

Revista Española de Lingüística

Órgano de la Sociedad Española de Lingüística

RSEL

54/1

enero-junio-2024

REVISTA ESPAÑOLA DE LINGÜÍSTICA
(RSEL)

54/1

Edita
SEL

<https://doi.org/10.31810/RSEL.54.1>

REVISTA ESPAÑOLA DE LINGÜÍSTICA (RSEL)

ISSN: 0210-1874 • eISSN: 2254-8769

Depósito Legal: M-24.769-1971

DIRECTORA: Victoria Escandell-Vidal (Universidad Complutense de Madrid)

SECRETARIA: Marianela Fernández Trinidad (Universidad Complutense de Madrid)

RESPONSABLE DE RESEÑAS: Ventura Salazar (Universidad Jaén)

CONSEJO EDITORIAL

Pilar Barbosa (Universidade do Minho)

Martin Becker (Universität zu Köln)

Virginia Bertolotti (Universidad de la República, Uruguay)

José A. Camacho (University of Illinois Chicago)

Ángeles Carrasco Gutiérrez (Universidad Castilla-La Mancha)

Concepción Company Company (Universidad Nacional Autónoma de México)

Bert Cornillie (KU Leuven)

Sonia Cyrino (Universidade Estadual de Campinas)

Eladio Duque Gómez (Universidad Complutense de Madrid)

Ricardo Etxepare (Centre National de la Recherche Scientifique)

Antonio Fábregas (University of Trondheim)

Aitor García Moreno (Consejo Superior de Investigaciones Científicas)

Eveling Garzón Fontalvo (Universidad de Salamanca)

Chiara Gianollo (Università di Bologna)

Kees Hengeveld (Universiteit van Amsterdam)

M.^a Dolores Jiménez López (Universidad Rovira i Virgili)

Johannes Kabatek (Universität Zürich)

Brenda Laca (Universidad de la República, Uruguay)

Jingsheng Lu (Shanghai International Studies University)

Rafael Marín Gálvez (Centre National de la Recherche Scientifique)

Pedro Martín Butragueño (El Colegio de México)

José Ignacio Hualde (University of Illinois Urbana-Champaign)

Victoria Marrero Aguiar (Universidad Nacional de Educación a Distancia)

Emilia Ruiz Yamuza (Universidad de Sevilla)

Begoña Sanromán Vilas (Helsingin Yliopisto / University of Helsinki)

Michelle Sheehan (Newcastle University)

Rui Sousa-Silva (Universidade do Porto)

Esperanza Torrego Salcedo (Universidad Autónoma de Madrid)

Luis Unceta Gómez (Universidad Autónoma de Madrid)

Axelle Vatrican (Université de Toulon).

COMITÉ DE HONOR

Juan Antonio Álvarez Pedrosa (Universidad Complutense de Madrid)

José Antonio Berenguer (Consejo Superior de Investigaciones Científicas)

Alberto Bernabé (Universidad Complutense de Madrid)

Margarita Cantarero (Sociedad Española de Lingüística)

Ramón Cerdá (Universidad de Barcelona)

Salvador Gutiérrez Ordóñez (Universidad de León y Real Academia Española)

Emma Martinell (Universidad de Barcelona)

José Carlos de Torres (Sociedad Española de Lingüística)

Los textos enviados para su publicación han de presentarse a través de la web: <http://revista.sel.edu.es/index.php/revista/about/submissions>. Deben ser originales e inéditos y ajustarse a las normas que aparecen en la página web de la *Revista Española de Lingüística*. Todos los artículos son sometidos al dictamen de al menos dos evaluadores (designados por el Consejo Editorial) que deben emitir informes de carácter confidencial.

Los derechos de publicación y difusión, bajo cualquier forma, son propiedad de la *RSEL*. Todo texto publicado en la revista obliga a sus autores a no cederlo a terceros, sin autorización previa de la revista. Todos los números de la revista se pueden encontrar en abierto en la web <<http://revista.sel.edu.es/>>.

SERVICIOS DE INFORMACIÓN: Los contenidos de la *RSEL* son recogidos sistemáticamente en *Bibliographie Linguistique/Linguistic Bibliography*, *Cindoc-Base de datos Sumarios Isoc*, *Dialnet*, *Doaj*, *Dulcinea*, *Circ*, *Latindex 2.0*, *Erih Plus*, *Dice*, *Cabells*.

REDACCIÓN: Sociedad Española de Lingüística, Centro de Ciencias Humanas y Sociales del CSIC, c/ Albasanz, 26-28, 28037 Madrid.

CORREO ELECTRÓNICO Y CORRESPONDENCIA: RSEL.secretaria@gmail.com

DISEÑO Y COMPOSICIÓN: Carmen Chíncoa: produccionRSEL@gmail.com



FEICYT-S2023/2023
Fecha de otorgación: 20 de julio de 2023 (7ª convocatoria)
Vigencia: 28 de julio de 2024

ÍNDICE 54/1 (2024)

<i>De la variación tipológica a la variación intralingüística (y viceversa): el caso de los familectos</i>	9
<i>From Typological Diversity to Intralinguistic Variation (and Vice Versa): Familects as a Case Study</i>	
ANTONIO BENÍTEZ-BURRACO Y ELENA FELÍU-ARQUIOLA	
<i>Términos genéricos y función léxica Gener en la teoría Sentido-Texto</i>	29
<i>Generic Terms and Lexical Function Gener in Meaning-Text Theory</i>	
BEGOÑA SANROMÁN VILAS	
<i>Estructuras seudocoordinadas: perífrasis y significados aspectuales</i>	59
<i>Pseudo-coordinative Structures: Periphrasis and Aspectual Meanings</i>	
LAURA MALENA KORNFELD	
<i>Evaluación morfológica de los vocabularios de subpalabras utilizados por los grandes modelos de lenguaje</i>	103
<i>Morphological Evaluation of Subword Vocabularies Used by Large Language Models</i>	
ÓSCAR GARCÍA-SIERRA, ANA FERNÁNDEZ-PAMPILLÓN CESTEROS Y MIGUEL ORTEGA-MARTÍN	
NOTAS	
<i>El panorama sociolingüístico actual en Senegal: interacciones y regulaciones desde la perspectiva ecolingüística</i>	133
<i>The Current Sociolinguistic Landscape in Senegal: Interactions and Regulations from the Ecolinguistic Perspective</i>	
OU MAR SEYDI	
RESEÑAS	147

ARTÍCULOS

DE LA VARIACIÓN TIPOLÓGICA A LA VARIACIÓN INTRALINGÜÍSTICA (Y VICEVERSA): EL CASO DE LOS FAMILLECTOS


Antonio BENÍTEZ-BURRACO¹
Universidad de Sevilla


Elena FELÍU-ARQUIOLA²
Universidad de Jaén

Resumen

Los estudios de tipología sociolingüística sugieren que diversos factores sociopolíticos podrían explicar parte de los rasgos estructurales de las lenguas y, en particular, en qué dominios resultan más complejas o más simples. Así, las lenguas habladas por sociedades cerradas (denominadas esotéricas) presentan una fonología y una morfología más compleja y una menor transparencia semántica, mientras que las usadas por sociedades abiertas (o exotéricas) manifiestan una mayor complejidad sintáctica y un mayor grado de composicionalidad semántica. En este artículo se defiende la utilidad de considerar este tipo de trabajos para una mejor comprensión de la variación intralingüística, el objeto principal de interés de la sociolingüística (sobre todo, la variacionista). En concreto, se propone extender el ámbito de aplicación de los conceptos esotérico y exotérico al estudio de la variación social del español. Para ilustrar un programa de análisis de este tipo se discute el caso de los familectos, esto es, las variedades de la lengua asociadas a las unidades familiares. Se propone un protocolo de análisis de estas variedades, que incluye una lista de rasgos estructurales de interés definidos a partir de la hipótesis de que los familectos pueden describirse e interpretarse satisfactoriamente en términos de usos esotéricos de la lengua estándar, con características distintivas que los diferencian hasta cierto punto de otras variedades del vernáculo, en particular, del español oral informal.

Palabras clave: Tipología sociolingüística; lenguas exotéricas; lenguas esotéricas; familectos; complejidad

1. abenitez@us.es;  <https://orcid.org/0000-0003-4574-5666>

2. efeliu@ujaen.es;  <https://orcid.org/0000-0002-3711-041X>

FROM TYPOLOGICAL DIVERSITY TO INTRALINGUISTIC VARIATION (AND VICE VERSA): FAMILLECTS AS A CASE STUDY

Abstract

Sociolinguistic typology studies have uncovered diverse social, cultural, and political factors that might account for aspects of the structural complexity of the world's languages. Accordingly, the languages spoken by close-knit (or *esoteric*) societies tend to feature more complex phonologies and morphologies, as well as increased semantic opacity. By contrast, the languages spoken by open (or *exoteric*) societies usually exhibit more syntactic complexity together with greater semantic compositionality. In this paper, we support the view that classical (i.e. variationist) sociolinguistics would benefit from applying this typological approach to the study of intralinguistic variation. Specifically, we argue for relying on the concepts of *esotericity* and *exotericity* for achieving a better characterization and understanding of Spanish familects (i.e. the private language varieties used within families). We advance a specific protocol for studying this linguistic variety, as well as a tentative list of features characterizing Spanish familects, which we have posited under the view that familects are esoteric varieties of the standard language, although with some distinctive features compared to the oral vernacular, particularly to informal speech.

Keywords: Sociolinguistic typology; exoteric languages; esoteric languages; familects; complexity

RECIBIDO: 20/05/2023

APROBADO: 12/03/2024

1. INTRODUCCIÓN

Los estudios de tipología sociolingüística han encontrado distintas correlaciones entre los parámetros que caracterizan la diversidad estructural de las lenguas y los factores que determinan la organización sociopolítica de los grupos humanos. Así, por ejemplo, el índice de aglutinación (que cuantifica el número de morfemas por palabra) se correlaciona negativamente con el número de hablantes de la lengua (Lupyan y Dale, 2010). Del mismo modo, una mayor complejidad sociopolítica parece correlacionarse positivamente con la obligatoriedad del marcaje del TAM (tiempo-modo-aspecto) (Gil, 2021). Más complicado resulta, sin duda, encontrar una posible causalidad que explique esta covariación entre diversidad lingüística y social. Una hipótesis al respecto es que los hablantes comunican, en general, la misma cantidad de información, pero dependiendo del contexto, y especialmente del grado de conocimiento mutuo, lo hacen de un modo más implícito o más

explícito, lo que afecta a las características estructurales del discurso que utilizan. Para hacer referencia a estos dos tipos de interacción se han acuñado los términos *comunicación esotérica* y *comunicación exotérica*, respectivamente (Wray y Grace, 2007), lo que ha llevado a hablar también de *lenguas esotéricas* y *lenguas exotéricas* (Thurston, 1987; 1994), o más recientemente, *lenguas S* y *lenguas X* (Chen *et al.*, 2024).

A nivel tipológico, una forma de testar esta hipótesis es buscar correlaciones entre parámetros de las lenguas sujetos a variación (con valores susceptibles de ser clasificados como más o menos complejos) y factores sociopolíticos (que permitan medir la complejidad organizativa de las sociedades). En la actualidad, se cuenta con bases de datos bastante completas tanto de diversidad lingüística (WALS, <https://wals.info/>, Dryer y Haspelmath, 2013; Grambank, <https://grambank.cldd.org/>, Skirgård *et al.*, 2023), como de diversidad sociopolítica (D-Place, <https://d-place.org/>, Kirby *et al.*, 2016; Ethnologue, <https://www.ethnologue.com>, Eberhard *et al.*, 2022; Glottolog, <https://glottolog.org/>, Hammarström *et al.*, 2022). Y se han llevado a cabo, de hecho, algunos intentos por realizar estudios de este tipo, que parecen corroborar la hipótesis de que las lenguas habladas por sociedades cerradas tienden a presentar una morfología más compleja, mientras que las habladas por sociedades abiertas tienden a poseer una sintaxis más elaborada (Chen *et al.*, 2023; 2024). No obstante, cabe esperar que, por las características sociodemográficas de los grupos que las hablan (población reducida, escasos contactos con otros grupos) y por encontrarse muchas de ellas, de hecho, en vías de extinción, las lenguas esotéricas se encuentren infrarrepresentadas en estas bases de datos y, sobre todo, insuficientemente descritas. Ahora bien, dado que los conceptos *esotericidad/exotericidad* son aplicables, como se indicó anteriormente, a cualquier forma de interacción lingüística, una manera de paliar este problema sería estudiar la comunicación esotérica entre hablantes de lenguas que, tipológicamente, tengan un carácter más exotérico.

Teniendo en cuenta lo anterior, en este trabajo (de carácter fundamentalmente programático) se defenderá extender el ámbito de aplicación de los conceptos *esotérico* y *exotérico* al análisis de la variación sociolingüística dentro de una misma lengua, en particular, el español. El objetivo será doble. Por un lado, se pretende aplicar lo que la tipología sociolingüística (y otras disciplinas, como la lingüística histórica o la evolución del lenguaje) ha averiguado sobre las relaciones entre diversidad estructural y sociopolítica (y, en particular, sobre la naturaleza de las lenguas exotéricas/esotéricas) al estudio y la interpretación de la variación intralingüística. Con ello, se espera llegar a entender mejor las causas de las diferencias

estructurales que las diversas variedades del español de interés sociolingüístico presentan entre sí y con el estándar: si bien estas variedades se han descrito de manera muy detallada (véase Silva-Corvalán, 2001; Díaz-Campos, 2011; Díaz-Campos *et al.*, 2020, entre otros muchos), su motivación funcional sigue sin estar del todo clara, tal como apuntamos anteriormente. Por otro lado, se busca aprovechar la caracterización que se ha hecho hasta la fecha de dicha variación para enriquecer también nuestra comprensión de las relaciones entre diversidad sociopolítica y diversidad lingüística, que, como apuntábamos anteriormente, se halla lastrada por el difícil acceso a las lenguas de tipo esotérico.

Para ilustrar la naturaleza de un programa de investigación de este tipo se discutirá el caso de los *familectos* en español. De un modo sucinto, los familectos pueden caracterizarse como las variedades de la lengua asociadas a las unidades familiares. Tales unidades presentan muchas de las características de las sociedades esotéricas. En el artículo se argumentará que las peculiaridades estructurales (aunque también pragmáticas y discursivas) de estas variedades intralingüísticas no solo pueden describirse e interpretarse satisfactoriamente en términos de usos esotéricos de la lengua estándar, sino que servirían como un laboratorio ideal para testar con detalle algunas de las predicciones de un modelo general de relación entre lengua y sociedad basado en el concepto de eso/exotericidad. El artículo se estructura de la siguiente manera. En la sección 2 caracterizaremos de forma sucinta las propiedades distintivas de las lenguas esotéricas y exotéricas. En la sección 3 haremos lo propio con los familectos. En la sección 4, que constituye el núcleo del trabajo y nuestra genuina aportación a esta cuestión, propondremos un protocolo de estudio de los familectos del español desde esta nueva perspectiva. Ello nos llevará a sugerir una relación de rasgos estructurales que esperamos encontrar en los familectos (los cuales hemos derivado de los que caracterizan las lenguas esotéricas), aunque también en las formas de comunicación esotérica en general. Incluiremos también algunas sugerencias, de carácter metodológico, destinadas a optimizar dicho protocolo de análisis. El trabajo finaliza con algunas conclusiones y sugerencias para futuras investigaciones en este campo.

2. COMUNICACIÓN ESOTÉRICA Y COMUNICACIÓN EXOTÉRICA

Siguiendo a Bolender (2007) y a Wray y Grace (2007), las lenguas esotéricas (o, de forma más neutra, las variedades lingüísticas asociadas a la comunicación esotérica) son las que emplean comunidades humanas que presentan las siguientes características:

- reducido tamaño;
- escaso contacto con hablantes de otras lenguas o variedades diferentes;
- redes sociales densas (esto es, contactos recurrentes entre los mismos individuos en diferentes dominios de interacción);
- escasa proporción de hablantes que han aprendido la lengua o la variedad de adultos.

Entre las características estructurales distintivas de las lenguas esotéricas se encontrarían las siguientes³:

- fonología compleja (inventarios fonológicos extensos, que incluyen a menudo sonidos poco frecuentes; combinaciones de sonidos complejas al formar las sílabas);
- morfología compleja (mayor grado de irregularidad; fenómenos morfofonológicos complejos; fenómenos supletivos);
- sintaxis más simple (menor frecuencia de oraciones compuestas; mayor grado de yuxtaposición o parataxis);
- menor composicionalidad semántica (incluyendo una frecuencia más elevada de modismos y locuciones idiomáticas, con menor transparencia semántica).

Se ha sugerido que estas características se explicarían por el mayor grado de conocimiento compartido entre los usuarios de este tipo de lenguas o variedades (al tratarse de individuos que se conocen bien y que comparten una misma cultura, y que interactúan entre sí de forma frecuente), por lo que buena parte de la comunicación se produciría de forma implícita, sin necesidad de recurrir a construcciones muy elaboradas y transparentes semánticamente (véase Wray y Grace, 2007 para más

3. Ciertamente, existe una viva discusión acerca de en qué sentido las lenguas pueden considerarse objetos complejos, de si difieren unas de otras en términos de su complejidad absoluta (de modo que pueda haber lenguas más simples que otras) o si dichas diferencias son relativas (de manera que unas lenguas puedan ser más complejas en ciertos niveles y más simples en otros) y, finalmente, del modo más apropiado de cuantificar dicha complejidad (sobre estas cuestiones, véase Szmeccsanvi, 2015; Miestamo, 2017; Joseph, 2021; Ehret et al., 2023, entre otros muchos). En las obras seminales de las que hemos extraído la caracterización básica de las sociedades y las lenguas esotéricas o exotéricas que presentamos en este apartado, el concepto *complejidad* no se trata de un modo tan técnico. En todo caso, y la línea de nuestro trabajo experimental sobre estas cuestiones (Chen et al., 2023; 2024), en este artículo asumimos que un determinado aspecto gramatical puede considerarse más complejo en una lengua que en otra si en la primera es necesario emplear más símbolos para su codificación, siguiendo a Li y Vitanyi (2008). Así, por ejemplo, hace falta más texto para describir las reglas morfológicas de una lengua aglutinante que las de una lengua aislante, por lo que consideramos más complejo el primer tipo de morfología que el segundo.

detalles). Otra razón, compatible con la anterior, es que estas lenguas son aprendidas, en general, en condiciones nativas y no como segunda lengua, por lo que sus características formales podrían ayudar de algún modo a su aprendizaje por parte de los niños (por ejemplo, una morfología más compleja puede conllevar un mayor grado de redundancia, lo cual puede ser beneficioso cuando se posee una memoria de trabajo reducida y menores capacidades inferenciales; véase Lupyán y Dale, 2010 o Benítez-Burraco y Kempe, 2018 para una discusión más detallada). Ejemplos de lenguas esotéricas serían muchas de las habladas por cazadores-recolectores actuales o buena parte de las que integran el grupo indopacífico, habladas en las zonas altas de Nueva Guinea, si bien un caso paradigmático podría ser el de la lengua pirahã, conocida por su extremada simplicidad sintáctica, incluyendo la ausencia de subordinación (Everett, 2005).

Por su parte, las lenguas exotéricas (o, de forma más neutra, las variedades lingüísticas asociadas a la comunicación exotérica) son las que emplean comunidades humanas que presentan las siguientes características:

- elevado número de hablantes;
- contactos significativos con otros grupos humanos culturalmente diferentes;
- organización preferente en forma de redes sociales laxas y/o a larga distancia;
- porcentaje significativo de hablantes que han aprendido la lengua o la variedad de adultos.

Las características estructurales que presentarían estas lenguas serían sustancialmente opuestas a las anteriores (aunque, como cabe imaginar, es mucho más exacto entender la oposición *esotérico/exotérico* como un gradiente y no como una dicotomía):

- fonología más simple;
- morfología más simple;
- vocabularios extensos;
- mayor transparencia semántica (lo que implica una mayor composicionalidad del significado y un menor recurso a expresiones idiomáticas);
- sintaxis más compleja (mayor grado de gramaticalización de información relevante para la comprensión de las oraciones, como las funciones sintácticas, el TAM, etc.; mayor uso de la recursividad).

En línea con el razonamiento anterior, estas características se explicarían por la menor dependencia contextual de las interacciones habladas (motivada, a su vez, por un menor grado de conocimiento mutuo entre los interlocutores) y/o el hecho de que estas lenguas son aprendidas en una alta proporción por hablantes adultos, que tienden a simplificar durante el aprendizaje justo aquellas características que están ausentes en ellas, como la flexión nominal o verbal (de nuevo, véase Lupyan y Dale, 2010, y Benítez-Burraco y Kempe, 2018, para una discusión más detallada). Ejemplos de lenguas exotéricas serían la mayor parte de las lenguas que cuentan con más hablantes en la actualidad, especialmente las que se han venido empleando como lenguas francas o lenguas oficiales de estados extensos y complejos, como el mandarín, el inglés o el propio español.

3. FAMILLECTOS

El término *familecto* (*familylect* en el original) fue introducido por Søndegaard (1991) para describir el sociolecto de una familia multilingüe. El familecto no sería, por tanto, otra cosa que la variedad lingüística compartida por un grupo de hablantes que son miembros de la misma familia. Así pues, al igual que un dialecto o un sociolecto, es esperable que un familecto presente características estructurales distintivas, así como usos distintivos de rasgos lingüísticos compartidos con el estándar o con otras variedades intralingüísticas. Como señala Hazen (2004), existe una interesante discusión en el campo de la sociolingüística acerca de si la familia constituye una unidad apropiada de estudio de la diversidad lingüística y, en particular, si hay que considerarla una comunidad de habla, un tipo de red social o una comunidad de prácticas (véase Hazen, 2004). Es cierto, no obstante, que, como señala también este autor, la mayor parte de los estudios lingüísticos sobre las interacciones en el seno familiar se han realizado desde la perspectiva del análisis del discurso (véase Varenne 1987; Beaumont 1995; Burrell 1995, entre otros). No es menos cierto tampoco que, en los estudios variacionistas, el interés por lo familiar se ha dirigido fundamentalmente a su posible papel causal en la variación observada en la comunidad a otros niveles (véase, igualmente, Hazen 2004).

Clancy (2016), quien emplea el término «discurso íntimo» para hacer referencia a la conversación entre parejas, miembros de una familia y/o amigos próximos, señala como uno de los rasgos distintivos de este tipo de discurso la idiomática, que correlaciona con el grado de intimidad y de complicidad entre los participantes, hasta el punto de acuñar el término *personal idiom* para caracterizar esta estrategia lingüística. Por su parte, Osovska (2019) señala como rasgo típico de los familectos

la existencia de un vocabulario propio, formado por apelativos particulares, palabras jergales o argóticas, coloquialismos y dialectalismos. En muchos casos, estos términos son de nueva creación, aunque se suelen generar mediante procedimientos ya establecidos en la lengua (diminutivos, sinónimos, antónimos, hipocorísticos). Existen además indicios abundantes de juego lingüístico. Ambos autores coinciden en la idea de que los rasgos que caracterizan a los familectos tienen como función principal la identitaria (aunque en ocasiones algunos de sus componentes puedan ser conocidos y usados por personas ajenas a la pareja o la familia). Por otro lado, los rasgos asociados al juego lingüístico podrían explicarse por la habitual presencia de niños en las unidades familiares, pero también por la infantilización inherente en muchos casos a las relaciones afectivas de pareja.

La caracterización más detallada de los familectos realizada hasta la fecha es la de Gordon (2009) para el inglés. En la línea de lo señalado por los autores anteriores, Gordon hace hincapié en la idea de que cada familia posee su propio repertorio léxico, cuyas formaciones se basan, con frecuencia, en fenómenos relacionados con el juego lingüístico y la imitación del habla infantil, como es el caso de la reduplicación (*wa-wa* para *water* ‘agua’, *ba-ba* para *bottle* ‘botella’, *go night-night* para *go to bed* ‘irse a la cama’). Otro rasgo distintivo de los familectos sería el carácter ritualizado del lenguaje que se emplea en determinadas actividades cotidianas propias del ámbito familiar, lo que refuerza la percepción de la familia como un grupo con una cultura propia y distintiva. Cabe destacar que Gordon se centra fundamentalmente en las características discursivas de la interacción entre los miembros del núcleo familiar y analiza aspectos tales como la intertextualidad, que se consigue mediante la repetición de fragmentos de discurso de otros miembros de la familia, combinada con fenómenos como la «ventrilocuización» (hablar como lo hace otro miembro de la familia, generalmente un niño). Se trata de estrategias que tienen como función que la familia se construya como un grupo social también desde el punto de vista discursivo y lingüístico. En suma, pese a la existencia de estudios que abordan los familectos desde una perspectiva discursiva, sigue pendiente una caracterización detallada de los rasgos distintivos de esta variedad lingüística desde un punto de vista estructural.

En línea con lo apuntado por estos autores, son interesantes (y significativos) los paralelismos formales y funcionales que existen entre los familectos y otras variedades no estándares de la lengua, en particular, los lenguajes secretos y el habla infantil (tanto la producida por los niños como la dirigida a ellos). Por *lenguaje secreto* u *oculto* se entiende cualquier sistema lingüístico que sólo resulta inteligible para un determinado (y restringido) grupo de individuos. Entre sus características distintivas

se encuentra, precisamente, un léxico idiosincrásico, que puede caracterizarse en su conjunto como un argot (o jerga con prestigio encubierto), compuesto generalmente por términos inventados o bien por palabras de la lengua estándar que se usan con un significado idiosincrásico o cuya forma se altera mediante diferentes procedimientos propios del juego lingüístico, como sustituciones, inserciones o metátesis (ejemplos bien conocidos son el lunfardo, el *Back Slang* o el *Pig Latin*) (para más detalles, véase Halliday, 1976, o Storch, 2017). Aunque en buena medida la función de los lenguajes secretos es transmitir información de forma segura, comparten con los familectos una importante función identitaria.

Por su parte, el *habla infantil* puede definirse como la variedad vernácula simplificada que usan los niños cuando están adquiriendo su lengua materna. Conlleva todo tipo de adaptaciones (en general, simplificaciones) fonológicas, morfológicas, semánticas y sintácticas (véase Fletcher y MacWhinney, 1995; Lust, 2006; Roberts, 2013, entre otros muchos). Entre las primeras cabe destacar la simplificación de la estructura silábica (reducción a sílabas abiertas con ataques simples CV, reducción del número de sílabas siguiendo un principio de focalización prosódica), las reduplicaciones o diferentes procesos de adaptación fonética (asimilaciones, sustituciones, armonía). Entre las de tipo morfológico destacan el uso de formas no flexionadas, la sobrerregularización o la generación de compuestos idiosincrásicos. En el plano semántico son especialmente llamativas la sobregeneralización y la superrestricción. Y en el ámbito sintáctico destacan la omisión de constituyentes o de palabras funcionales, la preferencia por órdenes de constituyentes no marcados o la ausencia de subordinación. Ciertamente, estas características vienen dadas, en general, por el menor desarrollo cognitivo del niño, que adapta la variedad adulta a sus limitadas habilidades perceptivas y de procesamiento. Sin embargo, el estilo discursivo de quienes pasan mucho tiempo interactuando con niños pequeños (denominado «habla del cuidador», «maternés» o «discurso dirigido al niño») presenta notables paralelismos con el discurso infantil, tal como se ha caracterizado anteriormente (véase Foulkes *et al.*, 2005; Lust, 2006; Roberts, 2013; Saint-Georges *et al.*, 2013 entre otros), de modo que cabría hablar de una suerte de variedad lingüística compartida. Se ha sugerido que las características de esta variedad empleada por los adultos en sus interacciones con los niños tendrían un valor funcional, en particular, la facilitación de la adquisición del lenguaje (Kempe *et al.*, 2007; Schick *et al.*, 2022). Al igual que ocurría con los lenguajes secretos, son notables los paralelismos que cabe advertir en el plano formal entre los familectos y el habla infantil y/o el discurso dirigido al niño.

Finalmente, muchas de las características de los familectos pueden relacionarse con algunas de las propiedades distintivas de las lenguas esotéricas, tal como se describieron en el apartado anterior, en particular, una fonología y una morfología especialmente ricas, así como el uso frecuente de modismos y locuciones idiomáticas. Como se apuntó anteriormente, las peculiaridades estructurales de los familectos parecen venir dadas, al menos en parte, por su función identificadora (destinada fundamentalmente a reforzar los vínculos familiares, más que a marcar distanciamiento u ocultar información privada, puesto que algunos de sus componentes suelen usarse en presencia de extraños) y en parte también por su función lúdica (cuyo fin sería semejante, aunque reforzado seguramente por la presencia de niños y/o la infantilización inherente a las relaciones de pareja). No obstante, dichas características también podrían explicarse en buena medida como una consecuencia de su mayor dependencia contextual, en el sentido de que los usuarios de los familectos comparten una gran cantidad de conocimiento enciclopédico (en forma de vivencias comunes, intereses compartidos, etc.) y de que las interacciones suelen versar sobre el aquí y el ahora, por lo que la comunicación depende en mayor medida de lo habitual de implicaturas conversacionales. Esta circunstancia aproxima a los familectos a las variedades lingüísticas asociadas a la comunicación esotérica y, en particular, a las lenguas esotéricas (es evidente, por lo demás, que dadas las características de las sociedades en las que surgen, las funciones identificadora y lúdica también podrían explicar los rasgos distintivos de estas lenguas y ello por las mismas razones aducidas en el caso de los familectos).

Como señalamos al comienzo de esta sección, y a pesar de su indudable interés, se han estudiado poco las características estructurales o gramaticales de los familectos. No conocemos además trabajos que versen específicamente sobre los familectos del español. Es obvio, por tanto, que un estudio descriptivo de los familectos de esta lengua, tal como se viene haciendo con otras variedades, sería particularmente deseable. Nuestro protocolo, tal como lo describimos en detalle en la siguiente sección, contempla una parte puramente metodológica, que aspira a ser de alguna utilidad en ese sentido. No obstante, la parte que juzgamos más novedosa e interesante es la relación de rasgos estructurales que esperamos encontrar en los familectos, pensados, en particular, para testar la hipótesis de que esta variedad lingüística representa una forma de comunicación esotérica. Por esta razón, señalamos también cuáles de ellos esperamos encontrar con mayor probabilidad. Para ello, nos hemos basado en nuestros trabajos previos sobre tipología sociolingüística (Chem *et al.*, 2023).

4. UNA PROPUESTA DE ANÁLISIS DE LOS RASGOS ESTRUCTURALES DE LOS FAMILLECTOS EN ESPAÑOL

Tal como indicamos al final del apartado anterior, nuestra propuesta incluye un protocolo general para el trabajo de campo destinado a recolectar datos de interés sociolingüístico en relación con los familectos (sección 4.1.) y una relación de rasgos estructurales que sería deseable buscar en dichos datos con el objetivo de determinar el posible carácter esotérico de esta variedad (sección 4.2).

4.1. Metodología de análisis

El protocolo de análisis contempla tres etapas:

- recogida de datos;
- transcripción de datos;
- análisis de los datos.

A. Recogida de datos

Técnicamente, la recogida de datos puede hacerse mediante cualquier dispositivo que permita grabar audio durante períodos prolongados, siempre que se coloque de modo que se recojan con nitidez las intervenciones de los diferentes participantes. Por esta razón, es preferible grabar conversaciones en las que los intervinientes no estén en movimiento (idealmente, que estén sentados a una mesa). Para maximizar la posibilidad de encontrar rasgos de interés se recomienda:

- grabar durante períodos prolongados (idealmente, conversaciones completas);
- grabar en diferentes momentos del día (porque, previsiblemente, se hablará de temas diferentes);
- aprovechar celebraciones familiares especiales (que suelen durar más y favorecer conversaciones sobre temas diversos);
- grabar conversaciones entre individuos que compartan un parentesco cercano y/o convivan desde hace tiempo (por ejemplo, es mejor evitar celebraciones multitudinarias, porque suelen participar personas con las que la relación es más indirecta);
- grabar preferiblemente conversaciones entre las mismas personas;
- no grabar conversaciones en las que participen niños menores de 8-10 años (para evitar interferencias con el lenguaje infantil y el lenguaje del cuidador).

En paralelo, sería oportuno solicitar a los intervinientes en las conversaciones la siguiente información:

- datos de interés sociolingüístico, cuando menos:
 - edad;
 - nivel de estudios;
 - ocupación actual;
 - lugar de nacimiento;
 - lugares de residencia en el pasado.
- una relación de términos que consideren como idiosincrásicos de su unidad familiar; habitualmente estos términos son más frecuentes en los siguientes campos:
 - apelativos;
 - objetos de uso cotidiano (cocina, baño, coche, etc.);
 - comida y preparación de los alimentos;
 - actividades compartidas (viajes, ocio, recuerdos familiares).

B. *Transcripción de los datos*

La transcripción de datos se hará en dos pasos:

- transcripción automática mediante un programa de reconocimiento de voz (como, por ejemplo, los que llevan incorporados los ordenadores portátiles);
- revisión de la transcripción automática (es necesario corregir los previsibles errores cometidos por el programa; además, se recomienda añadir algún comentario sobre el contexto en que transcurre la conversación, en particular, sobre aspectos paralingüísticos, como por ejemplo los tonos de voz, de interés para entender el sentido de las interacciones verbales cuando dicho sentido pueda ser opaco).

C. *Análisis de los datos*

Una vez transcritos, los datos se analizarán, teniendo especialmente presentes los rasgos de interés que se detallan en la siguiente sección.

4.2. *Rasgos familiares asociados a la comunicación esotérica*

En esencia, la propuesta que hacemos de los rasgos que caracterizan la comunicación esotérica en español (y que esperamos encontrar, por tanto, en los familiares) nace de los trabajos previos en tipología sociolingüística, tal como se

discutieron en la sección 2. Su mayor o menor probabilidad de ocurrencia viene dada por un trabajo previo de reanálisis en términos de menor o mayor complejidad de los rasgos estructurales de las lenguas tal como aparecen recogidos en WALS (véase Chen *et al.*, 2023; 2024 para más detalles).

A continuación, se enumeran los rasgos propios de la comunicación esotérica en español que esperamos encontrar en los familectos. Aparecen presentados por niveles lingüísticos e incluyen algunos ejemplos:

1. Nivel fónico

- c) modificaciones secundarias de sonidos habituales, como la palatalización (por ejemplo, *chí* por *sí*);
- d) alteraciones prosódicas, como cambios en el patrón acentual (por ejemplo, acortamientos que produzcan formas bisílabas de acentuación llana: *Miguel* > *Migue*).
- e) alteraciones de la fonotaxis habitual, en particular, el incremento de la complejidad silábica (como en *sip* por *sí*) o del número de sílabas mediante ampliación silábica o reduplicación (por ejemplo, *vayavaya* por *vaya* con valor enfático);
- f) juegos fónicos en general: aliteraciones, rimas, paronomasias, etc. (por ejemplo, en expresiones como *hola*, *caracola* o *de nada*, *monada*).

2. Morfología

- a) procesos derivativos:
 - empleo frecuente de sufijos apreciativos y de prefijos intensivos (por ejemplo, *fotaza* o *me requetechifla*);
 - empleo de sufijos propios del lenguaje juvenil (por ejemplo, *careto*), propios de ámbitos especializados (por ejemplo, *mamitis*) o de creación propia;
 - empleo de sufijos propios de la lengua general con bases inesperadas (por ejemplo, *peinaje* para designar la acción de peinar o *dormición* para la acción de dormir)
 - compuestos de nueva creación (por ejemplo, *posamóvil*, *cagarrisas*);
 - empleo de formas irregulares (por ejemplo, extensión del sufijo culto *-érrimo* a casos donde no es esperable, como en *malérrimo*);
- b) usos no habituales del género y el número gramaticales, como el empleo del masculino despectivo o desmerecedor (por ejemplo, *coso* o *palabro*).

3. Sintaxis

- a) orden de constituyentes no canónico, distinto de SVO, por razones que no sean las puramente informativas (por ejemplo, *Viene María tarde*);
- b) formación de interrogativas *in situ* (por ejemplo, *¿Has comido qué?*);
- c) construcciones con función antipasiva (por ejemplo, *Juan come*; por ejemplo, *Juan aprovecha tu propuesta* > *Juan se aprovecha de tu propuesta*);
- d) cambios de diátesis sin marca (por ejemplo, *Los garbanzos congelan bien*);
- e) predominio de las oraciones simples sobre las complejas (por ejemplo, *Mi amigo vino ayer. Se llama Juan*, frente a *Mi amigo, que se llama Juan, vino ayer*);
- f) preferencia por la yuxtaposición (por ejemplo, *Voy a abrir la ventana, hace calor* frente a *Voy a abrir la ventana porque hace calor*).

4. Léxico

- a) combinaciones léxicas poco frecuentes (por ejemplo, alteración de asociaciones léxicas habituales, como *rebanada de queso* o *loncha de pan* en lugar de *rebanada de pan* o *loncha de queso*);
- b) uso de modismos, especialmente frases hechas de creación propia y significado solo accesible para los miembros de la unidad familiar (por ejemplo, *te quiero de los montones* en lugar de *te quiero mucho* o *te quiero un montón*);
- c) palabras de uso general que se emplean con un significado idiosincrásico (por ejemplo, *el huevo* para referirse a un coche);
- d) coloquialismos (por ejemplo, *currar* por *trabajar*, *tío/tía* por *persona*);
- e) dialectalismos (por ejemplo, *chaleco* por *jersey* o *gañafote* por *saltamontes*, dos onubesismos);
- f) apodos e hipocorísticos (por ejemplo, *el tete* para el hermano, *yayo* o *yaya* para los abuelos).

Aunque la nuestra es una propuesta eminentemente programática y metodológica, las conclusiones de nuestros trabajos previos de carácter tipológico nos permiten predecir que determinados rasgos estructurales deberían ser especialmente frecuentes en los familectos: se trataría de aquellos para los que dichos estudios han encontrado correlaciones más significativas con los factores sociopolíticos que definen las

sociedades esotéricas. Según Chen y colaboradores (2023; 2024), quienes se han centrado en el ámbito de la morfología y la sintaxis, las lenguas habladas por sociedades esotéricas tienden a presentar un mayor índice de síntesis, incluyendo más afijación, un mayor número de casos y más formas sintéticas del verbo, así como una mayor diversidad de construcciones posesivas. Del mismo modo, suelen presentar un orden de constituyentes menos rígido, admiten la omisión de pronombres, y suelen presentar sintagmas determinantes con una estructura más simple (lo que incluye la ausencia de artículos definidos o un menor número de categorías nominales que marquen la pluralidad), así como un menor uso de construcciones acusativas, pasivas y de relativo. En consecuencia, en el caso de los familectos del español y teniendo en cuenta la naturaleza de los rasgos recogidos en este apartado, nuestra predicción es que los familectos estarán asociados preferentemente, al menos en el plano morfológico y sintáctico, a algunos de estos rasgos. Entre ellos se encontraría un empleo abundante de la morfología derivativa (especialmente de la morfología apreciativa), que, si bien no puede considerarse un correlato directo de un índice de síntesis más elevado (puesto que este parámetro mide la complejidad de la morfología flexiva), podría considerarse, no obstante, un indicio de una mayor tendencia a la afijación. En cuanto a las características sintácticas, cabe esperar, por ejemplo, un mayor recurso a las interrogativas *in situ*, que son estructuralmente más simples que las canónicas. El menor uso de construcciones acusativas y pasivas podría corresponderse con características como la mayor frecuencia de construcciones antipasivas y de cambios de diátesis sin marca, y, sobre todo, con una menor frecuencia de pasivas con *se*. Finalmente, también es esperable un menor empleo de oraciones de relativo y de subordinación en general, tal como se ha señalado previamente, lo que se traduciría en un mayor recurso a las estructuras paratácticas.

5. CONCLUSIONES

Más allá de las bondades que nuestra propuesta pueda tener desde el punto de vista metodológico, desde una perspectiva más teórica, queremos ver en nuestro trabajo una vindicación de la importancia, para el estudio de la variación intralingüística motivada por factores sociales, de las variedades lingüísticas asociadas a contextos más privados, que han sido objeto de un menor interés por parte de la sociolingüística variacionista. En buena medida, esto sigue siendo aplicable a las nuevas corrientes sociolingüísticas, que, si bien prestan más atención a las interacciones interindividuales (como sucede con los enfoques basados en las propias

redes sociales o las comunidades de prácticas), siguen centrándose fundamentalmente en los comportamientos públicos del individuo. Una virtud adicional que presenta el estudio de los familectos es que las características sociológicas de las unidades familiares son muy parecidas a las formas de organización sociopolítica que han sido predominantes durante casi toda la historia humana (bandas y clanes no son, en esencia, sino expansiones limitadas de dichas unidades), hasta el punto de que se ha defendido que las lenguas esotéricas no solo aparecen antes que las exotéricas, sino que han sido las predominantes (y seguramente las únicas) hasta el Neolítico (Benítez-Burraco y Progovac, 2020). En consecuencia, entender las características de los familectos y cómo dichas características son el resultado de las funciones que satisfacen (la transmisión de información entre iguales, el juego, la adquisición del lenguaje y la marcación de la identidad) y de las peculiaridades del entorno social en que aparecen (redes sociales densas) debería ayudar a comprender también el modo en que ha evolucionado el lenguaje y han surgido las lenguas actuales. Desde luego, y tal como hemos argumentado suficientemente, los familectos pueden constituir un laboratorio ideal para estudiar en detalle las interacciones lingüísticas esotéricas, pero también para comprender mejor la naturaleza de las lenguas esotéricas. Una parte de los grupos humanos actuales (ciertamente, cada vez una menor proporción de ellos) sigue hablando estas lenguas. Estudiarlas debería contribuir a enriquecer nuestra comprensión de los fenómenos lingüísticos, que sigue bastante condicionada por el hecho de que la lingüística se ha centrado fundamentalmente en el análisis de las lenguas habladas por sociedades occidentales, alfabetizadas, industrializadas y desarrolladas (las denominadas sociedades WEIRD, Heinrich *et al.*, 2010) y, sobre todo, de las lenguas con escritura y tradición literaria, estatus de lengua oficial y muchos hablantes (las conocidas como lenguas LOL, Dahl, 2015). Al igual que se está tratando de hacer en otras disciplinas, como las ciencias cognitivas (véase Majid, 2023 a este respecto), también la sociolingüística debe ampliar la nómina de fenómenos objeto de su interés si queremos alcanzar una comprensión adecuada de las complejas relaciones entre lengua y sociedad, especialmente desde el punto de vista histórico.

En último término, nuestro trabajo viene a relativizar la importancia que, desde el punto de vista lingüístico, tiene la tradicional frontera que separa a las lenguas de sus diferentes variedades. En nuestra opinión, las mismas condiciones que favorecen la aparición de los familectos son las que promueven el desarrollo de las lenguas esotéricas... y viceversa. Es posible (y deseable), por tanto, un diálogo más fluido entre las disciplinas que, como es el caso de la tipología y la sociolingüística, han trabajado tradicionalmente en lados diferentes de dicha frontera.

DECLARACIÓN DE CONTRIBUCIÓN DE AUTORÍA

Antonio Benítez-Burraco: Conceptualización, Adquisición de fondos, Investigación, Metodología, Administración del proyecto, Redacción – borrador original, Redacción – revisión y edición.

Elena Felú-Arquiola: Investigación, Metodología, Redacción – borrador original, Redacción – revisión y edición.

FINANCIACIÓN

Esta publicación es parte del proyecto de I+D+i PID2020-114516GB-Ioo, financiado por MCIN/ AEI/10.13039/501100011033/ (concedido a ABB).

REFERENCIAS

- Beaumont, S. L. (1995). Adolescent girls' conversations with mothers and friends: A matter of style. *Discourse Processes*, 20(1), 109-132.
DOI: <https://doi.org/10.1080/01638539509544933>
- Benítez-Burraco, A., y Kempe, V. (2018). The emergence of modern languages: Has human self-domestication optimized language transmission? *Frontiers in Psychology*, 9, Article 551. DOI: <https://doi.org/10.3389/fpsyg.2018.00551>
- Benítez-Burraco, A., y Progovac, L. (2020). A four-stage model for language evolution under the effects of human self-domestication. *Language & Communication*, 73, 1-17. DOI: <https://doi.org/10.1016/j.langcom.2020.03.002>
- Bolender, J. (2007). Prehistoric cognition by description: a Russellian approach to the upper Paleolithic. *Biology & Philosophy*, 22, 383-399.
DOI: <https://doi.org/10.1007/s10539-006-9058-2>
- Burrell, N. A. (1995). Communication patterns in stepfamilies: Redefining family roles, themes, and conflict styles. En M. Fitzpatrick y A. Vangelisti (Eds.). *Explaining family interactions* (pp. 290-309). Thousand Oaks, CA: Sage.
- Conti Jiménez, C. (2018). *Complejidad lingüística: Orígenes y Revisión Crítica del Concepto de Lengua Compleja*. Berna: Peter Lang.
- Chen, S., Gil, D., Gaponov, S., Reifegerste, J., Yuditha, T., Tatarinova, T. V., Progovac, L., y Benítez-Burraco, A. (2023, May 17). Linguistic and memory correlates of societal variation: A quantitative analysis. DOI: <https://doi.org/10.31234/osf.io/bnzzs>
- Chen, S., Gil, D., Gaponov, S., Reifegerste, J., Yuditha, T., Tatarinova, T. V., Progovac, L., y Benítez-Burraco, A. (2024). Linguistic correlates of societal variation: A quantitative analysis. *PLoS ONE*, 19(4). DOI: <https://doi.org/10.1371/journal.pone.0300838>
- Clancy, B. (2016). *Investigating intimate discourse. Exploring the spoken interaction of families, couples and friends*. Londres: Routledge.
- Dahl, Ö. (2015, Mayo 1-3). How WEIRD are WALS languages? [Conference presentation abstract]. Diversity Linguistics – Retrospects and Prospects, Leipzig.
https://www.eva.mpg.de/fileadmin/content_files/linguistics/conferences/2015-diversity-linguistics/Dahl_slides.pdf

- Díaz-Campos, M. (Ed.). (2011). *The Handbook of Hispanic Sociolinguistics*. Malden, MA: John Wiley & Sons.
- Díaz-Campos, M., Escalona Torres, J. M., y Filimonova, V. (2020). Sociolinguistics of the Spanish-Speaking World. *Annual Review of Linguistics*, 6(1), 363-388.
DOI: <http://dx.doi.org/10.1146/annurev-linguistics-011619-030547>
- Dryer, M. S., y Haspelmath, M. (Eds.). (2013). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
<http://wals.info>
- Eberhard, D. M., Simons, G. F., y Fennig, C. D. (Eds.). (2022). *Ethnologue: Languages of the World* (twenty-fifth edition). Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>
- Ehret, K., Berdicevskis, A., Bentz, C., y Blumenthal-Dramé, A. (2023). Measuring language complexity: challenges and opportunities. *Linguistics Vanguard*, 9(1), 1-8.
DOI: <https://doi.org/10.1515/lingvan-2022-0133>
- Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology*, 46(4), 621-646.
- Fletcher, P., y MacWhinney, B. (Eds.). (1995). *The Handbook of Child Language*. Oxford: Blackwell.
- Foulkes, P., Docherty, G. J., y Watt, D. (2005). Phonological variation in child-directed speech. *Language*, 81(1), 177-206. <http://www.jstor.org/stable/4489857>
- Gil D. (2021). Tense-aspect-mood marking, language-family size and the evolution of predication. *Philosophical Transactions of the Royal Society B* 376, 20200194.
DOI: <http://doi.org/10.1098/rstb.2020.0194>
- Gordon, C. (2009). *Making meanings, creating family: Intertextuality and framing in family interaction*. Nueva York: Oxford University Press.
- Halliday, M. A. K. (1976). Anti-languages. *American Anthropologist*, 78(3), 570-584.
<http://www.jstor.org/stable/674418>
- Hammarström, H., Forkel, R., Haspelmath, M., y Bank, S. (2022). *Glottolog 4.6*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
DOI: <https://doi.org/10.5281/zenodo.6578297> (Available online at <http://glottolog.org>).
- Hazen, K. (2004). The Family. En K. Chambers, P. Trudgill, y N. Schilling-Estes (Eds.). *The Handbook of Language Variation and Change* (pp. 500-525). Malden, MA.: Blackwell. DOI: <https://doi.org/10.1002/9780470756591.ch20>
- Henrich, J., Heine, S. J., y Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61-83.
DOI: <https://doi.org/10.1017/S0140525X0999152X>
- Joseph, J. E. (2021). Why does language complexity resist measurement? *Frontiers in Communication*, 6, 624855. DOI: <https://doi.org/10.3389/fcomm.2021.624855>
- Kempe, V., Brooks, P. J., Gillis, S., y Samson, G. (2007). Diminutives facilitate word segmentation in natural speech: cross-linguistic evidence. *Memory & Cognition*, 35(4), 762-773. DOI: <https://doi.org/10.3758/bf03193313>
- Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H. J., et al. (2016). D-PLACE: A global database of cultural, linguistic and environmental diversity. *PLoS One*, 11(7), e0158391. DOI: <https://doi.org/10.1371/journal.pone.0158391>

- Lupyan, G., y Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One* 5(1): e8559. DOI: <https://doi.org/10.1371/journal.pone.0008559>
- Lust, B. C. (2006). *Child Language: Acquisition and Growth*. Cambridge: Cambridge University Press.
- Majid, A. (2023). Establishing psychological universals. *Nature Reviews Psychology*, 2, 199–200. DOI: <https://doi.org/10.1038/s44159-023-00169-w>
- Miestamo, M. (2017). Linguistic diversity and complexity. *Lingue e Linguaggio*, 16(2), 227–253. DOI: <https://doi.org/10.1418/88241>
- Osovska, I. M. (2019). Lexical-semantic space of the German family discourse. En V. D. Bialyk, L. I. Didukh, M. Yu. Ivanchenko, et al. (Eds.). *Efficiency level and the necessity of influence of philological sciences on the development of language and literature: collective monograph* (pp. 123–144). LvivToruń: Liha-Pres. DOI: <https://doi.org/10.36059/978-966-397-170-4/123-144>
- Roberts, J. (2013). Child language variation. En J. K. Chambers y N. Schilling (Eds.). *The Handbook of Language Variation and Change* (pp. 263–276). Oxford: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781118335598.ch12>
- Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M. C., y Cohen, D. (2013). Motherese in interaction: at the cross-road of emotion and cognition? (A systematic review). *PLoS One*, 8(10), e78103. DOI: <https://doi.org/10.1371/journal.pone.0078103>
- Schick, J., Fryns, C., Wegdell, F., Laporte, M., Zuberbühler, K., van Schaik, C. P., et al. (2022). The function and evolution of child-directed communication. *PLoS Biology*, 20(5), e3001630. DOI: <https://doi.org/10.1371/journal.pbio.3001630>
- Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Washington, D.C.: Georgetown University Press.
- Skirgård, H., Haynie, H. J., Blasi, D. E., Hammarström, H., Collins, J., Latache, J. J., et al. (2023). Grambank reveals the importance of genealogical constraints on linguistic diversity and highlights the impact of language loss. *Science Advances*, 9(16), eadg6175. DOI: <https://doi.org/10.1126/sciadv.adg6175>
- Søndegaard, B. (1991). Switching between seven codes within one family—a linguistic resource. *Journal of Multilingual and Multicultural Development*, 12(1–2), 85–92. DOI: <https://doi.org/10.1080/01434632.1991.9994448>
- Storch, A. (2017). Typology of secret languages and linguistic taboos. En A. Y. Aikhenvald y R. M. W. Dixon (Eds.). *The Cambridge Handbook of Linguistic Typology* (pp. 287–322). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/9781316135716.010>
- Szmrecsanyi, B. (2015). Recontextualizing language complexity. En J. Daems, E. Zenner, K. Heylen, D. Speelman, y H. Cuyckens (Eds.) *Change of paradigms - new paradoxes: Recontextualizing language and linguistics* (pp. 347–360). Berlín: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110435597-020>
- Thurston, W. R. (1987). *Processes of change in the languages of north-western New Britain*. Canberra: Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.

