamiento a los contenidos de *Instagram* ha permitido destacar una notable presencia de publicaciones relacionadas con la variación lingüística de la lengua española, que se convierten en potenciales muestras reales del idioma a disposición de los

usuarios. En concreto, estas publicaciones permiten una mayor exposición de los internautas a la riqueza lingüística, hecho que puede conllevar una mayor difusión y aceptación de la misma, al tiempo que consiente la promoción de la identidad cultural y del orgullo nacional.

Esta tipología de publicaciones puede no solo jugar un papel fundamental en la difusión de contenidos dedicados a la identidad hispana utilizando el humor y la diversión, sino que también proporciona interesantes posibilidades didácticas, puesto que representa una fuente de información y formación acerca de la diversidad lingüística y brinda la posibilidad de crear actividades didácticas, promoviendo la reflexión metalingüística.

La incorporación de la diversidad del español en la didáctica de ELE ha sido objeto de estudios diversificados, como los de Andión Herrero (2007; 2008; 2009; 2019), Moreno Fernández (2017; 2021) y la reciente tesis doctoral de Gozzi (2024). En estos, se subraya la importancia de incorporar fundamentos de variación a la clase como parte del desarrollo de las competencias del estudiante; sin embargo, se detectan numerosos obstáculos al proceso de inclusión, entre los que destacan la ocasional escasa formación del profesorado y las faltas en los manuales de ELE. Por ende, las redes sociales podrían emplearse como medio para llenar el vacío de los materiales tradicionales.

Como señalan Carpenter, Morrison, Craft y Lee (2020, citado en Figueredo-Benítez, González-Quiñones y Machin-Mastromatteo, 2021), la literatura relacionada con el uso didáctico de las redes sociales está todavía desarrollándose. Sin embargo, como detallan Figueredo-Benítez, González-Quiñones y Machin-Mastromatteo (2021), resumiendo una serie de estudios relacionados con el uso de *Instagram* en los ámbitos educativos, algunas investigaciones han revelado que la red social fomenta las habilidades de los alumnos, que están comprendidos en el grupo de edad que hace mayor uso de la plataforma, y que su empleo presenta beneficios en el proceso educativo. También el estudio de Martí Climent y García Vidal (2018) hace hincapié en las numerosas posibilidades didácticas brindadas por la aplicación, aportando ejemplos de fructíferas utilidades de *Instagram* en el aula; entre estos, figura la *Valencianada* (#VLCda), organizada por la Asociación de profesorado *Ientretants*, que propone a estudiantes de segundo ciclo de Secundaria y de Bachillerato el uso de esta y otras aplicaciones para explorar los sitios de interés cultural de Valencia. Otro ejemplo aportado por Martí Climent y García Vidal (2018) es el de #*Instacacera*, propio de algunos Másteres de Secundaria, que propone a los alumnos de segundo ciclo de Educación Secundaria Obligatoria crear una publicación que analice algún neologismo, barbarismo y préstamo léxico encontrado en la fase de búsqueda prevista en la actividad propuesta.

4. LA VARIACIÓN DEL ESPAÑOL EN *INSTAGRAM*: OBJETIVOS Y METODOLOGÍA DE INVESTIGACIÓN

A partir de lo afirmado en la revisión de literatura, este estudio exploratorio tiene como objetivo observar cómo se incorporan las variedades lingüísticas del español en la plataforma *Instagram*. Específicamente, se pretende verificar las siguientes preguntas de investigación:

- ¿Qué actitudes predominan en las publicaciones consideradas? ¿Se presentan desde una perspectiva humorística, didáctica, o combinan ambas dimensiones?
- ¿Se detectan estereotipos o prejuicios en los contenidos analizados?

La metodología empleada para este estudio, de enfoque cualitativo, se estructuró en dos fases: en la primera etapa, se utilizó la función “buscar y explorar” de *Instagram* para identificar contenidos relacionados con los temas de interés organizados temáticamente mediante etiquetas (*hashtags*) (Androutopoulos, 2013), sin que fuera posible determinar la periodicidad de las publicaciones. Los resultados sugeridos por la plataforma aparecieron en la sección de “popu-

lares”¹, lo que indica que se trataba de contenidos que habían alcanzado una mayor visibilidad según los criterios de búsqueda aplicados. Cabe destacar que la búsqueda se realizó mediante un enfoque *screen-based*, es decir, considerando los datos generados por los usuarios de la plataforma, y el rol del investigador se limitó a la selección y recopilación de los elementos pertinentes (Androutsopoulos, 2013).

Los *hashtags*, utilizados como base para la búsqueda, “son metadatos explícitos integrados en el mensaje, mediante los cuales se señala el tema o temas [...] y se conecta ese mensaje a otros que contienen esa misma etiqueta. En general, están formados por palabras o secuencias de palabras precedidas de la almohadilla (#), que se convierten en hiperenlaces una vez insertados en el mensaje” (Pano Alamán, 2021: 48). En el presente estudio, las etiquetas² empleadas para la búsqueda fueron las siguientes:

- 1) #learns spanish
- 2) #argentinos en espana
- 3) #latinamerica
- 4) #colombianos en espana
- 5) #llanito
- 6) #canarias

La segunda etapa se apoyó en el algoritmo de *Instagram* que, después de la búsqueda inicial, propuso contenidos relacionados con la misma. Los criterios de selección de los *posts* en esta segunda etapa fueron:

- 1) Nivel de interacción con los usuarios.
- 2) Tendencia y popularidad.
- 3) Pertinencia con los temas investigados.

Para facilitar el análisis de los contenidos seleccionados, se empleó una plantilla creada específicamente para el estudio en formato Excel. En esta, se crearon las siguientes secciones:

- 1) Tipo de contenido: imagen, vídeo, carrusel, *reel*.
- 2) Contenido: breve resumen de los contenidos presentados en el archivo multimedia.
- 3) Actitud: humorística, didáctico-divulgativa, humorístico-didáctica.

Esta estrategia metodológica permite abordar con rigor y detalle los objetivos planteados, proporcionando una perspectiva integral sobre la representación de las variedades lingüísticas del español en una de las plataformas digitales más influyentes de la actualidad.

5. RESULTADOS

La muestra final recogida se compone de un total de 80 publicaciones relacionadas con los temas de investigación. Más en detalle, se distribuyen de la siguiente forma, según los cuatro ejes de estudio considerados (véase Tabla 1).

¹ La popularidad de la publicación depende del número de interacciones alcanzado por esta. El algoritmo de la *Instagram* selecciona estas publicaciones relacionadas con un criterio de búsqueda concreto y las reúne en una página dedicada.

² Los *hashtags* que se mencionan se eligieron como criterio de selección después de un análisis exploratorio inicial que destacó la presencia de un número relevante de contenidos relacionados con los mismos. Por ende, se establecieron como pauta de elección de las publicaciones para analizar.

Tabla 1: *División temática de las publicaciones de Instagram analizadas*

Tema	N. Publicaciones
Variación léxica	24
Choques culturales y lingüísticos (España vs Argentina/Colombia)	27
Orgullo canario	13
Llanito	16
<i>Total</i>	80

Según el análisis de los datos generales, cada enfoque de investigación ha capturado una muestra de publicaciones que varía entre 13 y 27 elementos. En cuanto a la variación léxica y los choques culturales, se observa que los contenidos provienen de diversas cuentas, aunque algunas han contribuido con más publicaciones que otras. En el caso del geolecto canario, todas las publicaciones se concentran en una sola cuenta, de la que se proporcionan mayores detalles a continuación. Respecto al llanito, los contenidos sugeridos por el algoritmo de la plataforma son más limitados y, aunque se originan en diferentes cuentas, se agrupan mayoritariamente en torno a un usuario específico.

A continuación, se presentan los datos observados para cada tema analizado.

5.1. *Variación léxica*

Por lo que respecta a la variación léxica, los contenidos seleccionados incluyen material audiovisual que explora las diferencias en el vocabulario utilizado en situaciones cotidianas dentro del mundo hispanohablante.

En cuanto al formato de los contenidos seleccionados, se evidencia una predominancia de *reels*, aunque también se ha recopilado una muestra considerable de carruseles con imágenes que presentan varios ejemplos de variantes léxicas, ocasionalmente acompañadas de traducciones al inglés o de imágenes explicativas sobre los términos abordados en las publicaciones. No se han identificado ejemplos de imágenes o vídeos individuales (véase Figura 1).

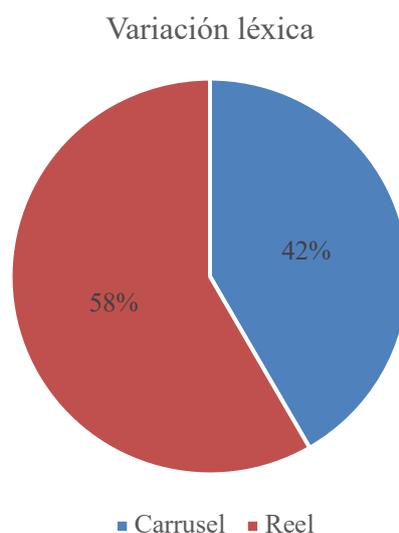


Figura 1: *Distribución de los contenidos según su tipo (variación léxica)*

En segundo lugar, respecto a la actitud adoptada en los *posts* hacia el contenido presentado, se observa que la mayoría (66,7%) adopta un enfoque didáctico-divulgativo, con el propósito de ofrecer información útil a los usuarios. Solo en 8 casos se ha identificado una actitud mixta, donde una intención didáctica se combina con un tono humorístico; este último es más común en los *reels* analizados, mientras que el enfoque exclusivamente didáctico prevalece en los carruseles (véase Figura 2).

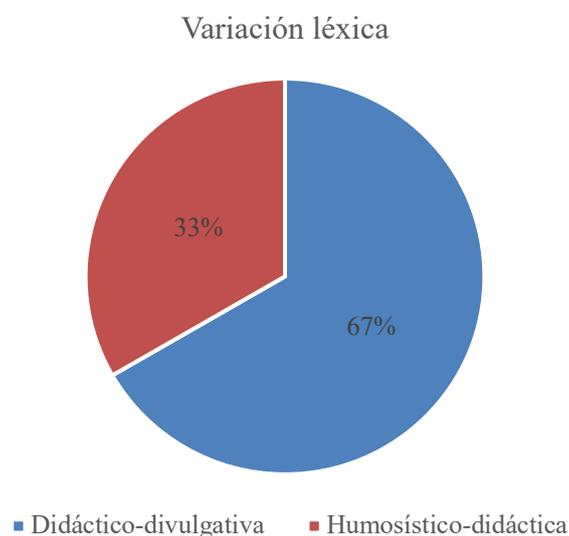


Figura 2: *Distribución de los contenidos según su actitud (variación léxica)*

Las publicaciones consideradas presentan diferentes enfoques hacia la variación léxica; en ocasiones, se proponen las diferencias geolectales en el léxico sin algún tipo de actitud o postura hacia las mismas, sino que se consideran parte de la riqueza de la lengua. Un ejemplo es el caso de un *reel* que presenta las diferentes formas de decir “chanclas” en el mundo hispano: en este, se explica que en Argentina se les llama “ojotas”, en España, México y Colombia “chanclas”, en Venezuela “chola”, en Chile “chalias”, en Honduras, Guatemala, Puerto Rico, Costa Rica “chancletas”, en Uruguay “chinelas” y en Paraguay “zapatillas” (contenido presentado por @chofitamanza en *Instagram*). Otra muestra es la del carrusel que opone los usos del español peninsular al americano, donde se presentan los siguientes vocablos:

Tabla 2: *Ejemplo de carrusel que muestra la variación léxica. Contenido adaptado de @spanishstefan en Instagram*

Español peninsular	Español latinoamericano
Coger (tomar)	Coger (tener relaciones sexuales)
Ahora	Ahorita
Papa (jefe de la iglesia cristiana)	Papa (patata)
Conducir	Manejar
Gafas	Lentes

El ejemplo resumido en la Tabla 2 se conecta a otro con enfoque más específico hacia la temática considerada, donde se pone la atención en esas variantes léxicas que, por su significado variable en el mundo hispano, pueden representar una fuente de choques culturales, al tener

un término concreto, en algún país hispanohablante, una acepción tabú. Es el caso de palabras como “coger”, citado anteriormente, que aparece en varias de las publicaciones analizadas, y también de “concha”, que, en algunas zonas como Argentina, indica vulgarmente las partes íntimas femeninas.

Asimismo, algunos *posts* presentan ejemplos de la jerga regional propia de un área concreta, para acercar a los usuarios a un español real que desvía de los usos estándares; por ejemplo, el mismo @spanishstefan propone un carrusel con los usos propios de la jerga colombiana y otro con los típicos de la mexicana, que se resumen en las tablas siguientes:

Tabla 3: *Ejemplo de carrusel que muestra la jerga mexicana. Contenido adaptado de @spanishstefan en Instagram*

Español peninsular	Jerga mexicana
Ahora	Ahorita
Mamá	Vieja
Trabajo	Chamba
No me digas	No manches
Chico	Chavo

Tabla 4: *Ejemplo de carrusel que muestra la jerga colombiana. Contenido adaptado de @spanishstefan en Instagram*

Español peninsular	Jerga colombiana
Amigo	Parcero
Asqueroso	¡Guácala!
Grupo de amigos	Parche
¿Cómo estás?	¿Quiubo?
Fiesta	Rumba

Por último, es preciso señalar que cinco de las publicaciones seleccionadas se han elaborado desde el enfoque del aprendiz de ELE, que siente confusión a la hora de enfrentarse con elementos léxicos variables de la lengua. En los *reels* seleccionados para la muestra del estudio que adoptan esta tipología de perspectiva, se aprecia cierta atención hacia los siguientes elementos:

- 1) Cómo se dice “tú eres” en español (en el *reel*, el protagonista lo pregunta a Alexa en inglés), considerando la variación morfológica de la expresión y presentando también las estructuras “vos sos” y “usted es”. También presenta la oposición entre los usos de “ser” y “estar”, que no es propia de otros idiomas, como el inglés.
- 2) Cómo se dice “pajita” en español.
- 3) Cómo se dice “palomitas” o *popcorn* en español.

Finalmente, por lo que respecta a la posible presencia de estereotipos hacia los elementos presentados, en este caso no se han detectado datos de relieve y no se evidenciaron actitudes negativas. La única parcial excepción está vinculada a la confusión que sienten los estudiantes de ELE ante el amplio abanico de posibilidades que proporciona el repertorio léxico hispano.

5.2. Choques culturales y lingüísticos: problemas de comunicación entre españoles y colombianos/argentinos

El segundo núcleo temático en el que se enfoca el estudio es el de los choques culturales y lingüísticos que se pueden crear entre españoles e hispanoamericanos por las diferencias existentes entre las diversas modalidades. Entre las numerosas posibilidades de choque que se pueden verificar en los varios países hispanohablantes, se han seleccionado dos regiones específicas, en concreto, la colombiana y la argentina. Esta elección se justifica por el hecho de que, durante la primera fase de exploración de *Instagram* necesaria para establecer los ejes del estudio, se evidenció una presencia significativa de contenidos relacionados con estas zonas.

Por lo que respecta al tipo de publicación, estas se encontraron exclusivamente en *reels*, por un total de 27 publicaciones analizadas. Asimismo, acerca de la actitud adoptada, se aprecia que, en este caso, la mayoría de ellas presenta carácter principalmente humorístico ante los choques derivados de la diversidad lingüística entre los países considerados. Solo en dos casos, correspondientes al 7.4% de la muestra, se ha destacado una actitud didáctico-divulgativa (véase Figura 3).

Choques culturales y lingüísticos

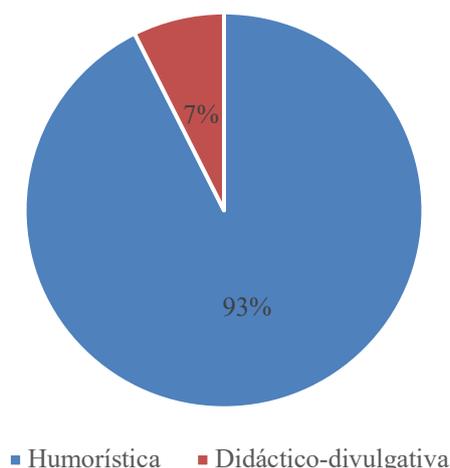


Figura 3: **Distribución de los contenidos según su actitud (choques culturales y lingüísticos entre España y Colombia/Argentina)**

En los enfoques adoptados por los autores de los contenidos, se destacaron dos perspectivas principales: la primera se centra en los posibles malentendidos entre los interlocutores debidos a las diferencias de significado de algunos elementos léxicos, la segunda considera las dificultades de comunicación derivadas de los diferentes usos léxicos y lingüísticos en el dominio territorial hispanohablante.

En el primer caso, se reunieron varias muestras, entre las que se citan las siguientes como ejemplo:

- 1) “Bruto”, en español peninsular, significa “brusco”, mientras que en la modalidad colombiana indica a una persona poco inteligente. Por consiguiente, en el *reel* considerado, cuando el emisor español le dice a la chica colombiana que es un poco bruta, esta se ofende (contenido presentado por @rolaenmadrid en *Instagram*).

- 2) En España, “currar” es sinónimo de “trabajar”, mientras que en Argentina significa “estafar”. Por consiguiente, la versión argentina de la actriz que interpreta los dos papeles en el *reel* queda un poco sorprendida cuando su versión española habla de su “curro”, estando muy orgullosa del mismo (contenido presentado por @anacarolina-parisi en *Instagram*).
- 3) Una chica colombiana dice que le gusta “jugar la polla”, refiriéndose a las apuestas deportivas, cuando mira los partidos de fútbol, pero la palabra “polla” tiene un significado vulgar en el español peninsular que deja desconcertado y confundido al interlocutor español (contenido presentado por @rolaenmadrid en *Instagram*).

En el segundo caso, que se focaliza en las dificultades comunicativas debidas a la variación, se citan las siguientes muestras como ejemplo:

- 1) En una tienda española, una chica argentina no sabe si utilizar “ordenador”, para que los dependientes la entiendan, o “computadora”, para mantenerse fiel a sus orígenes. Sin saber que hacer, la chica usa el anglicismo *laptop* (contenido presentado por @dosartistasviajeros en *Instagram*).
- 2) La camarera colombiana dice al cliente español que le va a “regalar” y “cancelar” la cuenta, para indicar que se la va a llevar, pero este entiende que no tiene que pagar y se va del restaurante antes de que la camarera pueda cobrarle (contenido presentado por @rolaenmadrid en *Instagram*).
- 3) Un dependiente argentino acaba de empezar a trabajar en una tienda de ropa española, sin embargo, al atender al primer cliente, no consigue entender lo que este le pide; en concreto, no comprende palabras como “sudadera”, “polo” y “sujetador”, entre otras, y termina escapándose (contenido presentado por @joaquin_castellano en *Instagram*).