- Thurston, W. R. (1994). Renovation and innovation in the languages of north-western New Britain. En T. Dutton y D. T. Tryon (Eds.), *Language Contact and Change in the Austronesian World* (pp. 573–609). Berlín: Mouton de Gruyter.
- Varenne, H. (1987). Talk and real talk: the voices of silence and the voices of power in American family life. *Cultural Anthropology*, 2(3), 369-394.
- Wray, A., y Grace, G. W. (2007). The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua*, 117, 543-578.
DOI: <https://doi.org/10.1016/j.lingua.2005.05.005>

TÉRMINOS GENÉRICOS Y FUNCIÓN LÉXICA *GENER* EN LA TEORÍA SENTIDO-TEXTO

Begoña SANROMÁN VILAS
*Universidad de Helsinki*¹

Resumen

Dentro de la teoría Sentido-Texto, la función léxica Gener forma parte de un sistema de codificación para formalizar las relaciones léxicas de una lengua. En concreto, Gener perfila una relación léxica paradigmática que da como valores términos que representan conceptos genéricos para la palabra llave a la que se aplica. Así, *asiento* sería un término genérico para *silla*, *sofá* o *taburete*, y *sentimiento*, para *tristeza*, *alegría* o *dolor*. Sin embargo, a pesar de la gran versatilidad de esta función léxica como herramienta teórica y práctica, su definición plantea algunas cuestiones: 1) ¿Son los valores de Gener del tipo *asiento* términos hiperónimos? En caso afirmativo, ¿cómo podemos verificarlo? 2) ¿Es Gener una relación léxica estrictamente paradigmática, a la luz de expresiones como *sentimiento de alegría*, en las que la palabra llave se combina con su valor? Si no es así, ¿cómo habría de clasificarse Gener, como una relación léxica mixta o como dos relaciones léxicas diferentes? El estudio ahondará en la naturaleza teórica de Gener para ofrecer una respuesta a las cuestiones planteadas, y establecer sus semejanzas y diferencias con otros conceptos de sentido general surgidos en el panorama lingüístico contemporáneo, los nombres ligeros y los nombres encapsuladores.

Palabras clave: funciones léxicas; Gener; términos genéricos; nombres encapsuladores; nombres ligeros.

GENERIC TERMS AND LEXICAL FUNCTION *GENER* IN MEANING-TEXT THEORY

Abstract

In Meaning-Text theory, the lexical function Gener is part of an encoding system to formalise the lexical relations of a language. Specifically, Gener outlines a paradigmatic lexical relation

1. begona.sanroman@helsinki.fi;  <https://orcid.org/0000-0003-3717-7283>

that provides as values terms corresponding to generic concepts for the keyword to which it is applied. Thus, *asiento* 'seat' is a generic term for *silla* 'chair', *sofá* 'sofa' or *taburete* 'stool', and *sentimiento* 'feeling' for *tristeza* 'sadness', *alegría* 'joy' or *dolor* 'pain'. However, despite the great versatility of this lexical function as a theoretical and practical tool, its definition raises some questions: 1) Are the values of Gener as *asiento* 'seat' hyperonymous terms? If so, how can we verify it? 2) Is Gener a strictly paradigmatic lexical relation, in the light of expressions such as *sentimiento de alegría* 'feeling of joy', in which the keyword is combined with its value? If not, how should Gener be classified, as a mixed lexical relation or as two different lexical relations? The study will delve into the theoretical nature of Gener in order to provide an answer to the questions raised, and to establish its similarities and differences with other generic concepts that have emerged on the contemporary linguistic scene, light nouns and shell nouns.

Keywords: lexical functions; Gener; generic terms; shell nouns; light nouns.

RECIBIDO: 13/06/2023

APROBADO: 01/05/2024

1. INTRODUCCIÓN

La teoría Sentido-Texto (TST) es una teoría lingüística de carácter integral que se caracteriza, entre otros aspectos, por la utilización de relaciones de dependencia en todos los niveles lingüísticos y por prestar una atención especial al léxico (Mel'čuk, 1997; 2016). Para la descripción del léxico de una lengua, la TST dispone de la herramienta lexicográfica de las funciones léxicas² (FL), un formalismo con el que pretende dar cuenta, en términos universales, de las relaciones léxicas que una unidad establece con otras unidades de la lengua, bien sea desde el punto de vista paradigmático, bien desde el sintagmático. Las FL paradigmáticas describen relaciones léxicas de sustitución y oposición entre los lexemas y las relaciones sintagmáticas, de combinación y contraste; en particular, las colocaciones. La información que proporcionan las FL es de carácter semántico y/o sintáctico.

Dentro del conjunto de las FL estándares del modelo, nos ocuparemos aquí de Gener, clasificada como una relación léxica de tipo paradigmático que, aplicada a un lexema (o palabra llave, L), da como resultado uno o varios términos (valores, L') que representan los conceptos genéricos más próximos al incluido en la palabra llave (Mel'čuk, 1996, p. 51; 2015, pp. 193-194; 2023, p. 219; Mel'čuk y Polguère,

2. Por limitaciones de espacio, no podemos describir en detalle las FL. El lector interesado puede consultar Mel'čuk (1998) y Mel'čuk y Polguère (2021), entre otras.

2021, pp. 29-31). De este modo, *asiento* contaría como valor genérico para *silla*, *sofá*, *banqueta* o *taburete*, mejor que *mueble*, que cubriría, además, *armario*, *cama* y *aparador*. Dentro de la misma FL se incluyen también valores como *sentimiento de alegría* o *proceso regenerativo*, que se utilizan en combinación con la palabra llave a la que se han aplicado (*alegría*) o con un derivado de ella (*regenerativo* < *regeneración*).

En vista de los usos descritos, se observan dos clases de valores de Gener: los del tipo *asiento* (aplicado a *silla*, *sofá*, *banqueta* y otros) o *flor* (válido para *rosa*, *tulipán*, *jazmín* y otros), que recuerdan a una relación paradigmática clásica de hiperonimia-hiponimia, y los del tipo *sentimiento de alegría* (*de dolor*, *doloroso*, etc.) o *proceso regenerativo* (*de regeneración*, *de gramaticalización*, etc.), que, dada la presencia de la palabra llave en el valor, se asemejan a las relaciones léxicas sintagmáticas. Como veremos más adelante, las propiedades semánticas y gramaticales de los dos tipos de valores difieren considerablemente unas de las otras, de manera que su estudio nos plantea, al menos, dos preguntas:

a) En cuanto a los valores del tipo *asiento* o *flor*, ¿es la relación léxica formalizada por medio de la FL Gener la misma que se describe bajo el concepto de hiperonimia-hiponimia? Desde dentro de la TST, la semejanza entre ambas relaciones se hace manifiesta en ocasiones (Mel'čuk, 1998, p. 35; Mel'čuk y Polguère, 2021, p. 29), pero se matiza en otras, rechazando la completa identidad (Mel'čuk, 2015, pp. 193-194). De tratarse de relaciones distintas, ¿qué pruebas nos permitirían dar cuenta de las diferencias?

b) En lo que respecta a los valores del tipo *sentimiento de alegría* o *proceso regenerativo*, el hecho de que la palabra llave, o un derivado adjetival, se combine con el valor, ¿es un indicio de que Gener no es estrictamente una FL paradigmática? De ser así, ¿se trataría de una relación de tipo mixto o habría que considerar la existencia de dos relaciones diferentes?

El objetivo de este artículo es ahondar en la naturaleza teórica de la FL Gener para dar respuesta a las preguntas anteriores, repercutiendo así en un mejor entendimiento y, en su caso, en una representación más clara de esta FL dentro de la TST. Por otra parte, trataremos de poner en relación el concepto de término genérico formalizado por la FL Gener con otros términos de sentido general, o genérico, que han hecho su aparición en el panorama lingüístico contemporáneo, en particular, los nombres ligeros y los nombres encapsuladores.

Hablaremos de «nombre ligero» para referirnos a aquellos nombres que, si bien funcionan como núcleo sintáctico de un sintagma, no constituyen su núcleo

semántico, limitándose a aportar un contenido general que clasifica o delimita aspectualmente a su complemento³. En (1a) el nombre ligero *circunstancias* denota una serie de acotaciones espaciotemporales no definidas de las que se puede predicar la propiedad de ‘ser difícil’. En (1b) se constata que *circunstancias* puede eliminarse de la oración al sustituir el complemento adjetival por un derivado nominal, *dificultades*, que incorpora el sentido aspectual que aportaba *circunstancias*.

- (1) a. Es cierto que Jimmy Carter atravesó *circunstancias* difíciles. (CORPES XXI, Perú: Jimmy Carter. *Caretas*, 2014/06/26.)
 b. Es cierto que Jimmy Carter atravesó dificultades.

El concepto de «nombre encapsulador», o «nombre cápsula», por su parte, se aplica a aquellos nombres que denotan segmentos de extensión variable dentro de un texto para identificarlos y condensar su sentido en una unidad (Schmid, 1998; 2000). Por una parte, el significado de estos nombres suele ser lo suficientemente general y amplio como para cubrir el de su referente; por otra parte, desde el punto de vista sintáctico, es frecuente que admitan un complemento argumental y/o apositivo. En (2a) *hecho*, válido para denotar cualquier evento o propiedad, va acompañado de un complemento apositivo, *durante décadas, el rol femenino quedara relegado a ser [...] de su cónyuge*, que constituye su referencia. También, en este caso, sería posible eliminar el nombre sin perjuicio del significado o de la gramaticalidad de la oración (2b).

- (2) a. El *hecho* de que, durante décadas, el rol femenino quedara relegado a ser una persona que cuidara del hogar y estuviera a disposición de su cónyuge tiene poco de genético. (CORPES XXI, España: Pradas, C. 2022. *Todo saldrá (bien)*. Barcelona: Nube de tinta.)
 b. El [Ø] que, durante décadas, el rol femenino quedara relegado a ser una persona que cuidara del hogar y estuviera a disposición de su cónyuge tiene poco de genético.

La estructura del estudio comprende cuatro apartados. Tras esta introducción, presentaremos los conceptos fundamentales que se desarrollarán a lo largo del artículo (§ 2). Comenzaremos por una revisión de la FL Gener en contraste con los conceptos de hiperonimia-hiponimia (§ 2.1), a la que seguirá un estado de la cuestión sobre los nombres ligeros (§ 2.2) y sobre los nombres encapsuladores (§ 2.3). A continuación,

3. Véanse otras definiciones de nombre ligero en Bosque (2001; 2016), Koike (2003) y Fábregas (2020, p. 35).

haremos una propuesta para la separación de dos tipos de FL Gener (§ 3), uno que formaliza una relación paradigmática y que tiene como valores lexemas plenos con sentido genérico (§ 3.1), y otro que describe una relación sintagmática cuyos valores son predicados semiléxicos con un sentido general (§ 3.2). La descripción de los dos tipos de Gener nos servirá asimismo para poner sus valores en relación con los conceptos de hiperonimia, de nombre ligero y de nombre encapsulador. Por último, se discutirán las repercusiones de la propuesta realizada y se harán unas reflexiones finales a modo de conclusión (§ 4).

2. LOS CONCEPTOS DE TÉRMINO GENÉRICO, NOMBRE LIGERO Y NOMBRE ENCAPSULADOR

En este apartado realizamos un breve recorrido por el concepto de término genérico (§ 2.1), con el que denominamos los valores de la FL Gener de la TST y los hiperónimos, así como por el de nombre ligero (§ 2.2) y el de nombre encapsulador (§ 2.3) a fin de contrastar sus semejanzas y diferencias.

2.1. *Los términos genéricos como valores de la FL Gener y como hiperónimos*

Una interpretación muy extendida de término genérico es aquella que lo asimila al concepto de hiperónimo, además de hacerlo a los valores de la FL Gener de la TST. Comenzaremos por presentar el concepto clásico de hiperonimia y, a continuación, examinaremos la FL Gener.

Un hiperónimo es un término cuyo significado está incluido en el de otros, implicando, por tanto, una relación de entrañamiento o inclusión entre un lexema más específico o subordinado (hipónimo) y otro más general o superordinado (hiperónimo) (Lyons, 1969 pp. 69-71; 1977, p. 291 y ss.)⁴. Así, *silla*, *sofá*, *taburete*, *butaca*, *banqueta*, etc., son hipónimos con respecto a *asiento*, término hiperónimo cuyo significado ('mueble para sentarse') está contenido en el de sus hipónimos —por ejemplo, *silla* puede definirse como 'asiento (= mueble para sentarse) con patas, con respaldo, sin brazos, para una persona, de material rígido, etc.' y *sofá*, como 'asiento (= mueble para sentarse) con patas, con respaldo, con brazos, para varias personas, de material blando, etc.'—. Para Lyons (1977, p. 291 y ss.), la hiponimia-hiperonimia es una relación paradigmática fundamental sobre la que se asienta la

4. Véanse también Leech (1974), Kempson (1977), Cruse (1986), Otaola Olano (2004), Escandell Vidal (2007), Espinal y Mateu (2014), entre otros.

organización semántica de la estructura jerárquica del léxico de una lengua. En la semántica estructural (véanse Greimas, 1966; Coseriu, 1977), el hiperónimo se corresponde con el archilexema de un campo léxico (p. ej., *asiento*) y los hipónimos, con los lexemas que forman el paradigma léxico del campo (*silla*, *sofá*, *taburete*, etc.). Sin embargo, cuando nos referimos al significado común de todos los lexemas de un campo, sin aludir a su realización léxica, hablamos de archisemema (‘mueble para sentarse’).

En tanto relación de implicación, la hiponimia-hiperonimia se ha estudiado ampliamente en la semántica lógica, en la que se destacan sus dos propiedades esenciales: la transitividad y la asimetría. La transitividad se manifiesta en que un lexema es hipónimo de su hiperónimo más inmediato y, a su vez, del hiperónimo de este, es decir, si *El País* es un *periódico* y un *periódico* es una *publicación diaria*, entonces *El País* es una *publicación diaria*⁵. Por otra parte, como la relación de implicación se da solo en una dirección, se dice que es asimétrica o unilateral. Para demostrar esta propiedad se suele recurrir a la prueba del condicional, con la que se obtiene una tautología, pero solo en uno de los sentidos (3).

- (3) a. Si algo es una silla, entonces necesariamente es un asiento → verdadero.
b. Si algo es un asiento, entonces necesariamente es una silla → falso.

Con la prueba del contraste se obtiene también un resultado semejante (4), en la que solo una de las versiones es posible y, en este caso, la verdad no es necesaria, sino contingente.

- (4) a. Esto es un *asiento*, pero no es una *silla* → verdadero.
b. Esto es una *silla*, pero no es un *asiento* → falso.

Dado que un hipónimo implica el sentido de su hiperónimo, en ocasiones se suele definir el hipónimo con el hiperónimo más un modificador que le aporte la especificidad de su contenido (Espinal y Mateu, 2014, p. 80) (5):

- (5) a. El País es un periódico español.
b. Una silla en un asiento con patas, respaldo y sin brazos, para una persona.

5. Partiendo de la existencia de diferentes niveles de jerarquización del léxico —nivel superordinado (ej., *producto lácteo*), nivel básico (*queso*) y nivel subordinado (*cabrales*)—, desde la semántica cognitiva se ha hecho hincapié en la importancia del nivel básico por ser el más frecuente y el más fácil de percibir y de aprender (Kleiber, 1995).

Como dijimos, dentro del conjunto de FL estándares de la TST, Gener es una FL de tipo paradigmático que se aplica a un lexema (o palabra llave, L) obteniendo uno o varios términos genéricos (o valores, L') para la palabra llave (Mel'čuk, 1996, p. 51; 2015, pp. 193-194; Mel'čuk y Polguère, 2021, pp. 29-31) (3):

- (6) a. Gener(*amor*) = *sentimiento, emoción*
 b. Gener(*astuto*) = *inteligente*
 c. Gener(*susurrar*) = *hablar*

Los ejemplos de (6) muestran que Gener puede aplicarse a cualquier categoría gramatical con la única restricción de que la palabra llave (L) y su valor (L') han de pertenecer a la misma categoría. Así, *sentimiento* y *emoción* (6a) son nombres genéricos con respecto a *amor* ('sentimiento/emoción de afecto intenso/a'); *inteligente* (6b), un adjetivo genérico en relación con *astuto* ('cualidad de inteligente aplicada al que se vale del engaño') y *hablar* (6c), un verbo genérico frente a *susurrar* ('hablar produciendo sonido sin voz'). Esta misma propiedad se da también en la relación de hiponimia-hiperonimia.

Si bien en algunas presentaciones más abreviadas se definen los valores de Gener como hiperónimos de la palabra llave (Mel'čuk, 1998, p. 35; Mel'čuk y Polguère, 2021, p. 29), en las exposiciones más detalladas se matiza que no se trata simplemente de un hiperónimo, o término que designa una clase más alta de entidades o hechos que la clase representada por la palabra llave. De acuerdo con Mel'čuk (2015, pp. 193-194), de tratarse de hiperónimos, los valores de Gener se podrían traducir de una lengua a otra por los correspondientes hiperónimos en la otra lengua, pero no siempre es el caso. No se proporcionan ejemplos, pero como explicación se aduce que el hiperónimo representa una relación semántica pura, universalmente válida, mientras que los valores de Gener constituyen una relación léxica, propia de una lengua particular⁶.

6. Un evaluador anónimo ha llamado nuestra atención sobre la distinción de la teoría del MSN (Metalenguaje Semántico Natural) entre los «superordinados taxonómicos» o hiperónimos, y los llamados «superordinados funcionales colectivos» (Wierzbicka, 1984, pp. 317-323; 1985, pp. 269 y ss.; Goddard 2017). Los primeros, como hemos visto, clasifican el léxico a modo de jerarquías con nodos y ramificaciones. El nodo constituiría el término hiperónimo (*árbol*) con respecto a sus ramificaciones o hipónimos (*abeto, abedul, álamo, arce, roble*, etc.), que incluirían en su semántica el significado del hiperónimo. De ello se deduce que los hipónimos presentan características similares. Frente a ellos, los superordinados funcionales colectivos no denotan clasificaciones jerárquicas sino agrupaciones de objetos, por lo general, de carácter funcional. Así, *ropa* agruparía nombres como *abrigo, pantalón, vestido, corbata, calcetín*, etc.; *cubtería* incluiría *tenedor, cuchillo, cuchara*, etc., y *armas* cubriría nombres como *pistola, cuchillo, arco*, etc. Los nombres agrupados bajo un mismo término superordinado no

En la TST se requiere que los elementos incluidos como valores de Gener satisfagan, al menos, una de las siguientes condiciones sintácticas: i) el valor de Gener es una unidad léxica que equivale semánticamente a la palabra llave si aparece modificada, bien por otra unidad en relación morfológica con la palabra llave (7), bien por un complemento preposicional que incluye a la propia palabra llave (8); ii) el valor de Gener es una unidad léxica que asigna la palabra llave y otras del mismo tipo a una clase (de objetos, hechos, propiedades, etc.); de ahí que pueda participar en una construcción del tipo $\langle L, L_1, L_2, \dots, L_{(N)} \text{ y otros (tipos de) Gener(L)} \rangle$ (9).

- (7) a. Gener_{ATR}(*amor*) = *sentimiento* [*sentimiento amoroso* = *amor*]
 b. Gener_{ATR}(*república*) = *estado* [*estado republicano* = *república*]
- (8) a. Gener₁(*amor*) = *sentimiento* [*sentimiento de amor* = *amor*]
 b. Gener₁(*regeneración*) = *proceso* [*proceso de regeneración* = *regeneración*]
- (9) a. Gener(*silla*) = *asiento* [*silla, sofá, banqueta... y otros asientos*]
 b. Gener(*español*) = *idioma* [*español, gallego, finlandés... y otros idiomas*]

Cualquiera de los términos genéricos ilustrados en (7) y (8), por ejemplo, *sentimiento* en (7a) y (8a), equivale a la palabra llave (*amor*) si va acompañado por un adjetivo derivado de *amor* (*amoroso*) o por un sintagma preposicional que incluye a la palabra llave (*de amor*), de manera que *amor* se identifica tanto con *sentimiento amoroso* como con *sentimiento de amor*. Los ejemplos de (9), en cambio, no responden a los mismos criterios. En este caso, *asiento* es un término genérico para *silla* en (9a) porque podemos decir *silla, sofá, banqueta... y otros asientos*. La diferencia entre los valores resultantes de la condición sintáctica (i), que se combinan con la palabra llave, y los de la condición (ii), que generalmente no lo hacen, ha sido criticada por varios autores. Entre ellos, Grimes (1990, citado en Alonso-Ramos y Tutin, 1996) rechaza la distinción entre FL paradigmáticas y sintagmáticas, debido a que los valores de algunas FL, en apariencia paradigmáticas, no lo son porque la palabra llave coocurre con su valor, como sucede en los ejemplos de (7) y (8). A la inversa, hay también casos de FL sintagmáticas, de las que se espera que la palabra llave se combine con su valor en una secuencia, como sucede con Magn ‘intenso’,

contienen en su definición al superordinado; de hecho, alguno de los nombres agrupados podría estar comprendido bajo más de un término superordinado, como *cuchillo* (elemento de la cubertería o tipo de arma). Desde el punto de vista gramatical, los superordinados colectivos suelen ser sustantivos colectivos (*cubertería, mantelería, mobiliario, ropa*) o *pluralia tantum* (*armas, joyas, muebles, ropas*). Según Wierzbicka (1984, p. 317), el hiperónimo podría representarse pictóricamente como un prototipo válido para todos sus hipónimos; sin embargo, no podría realizarse un dibujo que representase el prototipo de un superordinado colectivo.

que se aplica a una palabra llave para intensificar su sentido (10); sin embargo, en ocasiones, puede manifestarse por medio de una expresión fusionada (11)⁷.

(10) Magn(*miedo*) = *cerval, espantoso, horrible* [ej.: *miedo cerval*]

(11) Magn(*miedo*) = //*terror; //pánico* [ej.: *miedo cerval = terror*]

Alonso-Ramos y Tutin (1996, p. 151), desde dentro de la TST, se unen a las objeciones de autores como Grimes y, en el caso concreto de Gener, proponen distinguir los valores de los ejemplos (7) y (8) de los de (9), añadiendo a Gener el subíndice «p» cuando se trate de un valor paradigmático.

2.2. El concepto de nombre ligero

El concepto de nombre ligero (NL) surge a partir del de verbo ligero (Bosque, 2001; Butt, 2003; 2010). Fábregas (2020, p. 35) define nombre ligero como aquel que, aun presentando la forma de un nombre, no define la clase de objetos denotada, sino que aporta «información sobre la cantidad, número, clase o propiedades adicionales de otro sustantivo al que introduce». En consecuencia, *pila, rebanada, tipo, maravilla* y *situación* en (12) serían nombres ligeros. Todos ellos aportan algún dato relacionado con el nombre que les sigue y, si bien ocupan la posición de núcleo sintáctico del sintagma, no constituyen su núcleo semántico; de ahí que podamos eliminarlos y la oración continúe siendo gramatical, como se ve en (13).

- (12) a. Ordenó una *pila* de revistas. (CORPES XXI, Argentina: Maronna, J.; Pescetti, L. 2001. *Copyright...* Barcelona: Plaza & Janés.)
 b. Pellizcando distraídamente una *rebanada* de pan... (CORPES XXI, México: Murillo, A. 2018. *El niño que fuimos*. Ciudad de México: Penguin.)
 c. Y es muy difícil tomar este *tipo* de decisiones porque no existe una evidencia sólida. (CORPES XXI, España: Meyer, L. Darle una primera dosis... *Ethic*, 2022/01/26.)
 d. Gracias a Instagram nos topamos con esta *maravilla* de tienda... (CORPES XXI, México: Kohn, P. Las roscas de reyes más deliciosas... *Quién*, 2019/12/17.)
 e. ... [ambos] están en una *situación* de conflicto... (CORPES XXI, Chile: Macari, G. La última intriga de Joël Dicker. *La Cuarta*, 2022/07/25.)

7. El símbolo «//» ante *terror* y *pánico* indica que son expresiones fusionadas para 'miedo intenso'.

- (13) a. Ordenó revistas.
 b. Pellizcando pan distraídamente...
 c. Y es muy difícil tomar estas decisiones porque no existe una evidencia sólida.
 d. Gracias a Instagram nos topamos con esta tienda...
 e. ...[ambos] están en conflicto...

En este estudio nos centraremos únicamente en los nombres ligeros que, además de poder ir modificados por un sintagma preposicional como en (12), pueden acompañarse también de un adjetivo. Este sería el caso de *situación* (12e), que puede ir modificado tanto por un complemento preposicional (*en una situación de conflicto*) como por un adjetivo (*en una situación conflictiva*). Para el estudio de estos NL utilizaremos la clasificación de Sanromán Vilas (2023; en prensa) que divide este grupo de NL en predicados de nivel episódico (NLE) y predicados de nivel individual (NLI). Los NLE como *circunstancias* —véase (1) en § 1— o el propio *situación* (12e) denotan coordenadas espacio-temporales sin límites precisos en los que tiene lugar un evento o en los que sitúa una entidad. Otros NLE son *etapa*, *periodo*, *estado* o *momento* (14). Los ejemplos en (15) muestran que no pueden aparecer sin complemento. A la inversa, si el complemento es un nombre precedido de preposición, este nombre puede aparecer en lugar del NLE (16a-b) y si es un adjetivo, la nominalización correspondiente también puede sustituir al NLE (16c-d).

- (14) a. Tu mejor mes del año será junio, donde entras en una *etapa* de metamorfosis. (CORPES XXI, Ecuador: Mhni Vidente. Horóscopo... *Metro Ecuador*, 2022/05/30.)
 b. ...el «Barrio Chino», donde ella vivió en su *periodo* de formación... (CORPES XXI, España: Vidal, J. Teresa Lanceta teje su autobiografía... *El Cultural*, 2022/4/20.)
 c. ... que hoy [el palacio Baburizza]se encuentra en *estado* ruinoso... (CORPES XXI, Chile: Moreno, R. La Virgen Negra de Reñaca... *Tell Magazine*, 2021/02.)
 d. Empezar la terapia fue un *momento* complicado. (CORPES XXI, España: Jiménez, M. 2022. *Acepta y vuela*. Barcelona: Plan B.)
- (15) a. *Tu mejor mes del año será junio, donde entras en una *etapa*.
 b. ...*el «Barrio Chino», donde ella vivió en su *periodo*.
 c. ... *que hoy [el palacio Baburizza]se encuentra en *estado*.
 d. *Empezar la terapia fue un *momento*.
- (16) a. Tu mejor mes del año será junio, donde entras en una *metamorfosis*.
 b. ...el «Barrio Chino», donde ella vivió en su *formación*.
 c. ... que hoy [el palacio Baburizza] se encuentra en *ruinas*.
 d. Empezar la terapia fue una *complicación*.

Los NLI, por su parte, se refieren a las propiedades de un individuo en sentido absoluto como *carácter*, *naturaleza*, *índole*, *cualidad* o *temperamento* (17). En estos casos, el NLI tampoco puede aparecer sin un modificador (18) y, al igual que con los NLE, los NLI también pueden eliminarse siendo reemplazados por el nombre que forma parte de su complemento preposicional (19a-b) o por su complemento adjetival, que pasa a modificar al nombre que aparece en un nivel más alto (19c-d).

- (17) a. Nunca había convocado una comisión nacional de *carácter* de emergencia... (CORPES XXI, Bolivia: Llamam al Congreso... *El deber*, 2012/01/12.)
 b. ...los niños y niñas deben organizar reuniones para discutir la *naturaleza* del acoso escolar. (CORPES XXI, Perú: Cómo los países nórdicos previenen... *La República*, 2012/05/02.)
 c. Muchos fueron los problemas de *índole* política y ética... (CORPES XXI, México: Loaeza, S. Fin de fiesta en Gran Bretaña. *Nexos*, 2022/08/23.)
 d. ...convertir aquellas vagas aspiraciones suyas en proyectos de *cualidad* tangible... (CORPES XXI, Perú: Benavides, J. 2022. *Volver a Shangri-La*. Madrid: Alianza.)
- (18) a. *Nunca había convocado una comisión nacional de *carácter*.
 b. ...*los niños y niñas deben organizar reuniones para discutir la *naturaleza*.
 c. *Muchos fueron los problemas de *índole*.
 d. ...*convertir aquellas vagas aspiraciones suyas en proyectos de *cualidad*.
- (19) a. Nunca había convocado una comisión nacional de *emergencia*.
 b. ...los niños y niñas deben organizar reuniones para discutir el *acoso escolar*.
 c. Muchos fueron los problemas *políticos y éticos*.
 d. ...convertir aquellas vagas aspiraciones suyas en proyectos *tangibles*.

2.3. El concepto de nombre encapsulador

Como indica el término, los nombres encapsuladores (NE), o nombres cápsula, se usan para encapsular un segmento del discurso con el fin de identificarlo y condensar su sentido en una unidad que resulta más fácil de concebir y memorizar (Schmid, 1998; 2000). Los estudios dedicados a los nombres encapsuladores⁸ (Borreguero

8. Existe gran diversidad terminológica. López Samaniego (2011, p. 297) clasifica las denominaciones que han recibido estos nombres en tres grupos: a) las que se basan en el modo de significar como «nombres generales» o «nombres inespecíficos»; b) las que atienden a la función discursiva como «nombres anafóricos» «encapsuladores» o «nombres señaladores» y c) las que llaman la atención sobre la metáfora que relaciona el nombre y el segmento discursivo que contiene: «nombres contenedor», «etiquetas», «nombres portadores», «nombres envoltorio». Véase también Borreguero Zuloaga (2018, p. 180, nota 1).

Zuloaga, 2018; López Samaniego, 2011; 2015; Rodríguez-Espiñeira, 2015; 2018; Schmid, 1998; 2000; 2018; entre otros)⁹ remiten al ámbito discursivo, en particular, a las relaciones de cohesión que se establecen entre el llamado nombre encapsulador y su referente, así como a las funciones que desempeña este nombre dentro del discurso en cuestión. Se suele matizar que estos nombres no constituyen una clase por sí misma, sino una función (Schmid, 1998, p. 3; López Samaniego, 2011, p. 299), de manera que no se puede afirmar que un nombre en concreto es un encapsulador hasta no haber observado su comportamiento dentro de un discurso.

A diferencia de los hiperónimos y de los valores de Gener, los encapsuladores pertenecen siempre a la categoría nominal porque se asume que los nombres presentan una mayor capacidad para transformar los conceptos complejos y no delimitados en entidades individualizadas y atribuirles propiedades típicas de los objetos —operación que ha sido denominada «hipostatización» (Schmid, 1998), «reificación» (Borreguero Zuloaga, 2018, p. 183) o «cosificación» (López Samaniego, 2011, p. 321)—.

Los NE son contextualmente dependientes porque siempre hacen referencia a algún tipo de información que aparece en el discurso (Schmid 1998, p. 2). El referente de los NE denota entidades de segundo (actividades, procesos, estados) o de tercer orden (enunciados, afirmaciones), pero nunca de primer orden (objetos). De acuerdo con la posición que ocupe el segmento al que se refieren, los NE pueden clasificarse en dos grupos. Cuando el referente aparece antes del nombre, se habla de encapsuladores anafóricos o etiquetas discursivas (20); por el contrario, cuando el referente se sitúa detrás del nombre, se utiliza la denominación de encapsuladores catafóricos (21).

- (20) Pero nada es inmune a los vaivenes económicos y entre 1978 y 1982 el país atravesó una de sus peores crisis, con una caída anual del 7.8 % del PIB. *Este hecho* llevó a Costa Rica a transformar su economía esencialmente agrícola en una industrial y de servicios. (CORPES XXI, México: Hernández, F., El resurgimiento de la Suiza... *Expansión*, 2021/10/01.)
- (21) Es difícil creer que quienes reclaman más libertad y democracia puedan apoyar *la idea* de que Elizabeth II y su familia han sido escogidos por el propio Dios para dirigir los destinos de millones de personas. (CORPES XXI, Chile: Vélez-Plickert, M., Un jubileo con sabor a despedida. *DF Más*, 2022/06/03.)

9. En esta presentación seguimos fundamentalmente a López Samaniego (2011), Rodríguez-Espiñeira (2015) y Schmid (1998; 2000).

En (20) el fragmento *entre 1978 y 1982 el país atravesó una de sus peores crisis, con una caída anual del 7.8 % del PIB*, anterior al punto, es recuperado por el nombre *hecho*, que a su vez lo individualiza y lo presenta como la causa de la transformación económica del país. A diferencia de (20), en el que el NE y su referente están en estructuras sintácticas independientes, en (21), ocurren en la misma estructura, siendo el referente, desde el punto de vista sintáctico, una oración completiva dependiente del NE. En este caso, *que Elizabeth II y su familia han sido escogidos por el propio Dios para dirigir los destinos de millones de personas* se encapsula en el nombre *idea*, que, sin duda, imprime una perspectiva diferente a si se hubiera utilizado otro nombre, por ejemplo, *creencia* o *superstición*.

En principio, los NE constituyen una clase abierta y, si bien no se puede establecer de antemano qué nombres pueden o no ser NE, se considera que unos son más prototípicos que otros. Siguiendo una escala valorativa (Schmid, 2000, p. 85 y ss.), los miembros centrales de la categoría representan conceptos nominales primarios, tales como *concepto*, *hecho*, *asunto/cuestión*, *principio*, *problema* y *cosa*¹⁰, que no pueden descomponerse morfológicamente en unidades más simples. Llevan como referente entidades de tercer orden (Schmid, 2000, p. 66) y admiten ser agrupados en la clase de los nombres factuales como vimos en (20) con *hecho*, en la de los nombres mentales como *idea* (21) o en la de nombres lingüísticos como *mensaje* o *rumor*. En segundo lugar, se situarían los nombres que, si bien representan también entidades de tercer orden, tienen derivados verbales y adjetivales (ej., *orden* y *ordenar*; *probabilidad* y *probable*, etc.). Estos NE pertenecen a la clase de los nombres lingüísticos como *orden* y *supuesto*, de los mentales como *creencia* o *asunción*, y de los modales como *certeza* y *permiso*, pero pueden aparecer también nombres eventivos con un componente modal como *intento*, *esfuerzo* y *hábito*. En tercer lugar, se situarían los miembros periféricos de la categoría que ya no representan relaciones abstractas (entidades de tercer orden), sino eventos (entidades de segundo orden) como *movimiento* y *situación*. Por último, en una categoría todavía más periférica estarían los nombres circunstanciales de tiempo (*momento*, *estado*) y de lugar (*lugar*, *área*), cuya pertenencia al grupo de los NE es más dudosa.

10. Hemos adaptado al español los nombres en inglés de Schmid (2000, p. 85).

3. PROPUESTA PARA LA SEPARACIÓN DE DOS TIPOS DE GENER

A pesar de que las descripciones canónicas de la TST (Mel'čuk, 1996; 1998; 2015; 2021) mantienen que Gener es una FL de tipo paradigmático, aquí partiré de la propuesta de separar los valores paradigmáticos de los sintagmáticos (Alonso-Ramos y Tutin, 1996). Los paradigmáticos, o valores de GenerP, se pondrán en relación con los hiperónimos y con los nombres encapsuladores anafóricos (§ 3.1) y los sintagmáticos, o valores de GenerS, se compararán con los nombres ligeros y con los nombres encapsuladores catafóricos (§ 3.2). Hemos analizado datos extraídos de CORPES XXI.

Antes de proceder al análisis, fundamentaremos la propuesta de separación de los dos tipos de Gener apoyándonos en las siguientes premisas: los valores paradigmáticos (L', L'', L''', etc.) de un lexema o palabra llave (L) se asientan en las propiedades puramente semánticas de los lexemas que entran en relación (L y L', L'', L''', etc.), sin que medien entre ellos restricciones léxicas concretas; por el contrario, los valores sintagmáticos (L', L'', L''', etc.) de un lexema o palabra llave (L) se basan en la selección léxica de carácter restringido operada por el lexema L para expresar un sentido 'genérico'.

De acuerdo con las premisas anteriores, se espera que los valores de GenerP (= término genérico paradigmático) sean unidades léxicas plenas que denotan la clase de entidades o hechos a la que pertenece la palabra llave. El *genus proximum* o componente genérico de la definición de un lexema, si está lexicalizado, es un valor de GenerP, pero existen otros valores posibles, dependiendo de la perspectiva desde la que se esté considerando el significado del lexema analizado. De las dos condiciones sintácticas establecidas en Mel'čuk (2015, p. 194) y presentadas arriba (§ 2.1), los valores de GenerP se adscriben típicamente a la segunda condición, pero no la primera, de manera que podremos encontrarlos en la siguiente construcción: 'L, L₁, L₂ [...], L_(N) y otros tipos de GenerP(L)', p. ej., *silla, sofá, banqueta... y otros asientos* (v. 9a, arriba). Estos valores son típicos para los nombres de entidades físicas y para términos de la ciencia.

En cuanto a los valores de GenerS (= término genérico sintagmático), se espera que sean unidades léxicas semipredicativas o colocativos de la palabra llave. Siguiendo un paralelismo descriptivo, diremos que, de las dos condiciones sintácticas establecidas en Mel'čuk (2015, p. 194) y presentadas arriba (§ 2.1), los valores de

GenerS cumplen la primera condición, pero no la segunda, de manera que podrán aparecer en una de las siguientes estructuras¹¹:

- a. 'GenerS_{ATRIB}(L)-ATRIB→DER(L)' = 'L'
 «El valor genérico sintagmático de un lexema L (*sentimiento*) equivale semánticamente a L (*amor*) si este valor va modificado por un adjetivo derivado de L (*amoroso*); *sentimiento amoroso* = *amor*».
- b. 'GenerS_I(L)-I→L' = 'L',
 «El valor genérico sintagmático de un lexema L (*sentimiento*) equivale semánticamente a L (*amor*) si este valor va modificado por un complemento preposicional introducido por *de* y que incluye L (*amor*); *sentimiento de amor* = *amor*».

A continuación, trataremos de justificar la validez de esta propuesta utilizando como ejemplo el uso de *sentimiento*, *emoción*, *sensación* y *estado* como términos genéricos aplicados a los nombres que denotan 'emociones'. La prueba consiste en comprobar si los 32 nombres de 'emoción' que aparecen en la Tabla 1 pueden combinarse con los términos mencionados. Para llevar a cabo la comprobación se han hecho búsquedas en CORPES XXI a partir del patrón <un/a *sentimiento/emoción/sensación/estado* de X (nombre de 'emoción')>. Si encontramos ejemplos de la combinación, lo indicamos con un signo positivo (+) y, en caso contrario, con uno negativo (-). Cuando solo se ha documentado un ejemplo, dejamos constancia de ello. En las búsquedas solo hemos considerado las ocurrencias en las que el nombre de 'emoción' no lleva modificadores.

N de 'emoción'	SENTIMIENTO	EMOCIÓN	SENSACIÓN	ESTADO
<i>aburrimiento</i>	+ (1 caso)	-	+	+ (1 caso)
<i>afecto</i>	+	-	-	-
<i>alegría</i>	+	-	+	+
<i>amor</i>	+	-	+	-
<i>arrepentimiento</i>	+	-	+	-
<i>asco</i>	+	-	+	-
<i>asombro</i>	+	-	+	+ (1 caso)
<i>aturdimiento</i>	+	+ (1 caso)	+	+
<i>confianza</i>	+	-	+	+
<i>confusión</i>	+	-	+	+
<i>decepción</i>	+	-	+	+
<i>desprecio</i>	+	-	+ (1 caso)	-
<i>dolor</i>	+	-	+	-

11. En el apartado siguiente, presentaremos una aparente excepción a esta regla.

<i>duda</i>	+	-	-	+
<i>enfado</i>	-	-	-	-
<i>entusiasmo</i>	-	-	-	-
<i>envidia</i>	+	-	+ (1 caso)	-
<i>esperanza</i>	+	-	+	+
<i>felicidad</i>	+	-	-	+
<i>ira</i>	+	-	+	+
<i>nostalgia</i>	+	-	+	-
<i>odio</i>	+	+ (1 caso)	+ (1 caso)	-
<i>optimismo</i>	+	-	+	+
<i>pánico</i>	+	-	+	+
<i>pena</i>	+	-	-	-
<i>perplejidad</i>	+	-	+	+
<i>rabia</i>	+	-	+	+
<i>respeto</i>	+	-	+	-
<i>sorpresa</i>	+	-	+	-
<i>sospecha</i>	-	-	-	+
<i>temor</i>	+	-	+	+
<i>terror</i>	+	-	+	+
TOTAL	29 (91 %)	2 (6 %)	25 (78 %)	18 (56 %)

Tabla 1. Coocurrencia de los nombres de ‘emoción’ con *sentimiento*, *emoción*, *sensación* y *estado*

El resultado muestra que el 91 % de los nombres coocurren con *sentimiento*, el 78 % también lo hace con *sensación*, el 56 % puede combinarse con *estado*, pero solo el 6 % lo hace con *emoción*. A raíz de esta verificación, no podemos interpretar sin más que *sentimiento*, *sensación* y *estado* son válidos como términos genéricos para estos nombres, pero *emoción* no lo es. De hecho, una rápida búsqueda en Google nos permite comprobar que es precisamente *emoción* el término genérico que con más frecuencia se emplea en los vocabularios especializados para referirse a estos nombres. La conclusión que, en consecuencia, podemos extraer de los datos de la Tabla 1 es que *emoción* no se combina con los nombres particulares y, en consecuencia, no puede constituir un valor de GenerS; por el contrario, *sentimiento* y *sensación* coocurren con la mayoría de los nombres y *estado* lo hace con algo más de la mitad. En este caso, habría que especificar para cada caso individual qué nombre o nombres aceptan *sentimiento*, *sensación* y/o *estado* como valores de GenerS. En cambio, la prueba no dice nada de los valores GenerP, debido a que estos valores no se basan en la coocurrencia, sino que se extraen a partir de los componentes semánticos. A continuación, trataremos por separado los dos tipos de valores.

3.1. Términos genéricos paradigmáticos y su relación con los hiperónimos y los nombres encapsuladores anafóricos

En este apartado nos centramos en los valores paradigmáticos de la FL GenerP y en su relación con los hiperónimos y con los nombres encapsuladores anafóricos, en tanto que unidades que constituyen relaciones semánticas de inclusión. En particular, los valores paradigmáticos de la FL GenerP se corresponden con términos genéricos en sentido puramente semántico. Se trata de valores que siguen el patrón presentado arriba para *emoción*, es decir, denotan una clase de entidades en virtud de sus componentes semánticos comunes, pero no se combinan con los lexemas que designan las entidades particulares. Son frecuentes en los vocabularios especializados y terminologías. Desde este punto de vista, se equiparan a los hiperónimos, con los que no presentan diferencias, pues también estos se basan en relaciones semánticas, y no léxicas. En ambos casos, constituyen lexemas plenos con un significado estable y pueden presentarse en otras categorías aparte de la nominal. Encontramos, así, verbos como *engullir* o *devorar* que se agrupan bajo el verbo de sentido genérico *tragar*¹² y adjetivos tan variados como *añil*, *celeste*, *cobalto*, *índigo*, *marino* o *turquesa*, que sirven para especificar un tipo de color *azul*.

Como mencionamos arriba, los términos genéricos paradigmáticos entran en la construcción ' $L, L_1, L_2 [\dots], L_{(N)}$ y otros tipos de GenerP(L)', pero no en ' $\text{GenerS}_{\text{ATRIB}}(L)\text{-ATRIB}\rightarrow\text{DER}(L)$ ' = 'L' o ' $\text{GenerS}_1(L)\text{-I}\rightarrow L$ ' = 'L', propia de los términos genéricos sintagmáticos. Así ocurriría con *silla*, *sofá*, *banqueta*, *taburete... y otros asientos* (9a) (**asiento de silla* = *silla*) o con *alegría*, *decepción*, *esperanza*, *sorpresa... y otras emociones* (**emoción decepcionante* = *decepción*; **emoción de decepción* = *decepción*). No obstante, hemos advertido una excepción a esta regla que se daría con ciertos términos genéricos paradigmáticos como *idioma* (9b) que, además de encontrarse en la construcción *español*, *gallego*, *finlandés... y otros idiomas*, también pueden entrar en las construcciones típicas de los términos genéricos sintagmáticos, p. ej., *idioma español* = *español*. Hemos constatado que los nombres genéricos que se usan tanto paradigmática como sintagmáticamente son aquellos que se aplican a términos específicos que pueden funcionar como sustantivos y adjetivos¹³. Así, en *el español... y otros idiomas*, *español* es un

12. Ejemplo tomado de Escandell Vidal (2007, p. 66).

13. Algunos autores los llaman «categorías híbridas» (Varela Ortega, 2008). En particular, Bosque (1999, p. 62) considera que, en estos casos, se trata de adjetivos que se han recategorizado como sustantivos, es decir, que han sufrido un proceso semántico por medio del cual han pasado de designar una propiedad a aludir a una clase de entidades, lo que implica, desde el punto de vista sintáctico, que han

sustantivo, mientras que, en *idioma español*, se trata de un adjetivo. Otros ejemplos del mismo tipo serían los mostrados en (22-25), en donde los ejemplos de (a) representarían valores de GenerP y los de (b), de GenerS¹⁴:

- (22) a. GenerP(*azul*) = *color* [*azul, rojo, amarillo, verde... y otros colores*]
 b. GenerS_{ATR}(*azul*) = *color* [*color azul*_{ADJ} = *azul*_{SUST}]
- (23) a. GenerP(*recta*) = *línea* [*recta, curva, paralela, tangente... y otras líneas*]
 b. GenerS_{ATR}(*recta*) = *línea* [*línea recta*_{ADJ} = *recta*_{SUST}]
- (24) a. GenerP(*lavadora*) = *máquina* [*tostadora, secadora... y otras máquinas*]
 b. GenerS_{ATR}(*lavadora*) = *máquina* [*máquina lavadora*_{ADJ} = *lavadora*_{SUST}]
- (25) a. GenerP(*antidepresivo*) = *medicamento* [*calmante, sedante... y otros medicamentos*]
 b. GenerS_{ATR}(*antidepresivo*) = *medicamento* [*medicamento antidepresivo*_{ADJ} = *antidepresivo*_{SUST}]

A diferencia de los términos genéricos paradigmáticos y de los hiperónimos, que se especializan en el nivel semántico, los nombres encapsuladores anafóricos ejercen su función en el ámbito del discurso. Se trata de nombres que retoman un segmento textual anterior y lo actualizan en el presente discursivo de manera condensada y ágil, favoreciendo así su retención. Cualquiera que haya sido el sentido del fragmento que sirve de antecedente, al quedar «encapsulado» en un nombre concreto, es el sentido de este nombre el que sintetiza y generaliza el del segmento, imprimiéndole una nueva lectura que sirve los objetivos del discurso. Los ejemplos (26) y (27) ofrecen sendas muestras de nombres encapsuladores. En (26), *interpretación* encapsula todo el segmento precedente, en el que se explica que la biodiversidad para una gran mayoría se limita a dar cuenta de un número elevado de especies diferentes. En lugar de utilizar el nombre *interpretación*, el autor del fragmento podría haber usado *análisis*, *explicación*, *versión* o *teoría*, que se mantendrían en la misma línea, si bien con matices diferentes, o bien introducir juicios de valor negativos a través del empleo de nombres como *confusión*, *equivoco* o *error*. En (27), la secuencia que compara la longevidad de las mujeres con respecto a la de los hombres queda encapsulada en el nombre *ventaja (biológica)*, confiriéndole así un tono positivo. No obstante, otros nombres como *constatación* o *hecho* le otorgarían una visión más

pasado a funcionar como sustantivos y formar sintagmas nominales de interpretación no anafórica (*el español, el gallego, etc.*)

14. Los casos de recategorización en (23-25) están tomados de Bosque (1999, pp. 65-66).

neutra y otros, como *desigualdad*, *desproporción*, *diferencia* o *irregularidad* enfatizarían más el contraste con o sin un matiz negativo.

- (26) Para la mayoría, la biodiversidad es solo una lista de especies, una visión que reduce enormemente su complejidad ya que se queda en la parte meramente numérica y solo hace referencia a un nivel de organización de la vida. Según esta *interpretación*, un sitio más diverso sería aquel que tiene más especies y uno pobre, el que tiene menos. (CORPES XXI, España: Valladares, F. et al. 2022. *La salud planetaria*. Madrid: CSIC.)
- (27) [En Panamá] La esperanza de vida al nacer es mayor para las mujeres y es menor la tasa de mortalidad desde la infancia hasta la edad adulta. Esta *ventaja* biológica se pierde desde el punto de vista social, debido a que las mujeres son quienes presentan las peores condiciones de inequidad, una de cada cinco vive en pobreza extrema y una de cada dos vive en condiciones de pobreza. (CORPES XXI, Panamá: *Informe sobre la salud de la mujer de Panamá*. Panamá: OPS, 2006.)

(26) y (27) dejan también constancia de que el fragmento encapsulado y el nombre encapsulador mantienen una relación transitoria, limitada al texto en que opera, de ahí que se diga que el significado de los nombres encapsuladores no sea estable. Y no lo es porque, como dijimos, más que de una clase de nombres, se trata de una función para mejorar la cohesión del texto, de un lado, y persuadir al lector de la interpretación por la que debe optar, de otro.

El contraste entre la inestabilidad significativa de los nombres encapsuladores frente a la estabilidad de significado de los términos genéricos paradigmáticos y de los hiperónimos es otra de las propiedades que permiten distinguir a unos de otros, a pesar de que, como se ha notado en la bibliografía, los hiperónimos y los términos genéricos de la TST (Halliday y Hasan, 1976; Alonso-Ramos y Tutin, 1993; Lee y Evens, 1996) también pueden tener usos discursivos para potenciar la cohesión discursiva. De acuerdo con López Samaniego (2011, p. 413), incluso en esos usos, los hiperónimos siguen estableciendo vínculos semánticos sistemáticos con sus antecedentes, entidades de primer orden (objetos, individuos), sin añadir nuevos matices. Los encapsulares, como vimos, poseen la capacidad de reactivar un segmento del discurso (típicamente una proposición o entidad de tercer orden) y convertirlo en una unidad a la que se le añade una tonalidad significativa. Con todo, al lado de los hiperónimos descritos, que López Samaniego (p. 418) llama «convencionales», habría otros más propiamente «discursivos», «que mantienen con la entidad a la que clasifican una relación de inclusión que se crea en el discurso y

que limitan su validez al contexto en que aparecen» (p. 418). Se trataría de los hiperónimos valorativos o de aquellos en los que se designa a un participante a través del papel que desempeña en la actividad descrita. En estos casos (28), la relación entre el hiperónimo (*cliente*) y su antecedente (*Paloma*) no es estable, sino que varía de acuerdo con el texto y el referente.

- (28) Paloma confiesa haber tenido el «dudoso honor» de formar parte de un servicio de Internet ADSL de 256 KB a 1MB. El problema, según esta *cliente*, es que nadie la avisó de la migración. (CORPES XXI, España: Migración sin previo aviso. El País, 2004/09/02.)

En conclusión, podemos decir que cuando hablamos de valores, términos o lexemas genéricos paradigmáticos y de hiperónimos nos estamos refiriendo al mismo tipo de nombres, considerados en el nivel puramente semántico, pero presentados desde distintos marcos teóricos. Ambos aparecen en las mismas construcciones y tienen significado pleno. A pesar de las claras diferencias que estos términos genéricos mantienen con los encapsuladores, existe una categoría intermedia, la de los hiperónimos discursivos, que permite vincular a unos y otros.

3.2. Términos genéricos sintagmáticos y su relación con los nombres ligeros y los encapsuladores catafóricos

En este apartado nos ocuparemos de los valores de GenerS, o términos genéricos sintagmáticos, y de su relación con los nombres ligeros y con los encapsuladores catafóricos. Como hemos señalado más arriba, los valores de GenerS son términos genéricos de la palabra llave que se usan en combinación con un derivado de esa palabra —*estado* (*republicano*) cuenta como valor de GenerS para *república*— o con la propia palabra llave —*sentimiento* (*de amor*) es un valor de GenerS para *amor*—, de manera que entran en una de estas construcciones, ‘GenerS_{ATRIB}(L)–ATRIB→DER(L)’ = ‘L’ y ‘GenerS_I(L)–I→L’ = ‘L’, respectivamente, o en ambas. Tal y como hicimos notar, en las descripciones canónicas de las FL, GenerP y GenerS constituyen una única FL de carácter paradigmático. La razón por la que postulamos una FL GenerS de naturaleza sintagmática, aparte del hecho de que sus valores coocuran con la palabra llave, tiene que ver con su significado. Frente a los términos genéricos paradigmáticos, que presentan un significado pleno, los sintagmáticos, como veremos, repiten parte del ya contenido en la palabra llave.

Como ocurre en las colocaciones y en todas las FL sintagmáticas, la generalización sobre la que opera GenerS no se establece en términos puramente denotativos o semánticos, sino que es de naturaleza léxica. La palabra llave selecciona de manera léxicamente restringida otra palabra que, en ese contexto, funciona como su término genérico, es decir, se vacía parcialmente de su significado y adopta un sentido general o taxonómico, semejante a lo que ocurre con los valores verbales colocativos de la FL Oper₁, en la que un sustantivo selecciona un verbo, también parcialmente desemantizado, para verbalizar su sentido (*dar un paseo, hacer una promesa*). Tanto en el caso de GenerS como en el de Oper₁, la combinación de la palabra llave con su valor equivale a una unidad léxica plena, sea un sustantivo, en el caso de GenerS (*estado republicano = república*), o un verbo, en el de Oper₁ (*dar un paseo = pasear*). Otro aspecto en el que no se ha insistido lo suficiente y que justifica la separación entre los valores genéricos sintagmáticos y los paradigmáticos reside en que el nombre colocativo, valor de GenerS, al igual que sucede con los valores de Oper₁, no puede aparecer solo. En los ejemplos de (29) se observa que va acompañado por un modificador y que su eliminación (30) causa agramaticalidad¹⁵. Por el contrario, si se elimina *proceso* el resultado es una secuencia gramatical (31)

- (29) a. ...solicitando una ambulancia urgente para un hombre de 35 años que estaba sufriendo un *proceso* alérgico tras comer en un restaurante. (CORPES XXI, España: Un hombre fallece... *El Mundo*, 2012/12/08.)
 b. En 1949 Córdoba atravesaba un *proceso* de cambio social, iniciado antes ya de la guerra civil. (CORPES XXI, España: Castilla del Pino, C. 2004. *Casa del olivo. Autobiografía*... Barcelona: Tusquets.)
 c. No son pocos los que parecen convencidos de que nuestra época vive un *proceso* de degeneración material y moral... (CORPES XXI, España: Arias, M. 2020. *Nostalgia del soberano*. Madrid: Los Libros de la Catarata.)

15. Los nombres ligeros pueden tener variantes con significado pleno que se construyen sin necesidad de modificadores. Así, en (i), *proceso* significa 'desarrollo, evolución'; en (ii), 'juicio, sumario' y en (iii), 'procedimiento, tratamiento':

- i. La técnica de clonación mediante robots [...] inició su *proceso* en 2017... (CORPES XXI, Perú: Científicos chinos consiguen que sus robots clonen cerdos... *La República*, 2022/06/09).
- ii. [S]e le abrió un *proceso* a Hugo Blanco y estuvieron a punto de condenarlo a muerte. (CORPES XXI, Perú: Chávez, G. 2022. *Antes del olvido*. Lima: Alfaguara).
- iii. Ambos *procesos* son realizados generalmente con el empleo de una antena que rodea al paciente. (CORPES XXI, México: Rodríguez, A. Imágenes nítidas. *Revista Ciencia*, 2002/04/01).

- (30) a. *...solicitando una ambulancia urgente para un hombre de 35 años que estaba sufriendo un *proceso* tras comer en un restaurante.
 b. *En 1949 Córdoba atravesaba un *proceso*, iniciado antes ya de la guerra civil.
 c. *No son pocos los que parecen convencidos de que nuestra época vive un *proceso*.
- (31) a. ...solicitando una ambulancia urgente para un hombre de 35 años que estaba sufriendo una *alergia* tras comer en un restaurante.
 b. En 1949 Córdoba atravesaba un *cambio social*, iniciado antes ya de la guerra civil.
 c. No son pocos los que parecen convencidos de que nuestra época vive una *degeneración material y moral*...

La razón de este comportamiento se debe a que el nombre genérico y la palabra llave o su derivado forman un predicado complejo, de manera que el nombre constituye el núcleo sintáctico de la estructura, pero no el semántico, que es desempeñado por la palabra llave o su derivado. Se trata, en suma, del mismo comportamiento que atribuimos a los nombres ligeros (v. § 2.2), de manera que la descripción que acabamos de ofrecer para los términos genéricos sintagmáticos se corresponde con la de los nombres ligeros¹⁶ del tipo *situación peligrosa* y *carácter urgente*, es decir, la de los que pueden acompañarse de complementos adjetivales (Koike, 1998; 2003; Bosque, 2001), que son descritos como nombres ligeros episódicos e individuales en Sanromán Vilas (2023; en prensa). Las diferencias entre unos (los valores de GenerS) y otros (los nombres ligeros) dependen de la perspectiva teórica desde la que se analizan. La TST pone el énfasis en la naturaleza léxica de la combinación, mientras que la Gramática Generativa (GG) lo hace en sus propiedades sintácticas. Otra diferencia de enfoque se relaciona con la direccionalidad de la selección entre el nombre y su complemento (Bosque, 2004). En la TST se habla de selección léxica orientada desde la palabra llave (o base de la colocación) al término genérico (o colocativo); en cambio, en la GG, se parte de una selección semántica regular entre un predicado y sus argumentos, de manera que iría desde el término genérico hasta su complemento.

16. Dejamos fuera del grupo de los nombres ligeros, aquellos que solo se combinan con sintagmas preposicionales como modificadores, pero no con adjetivos. Se trata de nombres que aportan sentidos de medida, agregativos, aproximativos, etc. (Véanse Simone y Masini, 2009; 2014; Bosque, 2016; Fábregas, 2018, para los diferentes tipos). Conviene hacer notar que existen al menos dos FL específicas dentro de la TST que dan cuenta de los nombres de medida o de partes y nombres agregativos o de conjunto (Barrios Rodríguez, 2022). Se trata de Sing 'porción de' y Mult 'conjunto de'; por ejemplo, Sing(*chocolate*) = *tableta, onza*; Sing(*naranja*) = *gajo*; Mult(*abeja*) = *enjambre*; Mult(*uva*) = *racimo*.

En definitiva, en sus líneas fundamentales, puede decirse que la descripción de los términos genéricos sintagmáticos se solapa con la de los nombres ligeros. En uno y otro caso, el nombre —llamémosle término genérico o nombre ligero— más su complemento —la palabra llave o su derivado— equivalen a un nombre pleno, p. ej., *situación peligrosa* = *peligro* y *carácter urgente* = *urgencia*. Conviene hacer notar, además, que algunos de estos nombres gozan de cierta productividad, pues es posible rastrearlos con alguna regularidad en la lengua de las definiciones lexicográficas. El caso de *situación* es especialmente relevante porque lo encontramos en las entradas de algunos diccionarios —véanse en (32) algunos ejemplos extraídos del *DLE*— y aparece clasificado en Barrios Rodríguez (2010, p. 98) como una etiqueta terminológica que agrupa otros nombres, p. ej., *crisis*, *dificultad*, *éxito*, *peligro*, *suerte*, *ventaja*, etc. Además de *situación*, otro nombre que se repite con frecuencia en las definiciones es *cualidad* —(33) ofrece ejemplos tomados del *DLE*—.

- (32) a. adversidad. 3. f. *Situación* desgraciada en que se encuentra alguien.
 b. conflicto. 3. m. Apuro, *situación* desgraciada y de difícil salida.
 c. inferioridad. 2. f. *Situación* de algo que está más bajo que otra cosa o debajo de ella.
 d. paro. 4. m. *Situación* de quien se encuentra privado de trabajo.
- (33) a. accesibilidad. 1. f. *Cualidad* de accesible.
 b. blancura. 1. f. *Cualidad* de blanco.
 c. locuacidad. 1. f. *Cualidad* de locuaz.
 d. mediocridad. 1. f. *Cualidad* de mediocre.

Para dar cuenta de ejemplos como los de (33) sería necesario añadir una nueva construcción a las ya propuestas en la TST para representar los valores de GenerS, ‘GenerS_{ATRIB}(L)–ATRIB→DER(L)’ = ‘L’ y ‘GenerS_I(L)–I→L’ = ‘L’. Para estos casos, propondríamos ‘GenerS_I(L)–I→A_I(L)’ = ‘L’, que se leería como «el valor genérico sintagmático de un lexema L (*cualidad*) equivale semánticamente a L (*blancura*) si este valor va acompañado de un complemento introducido por *de* al que sigue un adjetivo derivado de L (*blanco*) que se predica del primer actante semántico de L», en otras palabras, «*blancura* [de X] = *cualidad de X de (ser) blanco*».

Por último, nos quedarían por considerar los nombres encapsuladores catafóricos y su relación con los términos genéricos sintagmáticos (y con los nombres ligeros). En principio, los encapsuladores catafóricos se diferencian de los anafóricos no solo por la posición en la que aparecen sus referentes —después o antes del nombre, respectivamente—, sino también por la forma como se vinculan con esos referentes.

Así, frente a los anafóricos, que mantienen un vínculo extraoracional con sus referentes, los encapsuladores catafóricos lo tienen intraoracional, pues el nombre y el referente forman parte de la misma estructura sintáctica. En particular, el encapsulador catafórico es el núcleo sintáctico de su sintagma nominal y el referente aparece en forma de oración subordinada completiva. En este sentido, se acercan a los términos genéricos sintagmáticos (y a los nombres ligeros), que también se construyen con un complemento obligatorio.

Desde el punto de vista semántico, los encapsuladores catafóricos están más restringidos que los anafóricos, que suelen considerarse una clase abierta. En particular, los encapsuladores catafóricos suelen ser nombres abstractos, de sentido amplio, que se completan con el de su referente, una entidad de segundo o, preferentemente, de tercer orden, que debe aparecer expresa. De acuerdo con Schmid (2002, pp. 85-86), los más prototípicos, o primarios, pertenecen a la clase de los nombres factuales (*hecho, razón*), de los mentales (*idea, concepto*) y de los lingüísticos (*mensaje, rumor*), siendo fácilmente reconocibles porque no pueden descomponerse morfológicamente en unidades más simples. Un poco más alejados del prototipo, pero todavía buenos ejemplares de la categoría, serían los nombres derivados de adjetivos y verbos, pertenecientes también a la clase de los nombres lingüísticos (*orden, supuesto*), mentales (*creencia, sentimiento*) o modales epistémicos o deónticos (*probabilidad, permiso*). Véase un ejemplo de *sentimiento* como encapsulador catafórico en (34a) frente a otro como nombre genérico (o nombre ligero) en (35a). Por último, los nombres eventivos como *situación* o los circunstanciales de tiempo como *momento* o *circunstancias*, ejemplos típicos de nombres ligeros, son, en cambio, más periféricos dentro de la categoría de los encapsuladores.

Desde una perspectiva sintáctica, los encapsuladores catafóricos, al igual que sucede con los nombres genéricos sintagmáticos —y, en suma, con los nombres ligeros—, exigen la presencia de su modificador —obsérvese la extrañeza o falta de completitud de (34b) y (35b)—. No obstante, ellos mismos parecen poder eliminarse en bastantes casos (34c), rasgo que también los acerca a los nombres genéricos sintagmáticos (nombres ligeros), que, como vimos, pueden eliminarse siempre (35c).

- (34) a. El divulgador debe ser capaz de comunicar el *sentimiento* de que solo la ciencia permite conocer la verdad... (CORPES XXI, España: FECYT, 2011. *10 años de divulgación científica*... Madrid: MIC.)
 b. #El divulgador debe ser capaz de comunicar el *sentimiento*.

- c. El divulgador debe ser capaz de comunicar [Ø] que solo la ciencia permite conocer la verdad.
- (35) a. Las momias históricas [...] ya no inspiran un *sentimiento* piadoso... (CORPES XXI, Chile: Vari, E., Los actos de libre disposición..., *actabioethica.uchile.cl*, 2019.)
 b. #Las momias históricas ya no inspiran un *sentimiento*.
 c. Las momias históricas ya no inspiran *piedad*.

En cuanto a las restricciones léxicas, por tratarse de la relación entre un nombre y un modificador oracional y no entre dos unidades léxicas, estas parecen quedar suspendidas. Sin embargo, es posible que la razón de que en algunos casos el nombre encapsulador no pueda eliminarse haya que buscarla en la relación léxica que el nombre mantiene con el verbo regente, como sucede en los siguientes ejemplos en los que *sentimiento* forma una colocación con *tener* (36), con *provocar* (37) y con *forjar* (38).

- (36) a. Otra vez *tuvo el sentimiento* de que Cornelius sabía que le había ocultado las llamadas.
 b. *Otra vez tuvo (el) [Ø] que Cornelius sabía que le había ocultado las llamadas.
- (37) a. Estar en un lugar extraño [...] *provoca un sentimiento* de que se puede manejar la situación.
 b. #Estar en un lugar extraño [...] *provoca* (el) [Ø] que se pueda manejar la situación.
 c. #Estar en un lugar extraño [...] *provoca* (el) [Ø] poder manejar la situación.
- (38) a. ...*se fue forjando el sentimiento* de que los distintos talleres deberían durar más tiempo...
 b. #...*se fue forjando* (el) [Ø] que los distintos talleres deberían durar más tiempo...

Se observa que llevan determinante porque actualizan un referente y tanto el nombre como la oración completiva aluden a la misma entidad.

4. DISCUSIÓN DE LA PROPUESTA Y CONCLUSIONES

En la clasificación estándar de las FL de la TST, Gener se describe como una FL cuyos valores son términos genéricos en relación paradigmática, o de sustitución, con la palabra llave. Algunas revisiones críticas posteriores rechazan la división entre FL paradigmáticas y sintagmáticas, bien porque responden a criterios de naturaleza diferente, bien porque es difícil mantener dicho contraste con nitidez través de los diferentes valores (Grimes, 1990, citado en Alonso-Ramos y Tutin, 1996). En particular, Alonso-Ramos y Tutin (1996, p. 151) proponen la separación

de dos FL Gener, una para dar cuenta de los valores que sustituyen a la palabra llave y otra para cubrir los valores que se combinan con la palabra llave.

En este estudio hemos reexaminado la FL Gener revisando la bibliografía existente y analizando sus valores con datos extraídos de CORPES XXI. En una línea similar a la de los autores mencionados, consideramos necesario separar los valores de Gener en dos grupos: uno para los términos genéricos paradigmáticos (o valores de GenerP), que cubre nombres del tipo *asiento*, *flor* o *emoción* y otro para los términos genéricos sintagmáticos (o valores de GenerS) como *sentimiento amoroso* o *proceso de regeneración*. Los del primer tipo presentan una equivalencia con los hiperónimos, dado que tanto unos como otros son unidades léxicas con significado pleno en relación paradigmática con la palabra llave con la que establecen vínculos de cuasisinonimia. El análisis de los términos genéricos paradigmáticos y de los hiperónimos se lleva a cabo desde una perspectiva puramente semántica, de manera que el sentido de la palabra llave, su hipónimo, incluye el del término genérico o hiperónimo. En ambos casos, la unidad léxica puede aparecer en la construcción L , L_1 , L_2 [...], $L_{(N)}$ y otros tipos de GenerP(L)', por ejemplo, *amor*, *miedo*, *enfado* [...], *sorpresa* y otros tipos de emociones.

Con respecto al segundo tipo, valores como *sentimiento amoroso* o *proceso de regeneración* muestran que el término genérico se combina con la palabra llave; de ahí que, en estos casos, Gener haya de considerarse una FL de tipo sintagmático. Los valores de GenerS son unidades semiléxicas. Funcionan como núcleo del sintagma en el que aparecen, pero no como núcleo semántico. Prueba de ello es que el término genérico puede suprimirse de la oración, pero no la palabra llave que funciona como complemento. Los valores de GenerS no equivalen a hiperónimos, sino a nombres ligeros, siendo la principal diferencia entre ellos el nivel lingüístico en el que ponen el énfasis: el léxico, en el caso de los valores de GenerS, y el sintáctico, en el de los nombres ligeros. En ambos casos, sin embargo, la unidad léxica puede aparecer en construcciones como 'GenerS_{ATRIB}(L)-ATRIB→DER(L)' = 'L' (*sentimiento amoroso* = *amor*) o 'GenerS₁(L)-I→L' = 'L' (*proceso de regeneración* = *regeneración*). A estas dos construcciones, hemos añadido una tercera, 'GenerS₁(L)-I→A₁(L)' = 'L', para dar cuenta de casos como *cualidad de inteligente* = *inteligencia*.

A lo largo del estudio, se ha tratado de justificar que la FL estándar Gener comprende en realidad dos FL diferentes, una paradigmática (GenerP) y otra de carácter sintagmático (GenerS). Las estructuras en las que pueden aparecer, así como las búsquedas realizadas en CORPES XXI, sirven como prueba de ello. Así mismo, se ha

mostrado que existe una equivalencia entre los valores de GenerP y los hipéronimos, de un lado, y entre los valores de GenerS y los nombres ligeros, de otro.

En lo que respecta a los nombres encapsuladores, que se han descrito como una función discursiva y no como una clase de nombres propiamente dicha, los vínculos con los valores de GenerP y GenerS no son tan directos. Así, vimos que los términos genéricos, tanto paradigmáticos como sintagmáticos, tienen un significado estable, de manera que *rosa*, *silla* y *amor* incluyen siempre *flor*, *asiento* y *emoción*, respectivamente; en cambio, los encapsuladores adoptan un sentido particular que puede variar de un texto a otro; por ejemplo, *la esperanza de vida al nacer es mayor para las mujeres* (27) puede traducirse en *ventaja biológica* en un texto, en *constatación* en otro y aun en *desigualdad* en un tercero. Se ha constatado también que los valores de GenerP, al igual que los encapsuladores anafóricos, pueden cumplir funciones discursivas, fenómeno que se ha denominado hiperonimia discursiva. No obstante, frente a los encapsuladores anafóricos, que llevan como referente entidades de segundo o tercer orden y actualizan un fragmento previo del texto condensándolo en una sola palabra, los valores de GenerP se refieren únicamente a entidades de primer orden y contienen un sentido genérico que incluye el sentido del referente.

Paralelamente, existe un segundo tipo de nombres encapsuladores, de naturaleza catafórica, que se puede relacionar con los valores de GenerS. En ambos casos, los nombres, que denotan conceptos de carácter general, van acompañados de un modificador que no puede eliminarse —una oración, en el caso de los encapsuladores, y un sintagma o un adjetivo, en el de los valores de GenerS—. Si bien las semejanzas entre ellos son considerables, el hecho de que el modificador esté representado por unidades diferentes y que no sea fácil encontrar nombres que puedan funcionar a la vez como encapsuladores catafóricos y como valores de GenerS dificulta el contraste, que deberá ser desarrollado con mayor amplitud en trabajos posteriores.

La comparación de los términos genéricos de la TST con los nombres ligeros, por un lado, y con los hiperónimos, por otro, nos permite comprobar la distancia a la que se encuentran unos de otros y justificar la separación de los valores de Gener en dos grupos. A la inversa, sería también útil contrastar el concepto de término genérico de la TST con el de nombre ligero usado en la bibliografía general porque podría arrojar alguna luz sobre las diferencias existentes entre términos del tipo *onza* (*de chocolate*), *rebaño* (*de ovejas*), *golpe* (*de suerte*), por un lado, y los nombres ligeros considerados en este estudio. Como se ha apuntado en la nota 16, este tipo de nombres se analizan en la TST como valores de otras FL —Sing,

Mult y Figur, respectivamente—. Por último, queda abierto un amplio campo de investigación para precisar con más detalle la relación de los encapsuladores con los nombres ligeros.

REFERENCIAS

- Alonso-Ramos, M., y Tutin, A. (1993). Les fonctions lexicales du *Dictionnaire explicatif et combinatoire* pour l'étude de la cohésion lexicale. *Linguisticae Investigationes*, 17(1), 161-188. DOI: <https://doi.org/10.1075/li.17.1.07alo>
- Alonso-Ramos, M. y Tutin, A. (1996). A Classification and Description of Lexical Functions for the Analysis of their Combinations. En L. Wanner (Ed.), *Lexical Functions in Lexicography and Natural Language Processing* (pp. 147-167). Amsterdam: John Benjamins.
- Barrios Rodríguez, M.^a A. (2010). *El dominio de las funciones léxicas en el marco de la teoría sentido-texto*. *Estudios de Lingüística del Español (ELiEs)*, 30. DOI: <https://raco.cat/index.php/Elies/article/view/195745>
- Barrios Rodríguez, M.^a A. (2022). *Las colocaciones nominales: una aproximación desde la teoría sentido-texto*. Santiago de Compostela: Universidade de Santiago de Compostela. DOI: <https://doi.org/10.15304/9788419155658>
- Borreguero Zuloaga, M. (2018). Los encapsuladores anafóricos: una propuesta de clasificación. *Caplletra*, 64, 179-203. <https://doi.org/10.7203/caplletra.64.11380>
- Bosque, I. (2001). On the weight of light predicates. En J. Herschensohn, E. Mallén y K. Zagona (Eds.), *Features and Interfaces in Romance* (pp. 23-38). Amsterdam: John Benjamins.
- Bosque, I. (2004). La direccionalidad en los diccionarios combinatorios y el problema de la selección léxica. En T. Cabré (Ed.), *Lingüística teórica: anàlisi i perspectives* (pp. 13-58). Bellaterra: Universidad Autónoma de Barcelona.
- Bosque, I. (2016). Sustantivo. En J. Gutiérrez-Rexach (Ed.), *Enciclopedia de lingüística hispánica* (vol. 2, pp. 111-122). Londres: Routledge. DOI: <https://doi.org/10.4324/9781315713441>
- Butt, M. (2003). The light verb jungle, *Harvard Working Papers Linguistics*, 9, pp. 1-9.
- Butt, M. (2010). The light verb jungle: Still hacking away. En M. Amberber, B. Baker y M. Harvey (Eds.), *Complex Predicates: Linguistics Perspectives on Event Structure* (pp. 48-78). Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511712234.004>
- CORPES XXI = RAE. *Corpus del Español del Siglo XXI*. <http://www.rae.es>
- Coseriu, E. (1977). *Principios de semántica estructural*. Madrid: Gredos.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Escandell Vidal, M.^a V. (2007). *Apuntes de semántica léxica*. Madrid: UNED.
- Espinal, M.^a T., y Mateu, J. (2014). Palabras y significado. En M.^a T. Espinal (Coord.), J. Macià, J. Mateu y J. Quer, *Semántica* (pp. 59-110). Madrid: Akal.
- Fábregas, A. (2018). On light nouns and what they tell us about the ingredients of N. En B. Sanromán Vilas (Ed.), *Semi-Lexicality: Studies on Light Verbs, Periphrases and Other Constructions* (pp. 55-81). Helsinki: Société Néophilologique.

- Fábregas, A. (2020). *Las categorías funcionales*. Madrid: Síntesis.
- Goddard, C. (2017). *Furniture, vegetables, weapons: Functional collective superordinates in the English lexicon*. En Z. Ye (Ed.), *The Semantics of Nouns*, (pp. 246-281). Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198736721.003.0010>
- Greimas, A. (1966). *Sémantique structurale. Recherches de méthode*. París: Larousse.
- Halliday, M., y Hasan, R. (1976). *Cohesion in English*. Londres: Routledge.
- Kempson, R. M. (1977). *Semantic Theory*. Cambridge: Cambridge University Press.
- Kleiber, G. (1995). *La semántica de los prototipos. Categoría y sentido léxico*. Madrid: Visor.
- Koike, K. (1998). Algunas observaciones sobre colocaciones sustantivo-verbales. En G. Wotjak (Ed.), *Estudios de fraseología y fraseografía del español actual* (pp. 245-255). Frankfurt/Madrid: Iberoamericana/Vervuet.
DOI: <https://doi.org/10.4994/hispanica1965.1996.14>
- Koike, K. (2003). Sustantivos «ligeros». *Moenia*, 9, 9-20.
- Lee, W., y Evens, M. (1996). Generating Cohesive Text Using Lexical Functions. En L. Wanner (Ed.), *Lexical Functions in Lexicography and Natural Language Processing* (pp. 299-306). Ámsterdam: John Benjamins.
- Leech, G. (1974). *Semantics*. Londres: Penguin Books.
- López Samaniego, A. (2011). *La categorización de entidades del discurso en la escritura profesional*. Tesis Doctoral. Universitat de Barcelona.
<https://www.tdx.cat/handle/10803/48757#page=1>
- López Samaniego, A. (2015). Etiquetas discursivas, hiperónimos y encapsuladores: una propuesta de clasificación de las relaciones de cohesión referencial. *RILCE*, 31(2): 435-462. DOI: <https://doi.org/10.15581/008.31.1551>
- Lyons, J. (1969). *Structural Semantics. An analysis of part of the vocabulary of Plato*. Oxford: Basil Blackwell.
- Lyons, J. (1977). *Semantics*. Vols. 1-2. Cambridge: Cambridge University Press.
- Mel'čuk, I. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. En L. Wanner (Ed.), *Lexical Functions in Lexicography and Natural Language Processing* (pp. 37-102). Ámsterdam: John Benjamins.
- Mel'čuk, I. (1997). *Vers une linguistique Sens-Texte*. París: Collège de France.
- Mel'čuk, I. (1998). Collocations and lexical functions. En A. P. Cowie (Ed.), *Phraseology. Theory, analysis, and applications* (pp. 23-53). Oxford: Clarendon Press.
- Mel'čuk, I. (2015). *Semantics: From meaning to text* (vol. 3). Ámsterdam: John Benjamins. DOI: <https://doi.org/10.1075/slcs.168>
- Mel'čuk, I. (2016). *Language: From meaning to text*. Moscú: Academic Studies Press.
- Mel'čuk, I. (2023). *General phraseology: Theory and practice*. Ámsterdam: John Benjamins.
- Mel'čuk, I., y Polguère, A. (2021). Fonctions lexicales dernier cri. En S. Marengo (Ed.), *La Théorie Sens-Texte. Concepts-clés et applications* (pp. 75-155). París: L'Harmattan.
- Otaola Olano, C. (2004). *Lexicología y semántica léxica*. Madrid: Ediciones Académicas.
- Rodríguez-Espiñeira, M.^a J. (2015). El sustantivo *hecho* como ejemplar de nombre encapsulador factual. En A. I. Álvarez, R. de Andrés Díaz, Á. Arias y F. Fernández de Castro (Eds.), *Studium Grammaticae. Homenaje al profesor José A. Martínez* (pp. 655-674). Oviedo: Universidad de Oviedo.

- Rodríguez-Espiñeira, M.^a J. (2018). Sustantivos con usos argumentativos testimoniales. En I. Bosque, S. Costa y M. Malcuori (Eds.), *Palabras en lluvia minuciosa: Veinte visitas a la gramática del español inspiradas por Ángela di Tullio* (pp. 315-332). Frankfurt/Madrid: Iberoamericana Vervuert.
DOI: <https://doi.org/10.31819/9783954877560-019>
- Sanromán Vilas, B. (2023). *Carácter, naturaleza e índole como nombres ligeros de nivel individual*. *Borealis. An International Journal of Hispanic Linguistics*, 12(2): 223-253. DOI: <https://doi.org/10.7557/1.12.2.7078>
- Sanromán Vilas, B. (en prensa). Nombres ligeros de estado episódico: análisis de *circunstancias, momento y situación*. *Verba*.
- Schmid, H.-J. (1998). Constant and ephemeral hypostatization: *thing, problem* and other «shell nouns». En B. Caron (Ed.), *Proceedings of the 16th International Congress of Linguists* (Paris, 20-25 July 1997, CD-ROM). Ámsterdam: Elsevier.
- Schmid, H.-J. (2000). *English Abstract Nouns as Conceptual Shells: from corpus to cognition*. Berlín: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110808704>
- Schmid, H.-J. (2018). Shell nouns in English. A personal roundup. *Caplletra*, 64, 109-128. DOI: <https://doi.org/10.7203/caplletra.64.11368>
- Simone, R., y Masini, F. (2009). Support nouns and verbal features: a case study from Italian. *Verbum*, 29(1/2), 143-172.
- Simone, R., y Masini, F. (2014). On Light Nouns. En R. Simone y F. Masini (Eds.), *Word classes: Nature, typology and representations* (pp. 51-73). Ámsterdam: John Benjamins. DOI: <https://doi.org/10.1075/cilt.332.04sim>
- Varela Ortega, S. (2008). Las categorías híbridas: comparación entre dos tipos. En M.^a J. Rodríguez Espiñeira y J. Pena (Coords.), *Categorización lingüística y límites intercategoriales*. Vol. 61, *Verba* (pp. 89-98). Santiago de Compostela: Edicións da Universidade de Santiago de Compostela.
- Wierzbicka, A. (1984). «Apples» are not a «kind of fruit»: The semantics of human categorization. *American Ethnologist*, 11(2), 313-328.
- Wierzbicka, A. (1985). *Lexicography and Conceptual Analysis*. Ann Arbor: Karoma.

ESTRUCTURAS SEUDOCOORDINADAS: PERÍFRASIS Y SIGNIFICADOS ASPECTUALES

Laura Malena KORNFIELD¹

Universidad de Buenos Aires

Consejo Nacional de Investigaciones Científicas y Técnicas (Conicet)

Resumen

Este artículo se propone indagar en las propiedades gramaticales y el significado de las estructurasseudocoordinadas: *agarra y les paga; va y se muere*, que se registran en variedades coloquiales de distintos países hispanoparlantes.

Laseudocoordinación da lugar a construcciones gramaticalizadas de dos formas verbales (flexionadas en las mismas categorías morfológicas y unidas por la conjunción *y*) que designan un único evento, dado que el primer verbo de cada secuencia está desemantizado o gramaticalizado. Un dato relevante es que los mismos verbos gramaticalizados que intervienen en estructurasseudocoordinadas (e.g., además de *agarrar e ir, venir, coger, pillar, saltar llegar* y también otros más novedosos: *volver, arrancar, largarse, lanzarse, ponerse, meterse, mandarse*) funcionan (o funcionaron) como auxiliares de perífrasis, a menudo con valor incoativo.

Empíricamente, nos enfocamos en determinar la distribución gramatical de las estructurasseudocoordinadas, diferenciándolas de las perífrasis incoativas, con el fin de ofrecer un análisis sintáctico y semántico de estas construcciones. El recorrido nos permitirá concluir que el significado primario de las estructurasseudocoordinadas gramaticalizadas es el aspecto aoristo (Smith, 1991; Camus Bergareche, 2004), en línea con Coseriu (1966) y en oposición a perspectivas recientes como García Sánchez (2007) y Jaque *et al.* (2018, 2019).

*Palabras clave:*seudocoordinación; gramaticalización; perífrasis; aspecto.

1. laura_malena@yahoo.com.ar;  <https://orcid.org/0000-0002-7998-8781>.

PSEUDO-COORDINATIVE STRUCTURES: PERIPHRAISIS AND ASPECTUAL MEANINGS

Abstract

This paper intends to explore the grammatical properties and meaning of pseudo-coordinative structures, as *agarra y les paga* or *va y se muere*, registered in different colloquial varieties of Spanish speaking countries.

Pseudo-coordination produces grammaticalized constructions of two verbal forms inflected on the same morphological categories and united by the conjunction *y*, which denote a unique event, since the first verb is desemantized or grammaticalized. A relevant fact is that the same grammaticalized verbs intervening on pseudo-coordinative structures (besides *ir* and *agarrar*, *venir*, *coger*, *pillar*, *saltar* and *llegar*, and the more novel *volver*, *arrancar*, *largarse*, *lanzarse*, *ponerse*, *meterse*, *mandarse*) function (or functioned) as periphrastic auxiliary verbs, often with an inchoative or ingressive meaning.

Empirically, the paper concentrates on determine the grammatical distribution of pseudo-coordinative structures, differentiating them to inchoative periphrasis, in order to offer a consistent syntactic and semantic analysis of these constructions. This will lead us to conclude that the primary meaning of grammaticalized pseudo-coordinative structures is aorist aspect (Smith, 1991; Camus Bergareche, 2004), in the same lines as Coseriu (1966) and in opposition to recent perspectives, as García Sánchez (2007) and Jaque *et al.* (2018, 2019).

Keywords: pseudo-coordination; grammaticalization; periphrasis; aspect.

RECIBIDO: 07/09/2023

APROBADO: 01/05/2024

1. INTRODUCCIÓN

Este artículo se propone indagar en las propiedades gramaticales y el significado de estructuras como las de (1), que se registran en variedades coloquiales de distintos países hispanoparlantes:

- (1) a. Esta amiga de mi amiga *va y se le queda* muerto el hombre en la cama (Cabrera Infante, *Tres tristes tigres*, Cuba)
- b. uno *agarra y les paga* con ser ingrato (César Bruto, «Hijo de madre viuda», Argentina)
- c. Para usía es gratis, mi general. Y él *coge y se va* (Mendoza, *La ciudad de los prodigios*, España)

d. El caporal *viene y le dice*: ... aquí está tu grupo (*Me llamo Rigoberta Menchú y así me nació la conciencia*, Guatemala)²

Las estructuras resaltadas en (1) son representativas de lo que puede concebirse como una forma específica de la coordinación falsa o asimétrica, a la que en este trabajo denominaremos «seudocoordinación» (cf. Ross, 2014 para una discusión terminológica). Ese procedimiento da lugar a construcciones gramaticalizadas de dos formas verbales flexionadas unidas por la conjunción *y*, que designan eventos únicos: el primer verbo de cada secuencia (*va*, *agarra*, *coge* y *viene*, respectivamente) está desemantizado y es el segundo verbo el que determina la selección argumental y léxico-aspectual de la oración.

El interés de este fenómeno es múltiple y ha sido abordado desde perspectivas teóricas muy diferentes, según repasaremos en forma sucinta en §2. Desde un punto de vista diacrónico, los verbos gramaticalizados de (1a-c) se registran muy tempranamente³, otros han sido relevados por la bibliografía especializada (cf. Kany, 1945), como sucede con *pillar*, *saltar* y *llegar* (cf. 2a-c), mientras que otras estructuras seudocoordinadas son aún más novedosas, como las de (2d-j):

- (2) a. Yo cuando no quiero molestias *pilló y me voy* al campo (web, España)
 b. Vea la cama duplex, el ventilador de techo, la limpieza que *salta y huele...* (Delgado Senior, *Sub-América*, Venezuela)
 c. Porque que una persona, digamos, *llegue y me diga* una vulgaridad en cierto medio puede faltarme el respeto (CREA oral, Costa Rica)
 d. Y me invitó una amiga a parrandiar/ y *vuelvo y la embarro* (Yeison Jiménez, Colombia)

2. En las secciones 1 a 3 empleamos casi exclusivamente (excepto 29 y 32-33) ejemplos recopilados, para los que indicamos el país de origen. Algunos provienen de fuentes literarias y artísticas (y se registra además el nombre de autor/a); en caso contrario, han sido relevados mediante búsquedas en Google y solo se han adaptado (mínimamente) en su ortografía. En cambio, en las secciones de análisis y discusión (§4-5) predominan los ejemplos contruidos, en particular cuando se trata de paradigmas de agramaticalidad. En este último caso, debe notarse que la variedad de referencia es el español rioplatense, por lo que pueden existir juicios divergentes con otras variedades dialectales.

3. Desde una perspectiva histórica, los primeros datos registrados de seudocoordinación en español parecen corresponder a los verbos de aprehensión: *toma y* +verbo (que actualmente casi no se utiliza) es consignada por primera vez por Juan de Valdés (1535), la RAE reconoce la expresión *coge y* + verbo a partir de la edición de 1780 del *Diccionario* y Cuervo (1872) menciona *agarra y* + verbo para el español americano. De todos modos, esa primacía debería ser tomada con cautela, puesto que las construcciones con verbos de aprehensión siempre han sido más censuradas (desde un punto de vista normativo) que los verbos de movimiento (cf. Ross, 2014). Por esa ausencia de registro es que Pérez Saldanya (2003, p. 85) anota que la estructura con *va y* + verbo está documentada recién en el siglo XIX, pese a que Jaque *et al.* (2019, p. 242) recopilan ejemplos en fuentes históricas del siglo XV.

- e. *Arranca y llueve* y se llenan los pozos, es una hermosura (prensa, Uruguay)
- f. *Me lancé y me animé* a escribir un librito (Twitter, Argentina)
- g. A capella, Albarn *se largó y cantó* el tema acompañado por todos (prensa, Chile)
- h. *Se puso* y prácticamente *la liquidó* tecleando durante dieciocho días (Sasturain, Argentina)
- i. alguien *se mete y escribe* una canción y de pronto se vuelve un hit (web, Colombia)
- j. alguien *se mandó y pasó* como si nada con nitrato de amonio (web, Argentina)

El artículo está organizado de la siguiente manera. En §2 hacemos un sucinto repaso por los datos esenciales y las ideas preliminares aportados por la bibliografía previa. Nos concentramos en §3 en la tarea de determinar las propiedades gramaticales básicas de las estructurasseudocoordinadas. Tomamos como punto de partida los casos más extendidos y gramaticalizados de *va y + verbo* y *agarra y + verbo* (1a-b) (§3.1); luego nos extendemos al resto de las estructuras de (1) y (2) (§3.2), para finalmente establecer los paralelismos con una serie de perífrasis aspectuales que comparten los verbos gramaticalizados (§3.3). En §4 apuntamos a detallar la distribución gramatical y semántica de las estructurasseudocoordinadas (centrándonos nuevamente en *va y + verbo* y *agarra y + verbo*), diferenciándolas de las perífrasis incoativas, en particular por su distinta combinatoria respecto del aspecto léxico o *Aktionsart* del verbo principal. Así, concluimos que las estructurasseudocoordinadas se ligan con el aspecto aoristo o perfectivo (cf. Smith, 1991; Camus Bergareche, 2004), en línea con el estudio clásico de Coseriu (1966). En §5 discutimos las consecuencias de ese análisis y sus diferencias con otras perspectivas, en particular la que se propone en Jaque *et al.* (2018, 2019).

2. SEUDOCOORDINACIÓN:

ANÁLISIS PREVIOS Y VARIACIÓN INTERLINGÜÍSTICA

Estructurasseudocoordinadas⁴ semejantes a las de (1-2) se registraban ya en lenguas clásicas (latín, griego y hebreo) y actualmente tienen una notable

4. La propia etiqueta de «seudocoordinación», que surge en los '70 en la bibliografía al describir los fenómenos del sueco y luego el inglés (cf. Teleman, 1974; Quirk y otros, 1985, *apud* Ross, 2014), es susceptible de debate, como ya anticipamos. En la bibliografía se encuentra una serie muy variada de términos y conceptos teóricos para describir este fenómeno, incluyendo «coordinación falsa y pleonástica», «coordinación asimétrica», «hendíadis verbal», «serialización verbal» (o «casi-serIALIZACIÓN»), «subcoordinación», «concordancia verbo a verbo», «doble verbo», «construcciones verbales paratácticas», «construcciones multiverbales (de formas finitas)», entre otras.

difusión en distintas familias indoeuropeas (bálticas, eslavas, germánicas, romances, más griego, albanés, persa, armenio), pero también en lenguas semitas (hebreo, árabe, maltés), ugrofinesas (finlandés, ingrio, vótico, mordvínico, cheremiso, sami, húngaro) y otras, como el vasco, el turco y las túrquicas en general (e.g., tártaro, yakut o kipchak) (cf. Ross, 2014)⁵. En la mayoría de esas lenguas se documentan construcciones que incluyen verbos con significado léxico de movimiento (e.g., *ir* o *venir*) o de aprehensión (e.g., *tomar*, *agarrar* o *coger*)⁶, en paralelo con nuestros datos iniciales del español.

Este registro coincidente en lenguas tan dispares es evidencia del estrecho vínculo entre las variantes de la coordinación asimétrica y la oralidad. El empleo de construcciones paratáticas como las estructurasseudocoordinadas es un recurso fundamental de las narraciones orales y, de hecho, es plausible considerarlas un caso de universales vernáculos en el sentido fuerte de Chambers (2004), esto es, «rasgos lingüísticos que se pueden encontrar más o menos universalmente en diferentes lenguas y en todo tipo de variedades» no estándares, incluyendo los dialectos rurales, el habla de las clases populares, las lenguas en contacto, las interlenguas, etc. (cf. Di Tullio y Pato, 2022: 9; Kornfeld, 2022). Así, se reconocen miles de casos de estructuras paratáticas ocasionales en las traducciones de la Biblia y otras obras antiguas y medievales (cf. Ross, 2014), al igual que sucede en recopilaciones actuales que pretenden documentar fielmente las narraciones orales y populares. El empleo de un primer verbo con algún grado de desemantización funciona como una transición retórica que asegura la continuidad de la narración para hablantes y oyentes y la vuelve más vívida o dinámica, al brindar un marco o una escena (cf. De Vos, 2006) a las acciones principales, que son introducidas por el verbo antecedido por la *y*. La primera forma verbal en esas estructuras ocasionales puede (o no) sufrir a posteriori un proceso de reanálisis, entendido generalmente como una gramaticalización (cf. *infra*).

El análisis gramatical de laseudocoordinación dependerá entonces, esencialmente, del marco teórico que se adopte y de la lengua bajo estudio. Así, por ejemplo, para las lenguas germánicas se ha planteado que laseudocoordinación es un caso particular

5. Debe señalarse que, pese a esta llamativa extensión, en (casi) todas las familias aparecen excepciones (poco explicables) de lenguas que carecen deseudocoordinación, como es el caso del francés entre las lenguas románicas (aunque el fenómeno sí existía en francés antiguo, como en la frase *prendre et partir* documentada desde *circa* 1170 (Tobler y Lommetzsch, 1969, en *Altfranzösisches Wörterbuch*), del islandés entre las escandinavas y del alemán (salvo algunos dialectos) entre las germánicas (cf. Ross, 2014).

6. Sin embargo, aparecen llamativos contrastes: así, el inglés admite que intervengan en las estructurasseudocoordinadas verbos de movimiento agentivos (*run* 'correr'), verbos de control (*try* 'tratar') o de postura (*sit* 'sentarse'), que en español no están documentados.

de conformación de un núcleo o predicado complejo (cf. De Vos, 2005 para el inglés y el afrikaans; Wiklund, 2008 para el sueco). Concretamente, De Vos (2005, p. 88) plantea que estructuras como *went and read* surgen de un procedimiento de formación de núcleos complejos por ensamble que se produce en el nivel del V° en la sintaxis, dado que son productivas y transparentes (y, por lo tanto, no están almacenadas en el léxico). Este análisis se ve favorecido por el hecho de que en esas lenguas existe un requerimiento de adyacencia estricta entre las dos formas verbales y la conjunción «coordinante» (inglés *and*, afrikaans *en* y sueco *och*).

Cardinaletti y Giusti (2001) utilizan los datos de estructurasseudocoordinadas del inglés americano (cf. 3) y el sueco (cf. 4) para compararlas con las de distintos dialectos del sur italiano (como el de Marsala, cf. (5)). Destacan el hecho de que en las tres lenguas se registran pares de oraciones virtualmente sinónimas donde aparecen verbos de movimiento (semi)gramaticalizados que participan de construcciones con preposiciones e infinitivos (cf. «construcción de infinitivo» en 3-5a) y de estructuras paralelas en las que se utiliza otro elemento conectivo, en lugar de la preposición, y ambos verbos están flexionados para las mismas categorías (cf. «construcción flexionada» en 3-5b):

- (3) a. *I go to buy bread* («construcción de infinitivo»)

yo voy a comprar-INF pan

b. *I go and buy bread* («construcción flexionada»)

yo voy y compro-PRES pan

'Voy a comprar el pan.'
- (4) a. *Jag går för att göra mig en grogg*

yo voy por a hacer-INF me un grogg

b. *Jag går och gör mig en grogg*

yo voy y hacer-PRES me un grogg

'Me voy a hacer un grogg.' (Cardinaletti y Giusti, 2001, p. 373)
- (5) a. *Vaju a pigghiari u pani*

voy-1S a traer el pan

b. *Vaju a piggiu u pani*

voy-1S y traigo-1S el pan

'Voy a traer el pan.'

El «coordinante» de (3-5b), al igual que las preposiciones de (3-5a), no introduce una cláusula nueva, sino que es esencialmente subordinativo (Cardinaletti y Giusti, 2001, pp. 393-394). Las autoras sostienen que en los dialectos italianos los verbos

de movimiento como *ir* o *venir* son elementos léxicos que tienen la posibilidad de ensamblarse en forma directa en un núcleo funcional bajo (i.e., «inmediatamente arriba» de la posición de materialización del verbo léxico), dando paso, así, a un significado gramatical que se asocia con ciertas propiedades gramaticales que los acercan a los auxiliares de perífrasis⁷, pero que conserva una serie de matices léxicos. La estructura que proponen para el dialecto de Marsala, que toma en cuenta la cartografía universal de Cinque (1999), es la de (6a). Esa configuración se basa en datos como (6b) donde coaparecen sucesivamente la «construcción flexionada» (i.e., seudocoordinada) del verbo de movimiento, el causativo y el «andativo» (i.e., la «construcción de infinitivo» en (3-5a)):

- (6) a. ...>verbo de movimiento >... causativo>... andativo>... V
 b. *U va(ju) a fazzu iri a pigghiari a Paola* (Cardinaletti y Giusti, 2001, p. 401)
 lo-CL ir-1s y hacer-1s ir-INF a traer-INF a Paola
 ‘Voy y hago a Paola ir a buscarlo’ [original: ‘I go to make Paola go to fetch it.’]