Con respecto a los estereotipos detectables en los contenidos considerados, en esta ocasión, se notó la presencia de algunos prejuicios. Entre estos, destaca la idea de los españoles de que los sudamericanos son más extrovertidos y amigables; asimismo, se señala la opinión peninsular según la que la variante argentina es difícil de entender, siendo este último un estereotipo válido también para la situación contraria, es decir, que para los argentinos es muy difícil entender la variedad castellana.

5.3. *El orgullo por la identidad canaria*

El tercer eje temático aborda los contenidos que expresan el orgullo de los canarios por su identidad geográfica, cultural y lingüística. La investigación de Guerrero Salazar (2020), dedicada al estudio de actitudes ante la variación lingüística, destaca que los hablantes del archipiélago valoran positivamente su variedad, defendiéndola frente al dominio de la modalidad estándar y reclamando una mayor presencia de esta en los materiales educativos. Asimismo, la autora subraya que la prensa canaria adopta una postura notablemente positiva hacia su variedad lingüística, la cual no suele asociarse con estereotipos negativos ni utilizarse para desacreditar a figuras públicas. Otro ejemplo del orgullo identitario se encuentra en el texto de Trapero Trapero (2024), donde el autor, de origen canario, describe la festividad típica de la Bajada de La Palma y expresa: “Verdaderamente sentí envidia de ser palmero. Orgulloso me sentiría yo, de serlo, de pertenecer a un pueblo que ha sido capaz de configurar unas fiestas como las que celebra en su Bajada. [...] Y bien quedó patente para mí que pocos pueblos habrá más orgullosos de sus fiestas que el palmero” (Trapero Trapero, 2024: 31).

En el presente estudio se examinaron 13 publicaciones, todas ellas en formato *reel*, creadas por @mrwiott. En esta cuenta, el creador de contenidos canario presenta publicaciones sobre temas variados con el objetivo de entretener, divertir e informar a su audiencia. En los breves vídeos analizados, el autor adopta predominantemente una actitud humorística, combinando solo en uno de los trece *reels* revisados el humor con un enfoque didáctico, en un esfuerzo por celebrar la identidad canaria y el orgullo por su cultura.

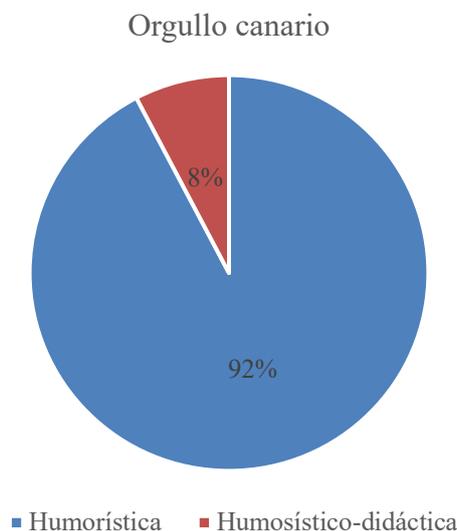


Figura 4: *Distribución de los contenidos según su actitud (el orgullo por la identidad canaria)*

Un ejemplo notable de este orgullo identitario es la irritación que manifiesta el protagonista del *reel*, un canario de Gran Canaria, al escuchar a un madrileño decir que tiene un amigo de Tenerife, como si ambos procedieran del mismo lugar, lo que denota un desconocimiento de la geografía del archipiélago por parte del hablante de Madrid. Otro caso es la molestia que sienten los canarios al escuchar a un compañero cambiar su acento o su forma de hablar después de pasar tiempo en la Península Ibérica, al viajar o al interactuar con hablantes peninsulares. Además, el creador de los contenidos muestra cierto grado de enfado hacia los españoles que tratan el acento canario con ironía, como algo cómico.

En cuanto a los estereotipos asociados con esta temática, se observa la presencia ocasional de algunos en los *reels* propuestos por la cuenta anteriormente mencionada. Por ejemplo, en un vídeo se asocia a las Islas Canarias la costumbre de poner nombres peculiares a los niños, como “Cathaysa” o “Chaxiraxi Chakira” (@mrwiott en Instagram). También se alude al prejuicio de que los canarios utilizan expresiones cariñosas como “mi amor” o “mi cielo” al interactuar con los demás, independientemente del nivel de confianza, y a otro según el cual los españoles de la Península Ibérica poseen un conocimiento limitado respecto a las islas. Asimismo, se hace referencia a cuestiones cotidianas como el hecho de que los pedidos *online* tardan mucho más en llegar a Canarias en comparación con otras zonas peninsulares.

5.4. *El caso del llanito*

El último grupo de publicaciones analizadas concierne al llanito, siendo esta una modalidad lingüística de contacto que mezcla el español con elementos propios de otros idiomas. Como ya se ha mencionado en la revisión de la literatura, el llanito es un producto lingüístico derivado del contacto entre inglés y español, combinándose asimismo con elementos propios de otras

lenguas como el portugués y el genovés, que caracteriza la región de Gibraltar (Moreno Fernández, 2020). Es, por lo tanto, una forma de hablar que alterna y mezcla componentes de los códigos lingüísticos de los que deriva.

En el caso de la presente investigación, se ha considerado oportuno examinar la presentación de esta modalidad en *Instagram*, al representar esta el resultado de un contacto lingüístico en territorio europeo y al compartir algunos rasgos fundamentales con otra modalidad de contacto, el *spanglish*.

Observando el formato de las 16 publicaciones seleccionadas, se nota que este es más variado respecto al de los ejes temáticos anteriormente considerados; de hecho, los contenidos se dividen igualmente entre imágenes y vídeos. La mayoría de las publicaciones son imágenes individuales (50%) y vídeos (31.3%), mientras que solo hacia el 20% son *reels*.

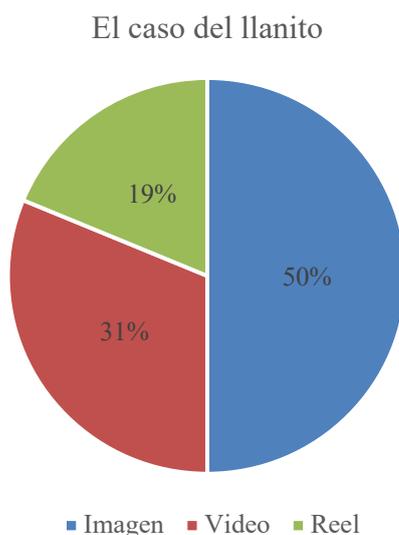


Figura 5: *Distribución de los contenidos según su tipo (el caso del llanito)*

Por lo que respecta, en cambio, a la actitud adoptada en las publicaciones, destaca en este caso que todas adoptan un enfoque didáctico-divulgativo, que busca proporcionar a los usuarios de *Instagram* nociones y “píldoras de conocimiento” acerca del llanito como modalidad propia y marca identitaria de Gibraltar.

Ejemplos de los contenidos propuestos son tanto la presentación del llanito como lengua y la celebración del orgullo hacia esta modalidad por la riqueza lingüística propia de la zona, así como ejemplos de palabras en llanito, como *sefbord* (tabla de surf), *drink* (copa) o *keki* (tarta).

En general, en estas publicaciones se ha detectado una actitud positiva hacia la modalidad lingüística, que valora el aporte cultural de la misma, ya que forma parte del patrimonio de la región. Asimismo, no se ha apreciado la presencia de estereotipos o prejuicios hacia la misma.

5.5. *Análisis de los resultados*

Los datos presentados en esta sección demuestran como la riqueza de la lengua española y su vínculo con la cultura propia de cada región ha pasado a ser objeto de numerosas publicaciones de *Instagram*. De hecho, como afirman Bautista, Cabezuelo-Lorenzo y de la Casa (2021), la red social brinda numerosas posibilidades de abarcar temas de carácter científico, alcanzando asimismo una cantidad significativa de interacciones. Este hecho resulta ser significativamente

relevante, puesto que permite una mayor difusión y promoción de las culturas hispanas, representando estos contenidos también una fuente de muestras reales de la lengua española en su variación y de contacto con hablantes procedentes de diferentes regiones y continentes.

En general, cabe destacar que las publicaciones analizadas buscan mantener la atención del usuario, siendo todas de fácil acceso y visualización y sin requerir estas demasiada concentración o esfuerzo por parte de este mismo. Esto se alinea con lo afirmado por Bautista, Cabezuelo-Lorenzo y de la Casa (2021), cuya opinión es que la red social representa un valioso instrumento por su facilidad de comprensión del contenido visual y funcionalidad. Se ha destacado que numerosos *posts* entre los seleccionados presentaban una actitud humorística ante el contenido tratado, siendo este un medio para conseguir mayor atención y difusión entre el público, por un lado, y una forma para que las nociones se queden en la memoria de los espectadores, por otro.

Observando las actitudes adoptadas por los creadores en sus contenidos, se nota que esta varía significativamente en función del tema de los *posts*. Como se aprecia en la Figura 6, temáticas como la variación léxica y el llanito se abordan desde un punto de vista principalmente didáctico, aunque en el primer caso también destaca un grupo significativo de publicaciones que combinan la didáctica y el humor. Por otro lado, los *posts* relacionados con los choques culturales y la identidad canaria suelen adoptar un enfoque principalmente humorístico, que solo en tres ocasiones se acerca a una perspectiva didáctica.

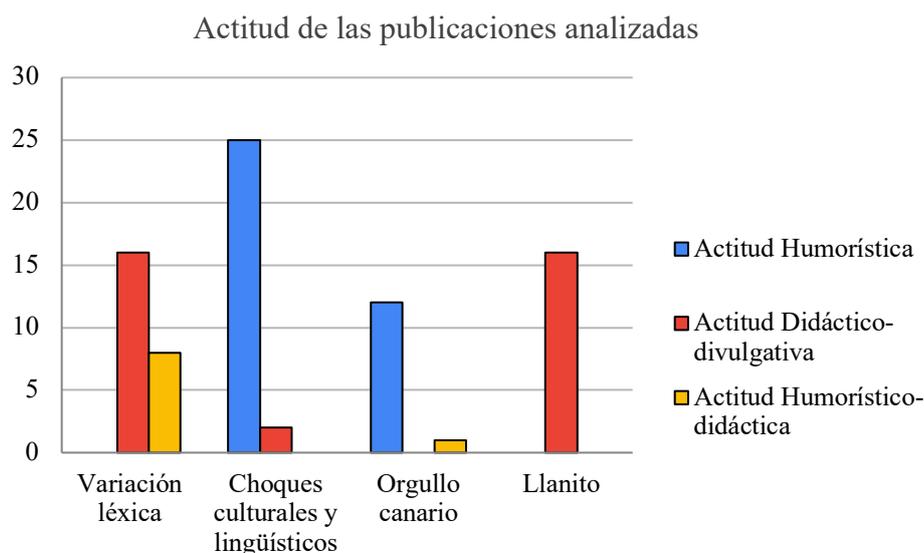


Figura 6: Tipo de actitudes detectada en los contenidos

Por lo que respecta, en cambio, a los estereotipos, se señala que estos aparecen solo en las publicaciones relacionadas con los choques culturales o la identidad canaria. Asimismo, destaca el hecho de que estos no son estrictamente negativos o perjudiciales, sino que representan más bien anécdotas sobre los países en cuestión.

En general, se valora positivamente la presencia de los numerosos recursos con perspectiva didáctica, puesto que estos representan una fuente de información directa para los que quieran expandir sus conocimientos, por un lado, y un posible recurso didáctico para los docentes, por otro. De hecho, respecto a este último punto, es preciso recordar que los profesionales de ELE de hoy en día tienen que relacionarse con un alumnado nativo digital, por lo tanto, pueden aprovechar las posibilidades brindadas por contenidos similares para actualizar sus métodos de enseñanza y acercarlos a sus estudiantes. “La comunicación digital global está cada vez más condicionada por el modelo que las redes sociales han venido imponiendo en torno a la inme-

diatez y la participación”, afirman Bautista, Cabezuelo-Lorenzo y de la Casa (2021: 145), por lo que se considera que las redes sociales representan un valioso instrumento para establecer una conexión exitosa con los alumnos en los ámbitos de ELE.

6. POSIBILIDADES DIDÁCTICAS

A la luz de lo expuesto en el apartado anterior, es evidente que *Instagram* ofrece múltiples oportunidades para interactuar con la variación lingüística y cultural que caracteriza al español en sus diferentes regiones. En este sentido, como se acaba de comentar, se considera que esta red social puede constituir un recurso didáctico valioso para abordar dicho tema en el ámbito de la enseñanza del español como lengua extranjera (ELE).

En este contexto, ya se ha mencionado como la integración de recursos tecnológicos como *Instagram* en el aula puede incrementar el interés y la motivación de los estudiantes en su proceso de aprendizaje. En particular, el uso de redes sociales se adapta con facilidad a las necesidades específicas de cada curso, promoviendo la interacción entre los estudiantes sin las restricciones impuestas por la geografía, permitiéndoles comunicarse simultáneamente y compartir intereses comunes (Rodríguez Gallego, López Martínez y Martín Herrera, 2017).

Asimismo, el estudio de Peña Hita, Rueda Lopéz y Pegalajar Palomino (2018) refuerza esta perspectiva, subrayando que los estudiantes de las generaciones más jóvenes perciben a las redes sociales como una herramienta valiosa para el proceso de enseñanza-aprendizaje. Los resultados de esta investigación destacan la relevancia atribuida por los estudiantes a estas plataformas, no solo como medio de comunicación, sino también como un espacio para encontrar y almacenar información. Además, los estudiantes opinan que promueve su socialización, que pueden conectarse fácilmente, así como intercambiar recursos didácticos. Por último, las redes sociales se valoran como una “ayuda valiosa para el enriquecimiento intercultural y la generación de nuevas ideas, permitiendo aprender a partir de la experiencia compartida con otros y adquiriendo habilidades de adaptación ante nuevas situaciones” (Peña Hita et al., 2018: 250).

En este marco, las posibles aplicaciones didácticas de *Instagram* en la enseñanza de ELE son numerosas, y la flexibilidad inherente a la plataforma la convierte en un recurso valioso tanto para los estudiantes como para los docentes. Para los primeros, se pueden plantear actividades como ejercicios de pronunciación basados en contenidos de la plataforma, juegos lingüísticos, desafíos colaborativos, experimentos y demostraciones prácticas e interactivas. Estas dinámicas fomentan una participación activa, así como la colaboración entre los estudiantes. Además, los aprendices pueden crear contenidos similares a los analizados en clase, o centrarse en temáticas específicas, lo que facilita el aprendizaje de conceptos clave.

Instagram también se presenta como una herramienta especialmente valiosa para aquellos estudiantes que aprenden de forma autónoma, dado que su uso no está vinculado a horarios específicos ni a entornos escolares formales. Esto hace que la red social sea accesible en cualquier momento, permitiendo a los aprendices interactuar con el contenido según su conveniencia.

Desde la perspectiva del docente, *Instagram* y otras redes sociales ofrecen la posibilidad de crear comunidades de profesores, donde estos pueden encontrar páginas o cuentas que ofrecen publicaciones educativas, consejos prácticos, y técnicas pedagógicas innovadoras. Estas plataformas también permiten a los docentes compartir experiencias, recursos y actividades con colegas, enriqueciendo así su práctica diaria.

Sin embargo, es importante reconocer las desventajas y los riesgos asociados al uso de redes sociales en el ámbito educativo. Tanto estudiantes como docentes pueden desarrollar adicción o experimentar dificultades en la gestión del tiempo, además de enfrentar riesgos informáticos relacionados con la exposición de información personal en línea (Rodríguez Gallego et al., 2017). Asimismo, otro desafío evidente está representado por las posibilidades brindadas por la misma plataforma, puesto que los formatos de contenido en *Instagram* están limitados

en términos de número de imágenes y duración de vídeos. Esto puede dar lugar a publicaciones incompletas o imprecisas las que, por ende, podrían requerir materiales adicionales para una comprensión más global.

Además, se suma el problema de la falta de verificación de la autoría y precisión de los contenidos en las redes sociales. En muchos casos, no hay garantía ni posibilidad de acreditar la formación adecuada de los creadores para abordar ciertos contenidos. Por lo tanto, es posible que los estudiantes que usan autónomamente la plataforma se expongan a información errónea.

En conclusión, si bien *Instagram* y otras redes sociales presentan un vasto potencial como recursos didácticos, es esencial utilizarlas de manera crítica y estratégica, considerando tanto sus ventajas como sus limitaciones. Este enfoque permitirá maximizar su impacto positivo en la enseñanza del español como lengua extranjera y en el desarrollo de habilidades interculturales en los aprendices.

7. CONCLUSIONES

La presente investigación exploratoria se ha centrado en observar cómo la diversidad y riqueza del español se refleja en *Instagram*, una de las redes sociales más influyentes y visibles en el mundo contemporáneo. El español, como idioma oficial de 21 países y varias organizaciones internacionales, presenta una realidad lingüística polifacética y diversa, marcada por la existencia de rasgos y variedades propios de cada región hispanohablante. Además, en varias áreas geográficas, el contacto con otras lenguas ha dado lugar a la creación de modalidades lingüísticas de contacto, como el llanito y el *spanglish*.

Debido a la complejidad inherente a la lengua de Cervantes, resulta particularmente interesante observar cómo esta se incorpora en redes sociales, permitiendo a los usuarios interactuar con muestras auténticas del idioma y entrar en contacto con su diversidad, superando las limitaciones impuestas por las distancias geográficas. *Instagram*, en particular, destaca como una plataforma valiosa debido a la simplicidad y accesibilidad de sus formatos, que emplean imágenes y vídeos como soportes visuales para facilitar el entendimiento.

La investigación, que se enfoca en cuatro ejes temáticos relacionados con la variación lingüística, ha puesto de relieve que la red social consigue vehicular con éxito una amplia gama de contenidos sobre el argumento. En numerosas ocasiones, los contenidos utilizan el humor para captar la atención del usuario, sin embargo, también destaca que en otras circunstancias se incorporan nociones de carácter didáctico y divulgativo.

En términos generales, los contenidos analizados son una fuente de información valiosa para estudiantes y docentes, ofreciendo nociones, palabras y situaciones útiles que, en casos como los dedicados a los choques culturales, emplean escenarios verosímiles que logran mantener la atención del espectador y facilitar la retención del conocimiento. Además de lo mencionado, la plataforma goza de amplia difusión tanto entre los jóvenes como entre los adultos. Por lo tanto, puede integrarse fácilmente a la enseñanza de ELE como recurso didáctico, ya que la mayoría del alumnado ya está familiarizado con su funcionamiento.

Sin embargo, es importante señalar que algunos contenidos, especialmente los relacionados con diferencias culturales o costumbres nacionales, pueden contener estereotipos o prejuicios basados en clichés muy difundidos. Además, es imperativo considerar los riesgos derivados del empleo de las redes sociales en el aula, como la posible presencia de imprecisiones debidas al espacio limitado del que disponen los autores en la plataforma y la falta de verificación de autoría en la mayoría de los casos. Asimismo, en el aula de ELE, podría resultar difícil para el docente vigilar el uso que los alumnos hacen de *Instagram* durante la clase.

En conclusión, se considera que *Instagram* es un medio preciado para la difusión de contenidos relacionados con las variedades del español, puesto que propone formatos simples y fácilmente comprensibles, que favorecen el aprendizaje y la asimilación de las nociones al aco-

gerse a soportes audiovisuales, tanto imágenes como vídeos. Asimismo, la plataforma analizada presenta ventajas notables, como el consumo rápido y sencillo de los contenidos propuestos, cuya fruición requiere poca energía o esfuerzo y que, al mismo tiempo, resultan interesantes y entretenidos para los usuarios. Por su formato, los contenidos se prestan fácilmente al uso en contextos de ELE; sin embargo, por los posibles riesgos anteriormente mencionados, se considera necesario asegurar una mediación del docente para maximizar su potencial educativo y minimizar los posibles efectos negativos.

Siendo este un estudio exploratorio, las futuras líneas de investigación que abre son muchas. De hecho, sería de extremo interés analizar la presencia de las temáticas analizadas en otras redes sociales como *Facebook*, *TikTok* o *X* (anteriormente, *Twitter*) y comparar los resultados obtenidos. Asimismo, sería relevante analizar el efectivo empleo de *Instagram* en contextos de ELE y examinar las finalidades con las que se utiliza, así como el tipo de contenido que se selecciona para la clase, observando posiblemente el uso de otras redes sociales también. A continuación, se considera interesante elaborar propuestas prácticas para emplear las publicaciones analizadas en el ámbito de la enseñanza, así como explorar posibles formas de incluir los recursos procedentes de las redes sociales en el diseño de materiales de ELE. Por último, se juzga extremadamente relevante analizar el contenido de los vídeos a nivel lingüístico, con el objetivo de identificar los recursos empleados para expresar la identidad regional, así como los contenidos del vídeo se tratan y utilizan como recursos para celebrar o construir la propia identidad frente a las demás. Finalmente, se podría observar la presencia de variantes lingüísticas asociadas a subculturas o grupos sociales específicos, como el lenguaje juvenil o el lenguaje propio de comunidades *online*.

CONTRIBUCIÓN DE LOS AUTORES

Los apartados 1, 2 y 4 están a cargo de Silvia Betti. Los apartados 3, 5 y 7 están a cargo de Lisa Gozi. El apartado 6 y las referencias están a cargo de las dos autoras.

REFERENCIAS

Androutsopoulos, J. (2013). Online data collection. En C. Mallinson, B. Childs, G. Van Herk (Eds.), *Data Collection in Sociolinguistics. Methods and Applications* (pp. 236-249). Londres: Routledge.

Andión Herrero, M. A. (2007). Las variedades y su complejidad conceptual en el diseño de un modelo lingüístico para el español L2/LE. *ELUA: Estudios de Lingüística. Universidad de Alicante*, (21), 21-33 [1-13]. <https://doi.org/10.14198/ELUA2007.21.02>

Andión Herrero, M. A. (2008). Modelo, estándar y norma ..., conceptos imprescindibles en el español L2/LE. *RESLA*(21), 9-26.

Andión Herrero, M. A. (2009). La variedad del profesor frente al modelo de enseñanza: convergencias, divergencias y actitudes. En Clavero, A. B., Camacho, J. C. M., Polo, V. R. D., Barjola, M. I. F. (Eds.), *El profesor de español LE L2: Actas del XIX Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE)* (pp. 167-183). Cáceres: Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE).

Andión Herrero, M. A. (2019). Del lectocentrismo al plurinormativismo. Reflexiones sobre la variedad del español como lengua segunda o extranjera. *Estudios filológicos*, (64), 129-148. <https://doi.org/10.4067/S0071-17132019000200129>

Bautista, P. S., Cabezuelo-Lorenzo, F. y de la Casa, J. M. H. (2021). Instagram como herramienta digital para la comunicación y divulgación científica: el caso mexicano de @pictoline. *Chasqui: Revista latinoamericana de comunicación*, (147), 143-162. <https://doi.org/10.16921/chasqui.v1i147.4472>

Betti, S. (2024). El español y spanglish en los Estados Unidos. Una reflexión desde Europa. En F. Rodríguez González (Eds.), *New Perspectives on Languages in Contact Festschrift for Félix Rodríguez González on the Occasion of his Retirement* (pp. 257-284). Turín: Lang Edizioni.

Carpenter, J. P., Morrison, S. A., Craft, M., y Lee, M. (2020). How and why are educators using Instagram?. *Teaching and teacher education*, 96, 103-149. <https://doi.org/10.1016/j.tate.2020.103149>

Figuerero-Benítez, J. C., González-Quiñones, F., y Machin-Mastromatteo, J. D. (2021). Instagram como objeto de estudio en investigaciones recientes. Una revisión de literatura con enfoque en revistas científicas. *Ámbitos. Revista Internacional de Comunicación*, (53), 9-23. <https://doi.org/10.12795/Ambitos.2021.i53.01>

Gozzi, L. (2024). *Las variedades de la lengua española: estudio de inclusión y creencias en contextos de ELE* [Tesis Doctoral]. Alcalá de Henares: Universidad de Alcalá. Recuperado de <https://ebuah.uah.es/dspace/handle/10017/62142>

Guerrero Salazar, S. (2020). Análisis comparativo de las actitudes lingüísticas sobre las modalidades andaluzas y canarias en la prensa española. *Moderna språk*, 108-138. <https://doi.org/10.58221/mosp.v114i1.7525>

Instituto Cervantes. (2024). *El español en el mundo. Anuario del Instituto Cervantes. 2024*. Madrid: Instituto Cervantes. Recuperado de https://cvc.cervantes.es/lengua/anuario/anuario_24/el_espanol_en_el_mundo_anuario_instituto_cervantes_2024.pdf

Mancera Rueda, A. y Pano Alamán, A. (2013). *El español coloquial en las redes sociales*. Madrid: Arco/Libros.

Martí Climent, A. y García Vidal, P. (2018). Redes sociales en la enseñanza superior. *IN-RED 2018. IV Congreso Nacional de Innovación Educativa y Docencia en Red*, 141-154.

Moreno Fernández, F. (2017). *Las variedades de la lengua española y su enseñanza* (2ª ed.). Madrid: Arco/Libros. <https://doi.org/10.4995/INRED2018.2018.8623>

Moreno Fernández, F. (2020). *La lengua española en su geografía* (5ª ed.). Madrid: Arco/Libros.

Moreno Fernández, F. (2021). *Qué español enseñar* (3ª ed.). Madrid: Arco/Libros.

Pano Alamán, A. (2021). Los «hashtags» en el discurso institucional español sobre la Covid-19 en Twitter. *Revista de Investigación Lingüística*, (24), 47-71. <https://doi.org/10.6018/rii.484901>

Peña Hita, M. D. L. Á., Rueda López, E. y Pegalajar Palomino, M. D. C. (2018). Posibilidades didácticas de las redes sociales en el desarrollo de competencias de educación superior: Percepciones del alumnado. *Píxel-Bit. Revista de Medios y Educación*, (53), 239-252.

Rodríguez Gallego, M. R., López Martínez, A. y Martín Herrera, I. (2017). Percepciones de los estudiantes de Ciencias de la Educación sobre las redes sociales como metodología didáctica. *Píxel-Bit. Revista de Medios y Educación*, (50), 77-93. <https://doi.org/10.12795/pixelbit.2017.i50.05>

Trapero Trapero, M. (2024). Orgullo de ser palmero. *Lustrum: gaceta de la Bajada de la Virgen*, (7), 30-33.

Yus, F. (2017). Los efectos de Internet y las redes sociales en el español. *Boletín de la Academia Puertorriqueña de la Lengua Española*, 55-76.



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1

Enero-diciembre 2024

Páginas: 169-193

Artículo recibido: 17/10/2024

Artículo aceptado: 11/12/2024

Artículo publicado: 31/01/2025

Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/691>

DOI: <https://doi.org/10.58859/rael.v23i1.691>

Competencias lingüísticas y comunicación inclusiva a través de la mediación en inglés para fines específicos

Linguistic Competences and Inclusive Communication through Mediation in English for Specific Purposes

ADRIANA BAUSELLS ESPÍN

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (UNED)/UNIVERSIDAD DE SAN JORGE

DIANA GINER ALONSO

UNIVERSIDAD DE ZARAGOZA

Este artículo presenta resultados de una intervención didáctica realizada en la asignatura de Inglés Científico para el Grado en Enfermería en la Universidad San Jorge (Zaragoza, España). La intervención integra actividades de mediación lingüística, un enfoque promovido por el Marco Común Europeo de Referencia para las Lenguas, para mejorar las habilidades comunicativas inclusivas en el alumnado. Esta intervención se abordó a través de ejercicios prácticos diseñados para fomentar la accesibilidad en contextos sanitarios. Los estudiantes participaron en interacciones simuladas con pacientes que presentaban diferentes requerimientos comunicativos, adaptando su discurso y estrategias comunicativas en función de los perfiles. Los resultados demuestran una percepción de mejora significativa en la conciencia y habilidades lingüísticas y comunicativas del alumnado y resaltan la viabilidad y efectividad de incorporar la mediación lingüística en la enseñanza de lenguas para fines específicos. Esta intervención subraya la importancia de la mediación para promover una comunicación equitativa en ámbitos sanitarios.

Palabras clave: *mediación lingüística; inglés para fines específicos; inglés para ciencias de la salud; accesibilidad; comunicación inclusiva*

This article shows the results of a pedagogical intervention carried out in the Scientific English course within the Nursing Degree at Universidad San Jorge (Zaragoza, Spain). The intervention integrates linguistic mediation activities, an approach promoted by the Common European Framework of Reference for Languages, to enhance inclusive communication skills in students. Linguistic mediation was addressed through practical exercises designed to foster accessibility in healthcare contexts. Students participated in simulated interactions with patients with different communicative

Citar como: Bausells Espin, A. y Giner Alonso, D. (2024). Competencias lingüísticas y comunicación inclusiva a través de la mediación en inglés para fines específicos. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 169-193. <https://doi.org/10.58859/rael.v23i1.691>

requirements, adapting their speech and communicative strategies accordingly. The study results show perceptions of a significant improvement in students' awareness and linguistic and communicative skills, while also highlighting the feasibility and effectiveness of incorporating linguistic mediation into the teaching of languages for specific purposes. This intervention emphasizes the importance of mediation in promoting equitable communication in healthcare settings.

Keywords: *linguistic mediation; English for Specific Purposes; English for Health Sciences; accessibility; inclusive communication*

1. INTRODUCCIÓN

Este artículo presenta resultados de una intervención didáctica realizada en la asignatura *Inglés Científico* del Grado en Enfermería en la Universidad San Jorge (USJ, Zaragoza),¹ consistente en la incorporación de actividades basadas en la mediación intralingüística como herramienta para el desarrollo de competencias lingüísticas y de comunicación efectiva, accesible, e inclusiva. Esta intervención sigue las recomendaciones del Volumen Complementario (VC) al Marco Común Europeo de Referencia de las Lenguas (MCER), que considera la mediación uno de los cuatro modos de comunicación (Consejo de Europa, 2018: 31).

En una sociedad globalizada y comunicativamente diversa, las diferencias lingüísticas y socioculturales pueden suponer obstáculos significativos para la integración y cohesión social. La mediación lingüística emerge como una herramienta vital para superarlos, pues facilita la comunicación y promueve la inclusión y la equidad (De Arriba y Cantero, 2004: 12; Beacco, Byram, Cavalli, Coste, Cuenat, Goullier y Panthier, 2016: 8). Como competencia comunicativa integradora, debería tenerse en cuenta en los procesos de enseñanza-aprendizaje de lenguas extranjeras (LE). Su incorporación es esencial para preparar al estudiantado para dicha diversidad comunicativa y para superar la visión reduccionista de la lengua como externa al aprendiente-usuario (Piccardo, North, y Goodier, 2019). Sin embargo, tradicionalmente esta incorporación se ha circunscrito a nociones de interculturalidad y ha ignorado o subordinado el valor de la mediación *intra*lingüística en intercambios cuyos interlocutores no comparten marcos de referencia esenciales—como el contexto social o el conocimiento previo—, o requieren adaptaciones comunicativas.² Esta concepción ampliada de la mediación y su incorporación al aula de LE sigue en fase inicial, especialmente en educación universitaria.

A fin de contribuir en el avance de la integración de la mediación *intra*lingüística en el proceso de enseñanza-aprendizaje de lenguas para fines específicos (LFE) en educación superior, este artículo presenta resultados de una experiencia de intervención didáctica en inglés para Enfermería. Con el objetivo de dotar a otras investigaciones similares de herramientas para poder implantar la mediación *intra*lingüística en otras asignaturas de LFE en el ámbito universitario, se proporciona una contextualización del papel de la mediación en LE y una descripción detallada de la intervención, así como unas reflexiones finales.

¹ Investigación asociada al Proyecto “Hacia una mejora de la enseñanza de idiomas en la Universidad San Jorge: un enfoque integral en el análisis de necesidades y enfoques pedagógicos en el contexto de la Universidad” (2023-2024); Grupo Lengua y Comunicación, Instituto de Lenguas Modernas (USJ), y respaldada por el Gobierno de Aragón [CIRES H16_23R].

² La redacción sigue las *Directrices para un lenguaje inclusivo* de las Naciones Unidas (2019; <https://www.unge-neva.org/es/about/accessibility/disability-inclusive-language>). Así, se utiliza *requerimientos* en lugar de *necesidades*, y *adaptaciones/servicios adaptados* en lugar de *necesidades especiales*.

2. MARCO TEÓRICO

2.1. La mediación lingüística en el MCER

Mediar es el proceso de facilitar el acceso a información/conocimiento a personas para quienes existen barreras lingüísticas, socioculturales o cognitivas. Mediar implica dos subprocesos complementarios: (co)construir significado(s), y facilitar la creación de condiciones adecuadas para el intercambio efectivo de información (Dendrinis, 2006; Consejo de Europa, 2020). Según el MCER, mediar, en enseñanza-aprendizaje de LE, implica superar la división entre los niveles individual y social del aprendizaje de modo que el alumnado actúe como agente social y priorice los requerimientos comunicativos del interlocutor (Consejo de Europa, 2020: 45). Además, la mediación lingüística³ no se restringe a la negociación ni a la traducción, sino que incluye la reformulación de mensajes, la adaptación del registro y los rasgos lingüísticos, o el desglose o la condensación de información. Esto la convierte en una herramienta didáctica indispensable para aprender a comunicarse en la LE y a facilitar espacios comunicativos inclusivos también en entornos profesionales, lo cual permite preparar al alumnado para las realidades diversas del mundo actual y para fomentar habilidades personales, profesionales, pluriculturales, de empatía, y de adaptabilidad lingüística y comunicativa (Beacco et al., 2016).

2.1.1. Evolución

La mediación *intercultural* fue introducida en el MCER en 2001 junto con la *interacción* para complementar a la *recepción* y la *producción* como las cuatro destrezas, posteriormente reformuladas como “cuatro modos de actividad” (North y Piccardo, 2016: 9). Esto enfatizó la dimensión social del uso-aprendizaje de la lengua, e introdujo los dos factores clave para la ampliación del concepto de mediación: la “coconstrucción del conocimiento” y la “conciencia de la naturaleza dinámica del significado” (North y Piccardo, 2016: 9). En los descriptores desarrollados por estos autores, *mediación* pasa a referirse al “papel del lenguaje para facilitar la comunicación (...) y la transmisión de información” (2016: 20).

Esta definición se acerca a un uso real de la lengua basado en la interacción, lugar de construcción de significados, donde la mediación lingüística es clave para generar espacios adecuados y superar barreras entre personas que no pueden entenderse “por cualquier motivo”⁴ (Consejo de Europa, 2020: 45). Así, se incorporan las dimensiones *intra*lingüística e *intermodal*, que exigen la reformulación o transcodificación de mensajes según el contexto social y la adaptación del lenguaje o registro a la audiencia en función de sus preferencias o requerimientos comunicativos (Dendrinis, 2018: 3; Consejo de Europa, 2020: 104). En conclusión, el VC proporciona una visión ampliada y desglosada de la mediación lingüística y destaca su relevancia en enseñanza de LE para el desarrollo de numerosas competencias comunicativas (Pazos Anido, 2018). Además, insta a descubrir la mediación lingüística como un proceso de aprendizaje cooperativo que facilita un enfoque basado en la inclusión (Sanz Esteve, 2020/2021: 196)

2.1.2. Actividades y estrategias

Como en las demás destrezas, el VC distingue entre mediación oral y escrita, y entre *actividades* y *estrategias* (Consejo de Europa, 2020/2021: 104). Las *actividades* de mediación se clasifican en función de *lo que es mediado* y se ejemplifican con acciones concretas:

- 1) Mediar textos: transmitir información, explicar datos, procesar texto, etc.
- 2) Mediar conceptos: colaborar para construir significado, fomentar el habla conceptual, etc.

³ *Mediación lingüística* y *mediación intralingüística* se alternan como sinónimos.

⁴ Conviene resaltar que no se menciona la discapacidad, la edad, o el desarrollo cognitivo como posibles fuentes de barreras comunicativas susceptibles de requerir mediación lingüística.

- 3) Mediar la comunicación: facilitar espacios pluriculturales, actuar como intermediario, etc.

Las estrategias de mediación lingüística se refieren a los *mecanismos* que se aplican en estos procesos:

- 1) Estrategias para explicar un concepto: adaptar el lenguaje, desglosar información compleja, etc.
- 2) Estrategias para simplificar un texto: ampliar un texto denso, destacar información, etc.