En cambio, para Cardinaletti y Giusti los verbos de movimiento en inglés y sueco se ensamblan en una posición más baja que la de (6a), dentro del dominio léxico del verbo, ya que tienen más propiedades léxicas y menos funcionales que los dialectos del italiano (cf. nota 7); de allí que en esas lenguas el análisis predominante sea en términos de núcleos complejos (cf. De Vos, 2006; Wilklund, 2008).

Al contrario de lo que ocurre en las lenguas germánicas, en español sí es posible que palabras o sintagmas intervengan entre las formas verbales y la conjunción de las estructuras seudocoordinadas, lo que permite descartar un análisis en términos de núcleos complejos. Ya hemos visto varios ejemplos con pronombres clíticos para la segunda forma verbal (cf. 1 y 7) (a diferencia del ascenso obligatorio en los dialectos del italiano, cf. nota 7), y también pueden intervenir adverbios (cf. 7a) e incluso el sujeto (7b):

7. Dentro de las propiedades que acercan las estructuras seudocoordinadas a las perífrasis en el dialecto de Marsala, Cardinaletti y Giusti mencionan que los verbos de movimiento son una clase cerrada, que tienen un orden fijo (ya que siempre deben preceder a los otros verbos, igual que en español: **leyó y fue*), que carecen de argumentos y adjuntos propios, que solo aparecen con las formas verbales menos marcadas del paradigma (i.e., presente, pretérito perfecto simple e imperativo), que obligan al ascenso de clíticos y que dan lugar a una estructura monoclausal (ya que ambas formas verbales tienen las mismas propiedades morfológicas de P+N y T+A+M y la interpretación es de único evento), entre otras. Como contraparte, los verbos de movimiento presentan bastante transparencia semántica y pueden tener restricciones léxicas ligadas con la animacidad o la agentividad.

- (7) a. *Se puso y prácticamente la liquidó* tecleando durante dieciocho días (Sasturain, «Sobre el arte de la espera», Argentina)
 b. *Va la mina y te denuncia* (web, Argentina)

Esta transparencia sintáctica sugiere que es más adecuado considerar a las estructurasseudocoordinadas como una variante de las perífrasis, que es, de hecho, el análisis predominante en la tradición gramatical española. Distintos autores han sugerido un reanálisis del primer verbo, en paralelo con el proceso de auxiliarización, si bien no hay acuerdo sobre la naturaleza semántica de las construccionesseudocoordinadas. Así, Kany (1945, p. 240) puntualiza que en *va y + verbo* «el verbo *ir* (sirve) meramente para expresar con mayor vigor el significado del verbo al que sirve de auxiliar. Keniston llama a *ir y* ‘auxiliar de aspecto unitario’; como ocurre a veces con *ir a*, pone de relieve el carácter unitario de la acción». De *coge y + verbo*, asegura que tiene una interpretación semejante, pero que indica una «mayor determinación», al igual que sus variantes hispanoamericanas *toma y + verbo* o *agarra y + verbo*.

Por su parte, Coseriu (1966) se centra en el caso de construcciones con un verbo gramaticalizado cuyo significado léxico es de aprehensión, como *tomo y me voy* (que, paradójicamente, casi ha dejado de utilizarse en el mundo hispano). Considera que las estructurasseudocoordinadas en las distintas lenguas indoeuropeas son perífrasis paratáticas y, desde un punto de vista semántico, rescata las intuiciones de Keniston, quien les atribuía el significado de «unidad e indivisión de la acción», ya que sostiene que los dos verbos expresan centralmente «la globalidad de la acción significada por el segundo verbo, es decir que implican tanto el punto inicial como el punto final» (Coseriu, 1966, p. 128), aunque pueden reflejar también otros valores o matices semánticos (e.g., intensidad, determinación, incoatividad, terminatividad, etc.) en función de la interacción con el significado del verbo principal (en §5 retomaremos y ampliaremos el análisis de Coseriu).

En una revisión relativamente reciente del clásico artículo de Coseriu, García Sánchez (2007) observa que hay una evidente filiación entre las estructurasseudocoordinadas (e.g., *tomo y escribo* o *va y se lava las manos*) y las que ostentan un significado léxico para el primer verbo (e.g., *tomo un lápiz y escribo*, *va al lavabo y se lava las manos*); sugiere que es precisamente la elipsis de los complementos la que «despoja al verbo de su referencia concreta y éste termina gramaticalizándose con el valor aspectual ingresivo que le es propio» (García Sánchez, 2007, p. 171-2). Plantea, entonces, que la interpretación primaria de las estructurasseudocoordinadas es aspectual y, en particular, ingresiva, ya que refuerzan el momento o la fase inicial

de la acción. También destaca el paralelismo entre estas construcciones paratácticas (oseudocoordinadas) y las hipotácticas (o perifrásticas), como en *voy a hacer* o *tomo a hacer* (cf. *infra* §3.3) Atribuye a la influyente perspectiva de Coseriu el hecho de que el análisis ingresivo haya sido tan poco explorado en español, ya que en otras lenguas numerosas investigaciones han asignado ese significado a la construcción paratáctica.

A su vez, Jaque *et al.* (2018) revisan las propiedades gramaticales de *llegar y + V*, a las que definen como «construcciones multiverbales de verbo finito coordinadas» (CMVFC) siguiendo a Aikhenvald (2011), y apuntan a establecer sus diferencias de las perífrasis (con las que han sido homologadas en la tradición española, según ya señalamos) y, simultáneamente, demostrar su carácter monopredicativo. Los autores proponen que las CMVFC expresan aspecto interruptivo, al que caracterizan como el que aparece cuando «la ocurrencia de un evento *e* corta de manera abrupta» el «desarrollo normal de los acontecimientos... y da origen a otra serie de eventos –expresada verbalmente o supuesta– que difiere de la esperada en la secuencia original» (Jaque *et al.*, 2018, p. 182-3). Además de retomar y expandir las conclusiones de su artículo previo, Jaque *et al.* (2019, p. 240-248) suman una indagación acerca de la posibilidad de que las CMVFC con significado interruptivo deriven en estructuras con una interpretación mirativa (cf. §5 para más detalles sobre ambos artículos).

3. CARACTERIZACIÓN GRAMATICAL

En esta sección exponemos una serie de propiedades gramaticales básicas de las estructurasseudocoordinadas en diferentes variedades del español. Para ello, nos concentramos primero en las construcciones con *ir* y *agarrar* (cf. 1a-b), las más extendidas en español, que, además, servirán de modelo para los casos menos difundidos de verbos gramaticalizados (cf. el resto de los ejemplos iniciales en 1-2), que analizamos en §3.2. Por último, en §3.3 presentamos una serie de perífrasis que toman como auxiliares a los mismos verbos gramaticalizados que se emplean en las estructurasseudocoordinadas.

3.1. Estructurasseudocoordinadas con *va* y *agarra*

Dos verbos particularmente representativos que participan de estructurasseudocoordinadas en español son el verbo inacusativo de movimiento *ir* y el verbo transitivo de aprehensión *agarrar*.

- (8) a. Y el beduino del desierto *va y le explica* que todavía tiene pendientes ciertas vueltas (Ramírez, *Un baile de máscaras*, Nicaragua)
 b. cuando un niño *agarra y rocía* con bencina a un profesor tiene que requerir un grado de mano dura (*El líbero*, Chile)

En (8) *va* y *agarra* están desemantizados (puesto que no se describe ningún evento de movimiento en (8a) ni un evento de aprehensión en (8b)) y no seleccionan ni argumentos ni adjuntos. Por lo tanto, la asignación de papeles temáticos estará a cargo solo del segundo verbo flexionado, que funciona como verbo principal de la oración: en cada oración de (8) hay un único evento en juego y el análisis de la oración es monoclausal. Estas propiedades son las que han permitido asimilar las estructurasseudocoordinadas con las perífrasis⁸, como ha ocurrido en la tradición gramatical española (cf. Keniston, 1936; Kany, 1945; Coseriu, 1966, entre los autores mencionados en §2). Sin embargo, al tratarse de dos formas verbales finitas, laseudocoordinación requiere que haya coincidencia en las categorías morfológicas de tiempo, aspecto y modo y también de persona y número (cf. Anderson, 2006; Cardinaletti y Giusti, 2001)⁹.

Los ejemplos de (8) contrastan, entonces, con construcciones superficialmente idénticas en que *ir* y *agarrar* conservan su significado léxico, tienen argumentos de alguna clase y entran en una verdadera relación de coordinación con otro verbo pleno. De hecho, la distinción entre una verdadera coordinación

8. Vale destacar que este análisis clásico de la tradición gramatical española se ve bien sustentado desde una perspectiva tipológica amplia, como la de Anderson (2006). Según este autor, tanto las estructurasseudocoordinadas como las perífrasis serían dos tipos de construcciones de verbo auxiliar (CVA) que comparten los rasgos esenciales relevantes: «Una construcción de verbo auxiliar es aquí definida como una estructura mono-clausal que, por lo menos, consiste en un verbo léxico que aporta el contenido léxico y un verbo auxiliar que aporta algún contenido gramatical o funcional a la construcción» (p. 7). Anderson (2006, p. 16) identifica las «formaciones coordinadas» como una fuente de auxiliarización en diversas lenguas. Si bien adherimos plenamente a su concepción, a lo largo del presente artículo empleamos la etiqueta más conservadora de «seudoauxiliar» (que Anderson solo usa en forma ocasional en su libro), para subrayar que no se trata de auxiliares propiamente dichos.

9. Autores como Bravo (2020) observan que la coincidencia en valores morfológicos puede no darse en algunos casos (cf. i.a) y, de hecho, el propio Kany (1945) ya citaba datos semejantes (cf. i.b-c):

- (i) a. ...las Perseidas, que *va* este año y *caen* con una Luna... (Bravo, 2020, p. 171).
 b. Si suelto los perros, *va y me muerden*.
 c. No llevamos los anteojos porque *va y los rompemos* (Gagini, 1919, p. 161, Costa Rica, cit. por Kany, 1945, p. 240)

A lo largo del artículo nos centraremos exclusivamente en los casos (abrumadoramente más frecuentes) en que sí hay coincidencia en persona, número, tiempo, aspecto y modo entre las dos formas verbales y procuraremos dar cuenta de sus propiedades gramaticales y semánticas. Recién nos referiremos brevemente a los casos de (i) en §5.

y una seudocoordinación puede no ser evidente, debido a que, según ya hemos apuntado, las estructuras seudocoordinadas surgen en el discurso oral como construcciones ocasionales que dan un marco a los eventos principales (cf. la noción de «coordinación de escena» de De Vos, 2006). Así, las oraciones de (9) muestran adyacencia entre las formas verbales y la *y*, por lo que parecen involucrar algún grado de desemantización del primer verbo. Sin embargo, advertimos huellas del carácter léxico de *vamos* en (9a) (i.e., el *donde* sugiere una idea locativa), mientras que en (9b) podríamos suponer que el objeto de *agarra* está (meramente) sobreentendido:

- (9) a. El colegio es como una casa donde *vamos y aprendemos* (web, Chile).
 b. *Agarra y empuja*: construye músculo (web, Perú).

En ese sentido, en los datos reales no siempre resulta nítida la frontera entre una coordinación «verdadera», en que *y* efectivamente es una conjunción coordinante, y la seudocoordinación, en que pasa a tener valor subordinativo (cf. Cardinaletti y Giusti, 2001, pp. 393-394).

Asimismo, cuando aparece un clítico pronominal junto con el «seudoauxiliar», como denominamos al primer verbo gramaticalizado en estas construcciones (cf. nota 8), tendremos que asumir que no hay una estructura seudocoordinada. Excepto en el caso del *se* anticausativo incorporado a ciertos pseudoauxiliares (*se puso / metió / mandó / lanzó / largó y + verbo*, cf. §3.2), los clíticos pronominales remiten automáticamente a una interpretación léxica del primer verbo, por lo que la secuencia se entenderá como una verdadera coordinación:

- (10) a. *Se va y se arrepiente*.
 b. *Lo agarré y leí de un tirón*.
 c. *Se agarra y camina con dificultad*.

En ese sentido, a diferencia de las perífrasis (y de las estructuras seudocoordinadas de los dialectos del italiano, cf. nota 7), las estructuras seudocoordinadas del español nunca admiten el ascenso de clíticos desde el dominio del predicado principal de la oración (cf. también Jaque *et al.*, 2018, p. 172).

Debe notarse que en una misma cláusula es posible tener más de dos verbos léxicos coordinados (cf. 11a) y también dos (o más) pseudoauxiliares (cf. 11b)¹⁰:

10. A veces encontramos, con una evidente intención irónica, la reiteración marcada de más pseudoauxiliares diferentes, separados por comas (i.a) o con la *y* también repetida (i.b):

- (11) a. ...*va* y *se preocupa* y *hace* acto de presencia (web, México)
 b. ...*va* y *agarra* y *se pone* a generalizar (web, Venezuela)

Además, se encuentran combinaciones de *va* y *agarra* entre sí, uno como seudoauxiliar y el otro como núcleo del predicado (i.e., con su significado léxico):

- (12) a. Del fondo del barco *va* y *agarra* un orondo sable berberisco (Romero Esteo, *Pasodoble: el barco de papel*, España)
 b. La gente ahora, reflexiona, anda con muchas vueltas, por cualquier cosa *agarra* y *va* al doctor (*El Chubut*, prensa, Argentina)

Señalamos antes que las construcciones seudocoordinadas requieren repetir la información morfológica en el seudoauxiliar y en el verbo léxico. Si bien se las registra mayoritariamente en primera y, sobre todo, tercera persona del singular (como en los diversos ejemplos previos), también aparecen otros valores de persona y número:

- (13) a. Estuve con un muchacho –dijo Teresa–. A lo mejor *vas* y *le pegas*, también (Vargas Llosa, *La ciudad y los perros*, Perú)
 b. De la nada *agarraste* y *te fuiste* (Kafu Banton, «Morena», Panamá)
 c. ¿Y si *agarramos* y *damos* el mejorazo? (Quino, *Mafalda*, Argentina)
 d. Las mujeres atrevidas que *van* y *viven* sin remordimiento y hacen viajes de Ulises... (Rossi, *María la noche*, Costa Rica)
 e. Ahora *vais* y *le pedís* a Zelenski que os aplauda (Twitter, España)

Kany (1945, p. 239) restringía los tiempos de las estructuras seudocoordinadas al presente y al pretérito perfecto (como en los ejemplos anteriores), además del imperativo, que en el caso de *ir* incluye a *ve* (14a) y la forma irregular *anda* (14b). Sin embargo, se las identifica, con menor frecuencia, en otros tiempos y modos: pretérito imperfecto (15a-b), futuro (15c), condicional (15d) y subjuntivo, ya que pueden aparecer en una cláusula subordinada (cf. 15e-f):

- (14) a. *Ve* y *cuéntale* que desde hace tiempo/ Ya no es lo mismo cuando estás con él (Antonio José, «Ve y dile», España)

-
- (i) a. *Va*, *coge*, *agarra*, *pilla* y *dice* que la utilización de cualquier anticonceptivo es pecado contra la Fe (*El País*, España)
 b. *Coge* y *agarra* y *va* y *escribe* que los libres e iguales pretendemos negar el expolio fiscal a Catalunya (*El País*, España)

- b. —*Anda y di* al hombre que en un solo surco siembre todas estas semillas («El mentiroso», *Literatura folklórica*, Bolivia)
- (15) a. Un amigo te decía que te reías y tú *ibas y te reías* (*La voz de Galicia*, España)
 b. Y luego yo *iba y te contaba* todo, sin saberlo, dice AnaCarla besándola en la frente (Adoum, *Ciudad sin ángel*, Ecuador)
 c. De mayor, *agarraré y me iré* a vivir a la selva (web, España)
 d. *Agarraría y me iría corriendo* como gacela, hasta perderme y no pensar (web, México)
 e. nunca más se hizo algún trámite para tener una entidad legal *que agarre y lleve* al ajedrez adelante en Junín (prensa, Argentina)
 f. aparentar lo máximo, que si tú tienes diez la gente *vaya y piense* que tú tienes doscientos (Vergés, *Sólo cenizas hallarás (bolero)*, Rep. Dominicana)

No se encuentran restricciones en la combinatoria léxica de *agarra y* + verbo y de *va y* + verbo respecto de la estructura argumental y temática del verbo principal. Los sujetos pueden ser agentes, como en la mayoría de los ejemplos anteriores, pero también temas o pacientes (cf. 16a-c) y experimentantes (cf. 16d), e incluso se combinan con verbos impersonales (cf. 16e):

- (16) a. ...*va y se me muere* a traición, antes que pudiera cumplir mis sueños (Allende, *La casa de los espíritus*, Chile)
 b. «Zitarrosa *agarró y se murió*» (título de una milonga uruguayá)
 c. Un vikingo *agarró y descubrió* América (sin saber que la estaba descubriendo) (web, España)
 d. Diciéndole a un fulano que es un tío de puta madre, *va y se alegra* (*El Mundo*, España)
 e. Apenas desatan el tiento de la inauguración, *agarra y llueve* (*La Prensa*, Uruguay)

Según se advierte en los datos relevantes de (16), *agarra y* + verbo suele agregar un matiz semántico de voluntad o intención, aun si esos rasgos no forman parte del significado intrínseco del verbo principal.

3.2. Otros pseudoauxiliares

En diversas variedades del español se registran otras estructurasseudocoordinadas distintas de *va y* + verbo o *agarra y* + verbo, como ya hemos anticipado para los ejemplos iniciales de (1-2). Dentro de las que tienen mayor difusión, se cuentan las

construcciones con los verbos de aprehensión *pillar* y *coger*, aunque no aparecen en todas las variedades hispanoamericanas:

- (17) a. Yo cuando no quiero molestias *pilló y me voy* al campo (web, España)
 b. La empresa concesionaria ya *pilló y dijo* que investiguen (prensa, México)
- (18) a. rápido *cogí y escribí* un poema (Kiko Veneno, «El deportista por la ventana», España)
 b. yo *cogí y le pagué* al chófer, que se fue (Cabrera Infante, *Tres tristes tigres*, Cuba)

También tiene gran difusión geográfica la estructuraseudocoordinada con el verbo de movimiento *venir*, que presenta un significado y una distribución sintáctica análogos a los de *va* y + verbo:

- (19) a. Supóngase usted sacerdote y que yo *vengo y le digo*: «Padre, acúsome de que amo» (Gamboa, *Suprema Ley*, México)
 b. Solo Lou *viene y se muere* un sunday morning (web, Argentina).
 c. Nosotros no *vinimos y dijimos*: «Eh, vamos a ser una banda de Internet» (web, España)

Además de los casos anteriores, Kany (1945, p. 242) consigna otras estructurasseudocoordinadas con una distribución más restringida: *llega* y + verbo (que limita a Chile) y *salta* y + verbo¹¹ (que atribuye a España, Chile y Puerto Rico). En la actualidad, ambos verbos están bastante más difundidos en los países americanos:

- (20) a. *Llega y me dice*: ‘eres Carrillo’ (web, Argentina)
 b. Entonces ella *llega y se mete* para adentro y me da café, entonces después cuando ella me da el café, *llego y me salgo* para afuera (CREA oral, Venezuela)
 c. Este año no sé si le alcanzará la cobija, pero si usted *llega y nos invita...* (*El Salvador Hoy*)
- (21) a. ella (Giuliana) *saltó y dijo* ‘pollo a la parrilla’ (prensa, Argentina)
 b. Al comparar el mapa₃ con el mapa₁, uno *salta y piensa*: «pobre ingleses» (web, Colombia)

11. Es conveniente notar que *saltar* se ha relexicalizado previamente como ‘reaccionar con viveza’ en diversas variedades, antes de dar lugar a ejemplos como (21).

Otras construcciones seudocoordinadas están menos gramaticalizadas y más limitadas en su combinación con tiempos verbales, ya que solo suelen aparecer en presente y pretérito perfecto, en las líneas de lo señalado por Kany (1945, p. 239). Se reconocen, así, pseudoauxiliares que provienen de verbos léxicos inacusativos, como *volver* (22) y *arrancar* (23) o diversas variantes anticausativas de verbos agentivos de movimiento: *largarse* (24), *lanzarse* (25), *ponerse* (26), *meterse* (27) y, en áreas geográficas más restringidas, *mandarse* (28), con el mismo significado de ‘meterse’:

- (22) a. *Vuelve y escribe* otra vez porque escribiste un disparate (Twitter, México)
b. Cuando *volví y leí* su autobiografía, me di cuenta de que no menciona a su hija (web, Colombia)
- (23) a. *arrancó y se metió* debajo de un palote (web, México)
b. En ese momento, Casán *ya arrancó y le dijo* «¿Y te disfrazás?» (web, Argentina)
- (24) a. *se larga y estudia* enfermería auxiliar (web, Argentina)
b. Y sobre la base de eso *me largué y aprendí* (web, Argentina)
- (25) a. IKEA *se lanzó y lo comparó* con su rallador (web, España)
b. un estadounidense *se lanza y le pregunta* (web, Colombia)
- (26) a. Ayer cuando llegué a casa *me puse y vi* el tráiler de Swamp Thing (web, España)
b. un día *te pones y creas* una nueva religión (web, España)
- (27) a. *yo soy así, me meto y voy para adelante* (web, Argentina)
b. Cuando uno *se mete y estudia* mucho... (web, Argentina)
- (28) a. *yo me mandé y compré* el mismo pasaje sin pensarlo ni analizar mucho (web, Argentina)
b. Llego al lugar, *me mando y le digo* al flaco de seguridad (web, Argentina)

Los verbos de (26-28) aportan un valor adicional de determinación, voluntad o brusquedad, al que *meterse* y *mandarse* agregan el de atrevimiento. Consecuentemente, tienden a combinarse con verbos agentivos y son anómalos (o fuerzan una lectura agentiva, cf. 29b) con inacusativos:

- (29) a. ??Se *lanzó/ largó/ mandó/ metió/ puso* y se *murió / nació / creció / llovió*.
b. *Se lanzó/ largó/ mandó/ metió/ puso y salió* por colectora.

En suma, vemos que el inventario de estructuras seudocoordinadas registradas en la bibliografía puede ser ampliado considerablemente si se toman en cuenta otras construcciones más novedosas y, por lo tanto, inestables.

3.3. *Perífrasis paralelas*

Un argumento que se esgrime a favor de un análisis ingresivo de las estructurasseudocoordinadas (como en García Sánchez, 2007 y también Kornfeld, 2019, 2022) es la existencia de perífrasis paralelas con los mismos (seudo)auxiliares. La mayoría de esas perífrasis, que ostentan un grado variable de gramaticalización, tienen, precisamente, un significado incoativo o ingresivo (cf. 30), mientras que otras ostentan algún otro valor aspectual (cf. 31):

- (30) a. Hoy *arrancamos a* consolidarnos y eso es importante!!! (Instagram, Argentina)
 b. Usted *salta a escribir* sandeces con tinta de amargura (web, Chile)
 c. *Se largó a ser* precandidato presidencial (web, Uruguay)
 d. La derecha *se lanzó a* hacer carne del tema catalán (Urdániz, declaraciones, España)
 e. Y para aprender, el cerebro se puso a jugar (web, España)
 f. Cuando *te metes a* ordenar la Ciudad, te metes contra muchos intereses creados (Twitter, Panamá)
 g. Un día que venía para el colegio, *se mandó a llover*, muy fuerte (web, Colombia)
- (31) a. Si vuelvo a nacer, elijo ser automovilista (Pernía, declaraciones, Argentina)
 b. «Llegué a desear la muerte de mi padre para que Moisés estuviera a nuestro lado» (Twitter, Chile).
 c. Flor Peña *fue a decir* lo que todos le queríamos decir a Claudia (Twitter, Argentina)
 d. Justo vino a tocarme a mí (Cabral, «Yo no quiero ser un ciudadano», Argentina)

En (31) encontramos perífrasis con valores aspectuales distintos del incoativo. Las que tienen una combinatoria más amplia son *volver a* + infinitivo, con significado iterativo, y *llegar a* + infinitivo, con significado culminativo (cf. Olbertz, 1998). Más restringidas resultan, en cambio, las perífrasis *ir a* + infinitivo¹² o *venir a* + infinitivo cuando llevan el auxiliar en aspecto perfectivo y expresan un valor escalar culminativo, con cierto matiz ligado a la contraexpectativa,

12. En cambio, la perífrasis que expresa aspecto prospectivo y/o futuro inmediato (en algunas regiones peninsulares) o tiempo futuro a secas (particularmente en América) carece de toda restricción argumental: se combina con verbos transitivos (i.a), inacusativos (i.b), inergativos (i.c) e impersonales (i.d):

- (i) a. Te crees que al mundo lo vas a arreglar vos...
 b. No te va a gustar (nombre de una banda uruguaya)
 c. ¡No me diga que va a barrer, Pereyra!
 d. Va a haber croquetas y bicicletas de verdad.

siempre que se combinen con predicados delimitados, es decir realizaciones y logros (cf. RAE-ASALE, 2010, p. 541)¹³.

Por su parte, todos los auxiliares de (30) con significado incoativo podrían ser parafraseados por medio de los más extendidos *comenzar* y *empezar*, aunque presentan más matices semánticos, al encontrarse (relativamente) poco gramaticalizados. El que tiene una combinatoria más amplia (gracias a su relexicalización previa como verbo inacusativo) parece ser *arrancar* (a veces con la variante *arrancarse*), que aparece con distintos tipos argumentales de verbos. Por esa causa, se lo encuentra con verbos con argumento externo (cf. 32a), pero también con verbos no agentivos, sean estados (32b) o inacusativos (32c):

- (32) a. Arrancó a leer / cocinar / gritar / quejarse / pegarle / toser / llorar / llover.
 b. Arrancó a ser paciente / estar tranquila / quererlo mucho.
 c. Arrancó a enfermarse / hundirse / quemarse.

Por el contrario, los auxiliares que son verbos transitivos con un *se* anticausativo suelen preferir verbos agentivos. Por esa causa, son compatibles con verbos que tomen un argumento externo (cf. 33a), mientras que suelen dar lugar a agramaticidades cuando se combinan con estados (cf. 33b) o verbos inacusativos (cf. 33c), en contraste con (32b-c)), excepto que se produzca alguna resignificación agentiva¹⁴:

- (33) (33) a. Se puso/ se metió / se mandó / se lanzó / se largó a leer / cocinar / gritar / quejarse / pegarle / toser / llorar / llover.
 b. ??Se puso/ se metió / se mandó / se lanzó / se largó a ser paciente/ estar tranquila / quererlo mucho.
 c. ??Se puso/ se metió / se mandó/ se lanzó / se largó a enfermarse / hundirse / quemarse.

13. Claramente, esta es una manifiesta simplificación de los complejos problemas que encierra esta perífrasis, cuya relación con la perífrasis de futuro (cf. nota anterior) no es fácil de establecer ni diacrónica ni sincrónicamente (para una discusión cf. Fernández de Castro, 1999; Bravo, 2014; Garachana, 2019, entre otros). Por su parte, Bravo (2014) y Garachana (2019) ofrecen una complejización de los valores que puede adoptar la secuencia, ya que distinguen una variante completiva, que expresa el final de una acción (como sería la de (31c) y también (31d) para *venir*), de una perífrasis focal, con matices modales «de desagrado o de disconformidad» (cf. Garachana, 2019, p. 136), que sí es posible combinar con actividades (e.g., *Fue a llover el día de su boda*).

14. A su vez, *ponerse*, mucho más antiguo (cf. Kany, 1945, p. 243), es también el más gramaticalizado de este grupo de auxiliares y se combina más frecuentemente con verbos no agentivos (cf., por ejemplo, el contraste *Se puso a brillar / repicar / tintinear* vs. **Se metió / mandó / lanzó a brillar / repicar / tintinear*).

Asimismo, resulta interesante destacar que participaban de perífrasis incoativas en distintas variedades del español los verbos de aprehensión *coger* y *agarrar* (cf. 34)¹⁵, según los datos de Kany (1945, pp. 242-243), aun si hoy ha disminuido o desaparecido ese uso como auxiliares:

- (34) a. La familia cogió a llamar médicos y más médicos (Romero García, *Peonía*, Venezuela)
 b. Agarró a caminar con el recaó al hombro (Lynch, *El romance de un gaucho*, Argentina)

4. ASPECTO, PERÍFRASIS INCOATIVAS Y ESTRUCTURAS SEUDOCOORDINADAS

Esta sección está consagrada a llevar a cabo una comparación entre las perífrasis incoativas y las estructurasseudocoordinadas, que, según acabamos de ver, comparten a menudo (seudo)auxiliares. Ese paralelismo podría ser considerado un argumento empírico a favor de la hipótesis de que las estructurasseudocoordinadas tienen significado ingresivo, como hace García Sánchez (2007) y repite Kornfeld (2019, 2022). Sin embargo, aquí esperamos desmentir esa idea a partir de un análisis sistemático de la combinatoria de ambas clases de construcciones gramaticalizadas. Nos concentraremos en los representantes más gramaticalizados de cada conjunto, esto es, las estructurasseudocoordinadas con *va* y *agarra* (cf. §3.1) y las perífrasis incoativas *comenzar a* + infinitivo y *empezar a* + infinitivo.

4.1. Aspecto léxico

Una serie de contrastes gramaticales significativos entre las perífrasis incoativas y las estructurasseudocoordinadas se vinculan con el aspecto léxico o *Aktionsart* (cf. Vendler 1957) de los verbos principales con los que aparecen.

Las perífrasis incoativas son incompatibles con ciertos estados permanentes, según se ilustra en (35a-b), excepto que puedan ser reinterpretados con una naturaleza

15. El mismo paralelismo se advierte entre las estructurasseudocoordinadas con el verbo *tomar*, actualmente desusadas (i.a), y la perífrasis correspondiente *tomarse a* + infinitivo (i.b):

- (i) a. *Tomo y me voy* (Coseriu, 1966)
 b. *tomáronse a quejar* (*Cid*, v. 852) (Kany, 1945, pp. 242-243)

eventiva al focalizarse la primera fase de la situación, como sucede con los estados transitorios de (35c-e)¹⁶:

- (35) a. ?? El gato empezó a tener cuatro patas.
 b. */?? La casa empezó a estar en la esquina de San Martín y Florida.
 c. Comencé a ser alta.
 d. Comenzó a estar mucho más tranquilo.
 e. María empezó a respetarlo.

En contraste, las estructurasseudocoordinadas no se combinan con estados permanentes ni transitorios, con los que dan lugar a oraciones agramaticales o al menos evidentemente anómalas (cf. también Jaque *et al.*, 2018, p. 182):

- (36) a. *El gato agarró y tuvo cuatro patas.
 b. *La casa fue y estuvo en un lugar tranquilo.
 c. *Agarró y fue alto.
 d. ?? Fui/ agarré y estuve más tranquila.
 e. ?? Va/ agarra y lo respeta.

Por su parte, como se observa en las oraciones de (37), las estructurasseudocoordinadas tienen compatibilidad plena con logros¹⁷, a diferencia de las anomalías que exhiben las perífrasis incoativas (cf. 38):

- (37) a. *Agarra y nace / muere.*
 b. *Va y queda pegada la mano.*
 c. *Fue y descubrió el virus.*
 d. *Agarró y encontró la remera.*
 (38) a. ??Comienza a nacer / morir.
 b. ??Empieza a quedar pegada la mano.
 c. ??Comenzó a descubrir el virus.
 d. ??Empezó a encontrar la remera.

16. Sin embargo, como ya hemos observado en §3.3, los auxiliares de perífrasis incoativas gramaticalizados a partir de verbos agentivos (e.g., *ponerse*, *largarse*, *lanzarse* y, sobre todo, *saltar* y *meterse*) manifiestan una incompatibilidad más radical con los estados.

17. Los pseudoauxiliares que tienen como origen verbos inacusativos son plenamente compatibles con logros (*va*, pero también *viene*, *vuelve*, *arranca*), mientras que *agarra* conlleva cierta reinterpretación más agentiva (cf. §3.1 y en particular los ejemplos pertinentes de 16). Por su parte, pseudoauxiliares como *se pone*, *se mete*, *se manda*, *se larga* y *se lanza* (cf. §3.2) son dudosas o, directamente, agramaticales con logros inacusativos (cf. 29a). En §5 retomamos esta diferenciación.

La explicación más natural para las oraciones anómalas de (38) es que los logros son eventos instantáneos, sin duración, por lo que no es posible reconocer etapas o fases, incluida la fase inicial del evento (para esta incompatibilidad entre logros y perífrasis fasales, cf., entre otros, Fábregas, 2019, p. 56). Sin embargo, la combinación puede coercionar la interpretación de los logros, por ejemplo si se les asigna cierta duración y las perífrasis incoativas señalan el inicio de ese evento (i.e., dejan de ser verdaderos logros instantáneos), como en (39). Aun si este es el caso, el significado de las estructurasseudocoordinadas será bastante diferente, ya que focalizan el evento en su totalidad, y no solo el comienzo (cf. 40):

- (39) a. María comienza a enfermarse.
 b. La casa empezó a hundirse/ quemarse.
- (40) a. María *va* / *agarra* y *se enferma*.
 b. La casa *fue* / *agarró* y *se hundió* / *quemó*.

En otras ocasiones, la combinatoria de las perífrasis incoativas con logros puede dar lugar a una interpretación iterativa de los eventos (cf. (41) y, también, alternativamente, (39a)); nótese que, en los mismos contextos, las estructurasseudocoordinadas conservan una lectura puntual y única del evento (cf. 42):

- (41) a. Empezó a patear la pelota/ cerrar la puerta.
 b. Comenzó a llegar / venir / volver a casa temprano.
- (42) a. *Va* y *patea* la pelota/ *cierra* la puerta.
 b. *Agarró* y *llegó* / *vino* / *volvió* a casa temprano.

Una preferencia inversa se verifica con las actividades. Las perífrasis incoativas no solo son siempre posibles sino además muy frecuentes con actividades (cf. 43), mientras que las estructurasseudocoordinadas se registran infrecuentemente con actividades (cf. también Jaque *et al.*, 2018, p. 182), aunque hemos encontrado algunos datos en los corpus (cf. 44):

- (43) a. Comienza / empieza a estudiar / cocinar/ laburar / correr.
 b. Comenzó / empezó a llover/ llorar / toser / sonar / chillar.
- (44) a. El que quiere estudiar, *va* y *estudia*, el que se quiere emborrachar, *va* y se emborracha (Twitter, Guatemala)
 b. Si está con criatura *agarra* y *corre* con el niño (ABC, Paraguay).
 c. El tipo *agarra* y *camina* como un cangrejo (web, Argentina).

- d. Ahora el señorito quiere paz / por qué no *vas y laburás* (Las Manos de Filippi, «El rey de la autocrítica», Argentina)
- e. Apenas desatan el tiento de la inauguración, *agarra y llueve* (*La Prensa*, Uruguay)

Con las realizaciones, en cambio, la compatibilidad es total y la frecuencia muy alta tanto para las estructurasseudocoordinadas como para las perífrasis. Sin embargo, pueden identificarse importantes diferencias semánticas. Cabe recordar que las realizaciones son eventos complejos que suponen dos subeventos: una actividad o proceso y un cambio de estado o locación (cf. Vendler, 1957; Morimoto, 1998; Bosque y Gutiérrez-Rexach, 2008, entre muchos otros). Según se corrobora en (45-47), las estructurasseudocoordinadas y las perífrasis incoativas se refieren de muy distinta manera a ese evento complejo:

- (45) a. Comenzó a romper la reja.
b. *Fue / agarró y rompió* la reja.
- (46) a. Empezó a construir la casa.
b. Ayer *fue / se puso y construyó* la casa.
- (47) a. Empieza/ Comienza a decirle que es una estupidez.
b. *Va/ Agarra y le dice* que es una estupidez.

Con las perífrasis incoativas (cf. 45-47a) solo se focaliza el inicio del subevento de actividad (y no se dice nada sobre el cambio de estado final). Por el contrario, en (45-47b) el evento se da por terminado y cumplimentado: se encuentra en foco el evento completo y, en particular, el cambio de estado final que supone la realización.

4.2. Otros factores gramaticales

Los contrastes entre perífrasis incoativas y estructurasseudocoordinadas no se limitan a la simple combinatoria con el aspecto léxico. Otro eje interesante de comparación está dado por los valores morfológicos que pueden tomar los (seudo) auxiliares. Hemos señalado que las estructurasseudocoordinadas son posibles, aunque muy poco frecuentes, en pretérito imperfecto (cf. 15a-b). Si analizamos con mayor detenimiento los casos relevados, comprobamos que el imperfecto se interpreta allí como modo hipotético (cf. 48a) o aspecto habitual (cf. 48b-c), pero nunca como progresivo o continuo.

- (48) a. Por suerte no me enteré, si no, *iba y lo mataba*.
 b. *Ibas y te reías*.
 c. *Agarraba y se enfermaba*.

De hecho, si buscamos contextos que induzcan o fuercen la lectura progresiva del imperfecto en la cláusula principal y en la subordinada, como los de (49a-b), la estructuraseudocoordinada resultará directamente agramatical (cf. 49c-d):

- (49) a. Juan miraba la tele cuando llegó María.
 b. María llegó cuando Juan miraba la tele.
 c. *Juan iba y miraba la tele cuando llegó María.
 d. *María llegó cuando Juan agarraba y miraba la tele.

En cambio, las perífrasis incoativas muestran plena compatibilidad con el valor progresivo del imperfecto:

- (50) a. Juan comenzaba/ empezaba a mirar la tele cuando llegó María.
 b. Cuando comenzaba/ empezaba a correr, me crucé a Pablo.

En este marco, resulta esperable el comportamiento dispar que exhiben las perífrasis incoativas y las estructurasseudocoordinadas al combinarse con perífrasis progresivas: mientras que las primeras son perfectamente compatibles (cf. 51), las segundas son agramaticales en los mismos contextos (cf. 52):

- (51) a. Juan estaba comenzando / empezando a mirar la tele (cuando llegó María).
 b. Pipo estaba comenzando / empezando a destruir el mueble (cuando llegó María).
 (52) a. *Juan estaba yendo y estaba mirando la tele (cuando llegó María).
 b. *Pipo estaba agarrando y estaba destruyendo el mueble (cuando llegó María).

Otros puntos de discordancia remiten, nuevamente, a la combinatoria de ambas clases de construcciones gramaticalizadas con las realizaciones. En §4.1 establecimos que las realizaciones se dan por concluidas (y efectivamente realizadas) con lasseudocoordinadas y por recién empezadas con las perífrasis incoativas (cf. 45-47). En ese sentido, resulta muy clara la oposición entre los dos tipos de construcciones gramaticalizadas ante una continuación oracional que implique que el evento en realidad no tuvo o no tendrá lugar. Esa continuación es natural con las perífrasis de (53a-54a) (que focalizan solamente la etapa inicial (interna) del evento), mientras que con las estructurasseudocoordinadas de (53b-54b) desencadena una anomalía:

se necesitaría muchísimo contexto para que esa «desmentida» resulte aceptable, lo que sugiere fuertemente que las estructurasseudocoordinadas presuponen que el evento fue completado:

- (53) a. Ayer empezó a arreglar el baño... pero al final no lo arregló.
 b. Ayer *fue y arregló* el baño... ??pero al final no lo arregló.
- (54) a. Esta tarde comienzan a podar tres árboles de mi cuadra (veremos si lo terminan de hacer alguna vez).
 b. Esta tarde *van y podan* tres árboles de mi cuadra (??veremos si lo terminan de hacer alguna vez).

Un conjunto de datos igualmente ilustrativos, a nuestro modo de ver, está constituido por el contraste que se produce al combinar ambas clases de construcciones con el denominado «*se* aspectual» (cf. Di Tullio, 2005) o «clítico télico» (cf. Sanz y Laka, 2002). Este tipo de *se* «es una marca sintáctica de la telicidad de un determinado tipo de evento; en particular, de la telicidad de los eventos con delimitación y medida, es decir, de las denominadas realizaciones» (Sanz y Laka, 2002, p. 331). Es siempre opcional y, cuando aparece con verbos «que denotan ‘consumición’ (física o intelectual), indica un ‘acto voraz y exhaustivo’» (cf. Di Tullio, 2005, p. 181), como en los ejemplos de (55). Por esa razón se combina exclusivamente con realizaciones¹⁸ que tengan un OD delimitado que sea, a su vez, un tema incremental (en el sentido de Dowty, 1991), según se corrobora en el inmediato contraste que se produce si los OD son no delimitados o si no hay OD (cf. 56):

- (55) a. Me fumé varios cigarrillos.
 b. Me caminé 15 cuadras.
 c. Me leí tres novelas de Camus.
- (56) a. ??Me fumé cigarrillos/ ??Me fumé.
 b. ??Me caminé un rato/ ??Me caminé.
 c. ?? Me leí novelas / ??Me leí.

La combinación del *se* aspectual con las construcciones gramaticalizadas bajo estudio arroja resultados disímiles. Con las perífrasis incoativas, los resultados son anómalos, igual que en (56) (cf. 57a); en contraposición, si el *se* aspectual no está presente, la perífrasis es perfecta con los mismos predicados, como se advierte en (57b):

18. De hecho, observan Sanz y Laka (2002, pp. 320-321), si el *se* aparece con estados transitivos coerciona la lectura aspectual como realizaciones, como en *Se sabe la lección*.

- (57) a. ??Comencé a fumarme varios cigarrillos/ caminar 15 cuerdas/ leerme tres novelas de Camus¹⁹.
 b. Comencé a fumar varios cigarrillos/ caminar 15 cuerdas/ leer tres novelas de Camus.

En cambio, las estructuras pseudocoordinadas resultan perfectamente naturales con el *se* aspectual en los mismos contextos, como muestra (58):

- (58) *Fui y me fumé* varios cigarrillos/ *Agarré y me caminé* 15 cuerdas / *Fui y me leí* tres novelas de Camus.

Debe notarse, sin embargo, que el contraste de aceptabilidad ilustrado entre (57a) y (58) no se produce cuando la oración que incluye el *se* aspectual contiene un objeto directo que presenta una interpretación singular²⁰: tanto la perífrasis incoativa (59a) como la estructura pseudocoordinada (59b) son posibles.

- (59) a. Comencé a fumarme un cigarrillo / leerme la última novela de Camus / tomarme la leche.
 b. *Fui y me fumé* un cigarrillo / *me leí* la última novela de Camus / *me tomé* la leche.

Veamos cómo pueden explicarse estos contrastes. La lectura ingresiva o incoativa propia de las perífrasis es compatible con realizaciones, según resaltamos, siempre que se pueda focalizar el inicio de la actividad que luego dará como resultado un cambio de estado. Esa interpretación sobre el subevento de actividad sigue habilitada, aun con el *se* aspectual, si el OD es singular (cf. 59a), pero no está disponible si las oraciones focalizan un resultado o un cambio de estado múltiple, como ocurre cuando el *se* aspectual (que enfatiza la telicidad de la realización) se combina con una pluralidad (delimitada) en el OD (cf. 57a). Las estructuras pseudocoordinadas, en cambio, resultan gramaticales con todas las variantes del *se* aspectual, precisamente porque no se limitan a focalizar el inicio del subevento de actividad, sino la realización en su totalidad, lo que incluye el cambio de estado²¹.

19. Debe señalarse que las oraciones de (57a) serían aún peores con ascenso de clíticos (i):

- (i) a. ?? Me comencé a mirar las últimas películas de Kubrick.
 b. *Me empecé a caminar 5 cuerdas.
 c. *Me comencé a fumar 20 cigarrillos.

20. Agradezco al revisor anónimo que llamó mi atención sobre este dato.

21. De hecho, pese a la pluralidad implicada en el OD, el evento en (58) se concibe como un evento unitario, tal como apuntan Keniston (1936), Kany (1945) y Coseriu (1966).

Los contrastes de comportamiento en relación con el aspecto léxico y gramatical (incluido el *se* aspectual) nos dan motivos para plantear nuestra hipótesis, que es compatible con todos los datos vistos hasta aquí: las estructurasseudocoordinadas expresan con énfasis el aspecto aoristo o perfectivo²². De acuerdo con Smith (1991), el aspecto gramatical aoristo (en sus términos, perfectivo) cubre el tiempo de todo el evento, incluyendo ambos límites (inicial y final), lo que supone que focaliza «el evento completo, desde su principio a su fin», de acuerdo con la descripción de Camus Bergareche (2004) para el español²³. Smith observa que, al interactuar con diferentes clases aspectuales, pueden relevarse lecturas más sesgadas, que incluyen una interpretación terminativa, en que el tiempo en foco incluye el límite del tiempo del evento (cf. Klein, 1992), y una interpretación ingresiva, en que el punto focalizado es el punto inicial del tiempo del evento (cf. también Camus Bergareche, 2004; Fábregas, 2016, con datos del español).

De este modo, para las realizaciones tendríamos tres variantes gramaticales estrechamente vinculadas entre sí. Por un lado, la variante «neutra», con el pretérito perfecto simple (cf. 60a), ya marca *per se* el aoristo: el evento se da por concluido y se adopta sobre él una perspectiva global que incluye su principio y su final. La variante (60b) agrega el llamado «*se* aspectual», que enfatiza el aspecto télico de la realización (que, obviamente, ya estaba presente en (60a)). Por último, la estructuraseudocoordinada, en la variante (60c), expresa con mayor énfasis el aspecto aoristo, es decir refuerza una perspectiva completa o global sobre el evento que aparecía implícita en las demás variantes²⁴:

- (60) a. Comió la torta.
 b. Se comió la torta.
 c. Fue y (se) comió la torta.

22. Privilegiamos la etiqueta de aoristo sobre perfectivo para no generar confusiones con el perfecto (cf. Camus Bergareche, 2004).

23. En la clasificación de Camus Bergareche (2004) el aoristo queda agrupado con el perfecto bajo la etiqueta de «evento concluido». El perfecto focaliza el momento posterior al fin del evento, que «quedará cerrado por la izquierda, pero abierto por su lado derecho». El aoristo, en cambio, supone un evento completo, acotado «desde su principio a su fin».

24. Es posible establecer un paralelismo entre el aspecto aoristo expresado por (60a) y, más enfáticamente, por (60c) con lo que ocurre con el aspecto progresivo, que puede expresarse mediante el pretérito imperfecto solo (i.a), pero resulta más explícito o enfático si aparece la perífrasis correspondiente (cf. i.b):

- (i) a. María miraba la tele (cuando llegó Juan).
 b. María estaba mirando la tele (cuando llegó Juan).

En suma, el *se* aspectual solo subraya la naturaleza télica de las realizaciones (siguiendo, una vez más, a Sanz y Laka, 2002), mientras que la estructura seudocoordinada expresa con énfasis el aspecto gramatical aoristo o perfectivo, que puede aplicarse a diversas clases léxico–aspectuales de eventos (i.e., realizaciones, logros o actividades).

Notemos, sin embargo, que la interpretación semántica de las estructuras seudocoordinadas difiere de la hasta aquí revisada cuando aparecen con actividades, una combinación que es poco frecuente pero posible. Si repetimos nuestros ejemplos de (44), comprobaremos que allí la interpretación de las estructuras es ingresiva y focaliza el límite inicial del tiempo del evento:

- (61) a. El que quiere estudiar, *va y estudia*, el que se quiere emborrachar, *va y se emborracha* (Twitter, Guatemala)
 b. Si está con criatura *agarra y corre* con el niño (*ABC*, prensa, Paraguay)
 c. El tipo *agarra y camina* como un cangrejo (web, Argentina).
 d. Ahora el señorito quiere paz / por qué no *vas y laborás* (Las Manos de Filippi, «El rey de la autocrítica», Argentina)
 e. Apenas desatan el tiento de la inauguración, *agarra y llueve* (*La Prensa*, Uruguay)

Estos datos, sin embargo, son perfectamente consistentes con la caracterización del aoristo. Ya hemos dicho que Smith (1991) reconoce una posible lectura ingresiva para el aspecto aoristo, en que el punto focalizado será el inicio del tiempo del evento, de acuerdo con la clase léxico–aspectual con la que se combine (cf. Camus Bergareche (2004) para el español y Bary y Egg (2012) para el griego)²⁵.

Por último, notemos que también es consistente con un significado aspectual de aoristo el hecho de que las estructuras seudocoordinadas no sean compatibles con estados, según hemos advertido para (36)²⁶.

25. El significado de las perífrasis incoativas y las estructuras seudocoordinadas también se aproxima en ciertas combinaciones con predicados delimitados. En los siguientes ejemplos, el predicado no puede darse por efectivamente completado, puesto que las oraciones tienen valor de futuro respecto del presente de la enunciación (i) o de otro tiempo de referencia (ii):

- (i) a. Ahora comienzo a cocinar la torta de cumple de Fede.
 b. Ahora *voy y cocino* la torta de cumple de Fede.
 (ii) a. Me dijo que esa noche comenzaba a escribir la composición.
 b. Me dijo que esa noche *agarraba y escribía* la composición.

Sin embargo, los ejemplos de (i–ii.b) siguen presentando una perspectiva marcadamente más global sobre el evento que los de (i–ii.a), que focalizan únicamente el inicio.

26. De hecho, algunos (pocos) hablantes sí aceptan la combinación de estructuras seudocoordinadas con ciertos estados no permanentes (siempre psicológicos y en presente histórico), como *Va y está*

En suma, los datos aportados por esta sección evidencian el interés de una comparación sistemática entre las estructurasseudocoordinadas y las perífrasis incoativas en relación con el aspecto léxico y otros factores gramaticales. Aun si las tendencias combinatorias que aquí esbozamos no fueran absolutas, alcanzan para subrayar un contraste gramatical significativo. Ese contraste resulta especialmente evidente en la conjunción con realizaciones, ya que las perífrasis incoativas solo pueden resaltar la primera fase (interna) del proceso, mientras que las estructurasseudocoordinadas expresan con énfasis la ejecución completa del cambio de estado final (cf. 45-47). Otras pruebas gramaticales que sustentan la misma distinción las proporcionan contextos como la (in) compatibilidad con el aspecto progresivo (cf. 49-52), con una continuación que suponga que el evento no se llevó a cabo (cf. 53-54) o con el *se* aspectual (cf. en particular 57a vs. 58).