Como se describe en el apartado 4, la intervención fue diseñada a partir de esta concepción amplia de mediación y sustentada sobre las nociones de *actividad y estrategia* establecidas en el VC.

2.2. Mediación en la enseñanza de LE

Este apartado recopila, cronológicamente, experiencias de integración de la mediación en enseñanza de LE que presentan aportaciones relevantes para este estudio. En 2013, Pazos Anido analiza la incorporación de la mediación *intercultural* en las planificaciones de profesores de español como LE en prácticas en el contexto portugués, explorando las actividades propuestas y las percepciones del profesorado. Concluye que la mediación como actividad comunicativa resulta discriminada en la planificación docente y limitada a cuestiones de interculturalidad, principalmente por la falta de conocimiento y formación en la materia. Este estudio revela tanto la reducida presencia de la mediación en la formación docente como su concepción restringida.

Trovato (2016) enfatiza que la competencia mediadora comprende habilidades interlingüísticas, pragmáticas, sociolingüísticas y socioculturales. Las actividades, ya sean *interlingüísticas* o *intra lingüísticas*, se dirigen a formar hablantes competentes que puedan actuar como mediadores facilitando el intercambio y la comunicación entre ciudadanos. Estas afirmaciones encajan con la concepción ampliada de mediación del VC e incorporan nociones clave como la accesibilidad comunicativa.

Alcaraz Mármol (2019) investiga la creciente importancia de la mediación *lingüística* en la enseñanza de LE en Educación Secundaria. Sus resultados indican un desconocimiento notable entre los estudiantes sobre lo que esta implica y una incorporación poco sistemática a las prácticas docentes, en contraste con la importancia que recibe en el MECR. La mayoría de participantes manifiestan inseguridad al realizar actividades de mediación y consideran necesaria una formación adicional para abordarlas correctamente. Esto subraya la necesidad de desarrollar estrategias educativas que incorporen la mediación lingüística como componente esencial del aprendizaje, en tanto que prepara para la participación en un mundo de constantes y diversas interacciones.

La integración más sistemática de la mediación lingüística ha sido implementada por las Escuelas Oficiales de Idiomas españolas (EOI), que han contribuido de forma esencial en su inclusión y consolidación en el currículo de LE (Calatayud Díez, 2019). El nuevo marco legislativo desarrollado⁵ actualiza el currículo conforme al VC adaptándolo a las realidades sociales del siglo XXI y destacando aspectos como la constante movilidad interpersonal y la diversidad comunicativa resultante, para cuyo abordaje es fundamental desarrollar múltiples competencias mediadoras.

⁵ Véanse: Real Decreto 1041/2017 (Boletín Oficial del Estado (BOE), 22/12/2017; <https://www.boe.es/buscar/doc.php?id=BOE-A-2017-15367>) y Real Decreto 1/2019 (BOE, 11/01/2019; https://www.boe.es/diario_boe/txt.php?id=BOE-A-2019-317).

Finalmente, Sanz Esteve (2020) confirma las observaciones de Alcaraz Mármol y concluye que la mediación debería ser más prominente en la planificación de cursos de idiomas. Añade que es una actividad altamente adaptable que puede mejorar significativamente las habilidades comunicativas del alumnado, pues supone una vía efectiva para alcanzar una competencia comunicativa amplia y profunda. Este énfasis en la versatilidad pedagógica de la mediación encaja nuevamente con la concepción ampliada del VC y permite pensar en aplicaciones en diversos contextos formativos.

2.3. Mediación y LFE

La mediación lingüística proporciona estrategias para facilitar el acceso a la información y su comprensión en interacciones comunicativas donde existan barreras de diversa naturaleza también en ámbitos profesionales, cuyos conceptos o lenguajes pueden resultar inaccesibles para personas de diferentes antecedentes o sin formación específica (De Arriba y Cantero, 2004; Beacco et al., 2016; Martín-Macho Harrison y Guadamillas Gómez, 2022). Resulta esencial que el especialista posea habilidades de mediación lingüística para asegurar que el interlocutor lego pueda acceder a información relevante, comprender sus implicaciones y actuar en consecuencia. Sorprende, pues, la todavía reducida presencia de estudios de aplicación de la mediación interlingüística en el ámbito de las lenguas para fines específicos (LFE) particularmente para ciencias de la salud.⁶

Chovancová (2018) proporciona un ejemplo de incorporación de la mediación en la enseñanza del inglés para fines legales, y explora cómo puede mejorar el aprendizaje de la LE y cómo puede integrarse en el currículo. En el ámbito legal, la mediación no se limita a la resolución de conflictos, sino que es esencial para la adecuada comunicación abogado-cliente. Chovancová concluye que la mediación lingüística es una habilidad fundamental notablemente ignorada en la enseñanza de LFE, y que la actualización del VC ilustra la necesidad de actualizar los planes de estudio para proporcionar al alumnado estrategias comunicativas esenciales para su futura profesión. Así, propone la realización de simulaciones de interacciones abogado-cliente que permitan practicar habilidades mediadoras diversas en situaciones comunicativas probables en un contexto laboral.

Esto es aplicable a la enseñanza de LFE para profesiones sanitarias, donde la mediación tampoco se restringe a la resolución de conflictos.⁷ Aplicar estrategias de mediación en interacciones con pacientes asegurará que comprendan instrucciones y procedimientos médicos complejos, diagnósticos, o tratamientos y sus consecuencias, algo esencial para garantizar la autonomía sanitaria (Dendrinis, 2018; Consejo de Europa, 2020). Además, un personal sanitario formado en estrategias de mediación lingüística podrá ayudar al paciente a articular sus preocupaciones, síntomas o requerimientos, y a promover una participación activa y equitativa en la gestión de la salud y los cuidados (Dendrinis, 2018).

Estas habilidades mediadoras son especialmente necesarias en el trato con pacientes que requieran servicios y comunicación adaptados —pacientes con discapacidades intelectuales o físicas, o de edad avanzada o temprana— pues proporcionarán al personal sanitario tanto la capacidad de identificar los requerimientos o adaptaciones pertinentes, como de adecuar su discurso y actitudes mediante distintas estrategias: uso de lenguaje sencillo, ayudas auditivas

⁶ Martín-Macho Harrison y Guadamillas Gómez (2022) aportan en su exhaustivo análisis propuestas de aplicación en LFE en cuatro áreas (turismo, negocios, relaciones internacionales y estudios artísticos) que consideran especialmente idóneas. El ámbito sanitario aparece dentro de la actividad propuesta para relaciones internacionales (2022: 91), pero el foco no recae sobre la comunicación profesional-paciente.

⁷ La mediación lingüístico-cultural en el ámbito sanitario se ha estudiado desde múltiples ángulos (véanse, entre otros, Baraldi y Gavioli, 2019; volumen 36, nº. 9, de 2021 de *Health Communication*). Sin embargo, la mayoría de estos estudios se han llevado a cabo con profesionales sanitarios, no con estudiantes ni en el ámbito concreto de las LFE. Por ello, incluirlas investigaciones en la revisión de la literatura queda fuera del alcance del presente artículo.

o visuales, técnicas de reformulación, etc. La aplicación de estos principios contribuye a crear un entorno de confianza y empatía, a priorizar el bienestar, la comodidad y la seguridad del paciente durante situaciones delicadas o estresantes, y a mejorar los procesos de anamnesis, diagnóstico, tratamiento y seguimiento (Trovato, 2016).

En conclusión, la mediación lingüística no solamente promueve el desarrollo de competencias lingüísticas y comunicativas, sino que transforma el aula de LFE en un entorno de aprendizaje inclusivo, proporcionando al alumnado oportunidades para actuar como agente social, concienciarse sobre diversas realidades comunicativas y desarrollar mecanismos para una comunicación accesible dentro y fuera del aula. Además, permite fomentar la concienciación sobre los propios sesgos y la necesidad de priorizar los intereses del interlocutor, y acerca la experiencia de aula a la realidad profesional.

3. OBJETIVOS

La relevancia del concepto ampliado de mediación dentro del aprendizaje de LFE y la necesidad de promover la accesibilidad e igualdad en la comunicación en todos los ámbitos, incluidos los profesionales como el sanitario, motivó el diseño de una intervención didáctica realizada en una asignatura de Inglés Científico para Enfermería basada en la integración de la mediación lingüística y enfocada en la interacción profesional-paciente.

3.1. Objetivos pedagógicos

El diseño distingue entre objetivos pedagógicos (OP) e investigadores. Los OP son:

- 1) OP1: Fomentar el desarrollo de habilidades y estrategias de mediación.
- 2) OP2: Fomentar la concienciación sobre accesibilidad e inclusión.
- 3) OP3: Fomentar la adquisición y producción de contenidos lingüísticos de la especialidad.

El grado de consecución se mide mediante los resultados en las pruebas de evaluación asociadas y las respuestas a los cuestionarios de percepción completados por el alumnado. Este artículo se centra en estas últimas.

3.2. Objetivos investigadores

Los objetivos investigadores (OI) están relacionados con los OP y son:

- 1) OI1: Explorar la viabilidad de incorporar actividades de mediación en la asignatura *Inglés Científico*.
- 2) OI2: Valorar su efectividad para la consecución de los objetivos de aprendizaje lingüístico.
- 3) OI3: Explorar las actitudes y percepciones del alumnado hacia la mediación lingüística.

El grado de consecución se mide mediante las mismas herramientas arriba mencionadas.

4. METODOLOGÍA

4.1. Contexto

Esta intervención se integra en la asignatura *Inglés Científico*, de primer curso del Grado en Enfermería (USJ), impartida a nivel B1 (MECR) en dos sesiones semanales de 1h 40 min, y orientada al desarrollo de competencias lingüísticas y comunicativas. El currículo cubre temas como “Admisiones”, “Dolor”, “Constantes vitales” o “Medicación”, con énfasis en el aprendizaje de léxico especializado y estructuras gramaticales relacionadas con funciones comunicativas clave (dar instrucciones/consejos, describir pasos, etc.) y el desarrollo de competencias de interacción oral y escrita con pacientes en situaciones diversas. El método de evaluación combina la evaluación continua y un examen final con componente oral y escrito.

4.2. Temporalización y procedimientos

En 2022-2023 se introdujeron por primera vez contenidos explícitos de mediación en la asignatura. Tras la valoración positiva de esta experiencia, se diseñó la intervención didáctica para el curso 2023-2024. Las fases seguidas fueron: (i) planificación, (ii) diseño, (iii) implementación y (iv) evaluación. En la fase (i) se confirmó el potencial pedagógico de la experiencia del curso anterior y se estableció un plan de trabajo. En la fase (ii) se completó el (re)diseño de actividades basado en la mediación lingüística, se compilaron materiales de apoyo, se crearon las pruebas de evaluación (redacción y examen oral; apartado 4.4.), se definieron los objetivos investigadores (apartado 3.2.) y se diseñaron los cuestionarios (apartado 4.5.). En la fase (iii) se realizaron las actividades de aula pertinentes y se completaron la evaluación y el *pre*cuestionario. Tras el periodo lectivo, durante la fase (iv), se analizaron los datos del *pre*cuestionario (datos en el apartado 4.3.) y se distribuyó el *pos*cuestionario. Finalmente, se analizaron los datos del *pos*cuestionario.

4.3. Participantes

En 2023-2024, la asignatura contó con 174 estudiantes, divididos en ocho subgrupos⁸ que siguieron el mismo programa temático y de evaluación, de modo que todos participaron en la intervención. El perfil es diverso, con presencia tanto de alumnado incorporado desde Bachillerato, como con formación académica o profesional en ciencias de la salud (CCS) en distintas especializaciones (63%). Un 31% compagina los estudios con trabajo en hospitales, residencias, o clínicas privadas. La competencia en inglés varía desde alumnado adulto con escasa formación previa y mínima exposición, hasta alumnado con formación en educación reglada (83,6%) ocasionalmente complementada en academias privadas (32,8%), lugar exclusivo de formación para el 7,5%. El 48% poseía una certificación oficial (B1 o B2, mayoritariamente).

En cuanto a conocimientos de mediación, el 58,2% no poseía formación previa y el 23,9% dudaba haberla recibido. Solamente el 14,9% respondió afirmativamente, mencionando contenidos relacionados con la mediación intercultural o entendida como “resolución de conflictos con pacientes o familiares complicados”, tratados en asignaturas de otras titulaciones sanitarias (9%), cursos específicos (4,5%) o ambos (1,5%). Únicamente el alumnado con experiencia en EOI (y, en menor medida, alumnado procedente de bachillerato) había trabajado la mediación *lingüística*. La mayoría relacionaba mediación con ideas como “negociar”, “arbitrar”, “conflictos” o “acuerdo” y, solo la minoría familiarizada con mediación lingüística, con “comunicación”, “traducción/explicación”, “saber expresarse” o “sensibilizar/aprender”.

⁸ Intervención realizada conjuntamente entre las autoras y otras dos docentes, a quienes se extiende el reconocimiento pertinente.

4.4. Diseño

La intervención consistió en integrar la mediación como objetivo y herramienta pedagógicos mediante la incorporación al temario de conceptos teóricos básicos —enlazados con contenidos lingüísticos relevantes y mediante un enfoque comunicativo— y de su tratamiento explícito en actividades de producción lingüística oral o escrita. Estas se complementaron con actividades de reflexión-concienciación para fomentar la capacidad de identificar requerimientos comunicativos y de realizar adaptaciones según el perfil de paciente. Los perfiles trabajados (pacientes con ceguera, con discapacidad intelectual y de edad avanzada o temprana) buscaban situar al interlocutor-paciente en el centro del proceso comunicativo enfatizando el valor de la lengua como herramienta de accesibilidad, y de la mediación *intra*lingüística como conjunto de estrategias de comunicación inclusiva.

La elección de estos perfiles responde a criterios didácticos y profesionales. En cuanto a los didácticos, la comunicación con personas de edad avanzada, con discapacidad intelectual, o con niños/as requiere adaptaciones lingüísticas concretas que se pueden formular como actividades y estrategias de mediación (condensar información, simplificar estructuras sintácticas, usar descripciones o comparaciones, repetir y reformular mensajes, etc.). Por su parte, la comunicación con personas con ceguera exige la explicitación lingüística de las acciones que lleva a cabo el profesional durante una consulta o exploración, dado que el paciente no tiene acceso a la información visual; llevado al aula, esto genera una motivación para prestar atención a la precisión lingüística al describir pasos o acciones. En cuanto a los criterios profesionales, estos perfiles representan pacientes con los que previsiblemente el alumnado tendrá que interactuar una vez se incorpore al mundo laboral.

Las actividades de aula reproducían las tres actividades de mediación identificadas en el VC (2020) —mediación de textos, de conceptos, y de la comunicación— y exigían la aplicación de estrategias de mediación (apartados 2.2.1. y 2.2.2.). En el diseño, se siguieron los pasos descritos en la *Figura 1*:

- 1) Establecimiento de objetivos pedagógicos (apartado 3.1.).
- 2) Identificación de temas relevantes.
- 3) Identificación de contenidos específicos.
- 4) (Re)diseño de actividades.
- 5) Búsqueda y creación de materiales de apoyo.
- 6) Diseño de pruebas de evaluación.

Los temas donde se identificó mayor potencial pedagógico (paso 2) fueron: “Dolor”, “Constantes vitales” y “Medicación”. En ellos, se identificaron contenidos específicos susceptibles de ser abordados mediante actividades y estrategias de mediación lingüística (paso 3). Posteriormente (paso 4), se adaptaron o crearon actividades para trabajar la mediación lingüística en relación con funciones comunicativas concretas (dar/solicitar información, transmitir resultados, etc.). Las actividades de reflexión grupal se realizaron por escrito en documentos colaborativos completados en pequeños grupos, y se comentaron oralmente entre toda la clase. Estas actividades se enfocaron en la identificación de requerimientos comunicativos por perfil de paciente y en la anticipación de acciones/estrategias de adaptación, a modo de preparación para las actividades de interacción, y la simulación del examen oral. Además, se proporcionaron materiales de apoyo (paso 5), incorporados al tema correspondiente o a la plataforma virtual. Estos incluían una tabla con conceptos básicos de mediación en relación con el temario y los objetivos pedagógicos; guías de buenas prácticas; guías para la elaboración de tareas de mediación escrita y para la preparación de interacciones orales; y testimonios de pacientes reales adaptados.

Las destrezas de mediación escrita (trabajadas principalmente en “Medicación”) se evaluaron mediante una prueba de redacción en la que el alumnado debía proporcionar instrucciones y consejos para la administración de medicamentos e información sobre su función a pacientes con requerimientos comunicativos: una niña de 8 años, o un hombre de mediana edad con discapacidad cognitiva. El enunciado proporcionaba contexto para la interacción, contenidos mínimos requeridos, foco lingüístico, y objetivos comunicativos y de mediación.

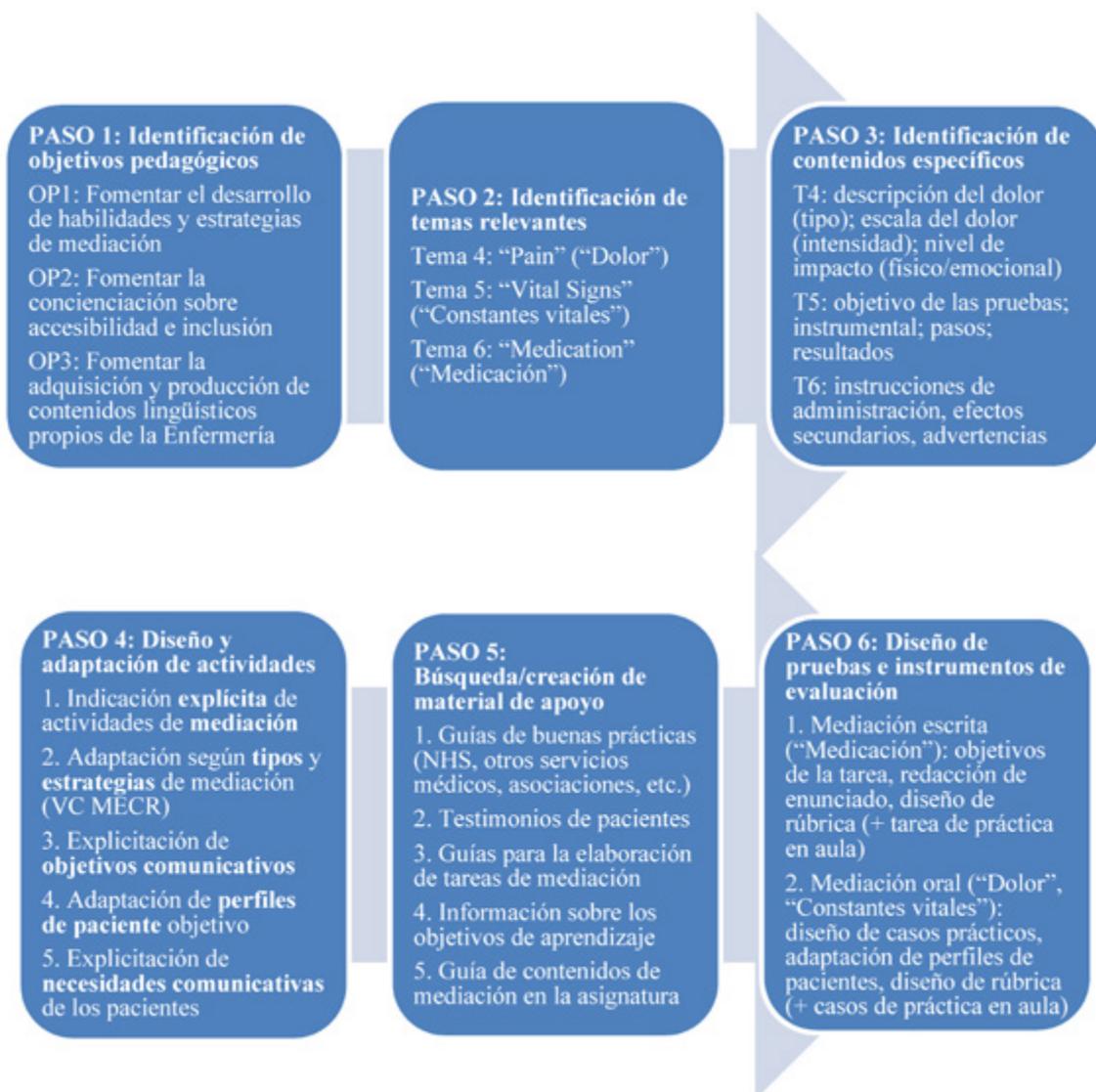


Figura 1. Pasos del diseño de la intervención

Las destrezas de mediación oral se evaluaron en el componente oral del examen final. Esta prueba consistió en la simulación de un diálogo entre un/a profesional y un/a paciente con alguno de los requerimientos comunicativos practicados en clase. Se diseñaron seis casos, tres para “Dolor” (donde el/la profesional debía realizar una valoración del dolor, asistiendo al/la paciente en la descripción) y tres para “Constantes” (donde el/la profesional debía tomar al menos dos constantes vitales, explicando el proceso detalladamente). Los casos incluían la contextualización de cada rol y los objetivos comunicativos y lingüísticos. Tras elegir aleatoriamente caso y rol, el alumnado disponía de dos minutos de preparación individual, con opción de tomar notas para consultar durante el diálogo.

Aquí, la inclusión de la mediación lingüística buscaba incrementar la producción de lenguaje objetivo mediante la necesidad de comunicarse con claridad y precisión según los requerimientos del/la paciente. En cuanto a las habilidades de mediación *intra*lingüística, se valoraba la capacidad del alumnado de detectar esos requerimientos y esforzarse por realizar las adaptaciones lingüísticas, conceptuales y comunicativas pertinentes para facilitar la comunicación accesible.

4.5. Instrumentos

Los datos de percepción sobre la intervención (fuente de los resultados de investigación aquí presentados) se recopilaron mediante un poscuestionario creado en *Microsoft Forms*,⁹ estructurado en seis apartados, con 8 preguntas obligatorias y 51 ítems en total, distribuidos según la *Figura 2*:



Figura 2. Estructura del poscuestionario

En los apartados (1) a (5), se utilizó una escala Likert de acuerdo/desacuerdo con cuatro valores (*Totalmente en desacuerdo*, *En desacuerdo*, *De acuerdo*, *Totalmente de acuerdo*) y la opción *No estoy seguro/a*. En el apartado (6) se combinaron el formato escala para la pregunta de satisfacción global, con valores del 1 (*No me ha gustado nada*) al 5 (*Me ha gustado mucho*), con el de respuesta abierta para las dos preguntas finales: "aspectos destacados", y "comentarios adicionales".

Los ítems se redactaron en enunciados afirmativos y negativos. El apartado (1) estudiaba la evolución en las percepciones del alumnado acerca del concepto "mediación" (respecto a las manifestadas en el precuestionario) mediante ítems como "La mediación consiste exclusivamente en la resolución de conflictos", "Existen distintos tipos de mediación", o "Mediación puede implicar adaptar el lenguaje". El apartado (2) buscaba valorar si esa evolución conllevaba la percepción de una aplicación directa de la mediación lingüística sobre la práctica profesional

⁹ Aprobado por el Comité de Ética USJ. Las respuestas fueron debidamente anonimizadas. Poscuestionario completo disponible en apéndice.

mediante ítems como “Tener conocimientos/capacidades de mediación ayuda a identificar y entender las necesidades de los/las pacientes”, o “Tener conocimientos/capacidades de mediación solo es importante en el trato con pacientes de otras culturas”.

El apartado (3) exploraba el impacto sobre el aprendizaje lingüístico general considerando distintas destrezas o elementos, tanto en la *adquisición* (“Los contenidos de mediación me han ayudado a adquirir conciencia sobre el uso del inglés”, o “me han ayudado a adquirir vocabulario en inglés”) como en la *producción* (“Los contenidos de mediación me han ayudado a aplicar/practicar el vocabulario en inglés”). El (4) exploraba el impacto en el aprendizaje de inglés para CCS, particularmente sobre la motivación o la relevancia (“Incluir la mediación resulta motivador para el aprendizaje del inglés”) y al desarrollo de habilidades transferibles (“Incluir la mediación contribuye a la adquisición de otras habilidades relevantes para la profesión sanitaria”). Finalmente, el (5) investigaba las percepciones sobre el impacto en la asignatura, con énfasis en la motivación/relevancia, la dificultad (“Incluir la mediación ha hecho que la asignatura fuera más complicada para mí”) o la apreciación (“Incluir la mediación ha hecho la asignatura más disfrutable para mí”).

5. RESULTADOS

5.1. Metodología de análisis

Para seleccionar los respondientes objetivo, se utilizó el muestreo no probabilístico por conveniencia,¹⁰ pues la selección venía predeterminada por el contexto de intervención (Cohen, Manion, y Morrison, 2007). Los respondientes objetivo eran el alumnado matriculado en la asignatura. Se obtuvieron 40 respuestas (de 174 posibles, 23%), número considerado aceptable dado que todos participaron en todos los aspectos de la intervención en mayor o menor medida—poseían una base experiencial suficiente—, y que se realiza un análisis descriptivo (Cohen et al., 2007). Se sigue una metodología mixta, combinando análisis cuantitativo (descriptivo-exploratorio) para las preguntas cerradas (apartado 5.2.) y cualitativo (apartado 5.3.) para las abiertas, basado este último en la exploración inductiva (Cohen et al. (2007: 472-473).

5.2. Análisis cuantitativo

El análisis cuantitativo presenta las tasas de acuerdo/desacuerdo con cada ítem,¹¹ mediante porcentajes acumulados de los dos valores a cada lado de la escala: *Totalmente en desacuerdo* más *En desacuerdo*; y *De acuerdo* más *Totalmente de acuerdo*. Las figuras muestran el número de respuestas para cada ítem y valor de acuerdo/desacuerdo, y etiquetas abreviadas del ítem, reproducidas en cursiva en el texto. Las tablas muestran la correspondencia entre número de respuestas y porcentaje sobre el total.

Los ítems referentes a la interpretación del concepto de mediación (*Figura 3/Tabla 1*) abarcan cuatro categorías: (i) concepción, (ii) interlocutores, (iii) rasgos y (iv) funciones. En cuanto a (i), aunque un porcentaje elevado (42,5%) sigue entendiendo mediación exclusivamente como resolución de *conflictos* o sinónimo de *negociación*, la mayoría manifiesta desacuerdo (57,5% y 52,5%, respectivamente) y el 82,5% muestra acuerdo con que existen distintos *tipos*. Esta tendencia se diluye al equiparar *mediación interlingüística* con *traducción*: el 47,5% las

¹⁰ Habitual en investigaciones a pequeña escala (estudios de caso, intervenciones didácticas). Usa audiencias “cautivas” (e.g., una clase) como respondientes y no asume generabilidad (Cohen et al., 2007: 114). *Respondientes objetivo* se refiere a quienes recibieron el cuestionario y *respondientes*, a quienes lo completaron.

¹¹ La tasa de respuesta para “No estoy seguro/a” no supera el 10% (alcanzado solo en 3 ítems, todos sobre el concepto “mediación”). Esto incrementa la fiabilidad del instrumento y los resultados, al reducirse el factor de incertidumbre.

considera equivalentes, frente al 42,5% que no. En cuanto a (ii), el 72,5% considera que en contextos sanitarios la mediación no se limita a la comunicación con *pacientes* y el 88% afirma que puede darse entre *profesionales*. Además, el 85% manifiesta desacuerdo con restringirla a interacciones entre personas de distintas *culturas*.

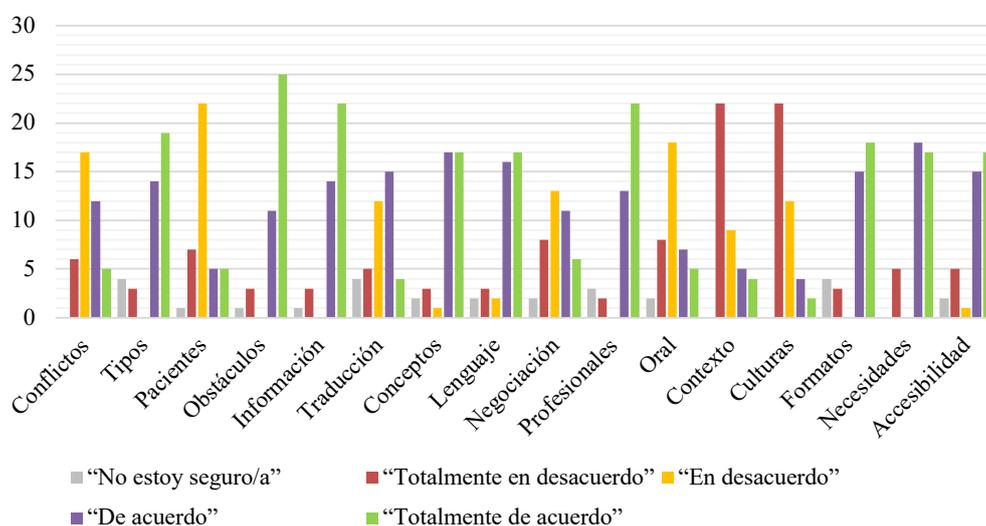


Figura 3. *Percepciones sobre el concepto de mediación*

Tabla 1. *Correspondencia datos numéricos y porcentuales: “Percepciones sobre el concepto de mediación” (Figura 3)*

Ítem	“No estoy seguro/a”	“Totalmente en desacuerdo”	“En desacuerdo”	“De acuerdo”	“Totalmente de acuerdo”
Conflictos	0	6 (15%)	17 (42,5%)	12 (30%)	5 (12,5%)
Tipos	4 (10%)	3 (7,5%)	0	14 (35%)	19 (47,5%)
Pacientes	1 (2,5%)	7 (17,5%)	22 (55%)	5 (12,5%)	5 (12,5%)
Obstáculos	1 (2,5%)	3 (7,5%)	0	11 (27,5%)	25 (62,5%)
Información	1 (2,5%)	3 (7,5%)	0	14 (35%)	22 (55%)
Traducción	4 (10%)	5 (12,5%)	12 (30%)	15 (37,5%)	4 (10%)
Conceptos	2 (5%)	3 (7,5%)	1 (2,5%)	17 (42,5%)	17 (42,5%)
Lenguaje	2 (5%)	3 (7,5%)	2 (5%)	16 (40%)	17 (42,5%)
Negociación	2 (5%)	8 (20%)	13 (32,5%)	11 (27,5%)	6 (15%)
Profesionales	3 (7,5%)	2 (5%)	0	13 (32,5%)	22 (62,5%)
Oral	2 (5%)	8 (20%)	18 (45%)	7 (17,5%)	5 (12,5%)
Contexto	0	22 (55%)	9 (22,5%)	5 (12,5%)	4 (10%)
Culturas	0	22 (55%)	12 (30%)	4 (10%)	2 (5%)
Formatos	4 (10%)	3 (7,5%)	0	15 (37,5%)	18 (45,5%)
Necesidades	0	5 (12,5%)	0	18 (45%)	17 (42,5%)
Accesibilidad	2 (5%)	5 (12,5%)	1 (2,5%)	15 (37,5%)	17 (42,5%)

En cuanto a (iii), la mayoría asocia mediación a *adaptaciones*: del *lenguaje* (82,5%), o de los *formatos* o medios de transmisión (83%). El 65% rechaza la idea de que se produzca exclusivamente en comunicación *oral*, y el 79,5% la de que el *contexto* comunicativo *no es importante*. Las tendencias respecto a las funciones de la mediación (iv) son claras: el 90% considera que ayuda a superar *obstáculos* comunicativos y facilitar la comprensión de *conceptos* o mensajes especializados, el 80% la relaciona con *accesibilidad* y el 87,5%, con identificación de *necesidades*.

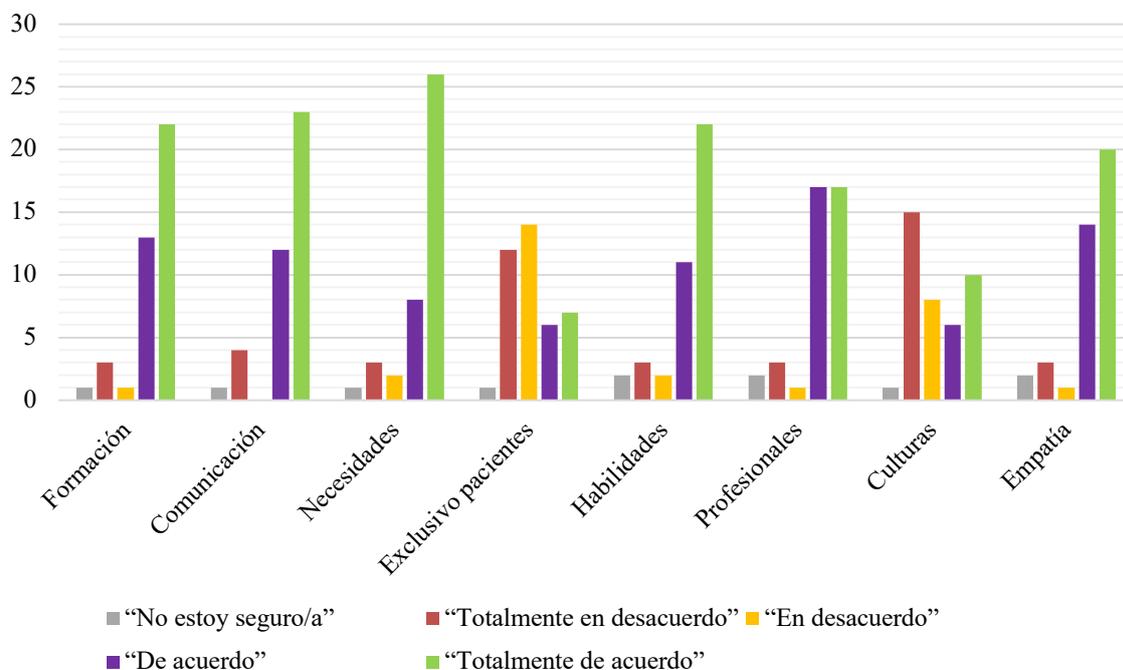


Figura 4. Percepciones: valor de la mediación en sanidad

Tabla 2. Correspondencia datos numéricos y porcentuales: "Percepciones sobre mediación en sanidad" (Figura 4)

Ítem	"No estoy seguro/a"	"Totalmente en desacuerdo"	"En desacuerdo"	"De acuerdo"	"Totalmente de acuerdo"
Formación	1 (2,5%)	3 (7,5%)	1 (2,5%)	1 (2,5%)	22 (55%)
Comunicación	1 (2,5%)	4 (10%)	0	0	23 (57,5%)
Necesidades	1 (2,5%)	3 (7,5%)	2 (5%)	2 (5%)	26 (65%)
Exclusivo pacientes	1 (2,5%)	12 (30%)	14 (35%)	14 (35%)	7 (17,5%)
Habilidades	2 (5%)	3 (7,5%)	2 (5%)	2 (5%)	22 (55%)
Profesionales	2 (5%)	3 (7,5%)	1 (2,5%)	1 (2,5%)	17 (42,5%)
Culturas	1 (5%)	15 (37,5%)	8 (20%)	8 (20%)	10 (25%)
Empatía	2 (5%)	3 (7,5%)	1 (2,5%)	1 (2,5%)	20 (50%)

Los ítems sobre percepción del valor de la mediación en sanidad (Figura 4/Tabla 2) abarcan dos categorías: (i) formación y (ii) comunicación. En cuanto a (i), la mayoría considera la mediación importante para la *formación* general del personal sanitario (87,5%); para el desarrollo de otras *habilidades* profesionales (82,5%); o para el fomento de la *empatía* hacia la dis-

capacidad (85%) y de la habilidad de identificar *necesidades* y realizar adaptaciones (85%). En cuanto a (ii), el 87,5% considera la mediación fundamental para la *comunicación* con pacientes, pero no exclusivamente (el 65% rechaza esta restricción). Una ligera mayoría (57,5%) rechaza la limitación de mediación a comunicación entre *culturas*.