5. DISCUSIÓN

En esta sección pretendemos sistematizar nuestra propuesta de que las estructurasseudocoordinadas expresan con énfasis el aspecto aoristo. Para ello, exploramos la interacción con otras construcciones gramaticalizadas que expresan aspecto, tiempo o modo, la configuración sintáctica involucrada (cf. Cinque, 1999, 2006; Cardinaletti y Giusti, 2001) y los procesos de gramaticalización en juego (cf. Bybee y Pagliuca, 1987; Company, 2004; Torres Cacoullós, 2016 y, nuevamente, Cardinaletti y Giusti, 2001). Asimismo, evaluamos nuestro análisis de las estructurasseudocoordinadas respecto de otras perspectivas previas en la bibliografía gramatical del español (cf. Coseriu, 1966; García Sánchez, 2007; Jaque *et al.*, 2018, 2019), ampliando la presentación preliminar hecha en §2 y aportando nuevos argumentos empíricos para discutir las otras propuestas.

Un argumento adicional a favor de un análisis aspectual para las estructurasseudocoordinadas proviene del hecho de que se registran datos reales en que los verbos principales de la oración forman parte de perífrasis incoativas con *comenzar* (62a-b) y *empezar* (62c-e) y terminativas (o completivas, cf. Cinque, 2006, p. 70 y ss), tanto con *dejar* (cf. 63a-b, e) como con *terminar* (63c-d):

tranquilo o *Va y lo ama* (en contraste con nuestros juicios de (36d-e)). Suponemos que esos hablantes están adoptando una interpretación de esos predicados como logros, en las mismas líneas planteadas por Fábregas (2016) para *Estuvo tranquilo* o *Lo amó*.

- (62) a. el consumidor *va y comienza a personalizar* su producto en línea (web, México)
 b. Todos los días me despierto y pienso ‘*hoy es un gran día para alcanzar mis metas*’, así que *voy y comienzo a hacer* algo para llegar a ellas (Instagram, España)
 c. El tipo *agarra y empieza a pincelar* el motor (web, Argentina)
 d. Uno *agarra y empieza a mirar* a todos lados menos la cámara (web, Bolivia)
 e. Massa *agarra y empieza a tomar* las medidas más populistas y berretas (web, Argentina)
- (63) a. Cuando imprimo unos 25 cms, *va y deja de imprimir* (web, España)
 b. Amazon *va y deja de vender* la versión digital de SimCity (web, Chile)
 c. Bueno, *agarré y terminé de escribir* todo lo que me pareció (web, Argentina)
 d. Ahora *vas y terminás de ver* el video (web, España)
 e. Si quieres dejar de jugar, simplemente *agarra y deja de girar* (web, EE.UU)

Por un lado, el hecho de que la estructuraseudocoordinada pueda tener alcance sobre una perífrasis aspectual incoativa o terminativa es otro indicio claro de que está ubicada entre las proyecciones funcionales que modifican al verbo, y no en un dominio léxico o semiléxico, como se ha propuesto para el inglés, el sueco o, en menor medida, los dialectos del italiano (cf. De Vos, 2006; Wilklund, 2008; Cardinaletti y Giusti, 2001 en §2). Por otro lado, esa posibilidad es consistente con los términos de Smith (1991) y Camus Bergareche (2004): puesto que el aoristo focaliza el evento completo «desde su principio a su fin», puede limitarse a uno solo de estos puntos, por lo que resulta esperable que tenga alcance, precisamente, sobre los significados incoativo y terminativo.

Por el contrario, otras perífrasis aspectuales, como la habitual (cf. 64a-b), la culminativa (cf. 64c-d) o la prospectiva (cf. 64e-f)²⁷, siempre tienen precedencia (y, por lo tanto, alcance) sobre las estructurasseudocoordinadas:

- (64) a. Tenía amigos y amaba la música y *solía ir y hablar* con los vecinos a veces (web, España)

27. Las estructurasseudocoordinadas parecen poco frecuentes con tiempos compuestos (cf. también Kany, 1945; Kornfeld, 2022), aunque se encuentran algunos ejemplos para *agarrar*:

- (i) a. *Hemos agarrado y hemos impuesto* que las empresas mantengan el precio (prensa, Bolivia)
 b. *Hemos agarrado y hemos hablado* con la gente de recursos humanos (prensa, Venezuela)
 c. Por qué *no hemos agarrado y hemos puesto* una flotilla de camiones cisterna (web, Colombia)
 d. ella *hubiera agarrado y hubiera comprado* chipá al morocho (Suárez, *Un fragmento de la vida de Irene S*, Argentina)
 e. Mientras *no hayas agarrado y olido* tu mano luego de eso... (web, España).

- b. En mi caso es uno de mis juegos favoritos que cada tanto suelo *agarrar y jugarlo* un buen par de horas (web, Argentina)
- c. Digo yo que gana Doom si *llega a agarrar y asestarle* un golpe al jefe (web, México)
- d. ...me termino drogando yo, como policía, o *termino agarrando y vendiéndosela* al «guacho» (Argentina)
- e. La verdad *estoy por agarrar y crear* mis propias medidas y clases (web, Argentina)
- f. En el matadero *estoy por ir y tirarles* una piedra (web, Argentina)

De este modo, mientras que los datos de (62-63) demuestran que las estructurasseudocoordinadas se ubican en las capas funcionales de la oración, los de (64) sugieren que no están muy altas. La posición relativamente baja del aoristo (que no aparece mencionado como proyección aspectual en la cartografía esbozada por Cinque²⁸) es consecuente, asimismo, con la sensibilidad de las estructurasseudocoordinadas hacia las propiedades léxico-aspectuales de los predicados (cf. §4.1). Adviértase que las perífrasis bajas de (62-63) manifiestan una sensibilidad análoga (según se ha discutido *in extenso* para las incoativas en §4.1), mientras que ninguna de las proyecciones aspectuales altas reflejadas en (64) presenta restricciones semejantes, ya que, de hecho, se combinan con cualquier clase argumental o léxico-aspectual de predicados.

En un trabajo previo propusimos, a partir de Cinque, Cardinaletti y Giusti, una estructura análoga a la Figura 1 (adaptada de Kornfeld, 2019). El pseudoauxiliar se ubica como núcleo de un Sintagma de Aspecto Aoristo y, más abajo, aparecen un SX (encabezado por la *y* subordinativa) y luego el Sv/SV correspondiente al verbo léxico²⁹:

28. En efecto, el Esquema 1, que esboza la cartografía de las proyecciones funcionales correspondientes al dominio del SFlex, según Cinque (1999, 2006), no contempla el Aspecto Aoristo. Sin embargo, por los datos de alcances y restricciones relevados, debería ubicarse entre el Aspecto Prospectivo (cf. 64e-f) y el Aspecto Incoativo (cf. 62):

Esquema 1

... Modo_{(posibilidad)(M7)} > Aspecto_{(habitual)(A1)} > Aspecto_{(demorado)(A2)} > Aspecto_{(preposicional)(A3)} > Aspecto_{(reiterativo)(A4)} > Aspecto_{(frecuentativo I)(A5)} > Modo_{(desiderativo)(M7)} > Aspecto_{(acelerativo I)(A6)} > Aspecto_{(terminativo)(A7)} > Aspecto_{(continuativo)(A8)} > Aspecto_{(perfecto)(A9)} > Aspecto_{(retrospectivo)(A10)} > Aspecto_{(aproximativo)(A11)} > Aspecto_{(durativo)(A12)} > Aspecto_{(progresivo)(A13)} > Aspecto_{(prospectivo)(A14)} > Aspecto_{(incoativo)(A15)} > Modo_{(obligación)(M8)} > Modo_{(capacidad)(M9)} > Aspecto_{(frustrativo)(A16)} > Modo_{(permiso)(M10)} > Aspecto_{(conativo)(A17)} > Aspecto_{(completivo I)(A18)} > Voz_(pasiva) > Aspecto_{(acelerativo II)(A19)} > > Aspecto_{(incoativo II)(A20)} > Aspecto_{(completivo II)(A21)} ...

29. Siguiendo a Cardinaletti y Giusti (2001, p. 400), suponemos que la coincidencia en las categorías morfológicas (tiempo-aspecto-modo y persona-número) entre el pseudoauxiliar y el verbo principal se debe a una operación de copiado de rasgos formales, que puede ser de naturaleza sintáctica o post-sintáctica

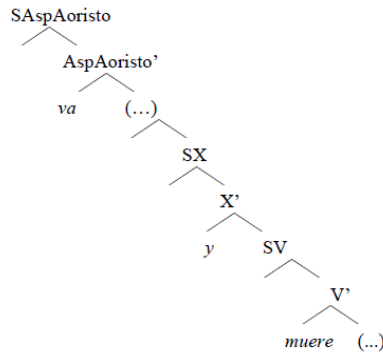


Figura 1

Es pertinente observar que el SX bien podría entenderse como un Sintagma de Intensificación, pese a que esa posibilidad no se encuentra formalizada en la Figura 1. Ya señalamos que las estructuras seudocoordinadas expresan «con énfasis» el aspecto aoristo; de hecho, esa sería su (única) diferencia semántica con el pretérito perfecto simple, que meramente expresa aspecto aoristo (cf. ejemplos (60) y nota 24). Sin embargo, es factible que el énfasis de las estructuras seudocoordinadas no solo se encuentre codificado en el núcleo del Sintagma de Aspecto Aoristo, sino también en el SX (o Sintagma de Intensificación), dado el carácter frecuentemente intensificativo de otros fenómenos que involucran la coordinación asimétrica³⁰ (cf. Kornfeld, 2022). Entre esos fenómenos pueden mencionarse la coordinación asimétrica de formas verbales gramaticalizadas como *dale* y *meta* en español rioplatense (cf. Di Tullio, 2006), con valor iterativo, durativo o de insistencia (cf. 65a-b), o la coordinación asimétrica de un mismo verbo léxico (cf. 65c), que parece tener, igualmente, una lectura semántica ligada con la iteración o la duración:

- (65) a. Dale y dale charlar y reír.
 b. Meta y meta vivir.
 c. Charla y charla; corre y corre; siguieron y siguieron.

(i.e., morfológica), habilitada por el hecho de que el verbo principal es una forma flexionada (al contrario de lo que ocurre con las perífrasis).

30. Eventualmente, el valor intensificativo puede ser resaltado si se acumulan pseudoauxiliares, tal como hemos señalado para ejemplos como (11b) (cf. también nota al pie 9 y Bravo, 2020). La Figura 1 admite, de hecho, la posibilidad de que se trate de recursos potencialmente recursivos.

Los fenómenos de (65) están, evidentemente, vinculados con la intensificación, aun si los significados aspectuales en juego son bien diferentes del que hemos atribuido aquí a las estructurasseudocoordinadas.

La discusión acerca de la naturaleza aspectual de las estructurasseudocoordinadas tiene una gran raigambre en la tradición gramatical española. En efecto, los términos que utilizamos en §4.2 para describir el aspecto aoristo enfatizado por esas construcciones rescatan en lo esencial la propuesta de Coseriu (1966) para *tomo y me voy*. Coseriu (1966, pp. 128-129) afirma que esa clase de construcción «expresa la globalidad de la acción significada por el segundo verbo, es decir que implica tanto el punto inicial como el punto final de la misma [...] Por la globalidad explícita, la acción verbal es separada de otras acciones (expresadas o solo sobreentendidas) y opuesta a estas, con lo cual se subraya su efectividad y el giro adquiere carácter intensivo [...] Si el verbo determinado por ‘tomar’ no implica perfectividad, la construcción paratáctica, implicando globalidad, le confiere carácter terminativo (así en griego moderno y en las lenguas románicas)», a diferencia de lo que ocurriría en lenguas eslavas (por ejemplo, el ruso), donde el verbo sí es perfectivo y se «destaca la efectividad y rapidez de la acción y su carácter repentino e inesperado». Si el verbo es durativo, finalmente, se «acentúa también el punto inicial de la acción... de aquí que haya podido ser interpretado como ‘ingresivo’». Coseriu (1966, p. 130) opone a esta perspectiva «global» propia de las estructurasseudocoordinadas el significado de las perífrasis verbales que son aspectualmente «cursivas», esto es, que señalan uno o dos puntos del desarrollo del evento, como la perífrasis progresiva (*estoy escribiendo*). De esta manera, la caracterización general de Coseriu se ajusta a la descripción gramatical y semántica que aquí hemos presentado para las construccionesseudocoordinadas del español³¹ y, en particular, a la definición del aspecto gramatical aoristo (cf. §4.2).

En cuanto a los estudios más recientes acerca de las estructurasseudocoordinadas, la propuesta que desarrollamos en §4 se distancia, evidentemente, de la postura de García Sánchez (2007), que habíamos adoptado en Kornfeld (2019, 2022). Recordemos que García Sánchez supone que las construccionesseudocoordinadas tienen valor ingresivo, basándose en los claros y sugerentes paralelismos con las perífrasis que indudablemente sí tienen ese significado y que comparten los mismos (seudo) auxiliares gramaticalizados (cf. §3.3). Sin embargo, creemos haber demostrado en

31. Disentimos en ciertos detalles, como, por ejemplo, su afirmación de que en las lenguas románicas las estructurasseudocoordinadas confieran un «carácter terminativo», que en nuestra opinión alude a un aspecto bien diferente y diferenciable.

§4 que el significado de las estructuras seudocoordinadas gramaticalizadas no puede ser ingresivo, puesto que sistemáticamente contrastan con las perífrasis incoativas en su distribución gramatical y/o interpretación semántica con diversas clases léxico–aspectuales de predicados y en otros fenómenos gramaticales³².

Otra propuesta reciente que atribuye un valor aspectual a las estructuras seudocoordinadas (a las que denominan «construcciones multiverbales de verbo finito coordinadas», cf. §2) es la de Jaque *et al.* (2018, 2019). Los autores se centran en el caso de *llegar y* + verbo y plantean que las estructuras seudocoordinadas expresan «interruptividad»: «dada una serie de eventos –expresada verbalmente o supuesta–, que define un desarrollo normal de los acontecimientos, la ocurrencia de un evento *e* corta de manera abrupta esa normalidad y da origen a otra serie de eventos –expresada verbalmente o supuesta– que difiere de la esperada en la secuencia original» (Jaque *et al.*, 2019, p. 182-3). Para caracterizar la noción de «desarrollo normal de los acontecimientos», acuden a la idea de «mundos inerciales» que Dowty (1977) empleó para analizar la perífrasis progresiva y afirman que las estructuras seudocoordinadas con *llegar y* + verbo siempre presuponen que el evento no es esperable, puesto que rompe con la «continuidad del mundo inercial previo» en términos de expectativas físicas (66a) o culturales (66b):

- (66) a. Estaba soleado y de repente *llega y se pone a llover*.
 b. Santos, de 67 años, ha recordado que el 3 de octubre de 2010 sirvió dos copas de orujo de hierbas a los acusados, que *cogieron y se marcharon* sin pagar.

Además de defender su perspectiva, Jaque *et al.* (2018, 2019) critican los estudios previos acerca de las estructuras seudocoordinadas en la gramática española, que las analizan como ingresivas³³, completivas o (ad)mirativas.

32. En nuestra opinión carece de sentido la afirmación de García Sánchez (2007, p. 174) de que «estos verbos [como auxiliares paratácticos e hipotácticos] conservan el mismo valor en ambos tipos de construcción, de manera que la diferencia se establece solo entre las construcciones consideradas en su conjunto», ya que es, precisamente, el significado de las construcciones en su conjunto el único que puede ser analizado en forma sistemática desde un punto de vista gramatical.

33. En cuanto al análisis incoativo de las estructuras seudocoordinadas (que atribuyen a Cuervo y García Sánchez), Jaque *et al.* lo refutan con bastante menos argumentación que la desarrollada en §4. Por un lado, mencionan la incompatibilidad con una continuación que señale que el evento en realidad no se ha llevado a cabo, a partir de ejemplos semejantes a los de (53). Por el otro, agregan el argumento de que los auxiliares incoativos no pueden encadenarse entre sí (cf. **se puso a empezar a llover*), mientras que las estructuras seudocoordinadas sí admiten la combinación con un auxiliar incoativo (*llegó y se puso a llover*, cf. también (62)) (Jaque *et al.*, 2019, p. 230-231).

En cuanto al análisis «completivo» de las estructurasseudocoordinadas (que coincidiría con el que aquí hemos defendido), los autores lo atribuyen a Keniston, Kany y Coseriu y lo consideran predominante en la tradición española, puesto que se retoma, por ejemplo, en el *Diccionario de perífrasis verbales* de García Fernández (2006). Para rechazarlo, se limitan a comparar la distribución de esas estructuras con la del *se* aspectual o télico (cf. §4.2). Constatan que no tienen la misma distribución, ya que el *se* aspectual es incompatible con las actividades, mientras que las estructurasseudocoordinadas sí pueden aparecer con ellas (aunque infrecuentemente), según ya señalamos (cf., en particular, los datos de (44/61)). Sin embargo, se trata de una comparación desafortunada, ya que, si bien ambas son construcciones enfáticas, remiten a dominios esencialmente diferentes: la aspectualidad léxica (en el caso del clítico télico) y el aspecto gramatical (en el caso de las estructurasseudocoordinadas), por lo que las coincidencias son parciales y, sobre todo, superficiales, como esperamos haber demostrado en §4.2 (cf., en particular, nuestra explicación sobre el paradigma de (60)).

Por su parte, en lo que hace a la propuesta concreta que defienden Jaque *et al.* (2018, 2019), notemos en primer lugar que la «interruptividad» supone solapamientos con nociones semánticas que han sido ligadas con la miratividad, como la contraexpectativa (cf. Aikhenvald, 2012). Sin embargo, Jaque *et al.* también descartan que las estructurasseudocoordinadas sean habitualmente mirativas (un análisis que atribuyen a Keniston, en nuestra opinión erróneamente³⁴) a partir de dos argumentos básicos. Por una parte, sostienen que las estructurasseudocoordinadas son frecuentes con la primera persona del singular y que DeLancey (1997) ha establecido que la miratividad es incompatible con esa persona gramatical. Debemos decir que este es un argumento falaz, puesto que Aikhenvald (2012, p. 441) observa que esa generalización de DeLancey era al menos apresurada, y ofrece numerosos ejemplos de (indudables) construcciones mirativas con primera persona en diversas lenguas del mundo. Por otra parte, Jaque *et al.* (2019, p. 234) destacan que el alcance de las estructurasseudocoordinadas tiene que ser bajo si se observa la posibilidad de que sean tomadas como complementos por distintas perífrasis temporales, modales y aspectuales (cf. también *supra* (64)):

- (67) a. ¿Cómo *vas a llegar y decirle* a un tipo que es un impresentable, o que es feo, o que es gilipollas? (Naveros, *Al calor del día*, España, CORPES XXI)

34. Keniston (1936) solo menciona la sorpresa o la irritación como matices que pueden adoptar (secundariamente) esas construcciones.

- b. Para poder triunfar simplemente *tienes que llegar y levantarte* cada vez que te caes en el camino. Solo así conseguirás tus propósitos. (frasesdelavida.wiki, Chile)
- c. Entonces *{debí / pude / tuve que} coger y decirle* a Pedro que no volviera. (García Fernández, 2006: s.v. coger y + verbo: (15a))
- d. Efectivamente, Juan *suele coger y dar* grandes paseos él solo. (García Fernández, 2006: s.v. coger y + verbo: (15b))

Dado que hay acuerdo en la bibliografía especializada en que la miratividad se liga con una proyección modal (muy) alta (cf. Aikhenvald, 2012; Cinque, 1999; Kornfeld, 2019), este segundo argumento de Jaque *et al.* sí resulta convincente para refutar la posibilidad de que las estructurasseudocoordinadas expresen habitualmente miratividad.

Ahora bien, dejando de lado la cuestión de las relaciones entre interruptividad y miratividad, nos interesa resaltar que numerosos casos de estructurasseudocoordinadas no parecen corresponderse con la idea de un evento disruptivo respecto de los precedentes. En efecto, en las siguientes oraciones con *agarrá y + verbo* (68) y *va y + verbo* (69) los eventos introducidos por las estructurasseudocoordinadas no solo no alteran el desarrollo normal de los acontecimientos sino que, de hecho, se presentan como consecuencias completamente lógicas de los eventos previos, gracias a la aparición de marcadores consecutivos como *entonces* (68a-c, 69a), *pues* (68d, 69b), *por lo tanto* (68e-f, 69c) o *así que* (69d-e):

- (68) a. Me di cuenta de que tenía cara de pocos amigos, entonces *agarré y aceleré* (web, Paraguay)
- b. Miré así y dije «y ahora por dónde empiezo», y dije «bueno vamos por el principio» y entonces *agarré y leí* los griegos y leí, leí, leí los griegos (web, Argentina)
- c. Por eso siento que más allá de lo que uno quiera hacer, tenés que pensar que tenés un tiempo, una energía, entonces *agarrá y ocupala* para hacer algo que te sirva (web, Argentina)
- d. Y cuando tengo hambre entre comidas pues *agarro y pelo* unas dos frutas... (web, Colombia)
- e. Estaba aburrido este domingo y dije «hace mucho tiempo no subo un blog». Por lo tanto, *agarré y vi* lo último que subí y fue este dibujo (web, Argentina)
- f. cuando yo compré la moto ya había sido reparado el cigüeñal y por lo tanto *agarré y compré* en Estados Unidos el conjunto completo (web, Argentina)

- (69) a. El ayuntamiento se convierte en un banco de horas: tú sabes arreglar coches, entonces *vas y dices* que estás dispuesto a canjear 3 horas de esa tarea por clases de inglés (Astelarra, declaraciones, España)
- b. Si debo ir y ayudar a mi compañero de la Sub 14 en una tarea, pues *voy y aporto* con mi feedback (La Pizarra del DT, Twitter)
- c. Ella queda muy preocupada y, por lo tanto, *va y habla* con Héctor y Silvia (web, Colombia)
- d. Todos los días me despierto y pienso ‘hoy es un gran día para alcanzar mis metas’, así que *voy y comienzo a hacer* algo para llegar a ellas (Allisonbeel, Instagram, España)
- e. Cuando me convertí en futbolista profesional, Internacional me dio la posibilidad de ir a la universidad, así que *fui y aprobé* el examen de ingreso (web, EE.UU.)

La misma relación de causa–consecuencia puede reconstruirse en los siguientes ejemplos, en que la estructuraseudocoordinada enuncia la consecuencia lógica y razonable sea del evento previamente introducido por una cláusula relativa (como en (70)), sea de uno o más eventos mencionados o enumerados en contextos paratácticos (cf. 71):

- (70) a. El que quiere estudiar, *va y estudia*, el que se quiere emborrachar, *va y se emborracha* (Twitter, Guatemala)
- b. Al que le interesa, *va y lo lee* (web, Argentina)
- c. Al que le interesa, *va y lo compra* (web, Uruguay)
- d. Quien viene aquí se entusiasma, *va y lo comenta* con familiares y amigos (web, México)
- e. Cuando quieras, *ve y memoriza* la traducción (web, México)
- (71) a. Un amigo me dijo que había pruebas para entrar en la escuela de teatro, la ENA, *fui y aprobé* (web, Cuba)
- b. El teórico muy fácil, haciendo tests en casa durante un par de semanas *fui y aprobé* a la primera sin fallos (web, España)
- c. Estoy viendo qué libro comprarme y *agarré y leí* Insuficiencia Cardíaca del Harrison y del Cecil (web, Argentina)
- d. La gente quiere escuchar un disco, entra a Spotify, le gusta, *va y lo compra* (web, México)

También se opone a un análisis interruptivo el hecho de que las estructurasseudocoordinadas con *va y* + verbo o *agarra y* + verbo puedan combinarse con adverbios modales evaluativos que ponderan el evento como plenamente esperable

(cf. *lógicamente* en 72a-b) u otras construcciones equivalentes con adjetivos (cf. *lógico, razonable, previsible, sensato* en 72c-g) o con verbos (cf. 72h-k):

- (72) a. En el caso 2, lógicamente, *agarra y llama* tú (web, España)
 b. ¿Qué se puede hacer seriamente con un país así? Y, lógicamente, *fue y se murió* (web, España)
 c. Lo más razonable es *ir y hablar* con Eddie y agradecerle (web, Chile)
 d. Lo más sensato es *ir y bloquear* solo aquellas aplicaciones que puedan dañar al chico (web, Colombia)
 e. Como es previsible, *va y gana* el gordo (*El país*, España)
 f. Parece lógico *ir y escoger* de nuevo la pirámide (web, España)
 g. La única solución razonable es *agarrar y twittearlo* (Twitter, Argentina)
 h. Como era de esperar, *va y la suelta* (*La Vanguardia*, prensa, España)
 i. Como era de esperar, *va y tira* otro papelito (Twitter, Argentina)
 j. La tía como era de esperar, *va y se salva* (web, España)
 k. Como esperaba, *va y dice* el muy cretino: «pues dile al panadero...» (web, España)

En todas las oraciones de (72) se evalúa en forma explícita el carácter previsible o esperable del evento introducido por las estructuras seudocoordinadas, lo que parece orientarse de manera exactamente contraria a la hipótesis de Jaque *et al.* de que expresan interruptividad como su significado principal.

Ahora bien, con los datos de (68-72) no apuntamos a demostrar que las estructuras seudocoordinadas con *va* y *agarra* sean incompatibles con una lectura interruptiva. De hecho, suelen emplearse con esa interpretación, como se observa en los diversos ejemplos relevados por Jaque *et al.* donde efectivamente el evento se considera disruptivo frente a otros eventos explícitos o implícitos; así, por ejemplo, en las oraciones de (66), *ir* o *agarrar* podrían funcionar perfectamente como pseudoauxiliares, reemplazando a *llegar* y *coger*. Lo que aquí sostenemos es que, dado que las estructuras seudocoordinadas con *va* y *agarra* también aparecen en contextos donde el evento se asume como esperable o lógico, como en (68-72), la interruptividad no puede ser su significado básico o primario. Como contraparte, vale resaltarlo, todos los ejemplos interruptivos registrados por Jaque *et al.* se ajustan sin inconveniente a la definición de aspecto aoristo: así, en las mismas oraciones de (66) las estructuras seudocoordinadas introducen desde un punto de vista aspectual una perspectiva global de los eventos desde su principio a su final, más allá de que esos eventos se presenten, discursivamente, como disruptivos respecto del contexto previo.

En ese sentido, si bien negamos que constituya el significado primario de las estructurasseudocoordinadas con *va* y *agarra*, nos resulta factible que la noción de interruptividad intervenga (o haya intervenido) en una primera gramaticalización, como una inferencia discursiva sistemática o cristalizada (cf. Company, 2004; Torres Cacoullos, 2016). Notemos, en ese sentido, que la propia caracterización de interruptividad de Jaque *et al.* (que entremezcla elementos vagamente aspectuales y modales) parece más próxima a la descripción tentativa de una inferencia discursiva que a la definición de un significado verdaderamente gramatical. Proponemos como alternativa, entonces, que podrían asociarse a la interruptividad los pseudoauxiliares menos gramaticalizados en (la mayoría de) las variedades actuales del español, como *salta / se pone / se mete / se manda / se larga / se lanza y + verbo*, que muestran más restricciones léxico–aspectuales en su combinatoria (cf. (29a) y notas 16 y 17) y además solo se flexionan en presente y pretérito perfecto simple (§3.2). En esa primera gramaticalización con interpretación interruptiva, las estructurasseudocoordinadas se ubicarían, desde un punto de vista sintáctico, en una proyección funcional relativamente baja y cercana a la capa léxica del verbo, semejante a la posición que proponen Cardinaletti y Giusti (2001) para los verbos de movimiento del italiano (cf. (6) en §2)³⁵. En el caso de pseudoauxiliares como *va*, *agarra* y otros³⁶, ese primer proceso dio lugar a una segunda gramaticalización en que la estructuraseudocoordinada pasó a expresar un significado puramente gramatical y a ubicarse en una posición mucho más alta en la cartografía sintáctica: el aspecto aoristo.

Asimismo, parece plausible que también en español los pseudoauxiliares sean verbos léxicos ensamblados directamente en nodos funcionales, tal como sostienen Cardinaletti y Giusti para los dialectos del sur italiano. Eso explicaría que se filtren o transparenten matices léxico–conceptuales del significado original del verbo,

35. Evidencia a favor de esta posición más baja sería que, a diferencia de (62), los pseudoauxiliares menos gramaticalizados no pueden combinarse con verbos principales que intervengan en perífrasis incoativas (cf. i), pero sí tomar como complementos construcciones causativas (cf. ii), tal como sucede con los verbos de movimiento en los dialectos del sur italiano (cf. Cardinaletti y Giusti, 2001 y (6) en §2):

(i) a. *el consumidor se manda y comienza a personalizar su producto en línea.

b. *Uno salta y empieza a mirar a todos lados menos la cámara.

c. *Cuando imprimo unos 25 cms, se mete y deja de imprimir.

d. *Amazon se pone y deja de vender la versión digital de SimCity.

(ii) Salto / me meto/ me lanzo/ me largo/ me mando/ me pongo y lo hago retroceder a los golpes.

36. Entre los pseudoauxiliares más gramaticalizados que expresan aspecto aoristo podemos incluir también, tentativamente, a *viene y + verbo y*, con más oscilaciones, *coge y + verbo*, *pilla y + verbo* (que no se usan en todas las variedades del español americanos), *llega y + verbo* (en Chile o Venezuela) y, tal vez, *arranca y + verbo* (cf. §3.2).

particularmente en la primera gramaticalización interruptiva, cuando se ubican en una posición más baja, cercana a la capa léxica de la oración³⁷. Recordemos, sin embargo, que cierta persistencia de valores semánticos filtrados desde el significado léxico original es esperable en la gramaticalización en general (cf. Bybee y Pagliucca, 1987), como, de hecho, se ha observado para perífrasis como la progresiva (cf. Torres Cacoullos, 2016)³⁸.

Una última y necesaria observación es que el análisis aquí defendido no implica que no existan ulteriores procesos de gramaticalización para las estructuras seudo-coordinadas. La propuesta que nos resulta más interesante y mejor sustentada en esa dirección es la de Jaque *et al.* (2019, p. 240-248), quienes plantean que hay una gramaticalización con interpretación mirativa en las estructuras seudocoordinadas cuando las formas verbales no coinciden en persona-número (73a) (cf. también nota 9), en tiempo (73b) o, en general, cuando se combinan con estados (74), en contra de nuestras predicciones en §4.1³⁹:

- (73) a. va y nos dicen esto
 b. va y era cierto (Jaque *et al.*, 2019, p. 224)
- (74) a. Imagínate que venga una gallega resabiosa a tomarnos cuentas. Ja. Aunque va y es una cuarentona divorciada y falta de marido (Dovalpage, *Muerte de un murciano en La Habana*, Cuba, CORPES XXI)
 b. [...] dice que en la casa de mi tía Carmen, dicen que estaban jugando en el gallinero ¿ve? los hijos de ella, y agarra y estaba... estaba el duende detrás del árbol, después los llama. (scielo.org.ar, Argentina, 2002) (Jaque *et al.*, 2019, p. 243)

37. Cuanto más baja sea la proyección funcional en que se ubiquen, mayores interacciones deberían tener los seudoauxiliares con el significado léxico del verbo principal. Así, si *agarra y* + verbo transmite cierto matiz de determinación o brusquedad, especialmente cuando se combina con verbos no agentivos (cf. §3.1 y, en particular, ejemplos de (16)), esa característica semántica no bloquea su combinación productiva con verbos no agentivos, ya que se liga con el Sintagma de Aspecto Aoristo (cf. Figura 1). Por el contrario, una franca imposibilidad de combinarse con verbos no agentivos se constata con los seudoauxiliares menos gramaticalizados, como los que tienen *se* anticausativo (cf. §3.2 y ejemplos (29a)), suponemos que porque se sitúan en una posición más baja.

38. En efecto, recordemos que, cuando los mismos verbos están gramaticalizados como auxiliares de perífrasis incoativas, presentan restricciones léxicas no muy distintas (cf. §3.3); así, los ejemplos (33) muestran la falta de compatibilidad de las perífrasis *meterse / mandarse / lanzarse / largarse a* + infinitivo (y, en menor medida, también *ponerse a* + infinitivo) con verbos principales sin argumento externo.

39. Una alternativa al análisis de Jaque *et al.* (2019) es considerar que el proceso de gramaticalización que se advierte en datos como (73-74) da lugar a adverbios o marcadores discursivos (cf. Bravo, 2020 para más detalles), una hipótesis que no discutiremos aquí por motivos de espacio y también porque se aleja demasiado de nuestros supuestos teóricos y del análisis concreto de las estructuras tal como ha sido expuesto en este trabajo.

Sería factible que construcciones como las de (73-74) reciban, efectivamente, una interpretación mirativa, tal como proponen Jaque *et al.* (2019). Aun en ese caso, el significado aoristo seguirá siendo el predominante en las estructurasseudocoordinadas gramaticalizadas del español. De hecho, en el CORPES, la combinación con estados (que dispara la interpretación mirativa) tiene un porcentaje de aparición mínimo: apenas supera el 1% en los siglos XX y XXI, según el propio estudio de Jaque *et al.* (2019, p. 243)⁴⁰.

Por su parte, consideramos que son indudablemente mirativas las construcciones con negación expletiva, tal como hemos propuesto para (75a) en Kornfeld (2019) (cf. también Jaque *et al.*, 2022):

- (75) a. Pues *no va y llueve*.
 b. Y *no llega y le confiesa todo* (Jaque *et al.*, 2022)

En los ejemplos de (75) *no va y* + verbo y *no llega y* + verbo están expresando centralmente la sorpresa frente a un evento que se contrapone con expectativas previas y pierden, por lo tanto, su carácter aspectual, si bien retienen algunas restricciones propias de laseudocoordinación (en particular, el uso exclusivo del presente histórico, cf. Kornfeld, 2019).

6. CONCLUSIONES

A lo largo de este artículo hemos apuntado a establecer las propiedades gramaticales y semánticas relevantes de las estructurasseudocoordinadas, concentrándonos en los casos de *va* y *agarra*, los pseudoauxiliares más gramaticalizados en la mayoría de las variedades del español.

El recorrido nos ha permitido concluir que el significado primario de las estructurasseudocoordinadas es el aspecto aoristo (en las mismas líneas de Coseriu, 1966) y ofrecer un análisis sintáctico consistente con sus propiedades gramaticales. Además, en las últimas secciones, hemos evaluado en detalle dos hipótesis alternativas para las estructurasseudocoordinadas: el análisis incoativo o ingresivo, defendido por García Sánchez (2007) y por Kornfeld (2019, 2022) (discutido en §4), y el análisis interruptivo propuesto por Jaque *et al.* (2018, 2019) (discutido en §5). En

40. Debe recordarse, además, que está bien establecida la estrecha relación que se establece entre aspecto aoristo y miratividad en diferentes lenguas (cf., por ejemplo, Aikhenvald, 2012, p. 463 para el hindi). Por lo tanto, la gramaticalización ulterior como una estructura mirativa no constituye *per se* un argumento a favor de un análisis interruptivo, como Jaque *et al.* (2019) parecen asumir.

ambos casos hemos rechazado esos análisis a favor de la hipótesis de que las estructuras seudocoordinadas más gramaticalizadas expresan con énfasis el aspecto aoristo como su significado primario, pero también hemos esbozado una posible integración de esas otras propuestas. Por un lado, observamos que las estructuras seudocoordinadas sí manifiestan aspecto ingresivo en combinación con las actividades, como resulta esperable con el aoristo (cf. ejemplos (44/61). Por otro lado, sugerimos que el significado interruptivo podría estar involucrado con el primer proceso de gramaticalización de las estructuras seudocoordinadas, puesto que la propia descripción de la interruptividad de Jaque *et al.* (que incorpora nociones vagamente aspectuales y modales) se asemeja más a una inferencia discursiva cristalizada que a un verdadero significado gramatical. Así, se interpretarían como interruptivas las estructuras seudocoordinadas con seudoauxiliares poco gramaticalizados (por ejemplo, los que llevan *se* anticausativo), que no presentan las mismas propiedades formales que los seudoauxiliares bien gramaticalizados como *va* o *agarra* (cf. §3.2 y ejemplos de (29a), notas 16, 17 y 35). Esa posibilidad (aquí solo sugerida) debería ser explorada en detalle en futuros trabajos.

Esperamos, así, haber ofrecido un análisis de las estructuras seudocoordinadas que integre las diversas piezas sueltas en una imagen coherente, para dar paso a nuevas polémicas sobre este tema fascinante de la gramática del español.

AGRADECIMIENTOS

Agradezco a quienes realizaron las revisiones anónimas de la versión original de este artículo, cuyas sugerencias y correcciones me han permitido mejorarlo considerablemente. Cualquier error que persista es de mi entera responsabilidad.

REFERENCIAS

- Aikhenvald, A. (2011). Multi-verb constructions: setting the scene. En A. Aikhenvald y P. Muysken (Eds.). *Multi-verb constructions. A view from the Americas* (pp. 1-26). Leiden: Brill.
- Aikhenvald, A. (2012). The essence of mirativity. *Linguistic Typology*, 16(3), 435-485.
- Anderson, G. (2006). *Auxiliary verb constructions*. Oxford: Oxford University Press.
- Bary, C. y Egg, M. (2012). Variety in Ancient Greek aspect interpretation. *Linguistics and Philosophy*, 35, 111-134
- Bosque, I. y Gutiérrez-Rexach, J. (2008). *Fundamentos de sintaxis formal*. Madrid: Akal. DOI: <https://doi.org/10.1515/lity-2012-0017>
- Bravo, A. (2014). The Spanish auxiliary *ir* 'to go': from resultative motion verb to focus marker. En M. Devos y J. van der Wal (Eds.). *COME and GO off the beaten grammaticalization path* (pp. 305-351). Berlín: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110335989.187>

- Bravo, A. (2020). On pseudo-coordination in Spanish. *Borealis – An International Journal of Hispanic Linguistics*, 9(1), 125-180. DOI: <https://doi.org/10.7557/1.9.1.5365>
- Bybee, J. y Pagliuca, W. (1987). The evolution of future meaning. En A. Giacalone Ramat, O. Carruba y G. Bernini (Eds.). *Papers from the 7th International Conference on Historical Linguistics* (pp. 109-122). Amsterdam: John Benjamins.
DOI: <https://doi.org/10.1075/cilt.48.09byb>
- Camus Bergareche, B. (2004). Perífrasis verbales y expresión del aspecto en español. En L. García Fernández (Ed.), *El pretérito imperfecto* (pp. 511-572). Madrid: Gredos.
- Cardinaletti, A. y Giusti, G. (2001). 'Semi-lexical' motion verbs in Romance and Germanic. En N. Corver y H. Van Riemsdijk (Eds.). *Semi-lexical categories* (pp. 371-414). Berlín: De Gruyter. DOI: <https://doi.org/10.1515/9783110874006.371>
- Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Nueva York: Oxford University Press.
- Cinque, G. (2006). *Restructuring and functional heads*. Nueva York: Oxford University Press.
- Company Company, C. (2004). Gramaticalización por subjetivización como prescindibilidad de la sintaxis. *Nueva Revista de Filología Hispánica*, 52(1), 1-27.
DOI: <https://doi.org/10.24201/nrfh.v52i1.2226>
- Coseriu, E. (1966). *Tomo y me voy*. Un problema de sintaxis comparada europea. En *Estudios de lingüística románica* (pp. 79-151). Madrid: Gredos, 1977.
- Chambers, J. (2004). Dynamic typology and vernacular universals. En B. Kortmann (Ed.). *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective* (pp. 127-145). Berlín/ Nueva York: Mouton de Gruyter.
DOI: <https://doi.org/10.1515/9783110197327.127>
- DeLancey, S. (1997). Mirativity: The grammatical marking of unexpected information. *Linguistic Typology*, 1, 33-52. DOI: <https://doi.org/10.1515/lity.1997.1.1.33>
- De Vos, M. (2005). *The syntax of verbal pseudo-coordination in English and Afrikaans*. Tesis doctoral, Universidad de Leiden.
- Di Tullio, Á. (2005). *Manual de Gramática del español*. Buenos Aires: La Isla de la Luna.
- Di Tullio, Á. (2006). Auxiliares y operadores aspectuales en el español rioplatense. *Signo y Seña*, 15, 267-285. <http://repositorio.filo.uba.ar:8080/xmlui/handle/filodigital/14138>
- Di Tullio, Á. y E. Pato (2022). Los universales vernáculos y la gramática del español. En Á Di Tullio y E. Pato (Eds.). *En torno a los universales vernáculos del español* (pp. 9-24). Madrid: Iberoamericana/ Vervuert.
- Dowty, D. (1977). Towards a semantic analysis of verb aspect and the English 'imperfective' progressive. *Linguistics and Philosophy*, 1, 45-77.
<https://www.jstor.org/stable/25000028>
- Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, 67, 547-619.
DOI: <https://doi.org/10.2307/415037>
- Fábregas, A. (2016). A note on how and why 'state + aorist = achievement'. *Onomázein*, 33, 57-68. DOI: <https://doi.org/10.7764/onomazein.33.5>
- Fábregas, A. (2019). Periphrases in Spanish: properties, diagnostics and research questions. *Borealis*, 8(2), 1-82. DOI: <https://doi.org/10.7557/1.8.2.4944>
- Fernández de Castro, F. (1999). *Las perífrasis verbales en español actual*. Madrid: Gredos.

- Garachana Camarero, M. (2019). De cuando «*ir a + INFINITIVO*» no se dirige al futuro. Construcciones gramaticales de pasado, aspecto culminativo y sentido focalizador, una aproximación desde la gramática de construcciones. *Revista Española de Lingüística*, 49, 119-146. <http://revista.sel.edu.es/index.php/revista/article/view/2005>
- García Fernández, L. (dir.) (2006). *Diccionario de perífrasis verbales*. Madrid: Gredos.
- García Sánchez, J. (2007). *Tomo y me voy*: expresión plena y elipsis. En B. Mariscal y M. T. Miaja de la Peña (Eds.). *Actas del XV Congreso de la Asociación Internacional de Hispanistas «Las dos orillas»* (pp. 163-176). México: Fondo de Cultura Económica/Asociación Internacional de Hispanistas/ El Colegio de México.
- Jaque, M., González, C., Guerrero, S., Hasler, F. y Orqueda, V. (2018). *Es llegar y llevar*: construcciones multiverbales de verbo finito coordinadas en español. *Lenguas Modernas*, 52, 163-186.
- Jaque, M., González, C., Guerrero, S., Hasler, F., Orqueda, V. y Covarrubias, M. (2019). Subjetivización, alcance y gramaticalización: de la interruptividad a la admiratividad en construcciones del tipo llegar y + VF1. *Boletín de Filología*, 54(2), 219-256. <https://lenguasmodernas.uchile.cl/index.php/LM/article/view/52851>
- Jaque, M., González, C., Guerrero, S., Hasler, F., Orqueda, V. y Covarrubias, M. (2022). A la altura de las expectativas: interacciones entre la negación y construcciones multiverbales del tipo *llegar y + VF*. *Verba*, 49, 1-36.
- Kany, Ch. (1945). *Sintaxis hispanoamericana*. Madrid: Gredos, 1969.
- Keniston, H. (1936). Verbal aspect in Spanish. *Hispania*, 19(2), 163-176.
- Klein, W. (1992). The present perfect puzzle. *Language*, 68, 525-552.
DOI: <https://doi.org/10.2307/415793>
- Kornfeld, L. (2019). Expresión de la sorpresa, miratividad y gramaticalización de verbos inacusativos en español. *Borealis*, 8(2), 165-197. DOI: <https://doi.org/10.7557/1.8.2.4913>
- Kornfeld, L. (2022). Estructuras pseudocoordinadas: un universal vernáculo. En Á. Di Tullio y E. Pato (Eds.). *Universales vernáculos en español* (pp. 253-280). Madrid: Iberoamericana/ Vervuert.
- Morimoto, Y. (1998) *El aspecto léxico: delimitación*. Madrid: Arco Libros.
- Olbertz, H. (1998). *Verbal periphrases in a functional grammar of Spanish*. Berlín: Mouton de Gruyter.
- Pérez Saldanya, M. (2003). La gramaticalización del verbo *ir* en construcciones narrativas del español. *Medievalia*, 35, 62-89.
DOI: <https://doi.org/10.19130/medievalia.35.2003.138>
- Real Academia Española y Asociación de Academias de la Lengua Española (2010). *Manual de la nueva gramática de la lengua española*. Madrid: Espasa Calpe.
- Ross, D. (2014). El origen de los estudios sobre la pseudocoordinación verbal. *Diálogo de la Lengua*, VI, 116-132.
- Ross, D. (2017). Pseudocoordinación del tipo *tomar y* en Eurasia: 50 años después. Conferencia en el VI Congreso Internacional de Lingüística Coseriuana: Actualidad y futuro del pensamiento de Eugenio Coseriu. Lima (Perú), 3 de agosto de 2017.
DOI: <https://doi.org/10.7557/1.9.1.5365>
- Sanz, M. y Laka, I. (2002). Oraciones transitivas con *se*. El modo de acción en la sintaxis. En C. Sánchez López (Ed.). *Las construcciones con «se»* (pp. 311-337). Madrid: Visor.
- Smith, C. (1991). *The parameter of aspect*. Dordrecht: Kluwer.

- Torres Cacoullos, R. (2016). Gramaticalización. En J. Gutiérrez-Rexach (Ed.), *Enciclopedia de lingüística hispánica* (pp. 504-514). Londres/ Nueva York: Routledge, volumen 2.
- Vendler, Z. (1957). Verbs and Times. *The Philosophical Review*, 66(2), 143-160.
<http://links.jstor.org/sici?sici=0031-8108%28195704%2966%3A2%3C143%3AVAT%3E2.o.CO%3B2-2>
- Wiklund, A. (2008). Creating surprise in complex predication. *Nordlyd*, 35, 163-187.
DOI: <https://doi.org/10.7557/12.141>

EVALUACIÓN MORFOLÓGICA DE LOS VOCABULARIOS DE SUBPALABRAS UTILIZADOS POR LOS GRANDES MODELOS DE LENGUAJE

Óscar GARCÍA-SIERRA¹

Universidad Complutense de Madrid; dezzai

Ana FERNÁNDEZ-PAMPILLÓN CESTEROS²

Universidad Complutense de Madrid


Miguel ORTEGA-MARTÍN³


Universidad Complutense de Madrid; dezzai

Resumen

Con el auge de los grandes modelos del lenguaje neuronales, especialmente aquellos basados en *Transformers*, la tradicional segmentación en palabras y morfemas que empleaba reglas lingüísticas ha sido reemplazada por algoritmos de segmentación estadísticos. Estos algoritmos son mucho más eficientes y, sin necesidad de intervención humana, son capaces de, a partir de corpus de millones de palabras, construir el vocabulario de palabras y subpalabras que necesitan los grandes modelos del lenguaje monolingües o multilingües. Ocurre, sin embargo, que estas subpalabras no se corresponden siempre con morfemas y esto repercute negativamente en el funcionamiento de los modelos del lenguaje que utilizan estos segmentadores. Cuánto se alejan los vocabularios estadísticos de un vocabulario real de palabras y morfemas de una lengua –lo que denominamos calidad morfológica del vocabulario–, y cuánto repercute esta falta de calidad en la eficacia de los grandes modelos del lenguaje son cuestiones todavía sin resolver. Este artículo aborda la primera cuestión, la calidad morfológica de los vocabularios, aportando un método de evaluación basado en tres medidas de calidad –relevancia, coherencia y corrección morfológica–, y un procedimiento para evaluarlas. El método se aplica para medir la calidad de los vocabularios generados por tres algoritmos de segmentación en subpalabras, *BPE*, *WordPiece* y *Unigram*, utilizados mayoritariamente para la construcción de los grandes modelos del lenguaje. Los resultados

1. oscarg02@ucm.es;  <https://orcid.org/0000-0002-8828-7338>

2. apampi@filol.ucm.es;  <https://orcid.org/0000-0002-6606-0159>

3. m.ortega@ucm.es;  <https://orcid.org/0000-0002-1880-5048>

que hemos obtenido indican que la calidad morfológica de los mismos es muy baja, por lo que merece la pena buscar nuevas soluciones para mejorar la calidad de los vocabularios de los grandes modelos del lenguaje.

Palabras clave: segmentación; morfemas; subpalabras; grandes modelos del lenguaje; lengua española.

MORPHOLOGICAL EVALUATION OF SUBWORD VOCABULARIES USED BY LARGE LANGUAGE MODELS

Abstract

Traditional tokenization methods using linguistic rules have been replaced by statistical segmentation algorithms. Although these algorithms show a higher efficiency and are capable of building subword vocabularies from large corpora without human supervision, these subwords do not consistently correspond to morphemes. This paper addresses this issue by proposing an evaluation methodology and applying it to the morphological quality of Spanish vocabularies produced by three prominent subword tokenization algorithms –*BPE*, *WordPiece*, and *Unigram*– commonly used in Large Language Models (LLMs). Three gold standards were created to measure relevance, coherence, and morphological accuracy of vocabularies of six tokenizers trained on Spanish corpus, exploring different vocabulary sizes. Evaluation results indicate that none of the three algorithms is suitable for accurately representing Spanish morphology.

Keywords: tokenizing; morphemes; subwords; large language models; Spanish language.

RECIBIDO: 08/03/2024

APROBADO: 30/06/2024

1. INTRODUCCIÓN

La segmentación (*tokenization* en inglés) es una de las fases más relevantes en el procesamiento del lenguaje natural. Implica la división del texto en unidades más pequeñas llamadas *tokens* (Friedman, 2023). Estos *tokens* pueden ser morfemas, palabras, sintagmas, frases o incluso caracteres, dependiendo de la aplicación específica. La segmentación es también un componente clave en la preparación de los datos de entrada para entrenar las redes neuronales artificiales en la medida en que representan las unidades discretas de información a partir de las cuales se construyen los modelos del lenguaje (Friedman, 2023). Por ejemplo, el modelo *GPT-2* depende de un vocabulario de cincuenta mil *tokens*, aproximadamente.

Así, en el entrenamiento de las actuales redes neuronales de tipo *Transformer* (Vaswani *et al.*, 2017) para generar (también se utiliza el término *aprender*), los grandes modelos del lenguaje (en adelante, utilizaremos *modelos del lenguaje*) se utilizan vocabularios en los que los *tokens* son palabras y, mayoritariamente, subpalabras. Una subpalabra es una parte de la palabra que tiene una frecuencia relevante de aparición en el texto segmentado y que no se corresponde necesariamente con un morfema. También, para poder utilizarse (como generador de texto, clasificador, traductor o sistema de diálogo, entre otros), los modelos del lenguaje necesitan utilizar el vocabulario de palabras y subpalabras con el que han sido creados.

Los segmentadores que se emplean actualmente para generar los vocabularios de los modelos del lenguaje utilizan estrategias estadísticas que aplican a enormes corpus de texto, en vez del conocimiento lingüístico formalizado en las gramáticas tradicionales basadas en reglas. Principalmente, se fundamentan en la frecuencia con la que ciertas cadenas de caracteres (palabras o subpalabras) aparecen en el corpus de entrenamiento, lo que produce como resultado que las palabras con una frecuencia relevante en el corpus se segmenten como un solo *token*, y, sin embargo, las palabras menos frecuentes se dividan en varios *tokens* de tipo subpalabra con una mayor frecuencia de aparición. Por ejemplo, el segmentador *WordPiece* del modelo de lenguaje *BETO* trata la palabra *población* como un solo *token*, pero descompone *desprestigiar* en varios *tokens*: *despre*, *##st*, *##igi*, *##ar*.

Entre los segmentadores utilizados para generar los vocabularios de los modelos del lenguaje destacan tres por ser los que se utilizan mayoritariamente: *Byte-Pair Encoding (BPE)* (Sennrich, 2015), *WordPiece* (Schuster, 2012; Wu, 2016) y *Unigram* (Kudo, 2018). Así, por ejemplo, el modelo *BERT* (Devlin *et al.*, 2019) utiliza *WordPiece*, *GPT* (Radford *et al.*, 2018) o *RoBERTa* (Liu *et al.*, 2019) utilizan *BPE*, y *Albert* (Lan *et al.*, 2019) emplea *Unigram*. Son estos tres segmentadores los que hemos seleccionado para llevar a cabo la evaluación empírica de calidad que presentamos en este artículo.

La ventaja indudable de los actuales segmentadores estadísticos frente a los simbólicos basados en reglas lingüísticas es su eficacia: son capaces de construir vocabularios de decenas de miles de *tokens* a partir de corpus de texto de gran tamaño sin intervención humana alguna. Además, la segmentación estadística permite obtener vocabularios independientes de las lenguas, lo que los hace especialmente útiles cuando se trabaja con modelos multilingües.

El inconveniente es, como demuestran diversos estudios, que la segmentación en subpalabras no se corresponde, en un alto porcentaje, con morfemas, perdiéndose el conocimiento lingüístico básico sobre el que se construyen los modelos del lenguaje y, en consecuencia, perdiéndose la eficacia y fiabilidad de dichos modelos (Church, 2020; Bostrom y Durrett, 2020; Hoffman *et al.*, 2021; Park, 2020). En este sentido, Church (2020) analiza cualitativamente algunas palabras complejas en inglés y llega a la conclusión de que estas se descomponen en demasiadas subpalabras que no se corresponden ni con palabras ni con morfemas reales, a pesar de lo que señalan algunos autores (Song *et al.*, 2020). Por su parte, Bostrom y Durrett (2020) comparan los segmentadores *BPE* y *Unigram* en inglés y japonés, y concluyen que el segundo segmentador genera mejores vocabularios de morfemas en ambos idiomas que el primero. Hoffman *et al.* (2021) analizan el tratamiento que hace *WordPiece* de las palabras complejas en inglés, concluyen también que generan subpalabras que no se corresponden con morfemas reales, y comprueban que, en ese mismo idioma, un modelo con un vocabulario de morfemas mejora los resultados respecto de un modelo que utiliza subpalabras puramente estadísticas. Por último, Park (2020) realiza un estudio similar al de Hoffman *et al.* (2021), pero para el coreano.

Denominamos *calidad morfológica del vocabulario de un modelo del lenguaje en una lengua* a su grado de semejanza respecto de un vocabulario real de morfemas y palabras de esa misma lengua. El problema que surge, antes de abordar la mejora de la calidad morfológica de los vocabularios de los modelos del lenguaje, es que –hasta donde conocen los autores de este trabajo– no existe un método para evaluar y contrastar de forma objetiva la calidad lingüística de estos vocabularios.

En este artículo proponemos una solución a la cuestión de cómo evaluar la calidad morfológica, basándonos en los resultados de los trabajos previos, y la aplicamos a la evaluación de los tres segmentadores (*BPE*, *WordPiece* y *Unigram*) trabajando con textos del español. El artículo se ha organizado de la forma siguiente: en la sección segunda presentamos el método de evaluación de la calidad morfológica; en la sección tercera se describe su aplicación a los tres segmentadores; en las secciones cuarta y quinta se discuten los resultados, primero los valores cuantitativos obtenidos para cada criterio de evaluación y, después, cualitativamente mediante un análisis de errores. En la sexta y última sección se presentan las conclusiones y el trabajo futuro.

2. PROPUESTA DE MÉTODO DE EVALUACIÓN DE LA CALIDAD MORFOLÓGICA DE LOS VOCABULARIOS DE LOS MODELOS DEL LENGUAJE

El método de evaluación de la calidad morfológica se fundamenta en dos elementos: i) tres criterios de calidad: la relevancia, coherencia y corrección morfológicas; y, ii), tres conjuntos de datos de validación para medir el grado de cumplimiento de cada uno de los tres criterios: primero, una lista de pares palabra-segmentación que sea una muestra equilibrada de palabras de las diferentes categorías morfosintácticas y que contenga todos los morfemas de la lengua junto con sus segmentaciones morfológicas para evaluar la corrección morfológica; segundo, una lista de los morfemas de la lengua, para evaluar la coherencia morfológica; y, finalmente, a partir de las listas anteriores, una lista de pares palabra-morfema que servirá para evaluar la relevancia morfológica. La definición de cada uno de los criterios de evaluación se describe en la subsección 2.1, mientras que el procedimiento para medir su cumplimiento se describe en la subsección 2.2.

2.1. Los criterios de evaluación

Recogiendo y ampliando los trabajos anteriormente mencionados de Bostrom y Durret (2020), Church (2020), Hoffman *et al.* (2021) y Park (2020) proponemos utilizar tres criterios o parámetros para la evaluación de la calidad de los vocabularios de los modelos del lenguaje. Estos criterios son: (1) la relevancia morfológica, (2) la coherencia morfológica y, (3) la corrección morfológica.

(1) La relevancia morfológica

La relevancia morfológica mide cuánto de morfológico es el vocabulario. Se calcula como la intersección entre el vocabulario que genera el segmentador –y que posteriormente utilizará el modelo de lenguaje– y los morfemas reales de la lengua (Figura 1). Así, por ejemplo, en el vocabulario del segmentador del modelo del lenguaje *BETO* está incluida la subpalabra *códi* que no es un morfema real del español, no está incluido el sufijo del español *-érrimo* y sí está incluido el morfema del español *-idad*. Este último morfema sería el que contaría para calcular la relevancia morfológica.

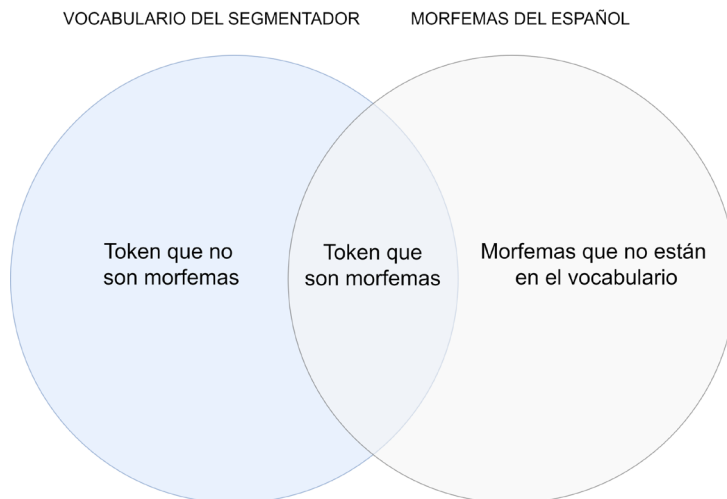


Figura 1. Esquema de concepto de relevancia morfológica

Para medir la relevancia proponemos el uso de las métricas tradicionales de precisión, cobertura y valor F_1 ⁴. Los valores se encuentran en un rango entre 0 (cuando no hay intersección entre vocabulario del modelo y conjunto de morfemas) y 1 (cuando la intersección es total). También pueden utilizarse valores porcentuales.

Las métricas de precisión, cobertura o valor F_1 han sido ya utilizados para analizar la calidad de los vocabularios en otras lenguas. Así, para la lengua coreana, Park *et al.* (2020) utilizan la precisión para medir qué porcentaje de los *tokens* del vocabulario del segmentador son morfemas correctos del coreano. Bostrom y Durrett (2020) utilizan las tres métricas, precisión, cobertura y F_1 , para medir la intersección entre los *tokens* generados por un segmentador *BPE* y una porción de morfemas reales del inglés. Por último, Hoffman *et al.*, (2022) emplean un corpus de segmentación morfológica que incluye la derivación y composición, llamado *CELEX* (Van de Wouden, 1990) para extraer una lista de morfemas del inglés, y posteriormente emplean la medida de cobertura para comprobar qué porcentaje de esos elementos están en el vocabulario de su segmentador.

4. La precisión, en este caso, es el número de *tokens* que son morfemas del vocabulario del segmentador (la intersección en la Figura 1) dividido entre el total de *tokens* del vocabulario del segmentador. La cobertura (*recall* en inglés) es el número de *tokens* que son morfemas del vocabulario del segmentador (la intersección) dividido por el total de morfemas de la lengua. El valor F_1 es la media armónica de la precisión y la cobertura.

En nuestra propuesta, a diferencia de los trabajos previos, la relevancia morfológica se medirá con las tres métricas para poder comparar de forma más objetiva este criterio en diferentes vocabularios.

(2) *La coherencia morfológica*

El segundo criterio de evaluación que proponemos es la coherencia morfológica, que mide con qué frecuencia las palabras con una misma estructura morfológica (palabras que comparten un determinado morfema) se segmentan, de acuerdo con alguna de las opciones siguientes (Figura 2):

- como un solo *token* (no hay segmentación)
- como varios *tokens* de modo que el morfema correcto se corresponde con uno de los *tokens* obtenidos en la segmentación
- como varios *tokens* de modo que el morfema correcto no se corresponde con ningún *token* de la segmentación.

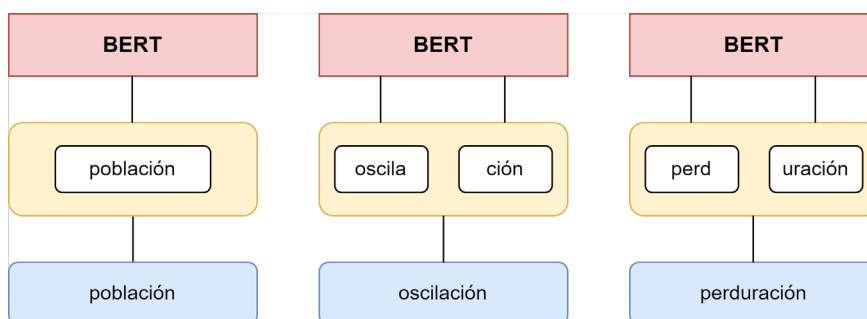


Figura 2. Falta de coherencia morfológica en vocabulario del segmentador utilizado por el modelo del lenguaje *BETO*

La Figura 2 muestra un ejemplo real de falta de coherencia morfológica de los *tokens* obtenidos por el segmentador *WordPiece* que utiliza el modelo del lenguaje *BETO* (Cañete *et al.*, 2023). En concreto, la figura muestra cómo tres palabras que comparten el sufijo *-ción* son segmentadas incoherentemente de tres formas diferentes:

- aparece sin segmentar en la palabra *población*, debido a que es frecuente en el corpus de entrenamiento;
- aparece correctamente segmentado como dos subpalabras en el caso de *oscilación* (*oscila* y *-ción*); y,

- en la palabra perduración, la segmentación es errónea en *perd* y *uración*, con lo que el sufijo *-ción* aparece perdido en la subpalabra *uración*.

En este sentido, el modelo de lenguaje *BETO* aprende que esas tres palabras no comparen el morfema, por lo que se pierde la posibilidad de modelar la similitud morfológica que realmente existe entre las tres palabras *población*, *oscilación* y *perduración*.

En los trabajos previos de Church (2020) y Hoffman *et al.* (2021) se apunta la existencia del problema de falta de coherencia morfológica al segmentar textos en inglés, pero ninguno de los trabajos lo evalúa cuantitativamente. Hoffman *et al.* (2021) analizan cualitativamente el sufijo *-ize*, y comprueban que muchas palabras que lo contienen no tienen en la práctica ningún token en común, puesto que se dan los tres casos ya explicados para el sufijo en español. Church (2020) analiza los casos en los que al añadir un prefijo a una palabra se pierde la coherencia entre las segmentaciones de la forma base y la forma derivada.

Nosotros proponemos medir, para cada tipo de morfema, qué porcentaje del total de palabras representan cada uno de los siguientes casos:

- palabras que contienen el morfema y que se segmentan como un solo token.
- palabras que contienen el morfema y se segmentan como varios tokens, pero el morfema no se corresponde con un token.
- palabras que contienen el morfema y se segmentan como varios tokens en las que morfema se corresponde con un token.

(3) Corrección morfológica

El criterio de corrección en la segmentación morfológica tiene como objetivo evaluar cómo funcionan los segmentadores a la hora de segmentar completamente una palabra en morfemas. A diferencia del criterio anterior, aquí se mide si el segmentador divide *correctamente* los morfemas de la palabra en cuestión. Así, puede ocurrir que un segmentador separe de forma totalmente coherente una palabra, pero los *tokens* que obtiene son erróneos porque no se corresponden con las palabras y morfemas correctos de la lengua. Por ejemplo, el segmentador del modelo del lenguaje *BETO* segmenta bien el prefijo de la palabra *contrapuntos*, pero no el resto de la palabra: [*contra*, *pu*, *n*, *tos*].

Para medir la corrección morfológica proponemos utilizar la métrica de corrección. El resultado de una segmentación es correcto si todos los *tokens* segmentados coinciden con los morfemas etiquetados. El rango de la corrección varía entre 0 y

100%, indicando con 0 que ninguna palabra se ha segmentado correctamente y con valor 100 que todas las palabras han sido segmentadas correctamente.

En los trabajos previos de Bostrom y Durret (2020) y Hoffman *et al.* (2021) se evalúa cualitativamente la corrección de los segmentadores de subpalabras a la hora de segmentar morfológicamente palabras en inglés, pero ninguno lo hace cuantitativamente.

2.2. Procedimiento de evaluación

Proponemos que, para la evaluación de la calidad de los segmentadores y sus vocabularios, los criterios anteriormente definidos de relevancia morfológica de los vocabularios, coherencia y corrección morfológica se calculen utilizando el procedimiento siguiente en cinco pasos:

Paso 1. Creación de tres conjuntos de datos de evaluación, uno para cada criterio:

- i) Para el criterio de relevancia morfológica de los vocabularios se debe crear un conjunto de datos que consiste en la lista de morfemas de la lengua, a ser posible divididos por su tipología: prefijo, sufijo y raíz.
- ii) Para el criterio de la coherencia morfológica se debe construir una lista de pares palabra y morfema. Esta lista se crea a partir del conjunto de datos anterior de morfemas y debe contener todos los morfemas de la lengua.
- iii) Para evaluar la corrección morfológica se crea una lista de palabras aleatoria etiquetadas con su(s) correspondiente(s) segmentación(es) morfológica(s). Para que sea una muestra representativa, esta lista debe contener todos los morfemas de la lengua al menos una vez y mantener la proporción de categorías gramaticales del *Diccionario de la lengua española* (en adelante, *DLE*; RAE-ASALE, en línea).

En la sección siguiente, se detallará un posible método de construcción, así como el contenido de los tres conjuntos de datos.

Paso 2. Selección de los algoritmos de segmentación de tokens en subpalabras que se van a evaluar.

Paso 3. Selección del corpus de entrenamiento de los segmentadores que se van a evaluar (seleccionados en el paso 2). Básicamente, se debe tener en cuenta que el corpus sea suficientemente representativo de la lengua y que se ajuste a los

recursos informáticos disponibles. Este corpus solo se utilizará para entrenar los segmentadores y generar los vocabularios.

Paso 4. Generación de los vocabularios con cada segmentador a evaluar mediante el entrenamiento de dichos segmentadores (seleccionados en el paso 2) con el corpus de entrenamiento creado en el paso 3.

Paso 5. Cálculo del nivel de cumplimiento de cada criterio utilizando los conjuntos de datos de evaluación creados en el paso 1. Además, en el caso del conjunto de datos de para evaluar la corrección morfológica, también es interesante almacenar el número de tokens utilizados por cada segmentador a la hora de segmentar, para posteriormente calcular la media de tokens utilizados por cada segmentador.

3. APLICACIÓN A LA EVALUACIÓN DE LOS VOCABULARIOS *BPE*, *WORDPIECE* Y *UNIGRAM* EN ESPAÑOL

Con el fin de probar la viabilidad del método de evaluación de la calidad morfológica de los segmentadores hemos llevado a cabo una prueba de concepto utilizando los tres segmentadores, *BEP*, *WordPiece* y *Unigram*, y generando vocabularios de dos tamaños diferentes por cada segmentador. En esta sección detallamos el proceso de aplicación del método y en la sección siguiente se muestran los resultados y su discusión. Así, el proceso de evaluación se llevó a cabo en los cinco pasos propuestos en la sección 2 de la forma siguiente:

Paso 1. En este paso se construyen los tres conjuntos de datos que servirán para realizar la evaluación: conjunto de datos para la relevancia morfológica, para la coherencia morfológica y para la corrección morfológica. Se comienza construyendo el conjunto de datos para evaluar la corrección.

En primer lugar, se ha construido manualmente el conjunto de datos para evaluar la corrección morfológica, puesto que este conjunto de datos se ha usado también para evaluar un segmentador propio basado en reglas morfológicas del español, que ha permitido construir automáticamente los conjuntos de datos de evaluación restantes con la corrección suficiente.

Para ello, se han anotado manualmente 1.231 palabras del español con todas sus segmentaciones morfológicas posibles, siguiendo las reglas de la *Nueva gramática de lengua española* (en adelante, *NGLE*). La selección de palabras se ha realizado aleatoriamente teniendo en cuenta que todos los morfemas del español estén representados al menos una vez y se ha respetado la proporción de categorías

gramaticales del *DLE*. La Tabla 1 contiene los totales y un ejemplo de cada categoría gramatical.

Categoría gramatical	Total	Ejemplos
Nombres	481	[‘camiones’: {‘NOUN’: [‘camion’, ‘es’]}]
Verbos	445	[‘dejan’: {‘VERB’: [‘dej’, ‘an’]}]
Adjetivos	299	[‘adecuadas’: {‘ADJ’: [‘adecu’, ‘ad’, ‘a’, ‘s’]}]
Pronombres	61	[‘estos’: {‘PRON’: [‘est’, ‘o’, ‘s’]}]
Adverbios	59	[‘aquí’: {‘ADV’: [‘aquí’]}]
Determinantes	30	[‘mis’: {‘DET’: [‘mi’, ‘s’]}]
Preposiciones	22	[‘desde’: {‘ADP’: [‘desde’]}]
Conjunciones	14	[‘pero’: {‘CONJ’: [‘pero’]}]

Tabla 1. Conjunto de datos para evaluar la corrección morfológica

En segundo lugar, se ha construido el conjunto de datos para medir la relevancia morfológica de los vocabularios. Los morfemas han sido extraídos automáticamente de la *NGLE* y del *Diccionario de uso del español* (Moliner, 1967/2009; en adelante, *DUE*), mientras que las raíces han sido generadas automáticamente por un segmentador de autoría propia y que ha sido evaluado en el conjunto de datos de 1.231 palabras con una corrección superior al 99%. En concreto se han seleccionado las 5.000 raíces más frecuentes del *DLE*. La Tabla 2 contiene los morfemas totales y un ejemplo de cada tipo de morfema.

Tipo de morfema	Total	Ejemplos
Prefijos	61	[‘des’, ‘re’, ...]
Sufijos	175	[‘mos’, ‘ción’, ...]
Raíces	5,000	[‘dej’, ‘alt’, ...]

Tabla 2. Conjunto de datos para evaluar la relevancia morfológica

Por último, para evaluar la coherencia morfológica se ha construido un tercer conjunto de datos de evaluación. Partiendo del conjunto de prefijos, raíces y sufijos empleados para evaluar la relevancia morfológica (Tabla 2) y las palabras del *DLE* se ha creado la lista de pares (palabra, morfema). La Tabla 3 contiene los totales de pares (palabra, morfema) para cada tipo de morfema, así como una muestra de ejemplos. Hay que señalar que una misma palabra puede encontrarse varias veces en este conjunto de datos, puesto que aparecerá tantas veces como morfemas tenga. Así,

por ejemplo, en el caso de la palabra *remodelación* el conjunto de datos contendrá el par (*remodelación, re-*), (*remodelación, modela*) y (*remodelación, -ción*). Además, el conjunto de datos contiene 3.920 verbos con pronombres enclíticos que se han descargado del corpus *Spanish Web 2018* mediante la herramienta de análisis textual *Sketch Engine*. Se ha elegido este recurso por su total accesibilidad y por ser suficientemente representativo del vocabulario actual del español.

Tipo de morfema	Total de palabras	Ejemplos
Prefijos	205.792	(<i>remodelación, re-</i>)
Raíces	234.564	(<i>ubicásemos, ubic</i>)
Sufijos	1.575.207	(<i>fijación, -ción</i>)
Clíticos	3.920	(<i>dilo, lo</i>)

Tabla 3. Conjunto de datos de evaluación de la coherencia morfológica

Como resultado de este paso se dispone de los tres conjuntos de evaluación para evaluar en el paso 5.

Paso 2. Se seleccionan los algoritmos que se van a evaluar que, como hemos indicado anteriormente, son *WordPiece*, *BPE* y *Unigram* por ser los principales algoritmos de segmentación en subpalabras utilizados actualmente por los grandes modelos del lenguaje (Church, 2020) y por estar disponibles en la plataforma *Hugging Face*.

Paso 3. Para la selección del corpus de entrenamiento hemos revisado la sección de conjuntos de datos disponible en la plataforma *Hugging Face* que es la principal plataforma de modelos y conjuntos de datos de la inteligencia artificial, donde existen más de 400 corpus para modelar lenguaje en español. El corpus, en nuestro caso, debe ajustarse a las limitaciones de almacenamiento de *Google Collab*, la plataforma que tenemos disponible para el entrenamiento de los segmentadores (12gb de RAM y 100gb de disco duro).

Se ha elegido el corpus *Oscar* (Ortiz Suárez *et al.*, 2019; Ortiz Suárez *et al.*, 2020), que es el segundo corpus más utilizado para modelar lenguaje en español de esta plataforma solo ligeramente por detrás del de *Wikipedia*⁵. Es un corpus con el contenido de diferentes webs preparado para utilizarse (limpio y postprocesado). Se ha preferido *Oscar* frente a *Wikipedia* porque este último contiene

5. https://huggingface.co/datasets?task_ids=task_ids:language-modeling&language=language:es&sort=trending

presumiblemente más nombres propios que no interesan desde un punto de vista morfológico. En concreto, se ha utilizado la versión pequeña de *Oscar*⁶, que consta de unas 600.000 frases, tiene un tamaño de 10Gb y se adapta mejor a nuestras limitaciones computacionales.

El resultado de este paso son 600.000 frases en español, limpias y disponibles para entrenar con ellas los tres segmentadores. Las frases se agrupan en lotes de 1.000 frases para facilitar y optimizar su posterior procesamiento durante el entrenamiento, siguiendo el proceso propuesto por *Hugging Face*⁷.

Paso 4. Creación de los vocabularios. Los vocabularios se generan mediante el entrenamiento de los segmentadores conforme la guía propuesta en *Hugging Face*. Los únicos hiperparámetros ajustables son el tamaño de los lotes (que, como ya se ha mencionado se ha fijado en 1.000 frases), y el de los tamaños de los vocabularios (que, como también se ha explicado, se han fijado en 31.000 y 52.000 *tokens*).

Para cada segmentador se han creado, por lo tanto, dos versiones de vocabulario, versión 31 y versión 52, correspondientes a los tamaños 31.000 y 52.000 *tokens*. De esta forma, se puede estudiar si el tamaño del vocabulario mejora o empeora la calidad morfológica del mismo. Los tamaños se han elegido porque se corresponden con tamaños de los modelos del lenguaje reales. Por ejemplo, *BETO* tiene 31.000 *tokens* y *RoBERTa* cuenta con 50.000. Además, el vocabulario de todos los morfemas del español tiene un tamaño aproximado de entre 38.000 y 52.000 *tokens*, según los cálculos que hemos realizado al construir los conjuntos de datos en el paso 2. La Tabla 4 muestra los segmentadores entrenados.

Vocabulario	Corpus	Tamaño corpus (frases)	Tamaño vocabulario (tokens)
wordpiece_oscar_31	oscar small	600.000	31.000
wordpiece_oscar_52	oscar small	600.000	52.000
BPE_oscar_31	oscar small	600.000	31.000
BPE_oscar_52	oscar small	600.000	52.000
unigram_oscar_31	oscar small	600.000	31.000
unigram_oscar_52	oscar small	600.000	52.000

Tabla 4. Listado de los vocabularios evaluados

6. <https://huggingface.co/datasets/nthngdy/oscar-small>

7. https://colab.research.google.com/github/huggingface/notebooks/blob/master/examples/tokenizer_training.ipynb

Paso 5. En este paso se han evaluado los seis segmentadores con los tres conjuntos de datos de evaluación creados en el paso 2 y se ha procedido a analizar los resultados. Esta tarea se ha centrado en la comparación entre los tres segmentadores y en los diferentes tamaños de vocabulario. Los resultados se muestran y discuten en la sección 4.

Como se ha indicado en la sección anterior, para medir la relevancia morfológica se han utilizado las métricas de precisión, cobertura y valor F1 basados en la intersección entre el conjunto de datos de evaluación de la relevancia morfológica (Tabla 2) y los seis vocabularios generados (Tabla 4).

Respecto a la coherencia morfológica, además de calcular la coherencia morfológica, para estudiar los errores, para cada tipo de morfema, prefijos, sufijos, raíces y clíticos, se han creado tres listas llamadas *un token*, *varios - token correcto* y *varios - token incorrecto* (Tabla 5). Las listas se han creado de la forma siguiente:

- i) se procesan los pares de palabra, morfema, segmentando la palabra;
- ii) si la palabra se segmenta como un *token*, se añade la palabra a la lista *un token*; si se segmenta en varios y el morfema es uno de ellos, se añade a *varios - token correcto*; y, si se segmenta en varios y el morfema no está entre ellos, se añade a la tercera lista de *varios-token incorrecto*;
- iii) una vez procesadas todas las palabras, se calcula el tamaño de cada una de las tres listas y se normaliza en forma de porcentaje respecto al total de *tokens*.

La coherencia morfológica se corresponde con la lista *varios - token correcto*.

	<i>Un token</i>	<i>Varios-token correcto</i>	<i>Varios-token incorrecto</i>
Prefijos	<i>retirar</i> : [<u>retirar</u>]	<i>circunvalar</i> : [<i>circun</i> , <i>vala</i> , <i>r</i>]	<i>desprogramar</i> : [<i>despro</i> , <i>grama</i> , <i>r</i>]
Raíces	<i>generar</i> : [<u>generar</u>]	<i>cebar</i> : [<i>ceb</i> , <i>ar</i>]	<i>suturar</i> : [<i>su</i> , <i>tura</i> , <i>r</i>]
Sufijos	<i>calmar</i> : [<i>calmar</i>]	<i>destapar</i> : [<i>des</i> , <i>tap</i> , <i>ar</i>]	<i>aburrir</i> : [<i>abur</i> , <i>rir</i>]
Clíticos	<i>ayudarte</i> : [<i>ayudarte</i>]	<i>llamarme</i> : [<i>llamar</i> , <i>me</i>]	<i>sáquel</i> : [<i>s</i> , <i>áque</i> , <i>l</i> , <i>o</i>]

Tabla 5. Ejemplos de las listas generadas para el análisis de errores del segmentador *Wordpiece_31*. En negrita el morfema evaluado para cada palabra

Por último, para evaluar la corrección morfológica se utiliza, como se ha indicado anteriormente, la métrica de corrección morfológica. Para calcular

esta medida se utiliza el conjunto de datos de corrección morfológica (Tabla 1). Se compara cada par (palabra-segmentación) con el conjunto de datos y se cuenta una segmentación como correcta cuando se comprueba que todos los morfemas de la palabra se han segmentado correctamente.

Además, en este paso se calcula la media de *tokens* utilizados para segmentar las palabras de este conjunto de datos, con el fin de averiguar qué segmentadores utilizan un mayor número de *tokens* por palabra.

En la siguiente sección se muestran y discuten los resultados de este proceso de evaluación respecto a cada criterio.

4. RESULTADOS Y DISCUSIÓN

4.1. *Relevancia morfológica*

La Tabla 6 muestra los resultados para precisión, cobertura y valor F_1 de relevancia morfológica de los seis vocabularios según los prefijos, sufijos y raíces contenidos en los vocabularios, y según el total de morfemas para poder estudiar, si existen, las diferencias respecto de cada tipo de morfema.

Como se observa, los valores de precisión son, en todos los casos, muy bajos. Estos valores bajos confirman que, efectivamente, el funcionamiento estadístico de los segmentadores no es capaz de extraer las unidades lingüísticas básicas del lenguaje natural, en este caso del español. La cobertura, sin embargo, presenta mejores valores porque mide qué proporción de los morfemas reales están en los vocabularios, aun cuando presenta una gran cantidad de subpalabras que no son morfemas. El valor F_1 es muy bajo porque es la media armónica de la precisión y cobertura y la precisión es muy baja. Básicamente, lo que ocurre es que estos tres segmentadores mejoran la cobertura mediante la inclusión de más subpalabras frecuentes lo que aumenta la probabilidad de contener los morfemas de la lengua, pero a costa de disminuir su precisión.

Tipo	Totales	Vocabulario	Precisión (%)	Cobertura (%)	F1 (%)
Prefijos	61	wordpiece_31	0.17	88,52	0.35
		wordpiece_52	0.11	96,72	0.23
		BPE_31	0.18	90.16	0.37
		BPE_52	0.11	96,72	0.23
		unigram_31	0.15	77.05	0.30
		unigram_52	0.10	83,61	0.20
Sufijos	175	wordpiece_31	0.41	73.56	0.82
		wordpiece_52	0.26	77.01	0.51
		BPE_31	0.41	72.41	0.81
		BPE_52	0.25	73.56	0.49
		unigram_31	0.33	59.20	0.66
		unigram_52	0.21	63.22	0.42
Raíces	5.000	wordpiece_31	8.59	9.97	9.23
		wordpiece_52	7.00	16.02	9.74
		BPE_31	7.75	8.99	8.32
		BPE_52	5.57	12.74	7.75
		unigram_31	5.75	6.68	6.18
		unigram_52	4.15	9.50	5.78
Totales	5.236	wordpiece_31	4,15	24,55	7,09
		wordpiece_52	3,13	31,08	5,69
		BPE_31	3,25	19,24	5,56
		BPE_52	2,36	23,40	4,28
		unigram_31	2,56	15,15	4,38
		unigram_52	1,99	19,79	3,62

Tabla 6. Resultados de la evaluación de la relevancia morfológica de los vocabularios. Se han marcado en gris los valores más altos

De forma más pormenorizada respecto al tipo de morfema se comprueba lo siguiente:

- En el caso de los prefijos, los vocabularios del segmentador *BPE* son los que presentan mejor relevancia morfológica con un valor F1 de 0.37 y hasta un 90% de cobertura para los prefijos del español. Los vocabularios de *WordPiece* ocupan el segundo lugar con un valor F1 de 0.35% y los de *Unigram* resultan ser los de peor rendimiento.

- En el caso de los sufijos, los vocabularios de *WordPiece* superan ligeramente a los de *BPE* y, de nuevo, claramente a los de *Unigram*.
- En el caso de las raíces, los vocabularios de *WordPiece* siguen siendo los mejores, de tal manera que cuando se utiliza un tamaño de vocabulario de 52.000 *tokens* se almacena hasta un 16% de las raíces evaluadas.

Respecto al tamaño del vocabulario, observamos que, como era de esperar, al aumentar el tamaño aumenta también el porcentaje de morfemas reales presentes en ellos y, por tanto, la cobertura. Sin embargo, como recoge la Tabla 6, los valores de F_1 empeoran para todos los casos al ampliar los tamaños, puesto que la precisión no crece en la misma proporción que la cobertura. Así, al aumentar el tamaño del vocabulario, se observa que el ritmo de aumento del número de *tokens* que no son morfemas reales del español es superior al aumento de morfemas reales de la lengua.

A pesar de ello, cabe destacar que, al aumentar el tamaño, los vocabularios de *WordPiece* logran los mejores resultados, igualando a los de *BPE* en los prefijos (0.23% de F_1) y superándolos en los sufijos y en las raíces. Los vocabularios de *Unigram* son los peores cuando tienen 52.000 *tokens*.

En resumen, de los resultados de la evaluación de la relevancia morfológica se puede concluir que ninguno de los vocabularios destaca significativamente respecto a este criterio de relevancia morfológica, aunque parece claro que los vocabularios de *Unigram* son los que ofrecen peores resultados independientemente del tamaño de vocabulario utilizado. El aumento de tamaño parece favorecer a los vocabularios generados por *WordPiece*, mientras que tamaños más reducidos benefician a los vocabularios de *BPE*. En ninguno de los casos el aumentar el tamaño del vocabulario hace que mejore el valor F_1 , lo que implica que al aumentar el tamaño del vocabulario los *tokens* que no son morfemas crecen más que los que sí lo son. Los vocabularios generados por los tres segmentadores tienen una cobertura aceptable, siendo los mejores los generados por *WordPiece* y *BPE* con un vocabulario de 52.000 *tokens* en la inclusión de los prefijos (incluyen más del 96%), pero su cobertura disminuye en la inclusión de los sufijos (73-77%) y empeora en la inclusión de las raíces. Los valores de precisión son extremadamente bajos porque, como hemos indicado, al ser una construcción puramente estadística, los morfemas reales representan una parte pequeña del total de los *tokens*.

4.2. Coherencia morfológica

La coherencia morfológica se refiere a la cantidad de veces que las palabras con una estructura morfológica similar (es decir, las que comparten un morfema) tienen también un *token* en común, siendo el *token* dicho morfema. Para analizar los resultados de la evaluación de la coherencia morfológica se han recogido en la Tabla 7 los resultados por tipo de morfema y se han añadido los clíticos. Como se observa, respecto a los tipos de morfema, todos los segmentadores tienden a usar varios *tokens* para segmentar las palabras del corpus de evaluación, lo que ofrece valores de coherencia realmente bajos, que llegan, en el mejor de los casos, a casi un 66% de coherencia en *BPE_31*. Valores semejantes de coherencia morfológica se obtienen en el caso de los clíticos, siendo *Unigram_31* el que mejores resultados presenta (casi un 67%). Es también relevante observar que el número de palabras únicas que incluyen estos vocabularios es muy bajo.

Categoría	Vocabulario	Total palabras	La palabra es un solo <i>token</i> (%)	Varios <i>tokens</i> (subpalabras)	
				El morfema es un <i>token</i> (%)	El morfema no es un <i>token</i> (%)
Prefijos	wordpiece_31	205.792	0,71	13,45	85,84
	wordpiece_52		1,41	9,03	89,56
	BPE_31		0,04	65,86	34,10
	BPE_52		0,10	60,45	39,45
	unigram_31		1,11	42,73	56,16
	unigram_52		1,78	38,61	59,62
Raíces	wordpiece_31	234.584	1,64	16,03	82,33
	wordpiece_52		2,69	18,75	78,56
	BPE_31		0,40	4,66	94,94
	BPE_52		0,64	5,50	93,86
	unigram_31		1,79	9,08	89,13
	unigram_52		2,62	10,93	86,45

Sufijos	wordpiece_31	1.575.207	0,74	15,20	84,05
	wordpiece_52		1,34	14,07	84,59
	BPE_31		0,11	10,10	89,79
	BPE_52		0,20	7,80	92,00
	unigram_31		0,92	20,74	78,34
	unigram_52		1,43	18,49	80,08
Clíticos	wordpiece_31	3.920	5,61	61,40	32,99
	wordpiece_52		16,61	51,66	31,73
	BPE_31		0,69	48,57	50,74
	BPE_52		1,12	44,21	54,67
	unigram_31		4,64	66,87	28,49
	unigram_52		11,94	63,44	24,62
Totales	wordpiece_31	2.019.504	0,86	15,21	83,93
	wordpiece_52		1,54	14,17	84,29
	BPE_31		0,14	15,23	84,64
	BPE_52		0,24	12,97	86,79
	unigram_31		1,05	21,71	77,24
	unigram_52		1,62	19,75	78,63

Tabla 7. Resultados de la evaluación de la coherencia morfológica. Se indican en gris los valores más altos

Respecto al tamaño del vocabulario, se observa que al ampliar los tamaños de 31.000 a 52.000 *tokens* empeora la coherencia morfológica (es decir, para todos los segmentadores, disminuye el porcentaje de casos en los que el morfema evaluado se corresponden con un *token* correcto). Cabe recordar que, al ampliar los tamaños de los vocabularios, los valores de F1 de la tarea de relevancia morfológica también decrecían en todos los casos debido a la disminución de la precisión.

También es relevante señalar que, aunque en todos los segmentadores el porcentaje de palabras segmentadas como un solo *token* es muy bajo, en el caso del segmentador *BPE* estos valores son especialmente bajos.

Finalmente, parece que podría existir cierta correlación entre los valores de relevancia y coherencia morfológica de los vocabularios para prefijos y raíces. Así, *BPE* presenta los mejores resultados de relevancia morfológica en los prefijos, y eso se traduce en mejores resultados de coherencia morfológica respecto a los prefijos. En las raíces ocurre lo mismo con *WordPiece*. En los sufijos, en cambio,

se rompe esa tendencia, ya que *WordPiece* es el que mayor relevancia morfológica presenta y, sin embargo, no es el que tiene mayor coherencia morfológica (son los segmentadores de *Unigram*).

En resumen, los tres segmentadores segmentan de forma poco coherente –el valor máximo de coherencia es de 66% de BPE_31– y los vocabularios más grandes tienen peor coherencia morfológica. En los prefijos y las raíces parece que existe una correlación entre la relevancia morfológica respecto a prefijos y raíces de los vocabularios y la coherencia morfológica. En los prefijos *BPE* es el que mejor relevancia morfológica del vocabulario presenta, y también es el que mejores resultados de coherencia presenta para este tipo de morfema. En las raíces, *WordPiece* es el mejor en relevancia morfológica y también lo es en coherencia. Sin embargo, en los sufijos, *Wordpiece* es el mejor en relevancia del vocabulario, mientras que *Unigram* lo supera en coherencia. Esto se debe a que, como se verá en el análisis de errores, aunque un morfema forme parte del vocabulario de un segmentador, no significa necesariamente que el segmentador vaya a usarlo para segmentar palabras que lo contienen.

4.3. Corrección morfológica

En el caso de la corrección morfológica, los resultados de todos los segmentadores son, de nuevo, realmente bajos. La Tabla 8 muestra los resultados. Como se observa, ninguno de los vocabularios supera el 15% de corrección.

Vocabulario	Totales	Corrección (%)	<i>Tokens</i> utilizados de media por palabra
wordpiece_31	1231	14,54	1,75
wordpiece_52		14,94	1,54
BPE_31		8,69	2,39
BPE_52		10,23	2,19
unigram_31		14,70	1,83
unigram_52		15,10	1,59

Tabla 8. Resultados de corrección morfológica y media de *tokens* utilizados por palabra

De forma más detallada, se observa que el segmentador *Unigram* obtiene los mejores resultados y que *BPE* es el peor de los tres. Respecto al tamaño de vocabulario, aumentarlo hace que mejoren ligeramente los resultados.

El análisis de errores nos ha permitido encontrar una explicación a estos valores tan bajos. Se ha comprobado que todos los aciertos de segmentación se corresponden con palabras segmentadas como un solo *token*. Es decir, con las palabras que no requieren ser segmentadas porque son muy frecuentes en el corpus. En cambio, ninguno de los seis segmentadores segmenta correctamente en morfemas la mayor parte de las palabras que requieren ser segmentadas. Por esta razón, el segmentador *Unigram*, que tiene tendencia a no dividir las palabras, obtiene una mejor corrección. Esto se aprecia también en la Tabla 8, en la columna de *tokens* usados de media en cada segmentación. Los segmentadores *Unigram*, que son los que obtienen mejores resultados, utilizan pocos *tokens* de media para segmentar cada palabra.

Respecto al tamaño del vocabulario, al aumentar el tamaño se observa que contienen más palabras completas, lo que lleva a que aumente la corrección del vocabulario.

Con relación al tipo de segmentador, la evaluación de la corrección morfológica muestra resultados diferentes de los obtenidos en los criterios de relevancia y coherencia morfológica. En primer lugar, el algoritmo *BPE* es el que ofrece peores resultados a la hora de segmentar una palabra en todos sus morfemas. *Unigram* es, en este caso, el mejor, lo que confirma que no puede señalarse que un algoritmo sea superior a los otros en el global de la evaluación.

En segundo lugar, las palabras que no necesitan segmentación y que con éxito se segmentan como un solo *token* son el motivo de la superioridad de *Unigram* en este criterio. Todos los aciertos de todos los segmentadores se corresponden con este tipo de palabras monomorfemáticas, y todos fallan completamente a la hora de segmentar morfológicamente las palabras que requieren ser descompuestas en morfemas. En la sección siguiente se muestra el análisis de errores que justifica estos resultados.

5. RESULTADOS DEL ANÁLISIS DE ERRORES

Para completar la evaluación se ha realizado un análisis de errores utilizando los resultados de la evaluación de la corrección morfológica (Tabla 7). Estos resultados se obtuvieron segmentado las palabras del conjunto de datos de corrección morfológica (1.231 palabras) con cada segmentador (en el paso 5 de la sección 3). El análisis de estos errores ha permitido identificar cuatro tipos de errores (Tabla 9):

Segmentador\Error	Tipo 1	Tipo 2	Tipo 3	Tipo 4	Total
Wordpiece_31	436	16	352	248	1.052
Wordpiece_52	568	11	251	217	1.047
BPE_31	98	85	591	350	1.124
BPE_52	149	65	513	378	1.105
Unigram_31	459	15	424	152	1.050
Unigram_52	575	10	335	125	1.045

Tabla 9. Resultados cuantitativos del análisis de errores de los segmentadores. Se han marcado en azul y gris los dos tipos de errores más comunes en cada tipo de segmentador

Tipo de error 1. Infrasegmentación. Las palabras frecuentes que necesitan segmentación no se segmentan porque se identifican en el corpus como muy frecuentes y se incluyen en el vocabulario como un solo *token*. Esto hace que no se incluya la segmentación en morfemas en muchas palabras que sí lo requerirían. Esto ocurre por ejemplo con la palabra *población*, palabras que los seis segmentadores tratan como un solo *token*. En la Tabla 9 se puede observar que la mayor parte de los errores de los algoritmos de segmentación de *WordPiece* y *Unigram* son de este tipo.

Tipo de error 2. Sobresegmentación. Algunas palabras poco frecuentes en el corpus de evaluación que no necesitarían realmente ser segmentadas se descomponen erróneamente en varios *tokens*. Esto ocurre, por ejemplo, con la palabra *tortilla* que el segmentador *BPE* descompone en [*tor*, *tilla*]. Este tipo de error es el menos frecuente en todos los algoritmos de segmentación.

Tipo de error 3. Ausencia del morfema. El morfema correcto no está en vocabulario, por lo que hay que recurrir a otros *tokens* o subpalabras para realizar las segmentaciones. Esto ocurre, por ejemplo, en todos los segmentadores con el sufijo *-érrimo*: ninguno lo tiene en su vocabulario. Los dos *WordPiece* producen la segmentación [*pau*, *pé*, *rr*, *imo*], los dos *Unigram* [*pau*, *pé*, *rri*, *mo*], *BPE_31* [*pa*, *up*, *ér*, *ri*, *mo*] y *BPE_52* [*pa*, *up*, *ér*, *rimo*]. Este tipo de error parece ser el más crítico, puesto que es el que se comete con más frecuencia en los segmentadores de tipo *BPE*, pero también es el segundo tipo de error más común de *Unigram* y *WordPiece*.

Tipo de error 4. No uso del morfema. No se utiliza el morfema correcto en una segmentación, aunque dicho morfema está incluido en el vocabulario aprendido por el segmentador. Según el tipo de segmentador pueden darse varias situaciones:

1. Los segmentadores *BPE* (Seinrich, 2015) y *Unigram* (Kudo, 2018) contienen en sus archivos las probabilidades aprendidas durante el entrenamiento para las diferentes subpalabras. Así, en el momento de segmentar una palabra, se elige la segmentación cuyas subpalabras son las más probables, lo que a veces puede implicar que el morfema real no sea el elegido. Por ejemplo, al procesar la palabra *escogiendo*, el segmentador *Unigram_52* tendría varias opciones: [*es, cogiendo*] o [*es, cog, iendo*]. Todas estas subpalabras están en su vocabulario aprendido en el entrenamiento. La segunda segmentación sería la correcta desde un punto de vista morfológico. Sin embargo, según las probabilidades que guarda el segmentador en su configuración, la primera es más probable en el corpus de entrenamiento, por lo que es la segmentación elegida finalmente.
2. A diferencia de ellos, *WordPiece* no almacena en ningún archivo las probabilidades de aparición de las subpalabras en el corpus de entrenamiento. A la hora de segmentar una palabra, *WordPiece* va buscando la subpalabra más larga de su vocabulario que coincida con el comienzo de la palabra (Song et al., 2021). Por ejemplo, hemos encontrado que el segmentador *WordPiece* de 31.000 tokens (palabras y subpalabras) de vocabulario segmenta la palabra *desamor* como [*desa, mor*] a pesar de que *des* y *amor* son tokens de su vocabulario. La razón es que el segmentador busca en el vocabulario, en primer lugar, los tokens *desamor*, *desamo* y *desam*. Como ninguno de ellos está en el vocabulario, se busca, con éxito, *desa*. A continuación, se repite el procedimiento con el resto de la palabra. Puesto que *mor* también es un token del vocabulario, se elige [*desa, mor*] como segmentación final. Este tipo de errores son los que explican que en los prefijos *WordPiece* presente una coherencia morfológica tan baja, a pesar de la alta cobertura morfológica que presenta su vocabulario: muchos prefijos se pierden al buscar el token inicial más largo.

6. CONCLUSIONES Y TRABAJO FUTURO

Este trabajo aporta una solución original, hasta donde sabemos, para resolver el problema de evaluar cuán eficaces son los vocabularios que utilizan los actuales grandes modelos del lenguaje. Estos vocabularios son generados automáticamente por segmentadores basados en estrategias estadísticas a partir de grandes corpus de texto. El vocabulario es un componente clave para la eficacia y fiabilidad de un modelo del lenguaje porque contiene las unidades lingüísticas que el modelo es capaz

de reconocer y combinar. Uno de los problemas actuales de los grandes modelos del lenguaje es que sus vocabularios están formados por secuencias de caracteres (las denominadas *subpalabras*) que tienen una alta frecuencia de aparición en el corpus de texto y que no siempre se corresponden con unidades lingüísticas de tipo morfema o palabra.

Para poder medir cuán bueno es un vocabulario y, así poder mejorarlo, hemos desarrollado un método de evaluación de su calidad morfológica. Entendemos que la calidad morfológica es el grado de similitud de un vocabulario de morfemas y palabras comparado con el vocabulario real de la lengua. Para valorar/evaluar la calidad lingüística, hemos propuesto medir tres criterios: la relevancia, la coherencia y la corrección morfológica del vocabulario.

Además, hemos desarrollado un procedimiento para realizar estas mediciones. Hemos aplicado el método a la medición de la calidad de seis vocabularios de dos tamaños diferentes generados con los tres segmentadores principales utilizados en la creación de los actuales modelos del lenguaje: *WordPiece* (usado, entre otros, por el modelo *BERT*), *BPE* (utilizado, entre otros, por los modelos *GPT*) y *Unigram* (empleado, por ejemplo, por el modelo del lenguaje *Albert*). Además, se ha explorado la posible influencia del tamaño de los vocabularios en la calidad morfológica.