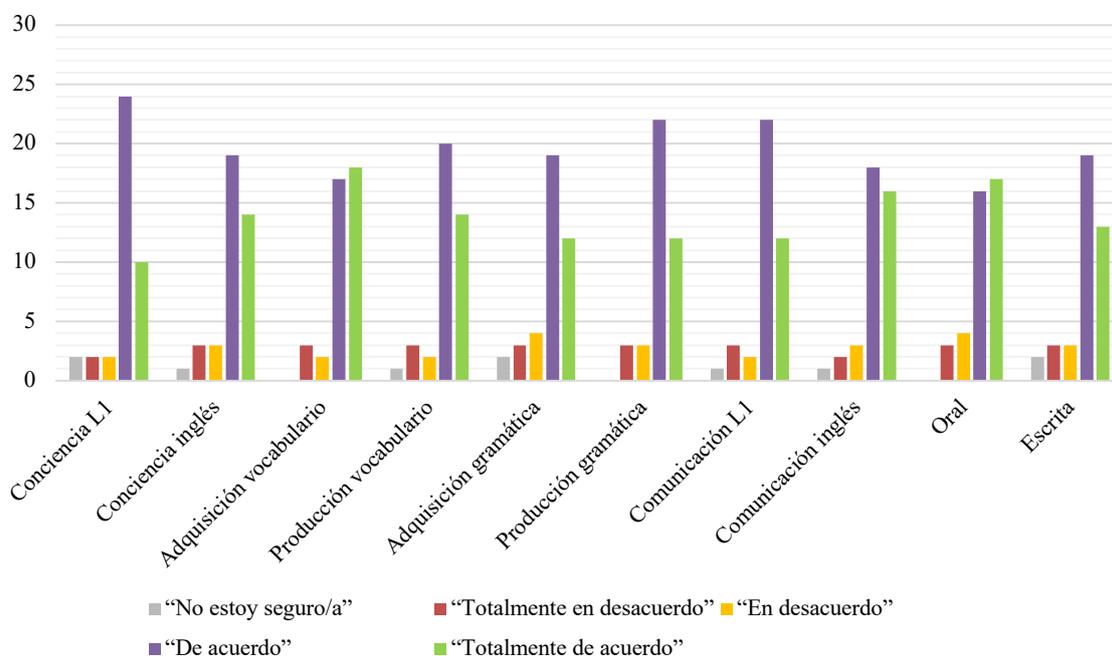


Figura 5. Percepciones: valor de la mediación para el aprendizaje lingüístico

Tabla 3. Correspondencia datos numéricos y porcentuales: "Percepciones sobre la mediación para el aprendizaje lingüístico" (Figura 5)

Ítem	"No estoy seguro/a"	"Totalmente en desacuerdo"	"En desacuerdo"	"De acuerdo"	"Totalmente de acuerdo"
Conciencia L1	2 (5%)	2 (5%)	2 (5%)	24 (60%)	10 (25%)
Conciencia inglés	1 (2,5%)	3 (7,5%)	3 (7,5%)	19 (47,5%)	14 (35%)
Adquisición vocabulario	0	3 (7,5%)	2 (5%)	17 (42,5%)	18 (45%)
Producción vocabulario	1 (2,5%)	3 (7,5%)	2 (5%)	20 (50%)	14 (35%)
Adquisición gramática	2 (5%)	3 (7,5%)	4 (10%)	19 (47,5%)	12 (30%)
Producción gramática	0	3 (7,5%)	3 (7,5%)	22 (55%)	12 (30%)
Comunicación L1	1 (2,5%)	3 (7,5%)	2 (5%)	22 (55%)	12 (30%)
Comunicación inglés	1 (2,5%)	2 (2,5%)	3 (7,5%)	18 (45%)	16 (40%)
Oral	0	3 (7,5%)	4 (20%)	16 (40%)	17 (32,5%)
Escrita	2 (5%)	3 (7,5%)	3 (7,5%)	19 (47,5%)	13 (32,5%)

Los ítems sobre percepciones del valor de la mediación para el aprendizaje lingüístico (Figura 5/Tabla 3) abarcan tres categorías: (i) uso, (ii) aprendizaje y (iii) competencias. En cuanto a (i), existe unanimidad sobre que la mediación contribuye a adquirir *conciencia* del uso del *inglés* (82,5%) y de la *L1* (85%). En cuanto a (ii), existen percepciones positivas para el

aprendizaje léxico y gramatical: el 87,5% la considera beneficiosa para la *adquisición* de *vocabulario* y el 77,5% para la de *estructuras gramaticales*; además, el 85% considera que fomenta la *producción* en ambos. En cuanto a (iii), el 85% considera que ayuda a desarrollar competencias de *comunicación* en ambas lenguas, tanto *escrita* (80%) como *oral* (72,5%).

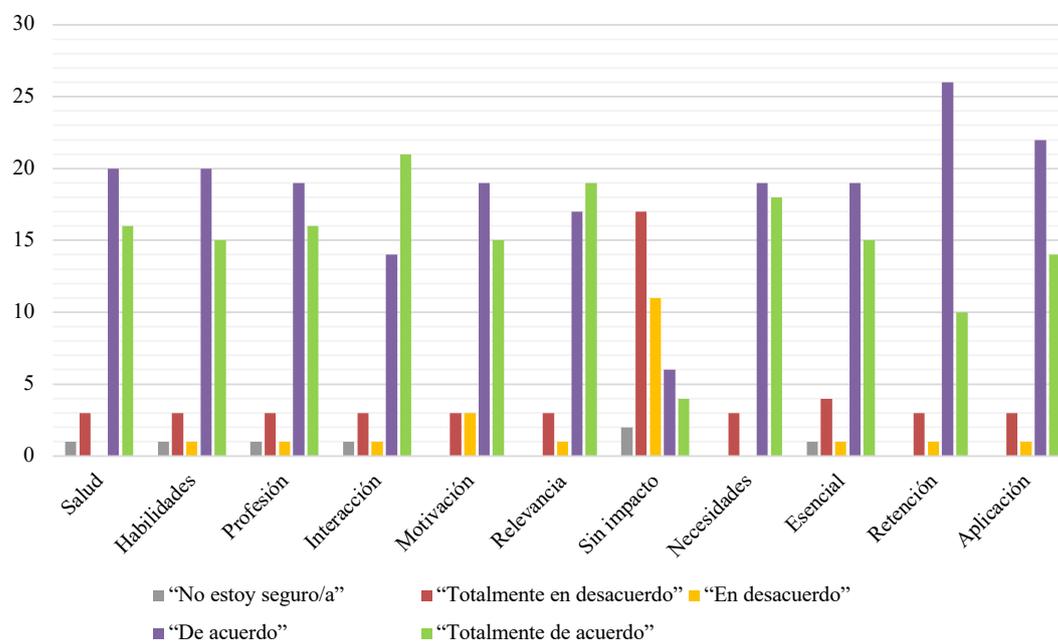


Figura 6. Percepciones: valor de la mediación en inglés para CCS

Tabla 4. Correspondencia datos numéricos y porcentuales: “Percepciones sobre la mediación para CCS” (Figura 6)

Ítem	“No estoy seguro/a”	“Totalmente en desacuerdo”	“En desacuerdo”	“De acuerdo”	“Totalmente de acuerdo”
Salud	1 (2,5%)	3 (7,5%)	0	20 (50%)	16 (40%)
Habilidades	1 (2,5%)	3 (7,5%)	1 (2,5%)	20 (50%)	15 (37,5%)
Profesión	1 (2,5%)	3 (7,5%)	1 (2,5%)	19 (47,5%)	16 (40%)
Interacción	1 (2,5%)	3 (7,5%)	1 (2,5%)	14 (35%)	21 (52,5%)
Motivación	0	3 (7,5%)	3 (7,5%)	19 (47,5%)	15 (37,5%)
Relevancia	0	3 (7,5%)	1 (2,5%)	17 (42,5%)	19 (57,5%)
Sin impacto	2 (5%)	17 (42,5%)	11 (27,5%)	6 (15%)	4 (10%)
Necesidades	0	3 (7,5%)	0	19 (47,5%)	18 (45%)
Esencial	1 (2,5%)	4 (10%)	1 (2,5%)	19 (47,5%)	15 (37,5%)
Retención	0	3 (7,5%)	1 (2,5%)	26 (65%)	10 (25%)
Aplicación	0	3 (7,5%)	1 (2,5%)	22 (55%)	14 (35%)

Los ítems de percepción del valor de la mediación en inglés para CCS (Figura 6/Tabla 4) abarcan tres categorías: (i) aprendizaje, (ii) valor añadido y (iii) concienciación. El 90% considera que favorece el aprendizaje del inglés para *salud*, tanto en la *aplicación* de contenidos lingüísticos (90%) como en su *retención* (85%). Además, el 87,5% afirma que promueven la

adquisición de *habilidades* relevantes para profesiones sanitarias. En cuanto a (ii), la mayoría considera que incorporarla incrementa la *relevancia* (90%) y la *motivación* (85%) y que debería convertirse en un elemento *esencial* en estas asignaturas. Sin embargo, ante la afirmación *Incluir la mediación no aporta nada al aprendizaje del inglés (sin impacto)*, el acuerdo disminuye hasta el 70%. En cuanto a (iii), el 92,5% afirma que la mediación lingüística fomenta la concienciación sobre *necesidades* comunicativas y conexiones entre lenguaje y otros aspectos de la *profesión* (87,5%) al hacerla *interacción* en el aula más realista (87,5%).

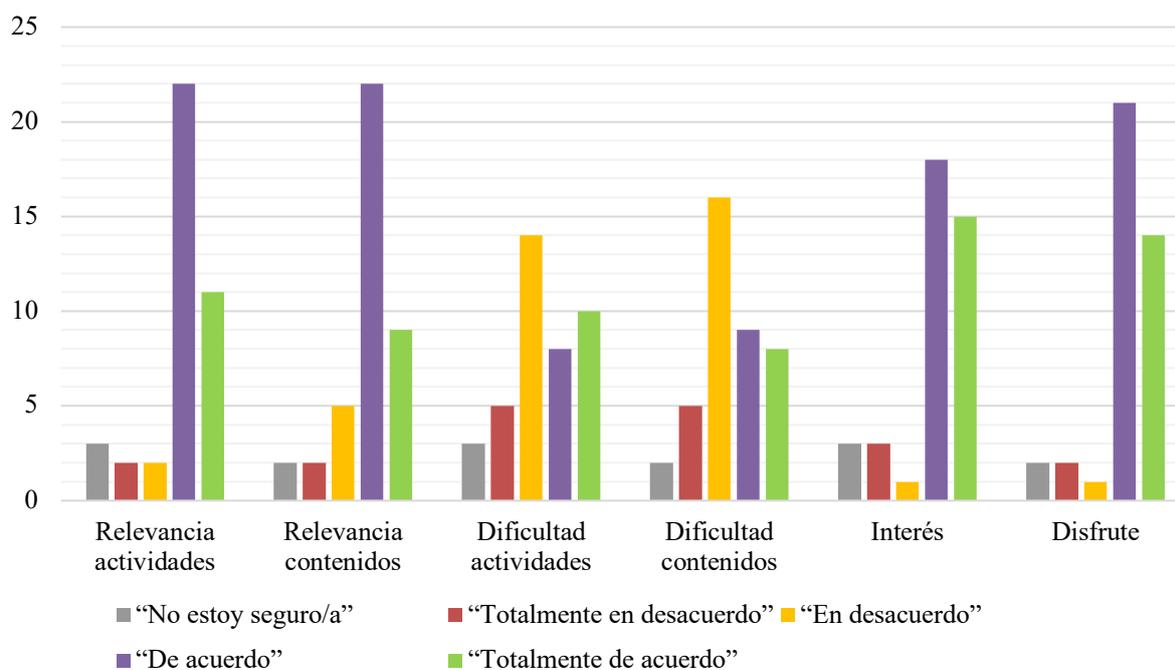


Figura 7. Percepciones: valor de la mediación en inglés para la asignatura

Tabla 5. Correspondencia datos numéricos y porcentuales: "Percepciones sobre la mediación para la asignatura" (Figura 7)

Ítem	"No estoy seguro/a"	"Totalmente en desacuerdo"	"En desacuerdo"	"De acuerdo"	"Totalmente de acuerdo"
Relevancia actividades	3 (7,5%)	2 (5%)	2 (5%)	22 (55%)	11 (27,5%)
Relevancia contenidos	2 (5%)	2 (5%)	5 (12,5%)	22 (55%)	9 (22,5%)
Dificultad actividades	3 (7,5%)	5 (12,5%)	14 (35%)	8 (20%)	10 (25%)
Dificultad contenidos	2 (5%)	5 (12,5%)	16 (40%)	9 (22,5%)	8 (20%)
Interés	3 (7,5%)	3 (7,5%)	1 (2,5%)	18 (45%)	15 (37,%)
Disfrute	2 (5%)	2 (5%)	1 (2,5%)	21 (52,5%)	14 (35%)
Relevancia actividades	3 (7,5%)	2 (5%)	2 (5%)	22 (55%)	11 (27,5%)

Los ítems de percepciones sobre el valor de incorporar la mediación a la asignatura (Figura 7/Tabla 5) abarcan tres categorías: (i) relevancia, (ii) dificultad y (iii) apreciación. La mayoría considera que incrementa especialmente la *relevancia* de ciertas *actividades* (82,5%), pero también de algunos *contenidos* (77,5%). Sobre (ii) existe más división, particularmente sobre el aumento de *dificultad* de algunas *actividades*: el 45,5% la considera incrementada, frente al

47,5% que no; para *contenidos*, la diferencia aumenta ligeramente: el 42,5% piensa que la mediación incrementa su dificultad y el 52,5%, que no. Sobre (iii), existe consenso en que aumenta el grado de *disfrute* (87,5%) y el *interés* hacia la materia (82,5%).

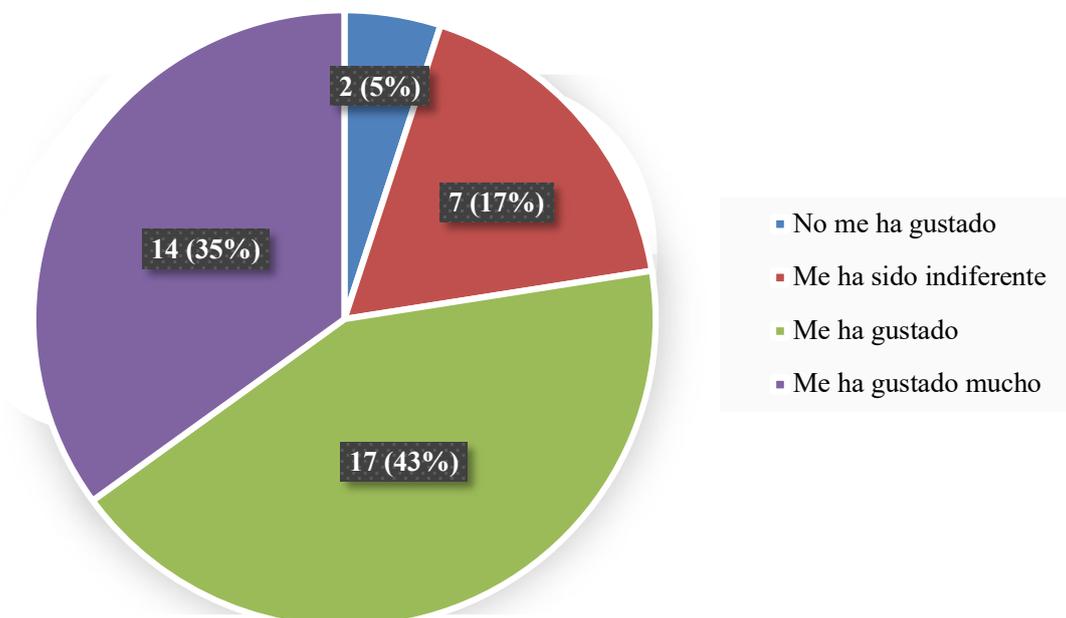


Figura 8. Satisfacción global con la intervención

Finalmente, la *Figura 8* muestra la satisfacción global con la intervención, que es elevada: un porcentaje acumulado del 78% manifiesta satisfacción alta (43%) o muy alta (35%), mientras solo un 5% opta por *No me ha gustado* y *No me ha gustado nada* no obtiene respuestas. Un 17% manifiesta indiferencia.

5.3. Análisis cualitativo

El análisis cualitativo considera las respuestas a las dos preguntas abiertas (PA): 8. *¿Qué valores más sobre la inclusión de la mediación en los contenidos de la asignatura?* y 9. *¿Hay alguna otra cosa que te gustaría comentar sobre la inclusión de la mediación en la asignatura?* El análisis comprendió los siguientes pasos: (i) análisis exploratorio inductivo de las respuestas para identificar temas; (ii) creación de etiquetas; (iii) codificación de los contenidos; (iv) revisión de etiquetas; y (v) recodificación final. La *Figura 9* muestra las etiquetas finales (categorías temáticas) para las PA. La PA8 recibió 34 respuestas válidas (85%); del resto, cuatro (10%) no aportaron contenido y dos (5%) no procedían. En la PA9, 15 (37,5%) resultaron válidas, 23 (57,5%) no aportaron contenido y dos (5%) no procedían.

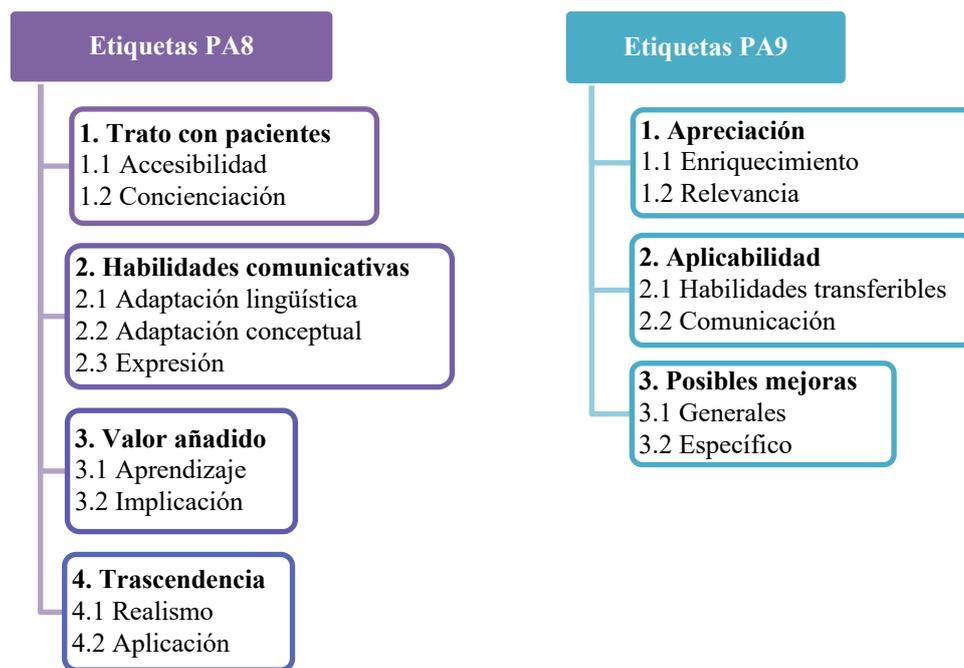


Figura 9. Etiquetas de codificación cualitativa

En las respuestas a la PA8 emergen cuatro temas: foco en el *trato con pacientes*, desarrollo de *habilidades comunicativas*, *valor añadido* para la asignatura y *trascendencia* de los contenidos. En el primero, destacan aspectos relacionados con la *accesibilidad* y la *concienciación*, especialmente con la facilitación del acceso a conceptos o contenidos, con “aprender a tratar con personas de manera adaptada a sus posibilidades” y con “la importancia de poder comunicarse con todos los pacientes”, mediante adaptaciones tanto lingüísticas como actitudinales. Una respuesta afirma que la mediación “permite una inmersión más real en las diferentes realidades de cada paciente” y otra valora haberse puesto “en esa situación futura” de dirigirse a pacientes realizando adaptaciones. Otras respuestas destacan la concienciación sobre las realidades de “personas con distintas capacidades” y la necesidad de comprenderlas y de adaptarse a ellas.