Los resultados obtenidos indican que:

1. La calidad morfológica de los vocabularios es muy baja. Estos segmentadores, por lo tanto, no parecen apropiados para conservar conocimiento morfológico del español. De forma más concreta, los resultados son los siguientes:
 - a. Respecto a la relevancia morfológica de los vocabularios generados, solo para la medida de cobertura se obtienen valores relevantes (el mejor vocabulario de todos, *Wordpiece_32*, tiene una cobertura media total de 31,08%) y, respecto a ellos, *BPE* y *Wordpiece* son superiores al segmentador *Unigram*.
 - b. Respecto a la coherencia morfológica, los resultados son malos y aún menos concluyentes respecto a cuál podría ser el mejor. Aunque no merece la pena elaborar un ránquin, se puede indicar que *BPE* es el mejor en la obtención de prefijos, *WordPiece* en raíces y *Unigram* en sufijos y clíticos.
 - c. Respecto a la corrección morfológica, los valores también son realmente bajos, con un valor máximo de un 15% de corrección. De nuevo, aunque con estos valores no merece la pena realizar una clasificación, *Unigram* y *WordPiece* parecen superiores a *BPE*. En todos los vocabularios, los

aciertos coinciden con las palabras formadas por un solo morfema y que, por tanto, no tienen segmentación en morfemas.

2. En cuanto a los tamaños de los vocabularios generados, no se puede concluir que aumentar el tamaño mejore los resultados en ninguno de los tres criterios.
3. Respecto al tipo de errores observados se puede concluir que son de cuatro tipos: i) no se segmentan las palabras muy frecuentes en el corpus de entrenamiento del segmentador; ii) sobresegmentación de las palabras poco frecuentes; iii) el morfema correcto no está en el vocabulario; y (iv) el funcionamiento de los segmentadores lleva a elegir otra combinación de tokens que no incluye al morfema correcto, a pesar de que este es un token del vocabulario.

El tipo de error más frecuente encontrado en las segmentaciones generadas por todos los algoritmos evaluados es el tipo 3 (no han aprendido alguno de los morfemas de la lengua), seguido del tipo 1 (infrasegmentación), que es frecuente en *WordPiece* y *Unigram* y el tipo 4 (no utilizan los morfemas, aunque los han aprendido), que es el más frecuente en *BPE*.

En conclusión, se ha encontrado que los algoritmos estadísticos *WordPiece*, *BPE* y *Unigram* utilizados de forma mayoritaria para la creación y funcionamiento de los grandes modelos del lenguaje actuales no son óptimos para segmentar correctamente los morfemas del español y generan vocabularios con subpalabras que no se corresponden con unidades lingüísticas en esta lengua. Para mejorar los resultados de estos algoritmos podría ser útil diseñar soluciones dirigidas a ayudarles a aprender los morfemas de la lengua de forma que disminuyan los errores más frecuentes (tipo 3) y a priorizar su utilización en las segmentaciones para disminuir el segundo tipo de error (tipo 4) que más frecuentemente cometen.

Para terminar, destacamos como contribuciones principales de este artículo:

1. La elaboración de un nuevo método de evaluación de la calidad morfológica de los vocabularios de los grandes modelos del lenguaje generados por segmentadores de naturaleza estadística.
2. La evaluación de la calidad morfológica de los vocabularios en español generados por los tres principales segmentadores actuales.
3. La constatación empírica de que la mejora de los tres principales segmentadores actuales puede dirigirse hacia el aumento y uso de su conocimiento morfológico de la lengua.

En qué medida la calidad morfológica de los vocabularios de los grandes modelos del lenguaje afecta a la eficacia y fiabilidad de dichos modelos es una cuestión todavía no resuelta. Nuestro trabajo actual y futuro se dirige a responder esta cuestión.

DECLARACIÓN DE CONTRIBUCIÓN DE AUTORÍA

Óscar García-Sierra: investigación, redacción–borrador original.

Ana Fernández-Pampillón Cesteros: supervisión, metodología, redacción–revisión y edición.

Miguel Ortega-Martín: supervisión, redacción–revisión y edición.

AGRADECIMIENTOS

A nuestros compañeros de Dezzai. A los revisores anónimos, por sus comentarios y sugerencias.

El trabajo ha contado con el apoyo del proyecto del Ministerio de Ciencia e Innovación PID2022-140897OB-I00 ROBOT-TALK.

REFERENCIAS

- Bostrom, K., y Durrett, G. (2020). Byte pair encoding is suboptimal for language model pretraining. *arXiv Preprint arXiv:2004.03720*.
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., y Pérez, J. (2023). Spanish pre-trained bert model and evaluation data. *arXiv Preprint arXiv:2308.02976*.
- Church, K. W. (2020). Emerging trends: Subwords, seriously? *Natural Language Engineering*, 26(3), 375–382.
- Devlin, J., Chang, M.-W., Lee, K., y Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv [Cs.CL]*. Retrieved from <http://arxiv.org/abs/1810.04805>
- Fang, H., Ostendorf, M., Baumann, P., y Pierrehumbert, J. (2015). Exponential language modeling using morphological features and multi-task learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(12), 2410–2421.
- Friedman, R. (2023). Tokenization in the Theory of Knowledge. *Encyclopedia*, 3(1), 380–386.
- Hofmann, V., Pierrehumbert, J., y Schütze, H. (2021). Superbizarre is not superb: Derivational morphology improves BERT's interpretation of complex words. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 3594–3608.
- Kudo, T., y Richardson, J. (2018). Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *arXiv Preprint arXiv:1808.06226*.

- Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., y Soricut, R. (2019). Albert: A lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv [Cs.CL]*. Retrieved from <http://arxiv.org/abs/1907.11692>
- Moliner, M. (1967/2012). *Diccionario de uso del español*. Madrid: Gredos.
- Park, K., Lee, J., Jang, S., y Jung, D. (2020). An Empirical Study of Tokenization Strategies for Various Korean NLP Tasks. *arXiv Preprint arXiv:2010.02534*.
- RAE-ASALE (=Real Academia Española-Asociación de Academias de la Lengua Española) (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- RAE-ASALE (=Real Academia Española-Asociación de Academias de la Lengua Española) (en línea). *Diccionario de la lengua española*. <https://dle.rae.es/>
- Radford, A., Narasimhan, K., Salimans, T., y Sutskever, I. (2018). *Improving language understanding by generative pre-training*. <https://paperswithcode.com/paper/improving-language-understanding-by>
- Sennrich, R., Haddow, B., y Birch, A. (2015). Neural machine translation of rare words with subword units. *arXiv Preprint arXiv:1508.07909*.
- Schuster, M., y Nakajima, K. (2012). Japanese and korean voice search. *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5149–5152. IEEE.
- Song, X., Salcianu, A., Song, Y., Dopson, D., y Zhou, D. (2020). Fast wordpiece tokenization. *arXiv preprint arXiv:2012.15524*.
- Suárez, P. J. O., Sagot, B., y Romary, L. (2019). Asynchronous pipeline for processing huge corpora on medium to low resource infrastructures. *7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7)*. Leibniz-Institut für Deutsche Sprache.
- Suárez, P. J. O., Romary, L., y Sagot, B. (2020). A monolingual approach to contextualized word embeddings for mid-resource languages. *arXiv Preprint arXiv:2006.06202*.
- Van der Wouden, T. (1990). Celex: Building a multifunctional polytheoretical lexical data base. *Proceedings of BudaLex*, 88, 363–373.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W. y Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv Preprint arXiv:1609.08144*.

NOTAS

EL PANORAMA SOCIOLINGÜÍSTICO ACTUAL EN SENEGAL: INTERACCIONES Y REGULACIONES DESDE LA PERSPECTIVA ECOLINGÜÍSTICA

Oumar SEYDI¹

Université Assane Seck de Ziguinchor / Senegal

Resumen

Este artículo se centra en el análisis del panorama ecolingüístico senegalés para explicar las interacciones y sus mecanismos de regulación y de pacificación sociolingüística. En efecto, Senegal, por ser uno de los países donde coexisten varias lenguas nacionales y extranjeras, se caracteriza por un multilingüismo complejo. Esta pluralidad se ha convertido en una riqueza sociocultural compartida por individuos de orígenes étnicos diversos, a los que se une la lengua colonial, el francés, que ostenta todavía el estatus de única lengua oficial. En las últimas décadas, los movimientos migratorios dentro y fuera de sus fronteras, y el contexto de la globalización, han propiciado importantes cambios sociolingüísticos, como la wolofización, la emergencia de las lenguas minoritarias y de otras lenguas como nuevos actores en el panorama sociolingüístico y cultural.

Palabras clave: Ecolingüística; plurilingüismo; regulación lingüística; lenguas de Senegal; pacificación lingüística.

THE CURRENT SOCIOLINGUISTIC LANDSCAPE IN SENEGAL: INTERACTIONS AND REGULATIONS FROM THE ECOLINGUISTIC PERSPECTIVE

Abstract

This article focuses on the analysis of the Senegalese ecolinguistic panorama to explain the interactions and their mechanisms of regulation and sociolinguistic pacification. Indeed, Senegal, being one of the countries where various national and foreign languages coexist, is characterized by complex multilingualism. This plurality has become a socio-cultural richness shared by individuals of diverse ethnic origins, joined by the colonial language, French, which

1. o.seydi@univ-zig.sn;  <https://orcid.org/0000-0002-5719-004X>

still holds the status of the only official language. In recent decades, migratory movements within and outside Senegal borders, and the context of globalization, have led to important sociolinguistic changes, such as Wolofización, the emergence of minority languages and other languages as new actors in the sociolinguistic and cultural panorama.

Keywords: Ecolinguistics; multilingualism; linguistic regulation; languages of Senegal; linguistic pacification.

1. INTRODUCCIÓN

Senegal, como la mayoría de los países subsaharianos, presenta un panorama sociolingüístico heterogéneo, multiconfesional, multicultural y multilingüe (Seydi, 2021). La complejidad de esa situación nos empuja a realizar un análisis de dicho panorama bajo el prisma de la ecolingüística. Esta perspectiva no pretende identificar el estado de convivencia secular de lenguas en Senegal como escenario idealizado y totalmente equilibrado, contrapuesto a la fuerza de las lenguas colonizadoras y neocolonizadoras como elementos de disrupción en una supuesta paz lingüística original. Nuestro objetivo es analizar las esferas de regulación pacífica, entendidas desde la ecolingüística como enfoque holístico, que siempre entrañan también violencia, pero que han resultado en un cierto equilibrio que se renueva a lo largo de la historia con nuevas configuraciones. Para ello, contemplaremos la convivencia, el contacto y conflicto lingüístico desde el punto de vista del equilibrio entre lenguas, la existencia de mecanismos para mantener los ecosistemas lingüísticos. Partiremos de las siguientes preguntas: ¿existe un ecosistema lingüístico en Senegal que asegure la armonía y el equilibrio entre lenguas y regule las relaciones entre los grupos sociales, facilitando la coexistencia de diferentes identidades culturales y étnicas?, ¿cuáles son las fuentes de regulación de los conflictos en este ecosistema?

Intentaremos responder a estas preguntas analizando la experiencia histórica de regulación lingüístico-cultural en Senegal antes y después de la colonización, centrándonos en los mecanismos de regulación lingüística que se están produciendo en la actualidad y determinando el papel regulador de las políticas lingüísticas u otras esferas de gestión de esos contactos y conflictos.

2. MARCO TEÓRICO

La aportación de este estudio es una crítica a la visión categorizada del panorama sociolingüístico senegalés bajo la perspectiva de la ecolingüística, contemplando la

convivencia, el contacto y conflicto lingüístico desde el punto de vista del equilibrio entre lenguas y la existencia de mecanismos para mantener los ecosistemas lingüísticos. Para ello tomamos las dos vertientes de los estudios ecolingüísticos, descritas por Fill (2001) según su modo de pensar la ecología del lenguaje de forma simbólica o metafórica (ecología lingüística) o no metafórica (lingüística ecológica). Ambas interpretaciones de la relación entre ecología y lingüística parten respectivamente de los estudios de Haugen (1972), que propugna la analogía entre lenguas y seres vivos o ecosistemas, y el sentido biológico del lenguaje en sí (Halliday, 2001).

Por una parte, la ecología lingüística se interesa por el estudio de discursos en torno al multilingüismo y de las situaciones sociolingüísticas complejas (Bastardas i Boadas, 1996; Lechevrel, 2010). Esta visión ecosistémica de la vida de las lenguas surge de la analogía de la ecología (biológica y/o medioambiental), que son la base de sus argumentos a favor de la protección de la diversidad lingüística, la revitalización y pacificación lingüística, el equilibrio y el desarrollo sostenible de los ecosistemas lingüísticos, de modo análogo a la defensa de la biodiversidad, la protección, el equilibrio y el desarrollo sostenible de los ecosistemas. En este sentido, se trata más bien de una ecología inter-lingüística (Steffensen y Fill, 2014, p. 8) que pretende analizar las relaciones e interacciones entre lenguas en contacto y entre lenguas y su contexto social (Haugen, 1972) y de los fenómenos derivados de la coexistencia lingüística: interferencia, cambio lingüístico, mezcla de código, pidginización y criollización (Mühlhäusler, 1986).

En una segunda vertiente, la lingüística ecológica trata de establecer de una manera renovada la relación entre las lenguas y su entorno medioambiental, es decir, cómo los fenómenos naturales se integran en las lenguas y culturas (Hagège, 1985), por lo que en principio se opta por una sociolingüística de corte etnográfico y antropológico para ahondar en esta relación natural. Sin embargo, estos estudios se han decantado preferentemente por la influencia inversa: cómo los discursos y prácticas lingüísticas están influyendo en el medio ambiente, en el marco de la preocupación ecológica creciente de finales del siglo XX y principios del siglo XXI.

Coincidimos con Steffensen y Fill (2014, pp. 19-21) en que ambas vertientes están fuertemente relacionadas con otras dos tendencias en la ecolingüística, que ponen el acento en la ecología sociocultural (las fuerzas sociales y culturales modelan las condiciones de los hablantes y de las comunidades lingüísticas), y en la ecología cognitiva (el lenguaje da lugar a capacidades cognitivas que generan conductas adaptativas al entorno). El modelo de ecolingüística adoptado no excluye ninguna de las opciones, por lo que el análisis de la situación sociolingüística de Senegal

contemplará las relaciones entre lengua, medio natural, sistema sociocultural y sistema cognitivo en una dirección y otra, en relación dialógica.

En este sentido, la propuesta de análisis se centra en la superación del debate entre naturaleza y cultura, entre esencialismo y relativismo lingüístico, con un estudio holístico de los procesos y actividades en los que los seres humanos tratan con el entorno biológico, en los que se incluyen las actividades lingüísticas, que forman parte de la complejidad del medio natural, y se mueven dentro de unos límites orgánicos, sociales y ecosistémicos (Steffensen y Fill, 2014, p. 21).

Por otra parte, el objetivo de este estudio no es solo describir el ecosistema lingüístico-cultural senegalés, sino también caracterizar los medios e instancias en las que se mantiene el equilibrio ecolingüístico en un contexto de convivencia de lenguas. La idea de regulación del ecosistema lingüístico y cultural entronca con los estudios sobre la paz y los conflictos en su vertiente de sociolingüística de la paz o sobre la paz (*Peace Sociolinguistics*, Friedrich, 2007), terreno transdisciplinar de intersección de la sociolingüística con los estudios sobre la paz y los conflictos, explícitamente orientada por valores, que estudia las regulaciones pacíficas y los factores lingüístico-culturales en los conflictos. En este campo se interesa, por tanto, por el estudio de la paz interlingüística, es decir, de los factores extralingüísticos de la regulación de conflictos, con el énfasis en la relación entre lenguaje y poder coercitivo. Este enfoque nos aporta la concepción del conflicto como nudo de regulación en el que siempre hay aspectos violentos, pero también pacíficos, es decir, hay rasgos de paz imperfecta, que presentan los conflictos en diferentes escalas, por ser categorías valorativas (Morillas, 2003).

También acudiremos a los estudios sobre lengua y poder y los conceptos de *violencia simbólica*, *capital simbólico* y *mercado de bienes simbólicos* aportados por Bourdieu (1980), que se argumentan en un sentido parejo a los de *violencia estructural* y *regulación de conflictos*. En este sentido, los fenómenos de interferencia, contacto y conflicto lingüístico se tratarán desde la perspectiva de los estudios de paz y conflictos, que coinciden con ciertas perspectivas interdisciplinares propuestas en el seno de la ecolingüística comunes, como las teorías de la complejidad y del caos (Steffensen y Fill, 2014). La actividad lingüística se contempla como una interactividad compleja y dialógica, en la que los recursos simbólicos socioculturales son potencialidades enraizadas en el medio natural, al igual que la teoría de paz y conflictos explica el nudo conflictivo como oportunidad o potencialidad de interacción compleja que se regula o gestiona con herramientas positivas y negativas. Una regulación satisfactoria y, por tanto, más pacífica,

implicaría la existencia de instancias y mecanismos que favorecen el equilibrio y la mutua satisfacción de necesidades.

La teoría de paz y conflictos también aporta el enfoque de análisis de la violencia, entre la que se ha descrito la violencia directa, indirecta y estructural, diferenciada o no de la violencia cultural. Estos conceptos son paralelos a la idea de la ecología lingüística y se pueden trasladar al ámbito sociolingüístico. Un ámbito donde se producen mediaciones entre regulación sociocultural violenta y pacífica, relaciones entre lenguas no solo en clave de enfrentamiento y conflicto negativo, sino también instancias de regulación interlingüística, como los fenómenos de interferencia, contacto lingüístico, préstamo, etc. que indican una adaptación al entorno biológico, tanto porque consideramos que las lenguas tienen un comportamiento similar a los seres vivos como por que las lenguas operan dentro de los marcos socioculturales y contribuyen a la interculturación y el mestizaje.

3. EL ECOSISTEMA LINGÜÍSTICO DE SENEGAL

Efectivamente, la sociedad senegalesa parece haber heredado una experiencia de regulación del ecosistema lingüístico entre diferentes grupos lingüístico-culturales desde tiempos remotos, porque su realidad sociolingüística se caracteriza por la heterogeneidad y la diversidad étnica (Diouf, Ndiaye y Dieme, 2017). Cissé (2005) asegura que Senegal es una encrucijada de tres grandes culturas: la africana, la árabe-islámica y la occidental.

La primera y más antigua influencia cultural llegó de la mano de pueblos procedentes del noreste de África que penetraron en la costa atlántica africana y se asentaron desde tiempos remotos. En la actualidad se hablan lenguas de diversas familias lingüísticas, entre las que predominan dos grupos del tronco nigero-congoleño: el atlántico occidental y el mande. Desde el punto de vista de la ecología lingüística, hay una especial correlación entre el paisaje geográfico, socioeconómico y cultural, de manera que, por ejemplo, se asocia a los *pulaar* con la ganadería y la agricultura, los *lebu* con la pesca, los *diola* como la horticultura y las plantaciones de frutales, los *serer* con la agricultura, y la comunidad wolof con el comercio.

Entre ellas la lengua wolof comenzó a extenderse gracias al comercio desde el siglo VII y alcanzaría gran importancia desde el siglo XVI en los antiguos reinos de Cayor (1549 - 1886), en el norte y centro del actual Senegal, y Walo (1287 - 1855), ubicado en el noroeste (Ndao, 1984; Dreyfus y Juillard, 2004; Ngom, 2004). El wolof se convirtió en lengua franca en las relaciones comerciales y probablemente lengua vehicular y de comunicación entre grupos, hasta los primeros contactos

con la colonización francesa en el siglo XVIII, cuando los colonizadores se apoyaron en la mediación de los wolof para asentar su poder, convirtiéndose en la primera lengua de contacto entre colonizadores e indígenas de orígenes diferentes (Ndao, 1996). En la actualidad, sigue siendo el vehículo de intercambio lingüístico en las relaciones comerciales intranacionales y en el sector informal (Mbaya, 2005), gracias a la fuerte influencia comercial de la cofradía sufi de los *Mouride*.

La segunda entidad cultural procedió del Magreb y penetró el territorio senegalés desde el norte del país en el antiguo reino de Tekrur hacia el siglo XI. En este componente se mezclan varias aportaciones lingüístico-culturales, árabe y bereber (o *amazige*), que comparten una misma entidad cultural amplia, lo islámico. El árabe consolidó un estatus de prestigio como lengua de cultura, frente a las lenguas habladas en Senegal, que eran mayormente ágrafas. Su presencia en Senegal está ligada mayormente a la religión islámica (en su modalidad formal y litúrgica, que es la presencia más palpable) y a la presencia minoritaria de los pueblos que la llevaron allí, en su modalidad coloquial o dialectal (*hasaniyya*).

La cultura europea occidental fue introducida en Senegal por los colonizadores franceses desde el siglo XVII durante un largo periodo de dominación cultural y sociopolítica de la metrópoli francesa, que duró hasta 1960. La irrupción del francés en la época colonial supone una nueva situación conflictiva que resulta en una nueva acomodación entre lenguas, de la que surge una nueva estratificación social ligada a la lengua y la cultura, una triglosia que eleva al francés como lengua de prestigio y lengua nacional, frente a las lenguas históricas consideradas autóctonas, donde se reposiciona el wolof como lengua de comunicación interétnica, y quedan en último lugar las lenguas minoritarias. La leve presencia histórica portuguesa, alemana, española, inglesa, y de otras lenguas europeas en algunas zonas del territorio nacional es susceptible de ser redimensionada en el contexto de la globalización.

El sistema colonial implantó una forma de regulación lingüística muy hostil con las lenguas y culturas locales de Senegal, con excepción del wolof. Tras la independencia, las decisiones políticas postcoloniales mantuvieron el francés como única lengua oficial (Seydi, 2021), sin corregir el desequilibrio ocasionado en el ecosistema lingüístico senegalés. Con los grandes cambios globales y la mundialización, Senegal afronta el siglo XXI una nueva situación conflictiva en la que se puede corregir estos desequilibrios de forma más constructiva. Desde 1971 se consideran como *lenguas nacionales* el *diola*, el *malinke* o *mandinga*, el *pulaar*, el *serer*, el *soninke* y el wolof. En los años 80 se produjo una ruptura sociolingüística ocasionada por la crisis económica y los movimientos de reivindicación que sacudieron el país (Daff,

1996). El sentimiento nacionalista y una toma de conciencia colectiva de la importancia de las lenguas y culturas locales propiciaron un cambio en la constitución de 2001, que amplió la condición de lengua nacional a 11 y a 24 lenguas años más tarde, aunque entre ellas el wolof sigue teniendo una especial preeminencia.

Este paisaje lingüístico presenta una realidad compleja: ocho comunidades de habla destacables (incluyendo el francés) entrelazadas social y geográficamente, entre las cuales es difícil establecer fronteras lingüísticas claras, sobre todo por el estudio todavía superficial e insuficiente de la historia lingüística y de las lenguas en Senegal. Los estudios socioculturales antes y después de la independencia solían ser simplificadores y tendían a la generalización. Por una parte, la tendencia *asociativa* establecía una equivalencia entre etnia y lengua (Cissé, 2005), aunque esta ecuación es difícil de cuadrar en muchos casos; y por otra, la tendencia *disociativa* se valía del criterio de inteligibilidad para definir las lenguas sin tener en consideración las estructuras étnicas, y resaltando la diversidad. Es el caso de la postura de Leclerc (2014) o la clasificación de *Ethnologue* (2009), que habla de la presencia de 40 lenguas locales en Senegal, 31 indígenas y 7 no propias del territorio, al catalogar diferentes hablas tradicionalmente consideradas como variedades del *diola* como lenguas distintas.

4. PLURALIDAD E INTERACCIONES EN EL MERCADO SOCIOLINGÜÍSTICO SENEGALÉS

En Senegal domina una imbricación constante de diferentes lenguas en la comunicación cotidiana de los hablantes bajo forma de mediaciones social, cultural y lingüística. Aunque caracterizada un multilingüismo desigual y una triglosia, la diversidad de las prácticas socioculturales influye en el desarrollo de las lenguas de Senegal. Asimismo, el mercado de bienes simbólicos, las lenguas tienen un valor y una función. En efecto, el francés, la lengua oficial, a pesar de no ser una lengua local sino una lengua extranjera impuesta, es la «lengua de la cresta» y tiene un valor elevado por su prestigio por ser la lengua de la administración, élite social y educación pública. El wolof, por su parte, es la lengua oficiosa y representa la «lengua de las masas». Se utilizada en el comercio y la comunicación interétnica. Va conquistando algunos dominios reservados al francés (Calvet, 1994; Rasoloniaina, 2000; Diatta, 2021; Ndao y Diène, 2022), respaldado por una conciencia social colectiva con fines ideológicos, económicos e identitarios. Por otro lado, están las lenguas locales y minoritarias que representan las «lenguas básicas» (Calvet, 1996; Daff, 1998; Larsson, 2006).

Por tanto, las lenguas, lejos de ser un simple medio de comunicación, representan un potencial ideológico, un capital económico y un poder social y político (Moure, 2011, p. 74). En el macromercado lingüístico característico de la modernización y globalización, sobre todo en los centros urbanos, domina el francés en pugna con otras lenguas. extranjeras como el inglés, el español, etc. Esta macroeconomía simbólica está sujeta a una ideología enraizada en el mundo occidental basada en la utopía de un «pueblo planetario» (Said, 1980 y 2012) y este ideal puede ser un obstáculo a la promoción de la diversidad, frenando el florecimiento de los pueblos minoritarios.

En el micromercado lingüístico, de forma análoga al nivel macro, el sector informal privilegia y favorece el florecimiento de las lenguas locales, con miras político-lingüísticas y económicas puestas en el nacionalismo. En las grandes ciudades como Dakar, Saint-Louis y Ziguinchor, el wolof tiene un valor inigualable y domina el mercado.

5. MECANISMOS DE REGULACIÓN DEL PANORAMA ECOLINGÜÍSTICO EN SENEGAL

El cambio de código constituye, hoy, uno de los mecanismos de regulación lingüística en Senegal. A través de la alternancia de código y código mixto wolof-francés y/o wolof-lengua local, ha emergido una forma de hablar popular también llamado wolof urbano o wolof mixto (Thiam, 1992; Auzanneau, 2006; Rasoloniaina, 2000). La emergencia de este lenguaje urbano se caracteriza por una interferencia lingüística marcada por la presencia de rasgos léxicos y estructurales de diferentes lenguas (*serer, pulaar, diola*) y/o del inglés (Dreyfus y Juillard, 2004). De allí nace el fenómeno de la wolofización a través de la koineización de la lengua común.

La aparición y el desarrollo del wolof urbano dio un cambio en la configuración sociolingüística senegalesa debido a dos principales mecanismos de regulación. El primero es la emergencia y aceptación colectiva de una nueva identidad urbana híbrida motivada por el éxodo rural y las migraciones masivas a los centros urbanos (Dreyfus y Juillard, 2004; Ngom, 2004; Mbaya, 2005; Rasoloniaina, 2000). Este giro fue facilitado también por algunos factores lingüísticos, como la adopción de una estructura gramatical simplificada y menos compleja y un léxico muy flexible y permeable a la incorporación de préstamos de otras lenguas. Otras características de esta habla común es el carácter transcultural, no «etnicizado» (Juillard y Calvet, 1996; Rasoloniaina, 2000) y estructurado en torno a un poder socioeconómico basado en la integración colectiva.

El segundo mecanismo de regulación es la indigeneización de la lengua colonial, gracias a la emergencia de un francés local senegalés, con nueva estructura léxica y sintáctica. Asimismo, hay que mencionar la emergencia de las lenguas locales y regionales que ocasionó la aparición de nuevas actitudes y comportamientos sociolingüísticos, a saber:

- La resistencia entre algunas comunidades de habla (*pulaar, mandinga, diola*), que no sólo persisten en utilizar su lengua en el ámbito familiar y próximo, sino que extienden su uso a los intercambios con personas desconocidas.

- El reconocimiento oficial de las lenguas minoritarias y normalización, que contribuyen a neutralizar los conflictos interlingüísticos, aunque de forma imperfecta.

6. REFLEXIONES FINALES

En suma, la interacción en el ecosistema sociolingüístico senegalés no está exenta de situaciones conflictivas entre determinadas lenguas, y el plurilingüismo social ha sido la forma de regulación histórica ligada a la adaptación a su medio natural. A pesar de que en otras partes del mundo la diversidad lingüística se contempla como desventaja y fuente de conflictividad, los diferentes pueblos de Senegal han sabido gestionar el multilingüismo y probablemente han formado un sistema de equilibrio dentro de una lógica *ecológica* que establece mecanismos en los que median aspectos más o menos violentos y más o menos pacíficos. Actualmente, uno de los factores de paz interlingüística más importantes en Senegal es la confluencia unificadora en una posible koiné en torno al wolof urbano. Este proceso puede asegurar una paz lingüística en la medida en que se aleje de una wolofización uniformadora y siga su curso evolutivo multiforme y unificador.

Por lo demás, las lenguas nacionales de Senegal no se consideran como simples medios de comunicación. Cada una tiene una dimensión identitaria, social, política y económica que la hace imprescindible, porque se encierra un determinado conocimiento del mundo, modos de ser y de hacer adaptados al entorno y contruidos históricamente que desaparecerían con ellas. Para dar un ejemplo típico de Senegal, la comunidad wolófona se asocia al comercio y a los negocios, la pulaarófona, a la ganadería de bovinos, la sererófona, a la agricultura y a la pesca, mientras que la *diola* se relaciona con las plantaciones de frutas en las selvas de Casamance.

Para terminar, las políticas lingüísticas deben contemplar este panorama sociolingüístico y socioeconómico antes de emprender cualquier forma de regulación. Porque la diversidad lingüística y cultural en Senegal asegura una paz colectiva y

sostenible y favorece la diversificación de ecosistemas naturales debido a que determinadas etnias y lenguas se relacionan con unas actividades socioeconómicas concretas.

REFERENCIAS

- Auzanneau, M. (2006). *Quelques indices pour l'étude de la dynamique sociolinguistique de Saint-Louis*. En A, Sarr, E. Makward, A.T. Fofana y C. Frederick (eds.), *The histories, languages, and cultures of West Africa: Interdisciplinary essays* (pp. 58-69). Lewiston: Edwin Mellen.
https://hal.archives-ouvertes.fr/haloo927295/file/Quelques_indices_pour_la_etude_de_la_dynamique_sociolinguistique_de_Saint-Louis.pdf
- Bastardas i Boadas, A. (1996). *Ecologia de les llengües. medi, contactes i dinàmica sociolingüística*. Barcelona: Proa.
- Bourdieu, P. (1980). Ce que parler veut dire. *Questions de sociologie* (pp. 95-112). París: Minuit.
- Calvet, L. J. (1994). Quel modèle sociolinguistique pour le Sénégal ? Ou il n'y pas que le véhiculaire. Le plurilinguisme au Sénégal. *Langage et société*, 69, 89-107.
https://www.persee.fr/doc/lsoc_01814095_1994_num_68_1_2658
- Calvet, L.-J. (1996). *Les politiques linguistiques*. París: Presses Universitaires de France.
- Cissé, M. (2005). Langues, Etat et société au Sénégal. *Sudlangues*, 5, 99-113.
- Daff, M. (1996). Situation et représentation du français au Sénégal. En C. Juillard, C. y L.-J. Calvet (eds.) *Les politiques linguistiques: Mythes et réalités* (pp. 143-148). Montréal: AUPÉLF-AUREF.
- Daff, M. (1998). L'aménagement linguistique et didactique de la coexistence du français et des langues nationales au Sénégal. *DiversCité Langues*, 3, 14.
<http://www.uquebec.ca/diverscite>
- Diatta, J. S. (2021). Plurilinguisme et gestion de la covid-19 au Sénégal: quelle contribution des langues locales à « l'effort de guerre»? *Akofena*, 3(1), 89-104.
- Diouf, I., Tidiane Ndiaye, C. y Binta Dieme, N. (2017). Dynamique et transmission linguistique au Sénégal au cours des 25 dernières années. *Cahiers québécois de démographie*, 46(2), 197-217. DOI: <https://doi.org/10.7202/1054052ar>
- Dreyfus, M. y Juillard, C. (2004). *Le plurilinguisme au Sénégal: langues et identités en devenir*. París: Karthala.
- Fill, A. (2001). Language and ecology: ecolinguistic perspectives for 2000 and beyond. *AILA Review*, 14, 60-75.
- Friedrich, P. (2007). English for peace: Toward a framework of peace sociolinguistics. *World Englishes* 26(1), 72-83.
- Hagège, C. (1985). *L'homme de paroles. Contribution linguistique aux sciences humaines*. París: Fayard.
- Halliday, M. A. K. (2001). New ways of meaning. The challenge to applied linguistics. En A. Fill y P. Mühlhäusler (eds.), *The Ecolinguistics reader. Language, ecology and environment* (pp. 175-202). Londres: Continuum.
- Haugen, E. (1972). The stigmata of bilingualism. *The ecology of language* (pp. 306-324). Stanford: Stanford University Press

- Juillard, C. y Calvet, L.-J. (eds). (1996). *Les politiques linguistiques: Mythes et Réalités*. Montréal: AUPELF-AUREF.
- Larsson, E. (2006). *Les politiques linguistiques au Sénégal et au Maroc. Comment ces pays-ont-ils réussi à réaliser les buts qui constituent la déclaration d'Harare?*. Universidad de Lund.
<https://lup.lub.lu.se/student-papers/record/1329456/file/1329457.PDF>
- Lechevrel, N. (2010). L'écologie du langage d'Einar Haugen. *Histoire, Epistémologie, Langage*, 32(2),151-166.
- Leclerc, J. (2014). *Sénégal. L'aménagement linguistique dans le monde*, Québec, CEFAN, Université Laval. <http://www.axl.cefan.ulaval.ca/afrique/senegal.htm> [15/04/ 2015].
- Mbaya, M. (2005). *Pratiques et attitudes linguistiques dans l'Afrique d'aujourd'hui: le cas du Sénégal*. Múnich: LINCOM EUROPA.
- Morillas, A. M. M. (2003) *La nada en el segundo Heidegger y el vacío en oriente. Hermenéutica contrastiva*. Tesis Doctoral. Universidad de Granada.
- Moure, T. (2001). *Universales del lenguaje y linguo-diversidad*. Barcelona: Ariel.
- Mühlhäusler, P. (1986). *Pidgin and Creole linguistics*. Westminster: University of Westminster Press.
- Ndao, P. A. (1984). Aspects linguistiques et sociolinguistiques de la situation sénégalaise: français et langues nationales. Tesis Doctoral, Université de Rouen.
- Ndao, P. A. (1996). *Contacts de langues au Sénégal: étude du code switching Wolof-Français en milieu urbain: approches linguistique, sociolinguistique et pragmatique*. Tesis Doctoral, Université Cheikh Anta Diop de Dakar.
- Ndao, P. y Diène, M. (2022). Écriture en wolof, pratiques glottopolitiques et stratégies de normalisation langagière. *Glottopol*, 36, 41-59.
 DOI: <https://doi.org/10.4000/glottopol.1883>
- Ngom, F. (2004). Ethnic identity and linguistic hybridization in Senegal. *International Journal of the Sociology Language*,170, 95-111.
- Rasoloniaina, B. (2000). *Etude des représentations linguistiques des serer*. París: L'Harmattan.
- Said, E. W. (1980). *L'Orientalisme. L'Orient créé par l'Occident*. París: Seuil.
- Said, E. W. (2012). *Culture and imperialism*. Nueva York: Vintage.
- Seydi, O. (2021). Análisis de las políticas y planificaciones lingüísticas postcoloniales de Senegal desde la ecolingüística. *Revista de Filología*, 43, 257-267.
- Seydi, O. (2021). Representaciones sociolingüísticas del wolof y del pulaar entre estudiantes de lenguas de la Universidad Gaston Berger de Saint-Louis (UGB). *Estudios interlingüísticos*, 9, 203-218.
- Steffensen, S. V. y Fill, A. (2014). Ecolinguistics: the state of the art and future horizons. *Language Sciences*, 41, 6-25.
- Thiam, N. (1992). Nouveaux modèles de parlars et processus identitaires en milieu urbain: le cas de Dakar. *Actes du Colloque International « Des langues et des villes de Dakar »* (pp. 495-512). París : ACCT Érudition.

RESEÑAS

RESEÑAS

HORNO CHÉLIZ, Mamen. *Un cerebro lleno de palabras*¹. Barcelona: Plataforma, 2024, 184 pp., ISBN: 978-84-10079-20-5.

Un cerebro lleno de palabras es el resultado de más de veinte años de investigación sobre el lexicon mental. En este libro, pues, confluyen de manera armoniosa la rigurosidad metodológica —producto de tantos años de trabajo— con la claridad y sencillez de su prosa, reflejo de su espíritu divulgativo, al servicio de la comunidad.

La relación mente/cerebro y lenguaje ha sido objeto de interés durante mucho tiempo no solo para investigadores en el campo de las neurociencias y la psicolingüística, sino también para el ser humano común y corriente. A su vez, durante los últimos años, se ha puesto especialmente en evidencia la necesidad de acortar, como mencionaba Kandel, la gran brecha entre las ciencias fundamentales y las aplicaciones prácticas, e incluso se ha llegado a ver el siglo XXI como «el siglo de la ciencia de la mente» (Arévalo y Abusamra, 2016). Ahora bien, en la actualidad, nos enfrentamos a una realidad donde la sobreabundancia de información al respecto es moneda corriente. Constantemente, nos encontramos frente a la presencia de datos que muchas veces son falsos (como en el caso de las *fake news*), imprecisos o carentes de rigor (como los «neuromitos», creencias, comunes pero erróneas, sobre cómo funciona el cerebro). En este contexto, *Un cerebro lleno de palabras* funciona como una «lente» que nos permite visualizar y comprender las relaciones entre cerebro, lenguaje y cognición de manera sumamente clara y precisa.

La autora expone su análisis y reflexión en torno al uso y almacenamiento de las palabras en el cerebro, a partir de la complementación y el contraste de diversas fuentes y en función de ejes temáticos específicos (Zunino y Muraca, 2012). En efecto, los cinco ejes tratados en este libro se estructuran en sendos capítulos presentados

1. Libro galardonado con el *Premio Archiletras* 2024 a la publicación del año.

por la autora en la introducción, los cuales conforman un camino que va desde el almacenamiento (capítulo 1) y la adquisición (capítulo 2) de las palabras hasta su pérdida (capítulo 3), sus efectos en la vida emocional y en la salud mental de los humanos (capítulo 4) y, finalmente, en la presencia de «cerebros sin palabras» (capítulo 5). Intercalados en el texto, cada capítulo incluye recuadros (compuestos por un título, una pequeña imagen y un texto breve) que permiten vincular cada uno de los temas con nuestro quehacer cotidiano. La autora emplea estos elementos paratextuales como una guía de lectura, un facilitador de la información para los lectores (Alvarado, 1994). Entre ellos, encontramos los siguientes:

- «Reflexiones sobre la vida cotidiana»: consejos prácticos vinculados a la información teórica revisada. Por ejemplo, en el capítulo 1 la autora menciona que, dado que se ha demostrado que cuantas más relaciones tenga una palabra en nuestro lexicón mental, más fácilmente acude a nuestra mente cuando la necesitamos, aumentar nuestro lexicón hará que tengamos más palabras disponibles y accedamos a ellas con mayor velocidad. Es decir, que seamos mejores oradores.
- «Aplicaciones prácticas»: información sobre aplicaciones concretas vinculadas con el (sub)tema abordado, como por ejemplo la mención a juegos de gamificación (p. ej., *Cerebriti*, *Educaplay*, *Classtools*, *Learning Apps* o *Quizlet*) para promover el aprendizaje del léxico (capítulo 2).
- «Desmontando mitos con información»: por ejemplo, la falsa creencia de que las lenguas de señas son inferiores y distintas a las orales (capítulo 1).
- «Charlas interesantes», «Lecturas recomendadas» y «Películas interesantes»: espacios en los que la autora recomienda y facilita recursos (vídeos de *Youtube*, referencias a libros, películas...) para que el lector pueda seguir informándose sobre los diferentes aspectos tomados en consideración, a la vez que se entretiene y aprende.

Antes de detenernos a presentar y describir cada uno de los capítulos que conforman esta obra, es preciso destacar que, a lo largo de ellos, lejos de abrumar al lector mediante la exposición de un cúmulo de información, la autora emplea diversos recursos para explicar con nitidez la complejidad de nuestro cerebro «lleno de palabras», tales como el uso de metáforas, analogías, ejemplos o reformulaciones. Incluso llega a interpelar al lector de manera directa mediante la propuesta de juegos con el lenguaje o la invitación a realizarse preguntas mediante las cuales establecer una analogía que permita comprender mejor la temática abordada. De esta manera,

logra con éxito el objetivo de todo texto predominantemente explicativo²: dar cuenta de un saber o conocimiento, volver nítido aquello que *a priori* se presenta incierto u oscuro (di Vincenzo, García y Podetti, 2019). Estos procedimientos permiten a la autora establecer un contrato cognitivo con el lector, apelar a su comprensión, buscando una simetría de saberes entre ambos (quien explica y quien recibe la explicación) (Arnoux, di Stefano y Pereira, 2002).

En el capítulo 1, «¿Cómo y dónde almacenamos las palabras en el cerebro?», la autora atiende la relación de las palabras en el léxico mental (las redes léxicas), deteniéndose en su funcionamiento y haciendo mención a diversos experimentos que han permitido conocer más a fondo su naturaleza. A continuación, se centra en explicar, de manera sencilla y clara —pero sin dejar de lado la inclusión de tecnicismos ni de las voces de expertos que han investigado sobre el tema— cómo accedemos a las palabras de nuestro léxico, las variables que influyen en este proceso y las aplicaciones prácticas de comprender este funcionamiento (p. ej., para la enseñanza de una segunda lengua). Posteriormente, destaca que, más allá de que el ser humano tienda a ser políglota, el léxico mental es uno solo, y brinda evidencia concreta al respecto, proveniente de experimentos neurocientíficos (como el estudio del efecto de *priming* o del procesamiento de cognados). Por último, finaliza con un apartado en el que brinda una reflexión acerca de cómo almacenamos las palabras en el cerebro, a partir de diversos hallazgos científicos como los de Paul Broca, en un inicio, o los de la neurobiología y sus técnicas de exploración del cerebro; por ejemplo, la tomografía por emisión de positrones (TEP) y, más recientemente, la resonancia magnética funcional (RMf).

En el capítulo 2, «¿Cómo han llegado hasta aquí? La adquisición del léxico nos ocupa toda la vida», la autora presta atención a la adquisición del léxico desde que somos bebés, deteniéndose en la explicación de lo que sucede en el caso de bebés bilingües y las ventajas de tener una infancia multilingüe. A continuación, reflexiona en torno a dos ideas clave del capítulo: (a) la adquisición del léxico no se limita a la primera infancia y (b) la posibilidad de continuar aprendiendo palabras a lo largo de toda la vida constituye una enorme ventaja para el ser humano. Aquí, Mamen Horno Chéliz se detiene en explicar, con su característica prosa clara y el empleo de recursos retóricos que facilitan la comprensión, conceptos esenciales en torno a

2. Si bien los textos de divulgación suelen ser catalogados como «explicativos», es preciso tener en cuenta que, generalmente, todo texto suele presentar segmentos con modos de organización diferentes: narrativos, argumentativos, explicativos, descriptivos o dialogales (García y Valente, 2019). En efecto, si bien este ensayo es explicativo, contiene fragmentos puramente argumentativos (la autora destaca, p. ej., la relevancia de considerar que la lengua de signos tiene las mismas características que la lengua oral y brinda razones para ello, como veremos a propósito del capítulo 1), así como también narrativos (p. ej., al narrar el argumento de las «Películas interesantes» que la autora recomienda).

los procesos de memorización, repetición y repaso. Destaca también el papel de los diccionarios y los corpus en este aprendizaje continuo. Finalmente, concluye con un apartado dedicado a desmentir el mito que sostiene que hay diferencias entre las mujeres y los varones en el uso y adquisición del léxico.

El capítulo 3 se titula «La pérdida de las palabras». Aquí, la autora brinda evidencia y reflexiona sobre el hecho de que, a veces, «podemos perder la capacidad mágica de hacer aparecer en nuestra mente todas las palabras que necesitamos» (p. 75). Se detiene en explicar qué sucede a los pacientes con anomia, sus causas y variantes, haciendo hincapié en la necesidad de comprender las dificultades específicas que tiene cada tipo de paciente, con el fin de poder ayudarlos más y mejor. Para ello es fundamental, a su juicio, la presencia de un equipo interdisciplinar. En este sentido, Horno Chéliz coincide plenamente con Igoa (2024). El segundo apartado aborda el caso de los migrantes de larga duración y cómo la pérdida de la lengua materna puede representar, para ellos, una pérdida de la identidad. Ahora bien, es en este punto donde la autora subraya, de manera acertada, que estos migrantes, lejos de sufrir un proceso de pérdida, están viviendo un proceso de cambio lingüístico y, por ello, constituyen «un verdadero tesoro», «un regalo para las sociedades multiculturales en las que vivimos» (p. 94). El capítulo se cierra con un apartado dedicado al mutismo selectivo, destacando el hecho de que se trata de un trastorno primario de ansiedad que puede revertirse, pero, para ello, es fundamental la detección e intervención temprana. De hecho, la autora enfatiza esta idea para los tres casos de pérdida de acceso al léxico mostrados en este capítulo: para evitar las graves consecuencias en la vida cotidiana de estas personas, «no hay mejor medida que una intervención temprana» (p. 101).

El penúltimo capítulo se titula «Palabras que dañan, palabras que sanan» y, tal como su nombre indica, aquí Mamen Horno Chéliz se detiene en explicar por qué la función comunicativa de las palabras es «la menos interesante de todas». En el primer apartado muestra cómo las palabras que elegimos al hablar pueden influir tanto en el estado de ánimo de los demás como en el propio, y se subraya la relevancia de controlar el modo en el que hablamos, evitando un habla tóxica que afecte nuestra salud o la de otros. Luego, la autora se detiene especialmente en el uso y función de las palabras malsonantes (es decir, palabrotas, blasfemias, maldiciones, entre otras groserías) y cómo el empleo de estos términos puede tener tanto un efecto nocivo como benéfico en nuestra salud. La autora combina nuevamente un lenguaje sencillo y sumamente claro con la presencia de términos y referencias precisas que dan cuenta de la evidencia científica disponible al respecto. Posteriormente, reflexiona en torno al valor de los nombres que elegimos para denominar a los demás —especialmente cuando

se trata de grupos minoritarios— y cómo esta elección puede generar inseguridad en el hablante o lastimar emocionalmente a las personas designadas (p. ej., a quienes presentan alguna discapacidad, denominados en principio *disminuidos* y luego *discapacitados*). Frente a esta problemática, la respuesta de la Mamen Horno Chéliz es «preguntarles a los colectivos cómo quieren que se les denomine. La experiencia me dice que ellos suelen tenerlo siempre bastante claro» (p. 131).

El quinto y último capítulo, «Un cerebro sin palabras», aborda la apasionante relación entre cerebro y lenguaje: ¿existe el pensamiento sin lenguaje, o son las palabras las que generan la existencia? Estas preguntas son respondidas en el primer apartado, a partir de la mención a las clásicas investigaciones al respecto (p. ej., la de Edward Sapir, Franz Boas y Benjamin Whorf), pero también a partir de la referencia a especialistas actuales como Dan Slobin, que han profundizado en los estudios en torno a la enorme influencia que tienen las palabras en nuestro pensamiento. Posteriormente, la autora se centra en explicar qué sucede con las personas que, por diversas razones, no han adquirido una lengua materna: los llamados «niños salvajes» o las personas sordas que no han adquirido una lengua de señas. Horno Chéliz desmiente enfáticamente los mitos en torno a la naturaleza de las lenguas de signos y del bilingüismo: «Todos los seres humanos, sin excepción, necesitamos alimentar nuestro cerebro con palabras. No permitamos que a ningún bebé se le prive de este derecho por desconocimiento por parte del entorno de las evidencias científicas» (p. 155). El capítulo concluye con un apartado dedicado a la eventual existencia de pensamiento pre lingüístico y a la ventaja que supone para los humanos tener «un cerebro lleno de palabras», pues esta condición nos brinda la valiosa posibilidad de «segmentar la realidad caótica en la que vivimos en conceptos aislados que podemos comprender» (p. 159).

El libro se cierra con un epílogo en el que se formula una pregunta crucial: ¿qué entendemos por *palabra*? Si bien puede resultar llamativo que se aborde este interrogante al final y no al principio de la obra, Mamen Horno Chéliz ofrece los argumentos por los que lo plantea a esta altura del recorrido.

En definitiva, este libro constituye un hermoso viaje de aprendizaje, reflexión y, sobre todo, concientización en torno a la enorme ventaja que supone, para nuestra especie, tener «un cerebro lleno de palabras».

REFERENCIAS

- Alvarado, M. (1994). *Paratexto*. Oficina de Publicaciones del CBC, Buenos Aires.
- Arévalo, A. y Abusamra, V. (2019). Neuromitos. En V. Abusamra y A.R. Ferreres *Neurociencias y educación*. Buenos Aires: Paidós
- Arnoux, E. N. de; di Stefano, M. y Pereira, M. C. (2002). *La lectura y la escritura en la universidad*. Buenos Aires: Eudeba.
- di Vicenzo, D, García, M. y Podetti, M. (2019) Los géneros en los que predomina la explicación. En E. Valente y M. García (Coords.) *Las prácticas de lectura y escritura en la comunidad académica* (pp. 53-86). Los Polvorines: Universidad Nacional de General Sarmiento.
- García, M. y Valente, E. (2019) Nociones en torno a la lectura y la escritura. En E. Valente y M. García (Coords.) *Las prácticas de lectura y escritura en la comunidad académica* (pp. 15-50). Los Polvorines: Universidad Nacional de General Sarmiento.
- Igoa, J. M. (2024, 22-25 de enero). ¿Cómo tiene que ser el lenguaje para que podamos aprenderlo y usarlo? Oportunidades y restricciones que ofrece la psicología a la lingüística [conferencia]. *LII Simposio de la SEL*, Madrid, España.
<http://sel.edu.es/wp-content/uploads/Resumen-conferencia-JM-Igoa-SEL2024.pdf>
- Zunino, C. y Muraca, M. (2012). El ensayo académico. En L. Natale (coord.). *En carrera: escritura y lectura de textos académicos y profesionales* (pp. 161-78). Los Polvorines: Universidad Nacional de General Sarmiento.