Destaca la valoración positiva de la mediación para el aprendizaje lingüístico, especialmente por contribuir a concienciar sobre la necesidad de adaptar el lenguaje “en distintas situaciones” para “facilitar la comprensión y comunicación de conceptos complejos”. Otras respuestas enfatizan el “conocimiento de cómo hablar y utilizar palabras más fáciles de entender para los pacientes”, o añaden “que la mediación... te ayuda a desenvolverte mejor con los... compañeros de oficio”. Destacan las numerosas referencias al fomento de la conciencia y la capacidad comunicativa oral y escrita (incluso en la L1), especialmente para “reaccionar de manera espontánea... y desenvolver[se] mejor”.

Como valor añadido para la asignatura, destacan el efecto sobre el aprendizaje y el aumento de la implicación. Subrayan la importancia global de adquirir estos conocimientos y competencias, independientemente del reto individual enfrentado. Recalcan el aporte de la mediación para la comprensión e interiorización del léxico u otros aspectos lingüísticos, para desarrollar habilidades de resolución de problemas y para cohesionar el aprendizaje. Una respuesta agradece la no limitación de los contenidos “simplemente [a] lo básico como el vocabulario”, y otra añade que esa mayor cohesión contribuye a involucrar más al alumnado haciendo las clases “más amenas”.

Finalmente, las respuestas reconocen la trascendencia para el futuro profesional. Muchas destacan que la mediación “acerca al estudiante al uso del inglés en la vida real”, al ser una “forma muy realista de aplicar[lo] al contexto laboral, planteando situaciones reales que pueden estar presentes en un futuro”, y al prepararles para “solventar los problemas que se pueden presentar”. Resaltan tanto “el conocimiento de cómo hablar” para comunicarse adecuadamente con pacientes, como que “la mediación sirve para saber hablar correctamente en los distintos ámbitos de tu vida”, y que lo aprendido proporciona “una base de la que tirar” para no sentirse “completamente “primerizos” ante estas situaciones”.

En las respuestas a la PA9 emergen tres categorías temáticas: *apreciación, aplicabilidad y posibles mejoras*. Enfatizan lo resaltado en la PA8, especialmente respecto al disfrute, y destacan que “aportar diferentes puntos de vista ha hecho más enriquecedor el aprendizaje”. Asimismo, valoran el planteamiento de las mediaciones y lo “fundamental” de desarrollar estas competencias. Subrayan el potencial de la mediación para el desarrollo de habilidades transferibles, como “habilidades críticas de interpretación... esenciales para la colaboración efectiva en contextos internacionales y multidisciplinarios”, o la “adaptabilidad y flexibilidad” para ajustarse al contexto. Finalmente, se aportan propuestas de mejora. La única valoración negativa demanda profundizar “en las explicaciones de la mediación... así como practicar [estructuras] más antes del examen”, al considerar insuficiente la preparación recibida. Otras sugieren mejoras generales (“más tiempo”), o específicas: incorporar pequeños diálogos con mediación al final de cada sesión, o “ver vídeos o más ejemplos de nativos en situación”.

6. DISCUSIÓN

Los resultados sugieren un elevado grado de consecución de los objetivos pedagógicos (OP, apartado 3.1.) y los investigadores (OI, apartado 3.2.). En cuanto a OP1, los datos cuantitativos sugieren una concepción ampliada del término *mediación* tras la intervención: aunque la asociación *mediación-resolución de conflictos/negociación* sigue siendo fuerte, la mayoría la asocia también a *accesibilidad, identificación de necesidades, adaptaciones o superación de obstáculos*. Además, las respuestas evidencian la adquisición de conocimientos sobre mediación al reconocer la existencia de diversos tipos, la relevancia del contexto para aplicarlos, y su valor para la comunicación general y especializada, hacia pacientes y entre profesionales. Esto queda respaldado por los datos cualitativos: las respuestas abiertas resaltan el valor de la mediación intralingüística para el desarrollo de la autoconciencia y del reconocimiento de su influencia sobre la comunicación; de la empatía; y de la comprensión hacia personas con distintos requerimientos. Esto sugiere concienciación sobre cuestiones de accesibilidad y sobre la importancia de promover una comunicación inclusiva (OP2).

En cuanto a OP3, los datos cuantitativos sugieren que integrar la mediación en los contenidos de inglés ha favorecido el aprendizaje lingüístico, especialmente en la adquisición y producción del léxico, tanto oral como escrita. Destaca el mayor porcentaje que subraya el impacto sobre la producción escrita, pues la intervención enfatizó el desarrollo de destrezas orales. Este dato podría explicarse porque a fecha de compleción del poscuestionario el alumnado había recibido *feedback* de la tarea de mediación escrita, pero no del examen oral.

Los datos cuantitativos muestran también una percepción elevada de retención de los contenidos lingüísticos gracias a la mediación. Los cualitativos añaden detalles: el alumnado aprecia particularmente el aprendizaje logrado por la necesidad de realizar adaptaciones lingüísticas según el contexto y los requerimientos del interlocutor, y enfatiza la percepción de mejora en las habilidades expresivas e improvisacionales. Destaca que estas percepciones son independientes del grado de dificultad enfrentado—es decir, son manifestadas incluso por alumnado que considera que la incorporación de elementos de mediación incrementa la dificul-

tad inherente al aprendizaje del inglés. Esto, junto con las percepciones de valor añadido para la asignatura, permite concluir que los beneficios didácticos superan los posibles riesgos, y que una incorporación sistemática es tanto deseable como viable (OI1) y efectiva (OI2).

Asimismo, sobresale la elevada apreciación de relevancia de la mediación lingüística para el ámbito sanitario. Tanto las valoraciones cuantitativas como las cualitativas reconocen su importancia como habilidad esencial para gestionar adecuadamente interacciones esperables en el ámbito profesional, y como herramienta promotora de competencias transferibles como el pensamiento crítico, la adaptabilidad, la flexibilidad, la capacidad de resolución y la cooperación (OP1, OP2, OP3). Particularmente, el alumnado destaca el potencial de las actividades de mediación para recrear contextos e interacciones realistas y para incrementar la trascendencia de contenidos lingüísticos y conceptuales de la asignatura cuya aplicabilidad consideran fundamental para el desarrollo profesional (OP3, OI2).

Finalmente, procede identificar algunas limitaciones. Respecto al poscuestionario, el ítem *Incluir la mediación no aporta nada al aprendizaje del inglés* podría haber generado confusión por estar redactado en negativo. Esto explicaría que solo el 70% de respondientes (porcentaje 10 puntos inferior al de los demás ítems con afirmaciones sobre el valor de la mediación) manifestaran desacuerdo. Comprobar las respuestas individuales reveló que respondientes que para este ítem seleccionaron las opciones *De acuerdo* o *Totalmente de acuerdo*, en otros valoraron muy positivamente el impacto de la mediación. Esto sugiere que algunas de las respuestas a *No aporta nada* podrían deberse a la inercia de seleccionar *acuerdo* para los ítems circundantes. Por otra parte, la temporalización del poscuestionario en periodo de exámenes pudo reducir la participación. La inferior tasa de respuesta respecto al precuestionario imposibilita la comparación individualizada de las respuestas, lo cual habría permitido profundizar en el análisis de las percepciones considerando factores como formación o experiencia profesional. Por último, en la redacción de preguntas e ítems se usó el término *necesidades* en vez de *requerimientos*. Esto se revisará para ajustarse a las recomendaciones de la ONU sobre lenguaje inclusivo. Respecto al diseño de la intervención, se observó la necesidad de ajustar mejor los perfiles de paciente a los objetivos y actividades de mediación para maximizar su beneficio didáctico, así como de escalonar los objetivos de aprendizaje para garantizar su consecución y el desarrollo de las habilidades asociadas. Todos estos aspectos se considerarán para el curso 2024-2025.

A modo de resumen, se puede afirmar que el alumnado participante demuestra un aprendizaje sobre la naturaleza del concepto de mediación, así como un aumento en su apreciación de la mediación *intra lingüística* como elemento clave para la accesibilidad comunicativa, así como para el desarrollo de su autoconciencia y empatía (habilidades fundamentales para profesionales sanitarios), y para favorecer el aprendizaje del inglés.

7. CONCLUSIONES

La mediación lingüística es una competencia esencial para futuros profesionales sanitarios, quienes desenvolverán su actividad laboral en un entorno de constante interacción interpersonal donde la atención individualizada es clave para promover un acceso equitativo a servicios básicos. La intervención didáctica presentada demuestra que la mediación lingüística puede integrarse de manera viable y efectiva en la enseñanza de inglés para ciencias de la salud, pues proporciona al alumnado herramientas fundamentales para comunicarse adecuadamente con pacientes con diversos requerimientos también aplicables en la L1.

Además, confirma la flexibilidad y adaptabilidad de la concepción de mediación lingüística defendida en el VC y demuestra que proporciona una base sólida para guiar una integración significativa en los currículos de LFE, a fin de dotarlos de enfoques metodológicos más dinámicos, holísticos e inclusivos. Este enfoque basado en la mediación no solo mejora las competencias lingüísticas del alumnado y su implicación en la asignatura y en el aprendizaje, sino

que promueve la inclusividad, reconoce la diversidad existente en la comunicación y fomenta el desarrollo de competencias esenciales para la participación efectiva en el ámbito profesional y en la sociedad en general.

En conclusión, este estudio subraya la importancia de la mediación lingüística en la formación de profesionales de la salud y recomienda su inclusión sistemática en la enseñanza de inglés para fines específicos en otras áreas de especialidad. Además, se demuestra que esta integración puede realizarse mediante actividades y herramientas fácilmente adaptables. Esto permitirá ampliar el foco de los programas de LFE al incorporar nociones de accesibilidad, lo cual enriquecerá la experiencia de aprendizaje en tanto que se incrementará la percepción de aplicabilidad a la práctica profesional, tanto para la comunicación entre especialistas como entre especialistas y legos. Así, el alumnado logrará adquirir un amplio repertorio de herramientas y estrategias para identificar y enfrentar los desafíos comunicativos de la profesión y para contribuir al desarrollo de modelos comunicativos basados en la inclusión, la accesibilidad y la equidad.

REFERENCIAS

- Alcaraz Mármol, G. (2019). Actitudes frente a la mediación lingüística en un contexto de educación secundaria. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 13(27), 112-128. <https://doi.org/10.26378/rnlael1327322>
- Baraldi, C. y Gavioli, L. (2019). *La mediazione linguistico-culturale nei servizi sanitari: Interazione ed efficacia comunicativa*. Milán: FrancoAngeli.
- Beacco, J.C., Byram, M., Cavalli, M., Coste, D., Egli Cuenat, M., Goullier, F. y Panthier, J. (2016). *Guide for the development and implementation of curricula for plurilingual and intercultural education* (English Edition). Council of Europe. ISBN 978-92-871-8234-0.
- Calatayud Díez, A. P. (2019). La mediación lingüística y las otras mediaciones: Aspectos clave para las escuelas del siglo XXI. *Supervisión*, 21(52), 1-18. <https://usie.es/supervision21/index.php/Sp21/article/view/379>
- Chovancová, B. (2018). Practicing the skill of mediation in English for legal purposes. *Studies in Logic, Grammar and Rhetoric*, 53(1), 49-60. <https://doi.org/10.2478/slgr-2018-0003>
- Cohen, L., Manion, L. y Morrison, K. (2007). *Research methods in education*. Londres/Nueva York: Routledge. <https://doi.org/10.4324/9780203029053>
- Consejo de Europa (2001). *Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación*. Estrasburgo: Publicaciones del Consejo de Europa.
- Consejo de Europa (2018). *Common European Framework of Reference for Languages. Companion Volume with New Descriptors*. Estrasburgo: Publicaciones del Consejo de Europa. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>
- Consejo de Europa (2020). *Marco Común Europeo de Referencia para las Lenguas: Aprendizaje, Enseñanza, Evaluación*. Estrasburgo: Publicaciones del Consejo de Europa. <https://rm.coe.int/marco-comun-europeo-de-referencia-para-las-lenguas-aprendizaje-ensenan/1680a52d53>
- De Arriba García, C. y Cantero Serena, F. J. (2004). La mediación lingüística en la enseñanza de lenguas. *Didáctica. Lengua y Literatura*, 16, 9-21.

Dendrinós, B. (2006). Mediation in communication, language teaching and testing. *Journal of Applied Linguistics*, 22, 9-35.

Dendrinós, B. (2018). Multilingualism language policy in the EU today: A paradigm shift in language education. *TLC Journal*, 2(3), 9-28. <https://doi.org/10.29366/2018tlc.2.3.1>

Martín-Macho Harrison, A. y Guadamillas Gómez, M. V. (2022). *Mediación lingüística en la enseñanza de lenguas: aportaciones del Volumen complementario y recursos para el aula*. Barcelona: Octaedro. <https://doi.org/10.36006/09127-4>

North, B. y Piccardo, E. (2016). *Developing Illustrative Descriptors of Aspects of Mediation for the CEFR*. Estrasburgo: Publicaciones del Consejo de Europa. <https://doi.org/10.1017/S0261444816000100>

Pazos Anido, M. (2013). La mediación lingüística en las clases de ELE de nivel inicial en el 'ensino básico' y el 'ensino secundário' en Portugal. En B. Blecua, S. Borrell, B. Crous, F. Sierra (Eds.), *Plurilingüismo y enseñanza de ELE en contextos multiculturales* (pp. 682-693). Gerona: ASELE.

Pazos Anido, M. (2018). Actividades y estrategias de mediación para el desarrollo de la competencia plurilingüe y pluricultural: aplicabilidad de las orientaciones del CEFR Companion Volume. En M. Sancho-Arnáiz y H. Otero-Doval (Eds.), *Internacionalización y enseñanza del español LE/L2: plurilingüismo y comunicación intercultural* (pp. 331-348). Gerona: ASELE.

Piccardo, E., North, B. y Goodier, T. (2019). Broadening the Scope of Language Education: Mediation, Plurilingualism, and Collaborative Learning: The CEFR Companion Volume. *Journal of e-Learning and Knowledge Society*, 15(1), pp. 17-36. DOI: [10.20368/1971-8829/1612](https://doi.org/10.20368/1971-8829/1612)

Sanz Esteve, E. (2020). La mediación lingüística en el aula de lenguas extranjeras. *El Español Por El Mundo*, 3(1), 193-220. <https://doi.org/10.59612/epm.vi3.91>

Trovato, G. (2016). *Mediación lingüística y enseñanza de español/LE*. Cuadernos de didáctica del español/LE. Madrid: Arco/Libros

VV.AA. (2021). When Clinicians and Patients do not Speak the Same Language - Interpreting in Health Care. *Health Communication*, 36(9). <https://www.tandfonline.com/toc/hhth20/36/9#>

Wright, R. y Spada Symonds, M. (2011). *English for Nursing 2*. Harlow: Pearson.

APÉNDICES

Apéndice 1

MEDIACIÓN EN INGLÉS CIENTÍFICO (POST-CUESTIONARIO)¹²

Objetivos, política de privacidad y tratamiento de los datos:

Este cuestionario tiene como objetivo recabar información sobre la **familiaridad del alumnado con el concepto de “mediación”** posteriormente a su **tratamiento en el aula**, así como sus **percepciones** acerca del valor de integrar la **mediación** dentro de los contenidos de la **asignatura de inglés científico**.

Los resultados **ANONIMIZADOS** podrán utilizarse para **publicaciones científicas**.

La **compleción de esta encuesta es VOLUNTARIA** y **no** está asociada a la calificación de la asignatura ni a la interacción alumnado-docentes.

Por favor, **selecciona las opciones que más se aproximen a tu realidad**. Ante cualquier pregunta, no dudes en contactar con tu profesor/a.

La encuesta tiene 8 preguntas y el tiempo estimado para completarla son **10-12 minutos**.

El responsable del presente proyecto de investigación es Universidad San Jorge con domicilio social en Autovía A-23 Zaragoza- Huesca, km. 299, 50830- Villanueva de Gállego (Zaragoza). La universidad cuenta con un Delegado de Protección de Datos que puede ser contactado a través de la dirección de correo electrónico privacidad@usj.es. Puede obtener más información sobre el responsable consultando nuestra política de privacidad a través del siguiente enlace: <https://www.usj.es/politica-de-privacidad>.

I) Percepciones generales sobre el CONCEPTO DE MEDIACIÓN

2. ¿Cuál es tu nivel de **acuerdo** o **desacuerdo** con las siguientes afirmaciones acerca del **concepto de mediación**?

0 = No estoy seguro/a; **1** = Totalmente en desacuerdo; **2** = En desacuerdo; **3** = De acuerdo; **4** = Totalmente de acuerdo

	0	1	2	3	4
a) La mediación consiste exclusivamente en la resolución de conflictos					
b) Existen distintos tipos de mediación					
c) En contextos médicos , la mediación se refiere solamente a la comunicación con pacientes					
d) La mediación puede consistir en ayudar a superar obstáculos de comunicación					
e) La mediación puede consistir en facilitar el acceso a contenidos o información					
f) La mediación interlingüística (entre distintas lenguas) es lo mismo que la traducción / interpretación					
g) La mediación puede consistir en facilitar la comprensión de conceptos o mensajes especializados					
h) Mediar puede implicar adaptar el lenguaje que se usa					
i) “Mediar” es sinónimo de “negociar”					
j) La mediación puede darse en la comunicación entre dos profesionales de la salud					
k) La mediación se da siempre en la comunicación oral					
l) El contexto de la comunicación no es importante para la mediación					
m) La mediación es necesaria solo para la comunicación entre personas de distintas culturas					
n) Mediar puede significar adaptar información a distintos formatos o medios de transmisión					
ñ) La mediación está estrechamente relacionada con la identificación de necesidades					
o) La mediación está relacionada con la accesibilidad					

¹² Cuestionario original creado con la aplicación *Microsoft Forms* y distribuido de forma virtual. Este Anexo presenta una versión adaptada para facilitar su visionado y lectura, respetando el formato de cada pregunta.

II) Percepciones sobre el valor de la mediación en el **ÁMBITO PROFESIONAL SANITARIO**

3. ¿Cuál es tu nivel de **acuerdo** o **desacuerdo** con las siguientes afirmaciones acerca del **valor de la mediación** para la **profesión sanitaria**?

0 = No estoy seguro/a; **1** = Totalmente en desacuerdo; **2** = En desacuerdo; **3** = De acuerdo; **4** = Totalmente de acuerdo

	0	1	2	3	4
a) Tener conocimientos y capacidades de mediación es fundamental para profesionales de la salud					
b) Tener conocimientos y capacidades de mediación es importante para la comunicación con pacientes					
c) Tener conocimientos y capacidades de mediación ayuda a identificar y entender necesidades de los/las pacientes					
d) Tener conocimientos y capacidades de mediación <u>solo</u> es importante de cara a la comunicación con pacientes					
e) Tener conocimientos y capacidades de mediación puede ayudar a desarrollar otras habilidades					
f) Tener conocimientos y capacidades de mediación es importante para la comunicación entre profesionales					
g) Tener conocimientos y capacidades de mediación solo es importante en el trato con pacientes de otras culturas					
h) Tener conocimientos y capacidades de mediación ayuda a empatizar con distintos tipos de diversidades funcionales o comunicativas					

III) Percepciones sobre el valor de la mediación dentro del **APRENDIZAJE DE LA LENGUA**

4. ¿Cuál es tu nivel de **acuerdo** o **desacuerdo** con las siguientes afirmaciones acerca del **valor de la mediación** para el **aprendizaje lingüístico general**?

0 = No estoy seguro/a; **1** = Totalmente en desacuerdo; **2** = En desacuerdo; **3** = De acuerdo; **4** = Totalmente de acuerdo

	0	1	2	3	4
a) Los contenidos de mediación me han ayudado a adquirir conciencia sobre el uso de mi propia lengua materna					
b) Los contenidos de mediación me han ayudado a adquirir conciencia sobre el uso del inglés					
c) Los contenidos de mediación me han ayudado a adquirir vocabulario en inglés					
d) Los contenidos de mediación me han ayudado a aplicar / practicar el vocabulario en inglés					
e) Los contenidos de mediación me han ayudado a adquirir estructuras gramaticales					
f) Los contenidos de mediación me han ayudado a aplicar / practicar estructuras gramaticales					
g) Los contenidos de mediación me han ayudado a adquirir y desarrollar capacidades comunicativas en la propia lengua					
h) Los contenidos de mediación me han ayudado a adquirir y desarrollar capacidades comunicativas en inglés					
i) Los contenidos de mediación me han ayudado a desarrollar habilidades orales en inglés					
j) Los contenidos de mediación me han ayudado a desarrollar habilidades escritas en inglés					

IV) ¿Cuál es tu nivel de acuerdo o desacuerdo con las siguientes afirmaciones acerca del valor de incluir la mediación en el aprendizaje de INGLÉS PARA CIENCIAS DE LA SALUD?

0 = No estoy seguro/a; **1** = Totalmente en desacuerdo; **2** = En desacuerdo; **3** = De acuerdo; **4** = Totalmente de acuerdo

	0	1	2	3	4
a) Incluir la mediación contribuye al aprendizaje de la lengua inglesa para el ámbito sanitario					
b) Incluir la mediación en las clases de inglés contribuye a la adquisición de otras habilidades relevantes para la profesión sanitaria					
c) Incluir la mediación ayuda a establecer conexiones entre el uso de la lengua y otros aspectos de la profesión sanitaria					
d) Incluir la mediación hace más realistas las actividades de interacción profesional-paciente					
e) Incluir la mediación resulta motivador para el aprendizaje del inglés					
f) Incluir la mediación hace más significativo el aprendizaje del vocabulario o la gramática					
g) Incluir la mediación no aporta nada al aprendizaje del inglés					
h) Incluir la mediación ayuda a desarrollar la conciencia sobre necesidades comunicativas					
i) Los cursos de inglés para la salud deberían incluir siempre contenidos de mediación					
j) Haber tratado la mediación me ayudará a la hora de retener ciertos contenidos lingüísticos					
k) Haber tratado la mediación me ayudará a la hora de aplicar ciertos contenidos lingüísticos					

V) Percepciones sobre el valor de la mediación dentro de LA ASIGNATURA de Inglés Científico

6. ¿Cuál es tu nivel de acuerdo o desacuerdo con las siguientes afirmaciones acerca del valor de incluir la mediación en la asignatura de Inglés Científico?

0 = No estoy seguro/a; **1** = Totalmente en desacuerdo; **2** = En desacuerdo; **3** = De acuerdo; **4** = Totalmente de acuerdo

	0	1	2	3	4
a) Incluir la mediación ha hecho que ciertas actividades fueran más relevantes para mí					
b) Incluir la mediación ha hecho que la asignatura fuera más relevante para mí					
c) Incluir la mediación ha hecho que ciertas actividades fueran más complicadas para mí					
d) Incluir la mediación ha hecho que la asignatura fuera más complicada para mí					
e) Incluir la mediación ha hecho que la asignatura fuera más interesante para mí					
f) Incluir la mediación ha hecho que la asignatura fuera más disfrutable para mí					

7. ¿Te ha gustado aprender sobre mediación/aplicar la mediación en la clase de Inglés Científico? (Marca tantos corazones como consideres, siendo 1 = “no me ha gustado nada” y 5 = “me ha gustado mucho”)



8. ¿Qué es lo que más valoras sobre la inclusión de la mediación en los contenidos de la asignatura de Inglés Científico?

9. ¿Hay alguna otra cosa que te gustaría comentar sobre la inclusión de la mediación en los contenidos de la asignatura de Inglés Científico?



RAEL: Revista Electrónica de Lingüística Aplicada

Vol./Núm.: 23/1
Enero-diciembre 2024
Páginas: 194-197
Artículo recibido: 02/08/2024
Artículo aceptado: 22/11/2024
Artículo publicado: 31/01/2025
Url: <https://rael.aesla.org.es/index.php/RAEL/article/view/653>
DOI: <https://doi.org/10.58859/rael.v23i1.653>

Barón, J., Celaya M. L. & Watkins, P. (2024). *Pragmatics in Language Teaching. From Research to Practice*. New York and London: Routledge.

**ISBN: 978-1-032-01821-8 (hbk), 978-1-032-01820-1 (pbk), 978-1-003-18021-0 (ebk)
(x + 124 pages)**

<https://doi.org/10.4324/9781003180210>

MARTA CARRETERO LAPEIRE
UNIVERSIDAD COMPLUTENSE DE MADRID

Pragmatics in Language Teaching. From Research to Practice (henceforth PiLT) belongs to a book series called *Research and Resources in Language Teaching*, aimed at integrating recent research in language teaching and learning with innovative classroom practice. All the books in the series follow a cyclical organisational principle, divided into four parts. Part I focuses on current research on the topic and outlines its implications for classroom practice. Part II proposes practical activities based on research results. Part III contains methodological suggestions for the implementation of the activities in syllabi and their integration in different teaching contexts, and Part IV sets forth proposals for classroom research. PiLT ends with a glossary, a section of references and an index that includes subjects and authors.

PiLT's Part I, "From research to implications" (pp. 1-20) is divided into two sections. Section 1, "An Introduction to Pragmatics", starts with a definition of this discipline as speaker's meaning in context, based on Yule (1996) and Crystal (1997), followed by a brief historical introduction focused on contrastive and cross-cultural pragmatics as well as interlanguage pragmatics. After some considerations on culture, society and context, (im)politeness is introduced, leaning mainly on Leech (2014) and Brown and Levinson's (1987) notions of positive face, negative face and face-threatening acts. The section ends with an informal analysis of various pragmatic phenomena in an excerpt of a 2009 comedy film. Section 2, "Pragmatics in second language learning and teaching", begins with a reflection on the difficulty involved in acquiring L2 pragmatics, evidenced by interlanguage pragmatics research. The next subsections provide an overview of L2 research on requests, a fruitful topic in this area. A useful table is also provided, with sample realisations, graded according to formality, of other speech act

Citar como: Carretero Lapeire, M. (2024). Book review: Barón, J., Celaya M.L., & Watkins, P. (2024). *Pragmatics in Language Teaching. From Research to Practice*. *RAEL: Revista Electrónica de Lingüística Aplicada*, 23, 194-197. <https://doi.org/10.58859/rael.v23i1.653>

types in which pragmatics plays a key role: agreeing / disagreeing, apologising, giving opinion, greetings, showing interest, suggestions and thanking. The section ends with an overview of approaches to teaching pragmatics, and a final summary.

Part II, “From implications to application”, occupies more than half of the book (pp. 20-88). It includes 37 activities aimed at teaching pragmatics, all accompanied by instructions about the aim, level, time, preparation, focus and procedure. The activities are organised into five sections. Section A is aimed at raising pragmatic awareness, including pragmatic elements of exchanges and moves as well as cultural differences. Section B concerns the exploitation of texts and materials with rich pragmatic input, including scenes of films as well as simple everyday texts such as car stickers, notes found on the fridge and voicemail. Section C is geared at learners’ elicitation of pragmatic input, focusing on moves such as topic changing, expressing interest or interrupting politely, and on the influence of register and relation to the addressee in elicitation. Section D provides students with opportunities to use pragmatics in social media, including hashtags, emojis, mobile texting and vaguebooking. Some activities in this section might well have included preparatory questions geared at bilingual and multilingual learners, in order to prompt them to reflect on their experience about crosslinguistic or cross-cultural differences in the use of these media. Section E proposes activities in which English is used as an international language, concentrating on awareness of intercultural and cross-cultural communication and of the risk of cultural miscommunication.

Part III, “From application to implementation” (pp. 89-105), aims to help teachers in the choice of pragmatic activities of the kinds displayed in Part II for their courses, considering the students’ needs and features of different teaching contexts. Some tests for assessing pragmatic awareness are proposed, such as discourse completion tests, role plays and pragmatic grids. The issue of naturalness plays a prominent role in two respects: first, a disadvantage of the kinds of tests mentioned above is pointed out, namely that students could say what they are expected to say instead of what they would naturally say; secondly, explicit teaching of pragmatic devices may lead to an artificial use in tests. In particular, the authors cite a personal communication reporting feedback from a Cambridge examiner, who advised to stop teaching stock answers to students (p. 90). However, this reviewer considers it worthwhile to raise pragmatic awareness through the explicit teaching of suitable expressions for different contexts, even at the expense of unnatural abuse of some of them. The high number of studies that evidence the preferability of explicit instruction over implicit instruction of pragmatics, some of which are cited in PiLT (see, for instance, pp. 17-19), lead to consider that such explicit teaching provides the scaffolding for future improvement of pragmatic adequacy in more advanced interlanguage stages.

Part IV, “From implementation to research” (106-115), encourages teachers to undertake action research on teaching pragmatics and provides suggestions for study designs, including discourse completion tests and role plays as possible bases for such research. A number of suitable topics for further investigation are proposed, such as intercultural contexts, resources and materials other than textbooks or the effects of pragmatic instruction on primary school students. The authors are clear in providing different avenues for classroom research, including mention or brief discussion of key references in the area. Although this part is succinct compared to the rest of the book, room should have been made for Rose and Kasper (2001), a pioneering edited volume solely devoted to classroom research on interlanguage pragmatics.

Throughout its pages, PiLT shows an implicit strong concern for accessibility and user-friendliness, which has a positive side and a negative side, the first clearly outweighing the second. On the positive side, PiLT will hopefully attract a high number of readers, which will contribute to an increased presence of pragmatics in ELT curricula; on the negative side, this concern for accessibility is probably the reason for the relatively limited theoretical contents on pragmatics, more of which could have been beneficial for teachers’ understanding how language – and pragmatics in particular – functions. To start with, the concept and scope of pragmatics are treated in a reductionist fashion. On the one hand, PiLT only covers part of the lin-

guistic phenomena falling within the area of pragmatics: the activities by and large concentrate on speech acts, indirectness and politeness. This restriction is justified, since the phenomena selected are those that most lend themselves to implementation in the EFL / ESL classroom. However, Part I could well have specified that pragmatics also encompasses other phenomena frequently included in introductions to pragmatics, such as the use of reference, deixis or pre-supposition in context (Levinson, 1983; Grundy, 2008; Birner, 2013), which pose few specific problems for ESL / EFL learners, thus being less suitable for implementation except for some academic learning contexts. On the other hand, PiLT highlights the pragmatic role of obvious linguistic devices for uttering messages attuned to interlocutors and context, but does not explicitly address the dimension of pragmatics as a pervasive feature of all naturally-occurring stretches of language, i.e. what Verschueren (1999) calls ‘pragmatic perspective’; however, this dimension is implicitly acknowledged through activities that address the influence of context on the choice of linguistic expressions in factual writing, such as Activities C2 and C3 (pp. 58-61).

The concern with accessibility is also the likely reason for the reduced glossary, which occupies one page and a half. Some well-known linguistic terms – recurrent in subjects and curricula on pragmatics in BA degrees all over the world – are absent; their inclusion in Part I and in the glossary would have enabled readers to understand pragmatics in greater depth. This is the case of Grice’s four maxims (Quality, Quantity, Relation, Manner), conversational implicature – replaced in PiLT with ‘implicit meaning’ – and flouting. On p. 22, the activities included in Section B are said to help students identify “pragmatic violations”, but flouting and violation are different kinds of failure to observe the maxims (Thomas, 1995: 64-74): floutings are blatant and aim at enriching communication, while violations are surreptitious and aim at misleading the addressee. Concerning politeness, PiLT provides no explicit definition of Brown and Levinson’s (1987: 68-71) superstrategies –bald on-record, redressive on-record, off-record–, which would have facilitated the distinction between degrees of indirectness in elicitations of FTAs. Similarly, no definition is found of Searle’s (1976) basic types of speech acts (representatives / assertions, expressives, directives, commissives, declarations, with the possible addition of questions as directives demanding information). This classification could have enriched the description of some activities: for example, in the description for Activities A.3.a. and A.3.b. (pp. 32-24), the teacher’s procedure could have specified that the speech acts in the left column of the tables are always indirect directives and those in the right column are indirect refusals to comply with these directives.

In other cases, the definitions are not precise. (Im)politeness is defined citing an early page of Leech (2014: 3) as “a form of communicative behaviour found generally in languages and among human cultures; indeed it has been claimed as a universal phenomenon of human society”. In the cited reference, the quoted stretch is aimed at providing a first approximation to (im)politeness but not as a definition. In fact, the cited words fail to specify what politeness actually is and would also be valid for other linguistic phenomena such as metaphor, irony or topic management, to name only a few. Still in the domain of politeness, the statement on p. 8 that “according to Brown and Levinson’s theory, FTAs should always be avoided, if successful communication is desired” may lead to confusion. FTAs are inevitably part and parcel of social interaction, as PiLT itself acknowledges by devoting substantial space and effort to tackle the thorny issue of teaching how to elicit FTAs with an adequate degree of smoothness. Another case in point is the term ‘culture’, whose entry in the glossary (p. 117) reads: “In the study of pragmatics, we refer to culture as the social norms and behaviours agreed by a particular community”. This definition, which corresponds to “little c culture”, is indeed of outmost importance in pragmatics, but it may confuse those readers who associate the word ‘culture’ with the “big C” culture, which pertains to the most visible manifestations of a culture such as major works of art, literature or music. For the sake of clarity, the difference between the two kinds of culture should have been explicit. As for requests, both the use of the term throughout the pages and the glossary definition as “[a] directive speech act that is used when somebody asks anoth-

er person to do something” (p. 118) show that, in PiLT, this term covers all kinds of directive speech acts. The word ‘directive’ would have been preferable as an umbrella term; in much of the literature, the term ‘request’ is associated with a concrete subtype of directives characterised as proposals for the addressee to do an action that benefits the speaker – thus differing from the subtype ‘advice’ – in contexts where the authority of the speaker over the addressee is weak or non-existent – thus differing from the subtype ‘command’ (Searle 1976: 3, 5-6, 12, 17).

Despite the aforementioned shortcomings, PiLT should be considered as a high-quality book, which provides teachers of English with first-class resources to get familiarised with pragmatics, implement the proposed activities and increment the body of classroom research on pragmatics. The adaptability for different learning contexts according to factors such as proficiency levels, age and interests, and the thoughtful inclusion of modern modes of communication in the materials besides traditional ones, lead to predict a widespread readership, thus contributing to raising consciousness that pragmatics should definitely play a major role in ELT curricula.

REFERENCES

- Birner, B. J. (2013). *Introduction to Pragmatics*. Oxford: Wiley-Blackwell. <https://doi.org/10.7202/1017210ar>
- Brown, P. & Levinson, S.C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511813085>
- Crystal, D. (Ed.) (1997). *The Cambridge Encyclopedia of Language* (2nd ed.). Cambridge: Cambridge University Press.
- Grundy, P. (2008). *Doing Pragmatics* (3rd ed.) London: Hodder Education.
- Leech, G. N. (2014). *The Pragmatics of Politeness*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195341386.001.0001>
- Levinson, S.C. (1983). *Pragmatics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511813313>
- Rose, K. R. & Kasper, G. (Eds.) (2001). *Pragmatics in Language Teaching*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524797>
- Searle, J. R. (1976). A classification of illocutionary acts. *Language in Society*, 5(1), 1-23. <https://doi.org/10.1017/S0047404500006837>
- Thomas, J. (1995). *Meaning in Interaction: An Introduction to Pragmatics*. London and New York: Longman.
- Verschueren, J. (1999). *Understanding Pragmatics*. London: Edward Arnold / New York: Oxford University Press. [https://doi.org/10.1016/S0378-2166\(98\)00097-6](https://doi.org/10.1016/S0378-2166(98)00097-6)
- Yule, G. (1996). *Pragmatics*. Oxford: Oxford University Press.