AILÍN PAULA FRANCO ACCINELLI
CHIPME-CONICET

BAÑOS, José Miguel, JIMÉNEZ LÓPEZ, María Dolores, JIMÉNEZ MARTÍNEZ, María Isabel, y TUR, Cristina (Eds.), *Collocations in Theoretical and Applied Linguistics: From Classical to Romance Languages / Las colocaciones en la lingüística teórica y aplicada: de las lenguas clásicas a las lenguas romances*. Madrid: Sociedad Española de Estudios Clásicos y Guillermo Escolar Editor, 2022, 404 pp., ISBN: 978-84-18981-87-6.

Esta excelente monografía se centra en las colocaciones verbo-nominales (CVN) del tipo *hacer una promesa, entrar en guerra o darle miedo a alguien*. Se trata de combinaciones léxicas restringidas en las que el sustantivo (la base de la colocación) sirve de principal predicado semántico, mientras que el verbo (el colocativo), «independientemente de su mayor o menor carga semántica, tiene sobre todo un valor funcional», contribuyendo a la expresión de «categorías gramaticales como persona, tiempo, modo, aspecto o voz» (p. 7, la traducción es mía) –unas expresiones a las que los autores del volumen también se refieren indistintamente como «construcciones con verbo soporte»¹.

Las propiedades sincrónicas de las CVN en las lenguas modernas han sido objeto de estudio durante varias décadas, pero la investigación sobre estas expresiones en las lenguas clásicas es escasa y bastante reciente. La monografía editada por Baños, Jiménez López, Jiménez Martínez y Tur, pues, contribuye a cubrir un hueco significativo. Aunque el centro de atención son las CVN del griego antiguo al bizantino y, sobre todo, del latín clásico y postclásico, también se discuten varias lenguas romances, en especial el español. Como señalan los propios editores, otra contribución clave es un enfoque interdisciplinario que pretende «enfatar la importancia de estas construcciones con respecto a varias áreas diferentes aunque relacionadas de la lingüística» (p. 9; la traducción es mía).

¿Qué características presentan las colocaciones verbo-nominales de las lenguas clásicas, y en qué se distinguen de la combinatoria libre y las unidades fraseológicas? ¿Cómo se comparan estas colocaciones con sus equivalentes en las lenguas modernas dentro y fuera de la familia romance? ¿Qué principios lingüísticos, estilísticos y cognitivos contribuyen a la selección de la base nominal y el verbo colocativo? ¿Y cómo surgen y evolucionan las CVN con el paso del tiempo? Estas son algunas de las preguntas clave que se abordan en el volumen.

1. En este uso, «verbo soporte» engloba no solo aquellos predicados cuya contribución semántica a la expresión en su conjunto es nula o redundante con respecto a la que aporta el sustantivo (p. ej. *tener*), sino también, como veremos más adelante, los que añaden información aspectual (incoativa, durativa o terminativa, p. ej. *entrar*) o diatésica (causativa o pasiva, p. ej. *dar*).

Tras una introducción muy efectiva (pp. 7-18), la monografía reúne once contribuciones (cuatro escritas en inglés, cinco en español, una en francés, y otra más en italiano) agrupadas en tres secciones. En la primera sección se examinan ciertas CVN del latín tanto desde un enfoque tipológico como dentro de la lingüística cognitiva. La segunda sección estudia las CVN del latín desde la perspectiva de la lingüística aplicada, incluyendo la pragmática, la traducción bíblica, la estilística y la pedagogía. Finalmente, la tercera parte comprende tres estudios diacrónicos: dos que investigan, respectivamente, la renovación de las CVN en griego postclásico y su evolución del latín a las lenguas romances, y un tercero que ofrece una panorámica general de los estudios diacrónicos de estas estructuras en español.

La primera sección se abre con la contribución de Roland Hoffmann («Latin Support Verb Constructions: A View from Language Typology» [Construcciones con verbo soporte en latín: una visión desde la tipología lingüística], pp. 21-56). Tras repasar las propiedades semánticas y sintácticas que distinguen las construcciones con verbo soporte (CVS) de las unidades fraseológicas, Hoffmann documenta tales construcciones en numerosas lenguas dentro y fuera de la familia indoeuropea. El autor propone una definición tipológica de las CVS basada no tanto en propiedades morfosintácticas específicas, sujetas a bastante variación interlingüística, como en su función, y ofrece un breve análisis comparativo de las CVS en latín, árabe, chino mandarín, alemán, japonés y suajili. Hoffman subraya que, como es típico de estas construcciones, las CVS del latín suelen corresponder a un verbo relacionado semántica y morfológicamente con la base nominal (p. ej. *spem habere* ‘tener esperanza’ y *sperare* ‘esperar, tener esperanza’), aunque también pueden cubrir huecos léxicos (p. ej. *odio esse* ‘ser objeto de odio’, que compensa la falta de pasiva del verbo defectivo *odi* ‘odiar’), o bien expresar distinciones aspectuales (p. ej., la fase inicial de un evento, como en *spem capere* ‘concebir esperanza’) o diatésicas (como la causación, p. ej. en *spem dare* ‘dar esperanza’); además, suelen formar paradigmas en torno a la base nominal (p. ej. *spem habere/capere/dare* ‘tener/concebir/dar esperanza’) o al verbo colocativo (p. ej. *spem/laetitiam dare* ‘dar esperanza/alegría’)².

En el siguiente trabajo («Combinatoria léxica y lenguaje figurado: algunas consideraciones sobre colocaciones latinas en la frontera con la fraseología», pp. 57-80), Cristina Tur investiga el contraste entre las colocaciones verbo-nominales

2. Dependiendo de su autor, las diferentes contribuciones al volumen citan los verbos del latín con su forma en infinitivo (p. ej. *habere* ‘tener’) o bien en la primera persona del singular del presente de indicativo (*habeo*, en el caso de ‘tener’). Aquí empleamos el infinitivo.

—que suelen permitir las mismas operaciones sintácticas de la combinatoria libre, aunque presentan diferentes grados de composicionalidad semántica— y las unidades fraseológicas, caracterizadas por un alto grado de fijación morfosintáctica y un significado no composicional. Para ilustrar dicho contraste, situado en un *continuum* combinatorio de límites difusos, la autora analiza dos tipos de CVN del latín formadas con el sustantivo *manus* ‘mano’: uno que emplea *manus* metafóricamente y donde el verbo colocativo codifica diferencias aspectuales y diatésicas, configurando un radio colocacional más extenso (p. ej. *in manus esse/incidere* ‘estar/caer en manos (de alguien)’, esto es, bajo su control, mediante la metáfora CONTROL ES MANIPULACIÓN DE OBJETOS); y otro tipo, más cercano a la fraseología, en el que la expresión en su conjunto se emplea metonímicamente (p. ej. *manum porrigere* ‘tender la mano’, esto es ‘ayudar’, a través de la metonimia EL GESTO POR LA ACCIÓN). Tur concluye, apropiadamente, que «para diferenciar entre combinatoria y unidades fraseológicas, resulta necesario considerar, además de las habituales pruebas sintácticas, otros mecanismos, como la metáfora y la metonimia conceptual de la Semántica Cognitiva» (p. 76).

También desde la lingüística cognitiva, Irene de Felice y Chiara Fedriani («Collocazioni verticali: metafore di orientamento *up/down* nella lingua latina» [Colocaciones verticales: metáforas de orientación *arriba/abajo* en latín], pp. 81-120) analizan el papel de las metáforas de orientación espacial (*arriba/abajo*) como base conceptual para un nutrido grupo de CVN del latín formadas con sustantivos que designan emociones. Las autoras identifican varios esquemas metafóricos, cada uno compuesto de diferentes subtipos, que contribuyen no solo a la elección del verbo colocativo (entendido siempre en un sentido figurado), sino también a los diferentes valores semánticos de la CVN, sobre todo de tipo aspectual y diatésico, así como a la expresión sintáctica de los participantes en el evento. Los esquemas principales conciben las emociones como (a) un objeto físico que, por ejemplo, el experimentador puede «agarrar desde abajo», en el sentido metafórico de ‘empezar a sentir’, como en *odium suscipere* ‘cobrar odio’; (b) seres vivos que pueden, por ejemplo, ‘levantarse’, esto es, ‘surgir’, como en *adsurgunt irae* ‘se despiertan (literalmente, se alzan) las iras’; (c) un lugar físico en el que, a modo de ilustración, alguien puede ‘caer’, como en *in metum incidere* ‘hundirse (lit. caer) en el miedo’; o (d) entidades que aumentan o disminuyen en altura o volumen y, por tanto, figuradamente, en intensidad (p. ej. en *amor crescit* ‘el amor crece’). Como bien señalan De Felice y Fedriani, la productividad de estas metáforas —que también encontramos en numerosas lenguas modernas— demuestra la importancia

de la experiencia corporeizada en la organización conceptual del campo semántico de las emociones y en su expresión mediante las CVN.

La Teoría de la Metáfora Conceptual de la Semántica Cognitiva (p. ej. Lakoff y Johnson, 1980) también inspira el trabajo que cierra la primera sección: «Colocaciones incoativas en latín y metáforas conceptuales» (pp. 121-160), de Guillermo Salas Jiménez. El autor investiga ciertas CVN incoativas del latín formadas con tres sustantivos de semántica afín: *bellum* ‘guerra’, *proelium* ‘batalla’ y *pugna* ‘lucha’. Su objetivo es dilucidar qué metáforas motivan la selección de los verbos que se combinan con dichos sustantivos para expresar figuradamente la fase inicial del evento (p. ej. *bellum inire* ‘entrar en guerra’, donde la metáfora EMPEZAR ES ENTRAR evocada por *inire* ‘entrar (en)’ se combina con la de los eventos como lugares físicos y la de un cambio de estado como un movimiento télico). Salas Jiménez demuestra que estas metáforas subyacentes, derivadas de nuestra experiencia encarnada de la realidad, determinan las diferentes clases semánticas a las que pertenecen los colocativos de las expresiones estudiadas (verbos de «coger», «entrar», «juntar» y «levantarse», entre otros). Lejos de ser arbitraria, pues, la elección de colocativos sigue patrones metafóricos que permiten reducir su imprevisibilidad. Otra conclusión importante es que «las colocaciones aspectuales son especialmente propensas a mostrar desarrollos metafóricos de sus colocativos y constituyen, por lo tanto, una categoría fértil para el estudio de la metáfora conceptual» (p. 157).

La contribución de Eusebia Tarrío Ruiz («Colocaciones verbales y pragmática: *gratiam inire* en la petición cortés», pp. 163-188) abre la sección dedicada a los enfoques desde la lingüística aplicada. El capítulo se centra en las propiedades sintácticas, semánticas y, sobre todo, pragmáticas de la CVN latina *gratiam inire* [*alicuius/ab aliquo*] ‘ganarse el agradecimiento [de alguien]’ (literalmente, ‘entrar en el agradecimiento [de alguien]’). Según Tarrío Ruiz, en dicha construcción –entre las más complejas dentro del abanico de expresiones formadas con *gratia* ‘agradecimiento’– el sustantivo aporta dos participantes semánticos que no siempre se expresan sintácticamente (un experimentante y un estímulo), mientras que *inire* contribuye no solo un sentido ingresivo, sino también un sujeto agente (el participante que empieza a disfrutar de la gratitud de otro debido a un beneficio prestado). Así, pues, *gratiam inire* complementa tanto diatésica como aspectualmente a la colocación estativa *gratiam habere* [*alicui*] ‘estar[le] agradecido [a alguien]’. Además, el hecho de que *gratiam inire* suela aparecer en contextos *irrealis* del tipo «si me haces este favor, te ganarás mi gratitud», sobre todo en diálogos y cartas, sugiere que esta CVN funciona como una fórmula de petición cortés con

la que «[e]l hablante se asegura el éxito de su demanda poniendo en primer plano el protagonismo y la ganancia de su interlocutor» (p. 185).

En el siguiente capítulo («Translation as a mechanism for the creation of collocations (II): the alternation *operor/facio* in the Vulgate») [La traducción como mecanismo de creación de colocaciones (II): la alternancia *operor/facio* en la Vulgata], pp. 189-234), José Miguel Baños y María Dolores Jiménez López investigan el papel de la traducción bíblica en la génesis y evolución de las colocaciones verbo-nominales del latín. Para ello, analizan el uso de *facere* ('hacer') y *operari* (originalmente, 'trabajar en, ocuparse de') con *abominatio* 'abominación', *iniquitas* 'injusticia', *peccatum* 'pecado' y sustantivos similares en la Vulgata, un texto heterogéneo donde se mezclan versiones revisadas y no revisadas de traducciones anteriores del griego con nuevas traducciones del hebreo. Su estudio demuestra que el idioma en el que se basa la traducción latina (el griego o el hebreo) es fundamental para explicar el uso de ambos verbos en las CVN estudiadas: mientras que las versiones basadas en textos griegos emplean *facere* y *operari* para traducir literalmente dos verbos cuyo uso se distingue en el texto fuente, las nuevas traducciones del hebreo por parte de San Jerónimo (más libres y más naturales) abandonan este contraste y utilizan *operari* como sinónimo de *facere* en la mayoría de contextos. Según Baños y Jiménez López, este nuevo uso de *operari*, que afecta tanto a las CVN como a expresiones con sustantivos concretos, es una innovación del latín bíblico que se extendió a la literatura cristiana y que se ha mantenido, aunque con ciertos cambios, en las lenguas romances (cf. *obrar milagros*). Una implicación clave es que «las colocaciones son un criterio objetivo para comparar la técnica de traducción de diferentes versiones de la misma obra» (p. 230; traducción mía).

La contribución de Iván López Martín («The use of verb-noun collocations as a criterion for attribution of authorship: the *Historia Augusta*») ([El uso de colocaciones verbo-nominales como criterio para la atribución de autoría: la *Historia Augusta*], pp. 235-255) se enmarca dentro de la estilística. El objetivo es analizar cuantitativa y cualitativamente el uso de las CVN en la *Historia Augusta* –una obra del latín tardío que consta de dos partes separadas por una laguna textual y cuya autoría ha sido objeto de debate durante décadas– para intentar dirimir si tiene uno o varios autores. López Martín documenta diferencias importantes entre las dos secciones de este texto tanto en la frecuencia general de las CVN como en el uso de los cinco colocativos más frecuentes en las mismas (*habere* 'tener', *esse* 'ser; estar; haber', *facere* 'hacer', *dare* 'dar' y *agere* 'mover hacia delante; hacer; actuar; pasar tiempo'), y concluye que «la *H[istoria] A[ugusta]* fue escrita por al menos

dos autores distintos, uno para cada una de las dos mitades de la obra» (p. 252; traducción mía). Su estudio, pues, sugiere que los patrones estilísticos asociados con las CVN, incluyendo tanto su productividad general como las combinaciones preferenciales, pueden servir como criterio para determinar la autoría de una obra.

En el último trabajo de la sección («Collocations et constructions à verbe support chez les grammairiens latins et dans quelques manuels antiques: un impensé didactique?» [Colocaciones y construcciones con verbo soporte entre los gramáticos latinos y en algunos manuales antiguos: ¿un aspecto didáctico pasado por algo?], pp. 257-301), Tatiana Taous aborda las CVN del latín desde el punto de vista de la enseñanza de idiomas. Como punto de referencia, Taous repasa brevemente las principales consideraciones pedagógicas que rodean la discusión sobre estas construcciones en las gramáticas didácticas del francés medieval, clásico y moderno, y a continuación investiga el papel que tales consideraciones habrían podido tener en el tratamiento de las CVN en textos hermenéuticos y de la tradición de las *artes grammaticae* del latín. Su estudio cuantitativo y cualitativo presta especial atención a las *Hermeneumata Pseudodositheana*, unos manuales elaborados entre los siglos I y III d.C. y utilizados por hablantes del latín que aprendían griego. Pese a la escasez de ocurrencias y la falta de referencias textuales a las prácticas docentes asociadas específicamente con estos manuales, la autora encuentra en dichos textos «indicios de un esfuerzo por desarrollar entre los aprendices una actitud reflexiva hacia la adquisición de una lengua», incluyendo el aprendizaje de las CVN y las CVS (p. 294; traducción mía).

La tercera y última sección, sobre la evolución histórica de las colocaciones, se abre con el capítulo de Alfonso Vives Cuesta y Lucía Madrigal Acero («Support-verb constructions in post-classical Greek and sociolinguistics: a diachronic study of εὐχὴν ποιέω as a level-of-speech marker» [Construcciones con verbo soporte en griego posclásico y sociolingüística: estudio diacrónico de εὐχὴν ποιέω como marcador de nivel de lengua], pp. 305-333). Su trabajo investiga la literatura hagiográfica cristiana escrita entre los siglos I y XIV en griego postclásico, caracterizado por una situación diglósica que condiciona la distribución de ciertas variantes lingüísticas según el nivel de habla. Como estudio de caso, se analizan dos CVS sinónimas formadas con el sustantivo εὐχή ‘oración’ y el verbo ‘hacer’ en voz media o activa (εὐχὴν ποιέσθαι y εὐχὴν ποιέιν ‘decir una oración’), respectivamente, de las que solo la primera se documenta en el griego clásico. El análisis sugiere que la distribución de estas variantes, cuya frecuencia de uso disminuye con el tiempo, está sujeta a las preferencias de cada autor y cambia con los siglos como resultado de la pérdida

de la voz media del griego y la evolución histórica del género hagiográfico desde la koiné baja o popular de los primeros textos, donde predomina εὐχὴν ποιεῖν, hacia la variedad más alta o prestigiosa de obras posteriores, donde se prefiere εὐχὴν ποιεῖσθαι. Según Vives Cuesta y Madrigal Acero, estos resultados implican no solo que el uso de las CVS puede servir de criterio para caracterizar diferentes niveles del habla, sino también que una descripción adecuada del griego postclásico requiere un enfoque «refilologizado» que preste atención tanto a la información filológica disponible como a los mecanismos sociolingüísticos que contribuyen al cambio histórico.

En «Continuidad y cambio en las colocaciones del latín a las lenguas romances» (pp. 335-370), María Isabel Jiménez Martínez y Chantal Melis investigan la evolución histórica de las CVN emocionales causativas del español (p. ej. *dar alegría*), el italiano (p. ej. *fare vergogna* 'dar vergüenza') y el portugués (p. ej. *meter medo* 'dar miedo'), donde el colocativo contribuye con el significado de 'hacer sentir'. Su estudio de corpus, que abarca desde el siglo XI al XX, examina las combinaciones formadas con cuatro verbos con equivalentes en las tres lenguas ('hacer', 'dar', 'poner' y 'meter') y seis sustantivos de «emoción» también compartidos ('alegría', 'placer', 'miedo', 'temor', 'vergüenza' y 'envidia'). Basándose en un trabajo anterior suyo, las autoras trazan los orígenes de estas construcciones en el latín clásico, donde ya encontramos los mismos cuatro colocativos: *facere*, el más común con los sustantivos analizados (p. ej. *pudorem facere* 'dar vergüenza'); *dare*, usado principalmente con sentimientos positivos y en textos poéticos (p. ej. *laetitiam dare* 'dar alegría'); y *mittere* y *ponere*, que se empleaban preferentemente con nombres de «miedo», pero no como causativos, sino con un sentido terminativo (p. ej. *timorem mittere/ponere* 'dejar de sentir temor'). Aunque las tres lenguas estudiadas han mantenido la preferencia de 'dar' con las emociones positivas y de 'poner/meter', ya en su nuevo uso causativo, con nombres de «miedo», una diferencia clave es que el italiano ha mantenido *fare* 'hacer' como colocativo más frecuente en las CVN analizadas, mientras que el portugués y, sobre todo, el español han expandido significativamente el uso de *dar* con el paso de los siglos. Las combinaciones que han sobrevivido en cada lengua, pues, obedecen tanto a una fuerte herencia latina como a patrones colocativos de gran arraigo histórico, incluyendo la prosodia semántica de 'dar' y el uso preferente de 'poner' y 'meter' con ciertos sustantivos.

El volumen concluye con el trabajo de Begoña Sanromán Vilas («Las colocaciones verbales en la historia del español», pp. 371-401). Sanromán Vilas hace un repaso

crítico de la investigación diacrónica sobre las CVN del español, las colocaciones más estudiadas en esta lengua. La autora distingue cuatro etapas que van desde una primera oleada de estudios con descripciones generales sobre los orígenes y la función de construcciones con *haber/tener*, *fazer/hacer* y *dar*, todavía no reconocidas propiamente como «colocaciones», a la bibliografía más reciente, centrada en la estructura semántico-sintáctica de las CVN (la segunda oleada), el papel de la lexicalización (la tercera) y la aplicación de modelos teórico-metodológicos de la lingüística sincrónica a la lexicografía y gramática históricas (la cuarta). La mayoría de estudios analizan las colocaciones como un fenómeno fraseológico (si bien algunos también enfatizan el papel de la selección gramatical) y privilegian las CVN formadas con nombres de «emoción» y verbos «ligeros» (*haber/tener*), causativos (*fazer/hacer*, *dar*, *poner*, *meter*) o incoativos (*caer*, *cobrar*, *entrar*, *prender*, *tomar* y *venir*, entre otros). Aunque la escasez relativa de trabajos y la variedad de marcos teóricos dificultan las generalizaciones, Sanromán Vilas destaca varias tendencias: tanto los verbos colocativos como las bases nominales experimentan distintos cambios históricos (por ejemplo, algunos verbos pueden predominar en una época, entrar en competición con otras alternativas y perderse o quedar restringidos a ciertas bases aisladas); la evolución de las colocaciones a veces está asociada con cambios internos y sociolingüísticos de alcance más general en la historia del español; la lexicalización parece jugar un papel más importante que la gramaticalización en el desarrollo de estas construcciones; y todavía se debate qué factores motivan la relación semántica entre la base y el colocativo y cuál de estos elementos selecciona al otro. En todo caso, «[1]as obras examinadas, aparte del valor de sus resultados empíricos y de las implicaciones conceptuales y metodológicas, presentan el gran mérito de haber iniciado este camino y de hacernos conscientes del gran potencial de la vertiente histórica de las colocaciones» (p. 395).

Fruto de varios proyectos de investigación bastante recientes, pero que ya han resultado muy productivos³, el volumen representa una contribución fundamental al estudio de las colocaciones no solo en las lenguas clásicas (y en especial el latín, que –como se subraya en varios capítulos– suele emplear las CVN más que el griego), sino también en el español y otras lenguas romances. En efecto, la monografía aborda aspectos de enorme relevancia para entender las colocaciones

3. «Interacción del léxico y la sintaxis en griego antiguo y en latín: construcciones con verbo soporte, diátesis y aspecto» (FFI2017-83310-C3-3), dirigido por José Miguel Baños y financiado por el Ministerio de Ciencia e Innovación, y el Programa Logos de ayudas a la investigación en Estudios Clásicos de la Fundación BBVA.

en cualquier idioma, desde su motivación y arbitrariedad a su génesis y desarrollo. El hecho de que se preste atención tanto al latín como a sus descendientes modernos dentro de una larga tradición textual que permite rastrear continuidades, rupturas y variaciones durante un período histórico muy dilatado es un gran valor añadido.

Otra virtud del volumen es el uso de ejemplos auténticos extraídos tanto de corpus electrónicos como de fuentes lexicográficas y filológicas tradicionales. El análisis cuantitativo y cualitativo de los datos emplea una metodología adecuada, si bien algunos aspectos podrían mejorarse. En efecto, en varias contribuciones no se establece una distinción explícita y sistemática entre frecuencia de tipos (combinaciones únicas) y frecuencia de *tokens* (ocurrencias repetidas de la misma combinación). Por ejemplo, en la discusión sobre la *Historia Augusta* se menciona que «[h]abere es el verbo soporte más productivo [...], con un total de 167 colocaciones» (p. 245; traducción mía), sin aclarar cuántas combinaciones diferentes tenemos ni cuántas ocurrencias hay de cada una. Por otra parte, en el análisis de los datos del español del capítulo sobre las colocaciones causativas emocionales, obtenidos del *Corpus del español* (*CdE*), no se especifica el manuscrito del que proceden las ocurrencias del Medievo ni se distingue la fecha de composición y la de copia de dichos manuscritos. A modo de ilustración, las autoras del capítulo datan los ejemplos del *Libro del Caballero Zifar* (pp. 345 y 346) en «ca. 1300», que corresponde a la fecha de composición, y no indican de cuál de los dos manuscritos incluidos en el *CdE*, ambos copiados entre el segundo y el tercer cuarto del xv, proceden. Aunque esto no invalida las conclusiones generales de un estudio por lo demás estupendo, se trata de un aspecto metodológico importante que debería reconocerse de forma explícita, no solo porque, al menos ocasionalmente, diferentes manuscritos del mismo texto pueden emplear distintos colocativos para las mismas bases nominales (cf. Alba-Salas, 2012, p. 8), sino también porque, en el caso de ejemplos del *Zifar* y de no pocos textos medievales incluidos en el *CdE*, seleccionar una u otra fecha determina si pertenecen o no al primer corte cronológico que las autoras establecen para su análisis diacrónico (los siglos xi-xiv).

Informadas tanto por los trabajos pioneros sobre las CVS del latín (p. ej. Rosén, 1981) como por varios estudios recientes en esta misma lengua (véanse las referencias citadas en la monografía), las contribuciones al volumen también reflejan nuevos enfoques en el estudio de las lenguas clásicas, como la Lingüística Clásica Cognitiva (p. ej. Mocciano y Short, 2019). Además, se anclan en un conocimiento sólido de las distintas propuestas teóricas que han predominado en los estudios sincrónicos sobre las colocaciones y las CVS del español y otras lenguas romances, sobre todo

dentro del Léxico Gramática (p. ej. Gross; 1981; Giry-Schneider, 1987), la Teoría Sentido-Texto (p. ej. Alonso Ramos, 2004), el generativismo (p. ej. Mendívil Giró, 1999; Bosque, 2001; De Miguel, 2008) y la fraseología (p. ej. Corpas Pastor, 1996; Koike, 2001; García-Page, 2008; Penadés Martínez, 2012)⁴. Esta variedad de enfoques complementarios –a la que, como ya hemos visto, se suman la pragmática, la sociolingüística, la estilística, la traductología, la tipología y la pedagogía– enriquece enormemente el análisis de un fenómeno de gran complejidad empírica y conceptual.

Teniendo en cuenta la atención que se presta en el volumen a los paralelismos y discontinuidades entre el latín y sus descendientes modernos, especialmente el español, sorprende que no se enfaticen más las conexiones con estudios anteriores sobre la evolución histórica de las CVN en español, algunos de los cuales también tratan brevemente tanto sus orígenes en el latín como sus equivalentes en otras lenguas romances. Dichos estudios se reseñan en el excelente capítulo de Sanromán Vilas, pero tienen un papel mínimo o nulo en las otras contribuciones. Donde más se advierte este aspecto es en el capítulo sobre las colocaciones formadas con ‘dar’, ‘hacer’, ‘poner’ y ‘meter’ más nombres de «emoción» en español, italiano y portugués. En efecto, al margen de otro trabajo de las propias autoras, dicha contribución solo menciona de pasada (p. 359) dos estudios del español que también investigan la evolución de estas mismas combinaciones desde el medievo al siglo XX con datos del *CdE* y que arrojan resultados similares: Alba-Salas (2007) y García Salido (2017b). Además, el capítulo deja de lado otros análisis diacrónicos de las causativas emocionales del español igualmente relevantes: Alba-Salas (2009) y García Salido (2017a). Tal situación podría obedecer a una falta de espacio (que, por otra parte, habría podido resolverse con referencias cruzadas al último capítulo). En todo caso, una breve discusión de estos y otros estudios afines habría contribuido a reforzar varias conclusiones clave sobre el desarrollo de las CVN analizadas, incluyendo aquellas relacionadas con la victoria de *dar* sobre *hacer*, el predominio inicial de *dar*, *hacer* y *meter* con ciertos sustantivos, el papel de la polaridad o prosodia semántica de las bases nominales en la selección del colocativo (cf. Alba-Salas, 2013), las diferencias de registro asociadas con el uso de *meter* en la lengua moderna y las posibles razones para el diferente rumbo de las causativas emocionales en español, italiano y portugués.

4. La bibliografía sincrónica sobre las CVN y las CVS del español y otras lenguas romances incluye también otras monografías y tesis doctorales que no se citan en el volumen, tanto dentro de la Teoría Sentido-Texto (Alonso Ramos, 1998; Sanromán Vilas, 2003), como en otros marcos teóricos (p. ej. Alba-Salas, 2002, desde la Gramática Relacional, y Bustos Plaza, 2005, dentro del enfoque de la lexicalización).

Otro aspecto encomiable del volumen es su organización y lo accesible que resulta para lectores que tal vez no estén familiarizados con el estudio de las CVN en las lenguas clásicas o romances. Las contribuciones suelen enfatizar el contexto histórico, lingüístico y cultural de los textos analizados, y la secuencia de los capítulos permite al lector aclimatarse gradualmente a medida que se cambia de enfoque y que el centro de gravedad pasa del latín y el griego a las lenguas romances. Es cierto que los diferentes idiomas en que están escritas las distintas contribuciones –y las citas en otras lenguas, como la que aparece en alemán, sin traducción, en la p. 46– podrían suponer un reto para quienes los desconozcan. Sin embargo, el resumen en inglés que precede cada capítulo, al igual que la conclusión que lo cierra, es de gran utilidad en este sentido. También son muy útiles las glosas y traducciones que acompañan los ejemplos citados, así como las tablas y figuras incluidas en casi todas las contribuciones. No obstante, algunas tablas carecen de glosas que facilitarían su lectura (por ejemplo, en el caso de las expresiones del griego que aparecen en la tabla 5 de la p. 322), y ciertas figuras resultan difíciles de leer (por ejemplo, en la figura 1 de la p. 315 cuesta distinguir visualmente $\pi\omicron\iota\omicron\upsilon\mu\alpha\iota$ $\epsilon\upsilon\chi\eta\acute{\nu}$ de $\pi\omicron\iota\eta\tau\acute{\epsilon}\omicron\nu$ $\epsilon\upsilon\chi\eta$, porque el sombreado de ambas categorías es casi idéntico). Además, se echa en falta un índice onomástico y de palabras clave.

En general, la argumentación de las diferentes contribuciones es coherente y matizada, con un estilo de presentación claro e informativo. No obstante, en varios capítulos la distinción conceptual entre participantes semánticos y dependientes sintácticos –fundamental para el análisis de las colocaciones verbo-nominales y las construcciones con verbos soporte– no siempre queda del todo clara, sobre todo porque ciertos términos (especialmente *argumento*) a veces se emplean de forma ambigua sin precisar si nos referimos a argumentos semánticos o sintácticos. En ocasiones, también encontramos afirmaciones un tanto sorprendentes. En particular, el estudio tipológico de las construcciones con verbo soporte nos advierte que «[...] tampoco se puede afirmar que las CVS presupongan un cierto nivel de cultura superior y elevado, sino que estas construcciones parecen encontrarse independientemente de esto en cualquier lengua, es decir, tanto en las lenguas indígenas como en las lenguas de culturas altamente desarrolladas» (p. 31; traducción mía). Aunque esta no sea la intención del autor, tal caracterización –sin otras matizaciones– parecería reforzar inadvertidamente el estereotipo de las lenguas indígenas como lenguas de culturas menos avanzadas.

El volumen está muy bien editado, pero ocasionalmente encontramos ciertos descuidos (p. ej. en la numeración de las referencias cruzadas a los ejemplos (41)-(43)

en las pp. 48-49). Algunas traducciones al inglés también podrían mejorarse, como en el caso del francés *quelques manuels antiques* ‘algunos manuales antiguos’ y un *impensé didactique* ‘un aspecto didáctico pasado por alto’, que se vierten, respectivamente, como *a few antique manuals* y *a didactic unthought* (p. 257) en lugar de con equivalentes más naturales (p. ej. *some ancient/old manuals* y *an overlooked pedagogical aspect*).

Por todas sus virtudes, esta monografía está destinada a convertirse en referencia obligada para quienes investigan la sincronía y la diacronía de las colocaciones no solo en las lenguas clásicas y sus descendientes modernos, sino también en otros idiomas.

REFERENCIAS

- Alba-Salas, J. (2013). *Cobrar miedo*: Sobre el uso histórico de *cobrar* con sustantivos que designan cualidades o estados negativos. *Scriptum Digital*, 2, 77-106.
- Alba-Salas, J. (2012). Colocaciones incoativas con *tomar* y *prender* en diacronía. *Revista de Historia de la Lengua Española*, 7, 3-38.
- Alba-Salas, J. (2009). Las estructuras tipo *meter miedo* en diacronía: Más detalles sobre la evolución histórica de las colocaciones causativas, en A. Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas: Nuevas aportaciones desde la lingüística de corpus* (pp. 345-365). Madrid: Vervuert / Iberoamericana.
- Alba-Salas, J. (2007). On the Life and Death of a Collocation: A Corpus-Based Diachronic Study of *dar miedo/hacer miedo*-type Structures in Spanish. *Diachronica: International Journal for Historical Linguistics*, 24, 207-252.
- Alba-Salas, J. (2002). *Light Verb Constructions in Romance: A syntactic analysis*. Tesis doctoral. Ithaca: Cornell University. https://www.academia.edu/7921007/2002_Light_Verb_Constructions_in_Romance_A_Syntactic_Analysis_Ph_D_Dissertation_Cornell_University
- Alonso Ramos, M. (2004). *Las construcciones con verbo de apoyo*. Madrid: Visor.
- Alonso Ramos, M. (1998). *Étude sémantico-syntaxique des constructions à verbe support*. Tesis doctoral. Montreal: Université de Montréal. https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/6777/Alonso_Ramos_Margarita_1998_these.pdf?sequence=1&isAllowed=y
- Bosque, I. (2001). Sobre el concepto de ‘colocación’ y sus límites. *Lingüística Española Actual*, 23, 9-40.
- Bustos Plaza, A. (2005). *Combinaciones verbonominales y lexicalización*. Frankfurt: Lang.
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid, Gredos.
- De Miguel, E. (2008). Construcciones con verbos de apoyo en español. De cómo entran los nombres en la órbita de los verbos, en I. Olza Moreno, M. Casado Velarde y R. González Ruiz (eds.), *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística (SEL)*, (pp. 567-578). Pamplona: Servicio de Publicaciones de la Universidad de Navarra.

- Corpus del español (CdE)* = Davies, Mark (2002). *Corpus del español*. <https://www.corpusdelespanol.org/>
- García-Page, M. (2008). *Introducción a la fraseología española*. Barcelona: Anthropos.
- García Salido, M. (2017a). Diacronía de colocaciones causativas con los verbos *meter*, *causar*, *producir* y *provocar*. *Hispanic Research Journal*, 18, 181-196.
- García Salido, M. (2017b). On causative *dar* and its alternatives in the history of Spanish. *Folia Linguistica Historica*, 38, 91-124.
- Giry-Schneider, J. (1987). *Les prédicats nominaux en français. Les phrases simples à verbe support*. Ginebra: Droz.
- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63, 7-52.
- Koike, K. (2001). *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Alcalá de Henares: Universidad de Alcalá de Henares.
- Lakoff, G. y M. Johnson (1980). *Metaphors We Live By*. Chicago: Chicago University Press.
- Mendívil Giró, J. L. (1999). *Las palabras disgregadas. Sintaxis de las expresiones idiomáticas y los predicados complejos*. Zaragoza: Prensas de la Universidad de Zaragoza.
- Mocciaro, E. y W.M. Short, (eds.) (2019). *Towards a Cognitive Classical Linguistics: The Embodied Basis of Constructions in Greek and Latin*. Berlín: De Gruyter.
- Penadés Martínez, I. (2012). *Gramática y semántica de las locuciones*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.
- Rosén, H. (1981). *Studies in the Syntax of the Verbal Noun in Early Latin*. Múnich: Fink.
- Sanromán Vilas, B. (2003). *Semántica, sintaxis y combinatoria léxica de los nombres de emoción en español*. Tesis doctoral. Helsinki: Universidad de Helsinki.
<https://core.ac.uk/download/pdf/14914838.pdf>

JOSEP ALBA-SALAS
College of the Holy Cross

LÓPEZ SERENA, Araceli (2019). *La lingüística como ciencia humana. Una incursión desde la filosofía de la ciencia*. Madrid: Arco Libros, 316 pp., ISBN: 978-84-7635-989-1.

Profundizar en los propios cimientos epistemológicos sobre los que asentamos nuestras pesquisas lingüísticas es un ejercicio clave de garantía científica. No solo en términos generales de proyección en el modo de valorar el cultivo de una lingüística rigurosa, sino, sobre todo, en la dimensión particular de asegurar desde enfoques meta-teóricos transversales que los planteamientos y las rutas metodológicas elegidos se hallan justificados. Es un ejercicio racional-estimativo como científicos, similar al que nos corresponde en la función reflexiva como hablantes. Solo la auto-mirada desde la atalaya del juicio epistemológico canaliza la precisión y la fiabilidad en las opciones de enfoque y de método en nuestras pesquisas; de otro modo, o falta objetividad y se elige lo habitual y conocido, o bien no hay compromiso probatorio establecido derivado de sopesar entre diferentes posibilidades. En un sentido parejo, únicamente la autoconsciencia, en exclusiva el llamado rasgo de diseño de la *reflexividad*, nos habilita como hablantes competentes que ejecutamos el procesamiento verbal para expresarnos; sin consciencia, no es posible ni el control en el decir ni, sobre todo, la capacidad de decisión para hacerlo.

Aun siendo procederes inexcusables, el de la valoración epistemológica (según el *conocimiento-de-observador* técnico como lingüistas), y el de la estimación consciente (según el *conocimiento-de-agente* como hablantes), sin embargo, no son ciertamente frecuentes las aproximaciones en esta esfera de mirada-meta. La trayectoria investigadora de Araceli López Serena es, entre nosotros, una brillante y muy destacada excepción. En todos sus trabajos los ingredientes epistemológicos son esenciales, sea para la consideración del problema, sea para elegir el tránsito metodológico idóneo. Sin duda, la publicación de su libro *La lingüística como ciencia humana. Una incursión desde la filosofía de la ciencia* (Arco Libros, 2019) es una excelente y nutritiva contribución a un área crucial de la Lingüística madura y bien fundada, que tanto promovemos en la labor científica de desarrollo del conocimiento, y que asimismo enseñamos en la vertiente de formación para la investigación en el lenguaje y las lenguas.

El libro contiene tres capítulos iniciales que dan fundamento a la Lingüística-comociencia (el capítulo inicial dedicado a «La Filosofía de la Lingüística como Filosofía de la ciencia hermenéutica»), y que contextualizan las coordenadas epistemológicas de nuestra disciplina sobre figuras de lingüistas-estela por sus aportaciones en este

dominio (los capítulos 2 y 3, centrados, respectivamente, en «Eugenio Coseriu y Esa Itkonen frente a frente. Lecciones de Filosofía de la Lingüística»; y en «La interrelación entre Lingüística y Filosofía en el pensamiento de Eugenio Coseriu»). A estas primeras secciones de cimientos se suman cinco capítulos más que subrayan temáticas de cariz descriptivo, pero tratadas desde tenores epistemológicos (el capítulo 4 sobre «La tensión entre teoría y norma en la Nueva Gramática de la lengua española. Una falsa disyuntiva epistemológica»; el 5 centrado en «Empiricidad y análisis del discurso. El estatus epistemológico de la lingüística de la (des)cortesía y de los sistemas de unidades de análisis del discurso»; el capítulo 6 dedicado a una espinosa y delicada cuestión como es el cambio lingüístico, «De la selección natural a la explicación racional en la aprehensión epistemológica del cambio lingüístico»; el capítulo 7 acerca de «La conformación diacrónica de marcadores del discurso: Teoría de la gramaticalización y explicación racional»; y, finalmente, el capítulo 8, en el que, a modo de corolario, fructifican las semillas epistemológicas esparcidas para abordar la historia de la lengua, «Conocimiento de agente, teoría y datos en historia de la lengua. Las hipótesis sobre la gramaticalización del por cierto epistémico en español a la luz de la Filosofía de la lingüística»).

Además de su incuestionable valor de rebato en el contexto actual del quehacer lingüístico, en donde el predominio del trabajo descriptivo-en-un-marco-teórico-concreto es abrumador, el libro de López Serena sobresale precisamente por su compromiso con la solidez epistemológica que camina a la par de los problemas analíticos que se abordan. Si bien la publicación, por su temática, se incardina en la línea de la estimación epistemológica instaurada en nuestra tradición peninsular por los trabajos de, por ejemplo, Fernández Pérez (1984, 1986, 1993), Martín Miguel (1998), Martí Sánchez (1998), o Bernardo Paniagua (1995), Moure (1996) y Jiménez Ruiz (2007), y al margen de estudios generales como Bunge (1969, 1983), Itkonen (1978, 1983), Botha (1973, 1981) o Perry (1979) y Carr (1990), entre otros, sin embargo, en el quehacer particular de López Serena han sido curiosamente investigaciones de corte descriptivo las que han exigido precisión epistemológica: un proceder parejo al de Rocío Caravedo, quien, en sus pesquisas sociolingüísticas y de lingüística de corpus, introduce bases solventes de estimación epistemológica que validen las rutas de método elegidas (Caravedo 1999).

La publicación de López Serena no se define en torno a la valoración del proceder investigador de escuelas y con objeto de apreciar sus bases de enfoque y sus métodos —como ha sido lo habitual en contribuciones sobre filosofía de la ciencia en gramática generativa, en gramática sistémica, en sociolingüística, en lingüística funcional—, sino

que el cometido de atribuir garantías y precisión a los temas de gramática descriptiva ha sido la espita que hace emerger la dimensión estimativa para enjuiciar objetivamente diferentes opciones. De modo que se constata la indudable idoneidad de meta-elementos en la aproximación y el tratamiento de los problemas. El juicio valorativo de cariz epistemológico para abordar la descripción se vuelve imprescindible, se verifica así que la reflexión sobre el proceder investigador no es una cuestión coqueta o menor en cualquier trabajo lingüístico. Ha de estar siempre focalizada porque los hechos lingüísticos tienen una determinada naturaleza, porque las técnicas y los métodos para profundizar en su análisis cambian y se reajustan y, en definitiva, porque los horizontes de las investigaciones varían en función de los objetivos que se definan.

En este sentido, la reciente aportación de López Serena sobresale porque hace visible la importancia de componentes estimativos centrales en cualquier investigación, sea incluso de cariz descriptivo y con cierta orientación previa de postulado teórico. Cuestiones como (a) la naturaleza de los hechos lingüísticos en esferas como la «norma» de la Academia, (b) el carácter real y empírico de los datos en inventarios y corpus verbales, o (c) el paralelismo e identificación entre *hechos* (= *eventos*) y *acciones* en el tratamiento evolutivo de las lenguas, se convierten en el libro de López Serena en verdaderos espejos de conocimiento susceptibles de ser contemplados en distancia y sin sesgos de escuela. Sirvan estos tres asuntos como botón de muestra de la pertinencia capital de su planteamiento epistemológico.

Los iluminadores trabajos de Renate Bartsch (1982, 1983) sobre la distinción entre *norma* y *regla*, así como las contribuciones más actuales de Esa Itkonen (2020), han clarificado con contundencia que, al margen de la diversidad de convenciones comunes entre usuarios-agentes de la lengua, toda pesquisa lingüística comporta un prisma «normativo-de-elección-prescriptiva» cuando se trata de abordar el estudio: el objeto de indagación comporta siempre selección sobre la heterogeneidad material —«es el punto de vista el que crea el objeto» (Saussure, 1916, p. 73; Fernández Pérez, 1993 § 4. 2)—. En calidad de «observador-técnico», el lingüista edifica el conocimiento teórico y descriptivo y delinea *reglas* sobre la base de normas-que-se-destacan como objeto de estudio mediante cauces de objetificación —imprescindibles en esferas *fáctico-normativas* como son las lingüísticas (Bunge, 1983, pp. 100 y ss.)—. De modo que las *reglas teóricas* se sustentan en normas (siempre objetivadas en términos de evidencias intersubjetivas) que se describen sobre criterios de selección (prescripción) justificados. En otras palabras, no hay descripción sin prescripción previa que cribe lo pertinente, como tampoco hay observación si no se dispone de formación teórica para saber ver. Las consideraciones de Elffers y de Hanson al respecto son rotundas:

The arbitrariness is then however not inherent to prescription, but a consequence of prescription not supported by argument (Elffers, 1988, p. 131).

la observación científica es una actividad ‘cargada de teoría’ (Hanson, 1971, p. 13).

El ojo nunca podría ver el sol, si no estuviera acostumbrado a él (Hanson, 1958 p. 77; epigrama de los *Zahme Xenien* de Goethe como epígrafe del capítulo 1 de *Patrones de observación*).

Así las cosas, el dilema «prescripción» *vs.* «descripción», cuando se trata de seleccionar *normas* como objeto de estudio, resulta ser impropio y engañoso: basta con suprimir el componente imperativo de la prescripción y convertirlo en un prisma idóneo de orientación justificada. Punto de vista que define la atalaya desde la que poder contemplar esas normas por su relevancia. Sea porque tienen realidad intersubjetiva en una comunidad (Coseriu), sea porque se han definido como patrón comunicativo en ciertos marcos (Hymes), sea porque se han sustentado en recuentos estadísticos de frecuencias (sociolingüística cuantitativa de Labov), sea porque se han verificado como recurrentes en un repertorio de muestras (lingüística de corpus).

En cuanto al sensible y áspero tema del llamado *empirismo* en el quehacer lingüístico, y, por concomitancia, a la eventual realidad factual de los objetos de estudio en nuestro campo —y que López Serena aborda sobre todo en el capítulo 5, en el que se discute la naturaleza ontológica de las unidades del discurso—, los sólidos argumentos que la autora desarrolla acerca del carácter hermenéutico y finalista de los elementos y métodos en el análisis del discurso no debieran, sin embargo, limitar otras rutas más allá de las categorías y reglas previas y que propicien realidades sustanciales susceptibles de ser observadas de modo cadencial en dinámicas de refinamiento paulatino: las categorías no están dadas, sino que hay que hallarlas y perfilarlas. Es lo natural en lingüística de corpus y es el planteamiento de la lingüística de construcciones. Como señala Caravedo (1999, p. 20) al considerar en la esfera de la lingüística de corpus la llamada «dependencia empírica»,

Se establece un estrecho vínculo entre teoría y datos, en el sentido de que los contenidos teóricos se construyen a partir de la percepción razonada sobre hechos ocurridos en la realidad lingüística.

Se trata de aproximaciones a los datos escasamente comprometidas con posiciones teóricas que condicionen posibles hallazgos conceptuales o también categorizaciones no previstas. Es lo propio de cauces retroductivos en cuanto al método, o de rutas *bottom-up* asociadas a la lingüística de usos/construcciones. Las dinámicas son epifenoménicas —Hopper (1998, p. 156) señala que la gramática y sus categorías no

son previas, sino que se hallan «always in a process but never arriving, and therefore emergent»— y requieren eludir lo que Langacker (1987) llama «la falacia de la lista de reglas» para así atribuir protagonismo a las observaciones mediante tácticas de ida-y-vuelta. Es lo que S. Kemmer & M. Barlow (1999, p. ix-x) definen en los siguientes términos:

The bottom-up property adds that the specific and idiosyncratic elements of the system are privileged over the general in the acquisition and operation of the system: the general arises out of the specific, and the specific is what is most directly taken from the experience.

Cierto que López Serena envidia el empirismo a la carta del positivismo y de la fase de justificación propia de las ciencias naturales, y circunscribe la pesquisa lingüística a bases exclusivamente hermenéuticas propias de las investigaciones teóricas en gramáticas-como-producto antes que gramáticas-en-proceso —lo que Itkonen (1978) calificaba de *autonomous linguistics*—. Ni siquiera considera en su fertilidad, sobre todo cuando se opera con repertorios de datos, el denominado *contexto de descubrimiento*:

En la fase de elaboración de las teorías (o, en el caso de la lingüística, en la fase de recapitulación de las normas de las que hay que dar cuenta en la gramática), es posible acudir tanto a la observación como a la intuición (...) En cualquier caso, la observación nunca se emplea en solitario, sino que siempre va acompañada necesariamente por juicios de intuición (López Serena, 2019, p. 156).

Conviene, no obstante, potenciar la noción de *ser empírico/ ser contrastable* en la actividad científica y sin necesidad de tomar como horma las ciencias naturales. Entre otros, Mario Bunge (1969, pp. 736 y ss., 1983, pp. 100 y ss.), Botha (1973) y el propio Esa Itkonen (1979, 1983) incorporaban la calidad de ser empíricas en investigaciones lingüísticas que manejan datos fáctico-normativos. Para ello, las evidencias contrastadoras han de ser datos observables (directa o indirectamente) y relevantes. Para que los datos normativos se conviertan en evidencia contrastadora, su realidad ha de comprobarse por medios indirectos o «indicadores» de su existencia: tienen realidad intersubjetiva porque soportan la comunicación (Wunderlich, 1974), son sistemáticos en su frecuencia, son *templates* en su presencia cadencial. De modo que, sin duda, cabe la contrastación en ciencias humanas como la lingüística, porque es posible hallar realidad intersubjetiva aun tratándose de *acciones* cuya ontología es fáctico-normativa (Chafe, 1992).

En cualquier caso, no debiera obviarse que hay grados de empirismo (Itkonen, 1979) y sí advertir que «ser empírico» se aplica a una investigación lingüística concreta: la cualidad ni es categórica ni tampoco se proyecta en la globalidad de una ciencia (Fernández Pérez, 1983). En Lingüística hay investigaciones —incluso gramaticales— que beben en repertorios de datos de corpus, pesquisas cuantitativas en sociolingüística, indagaciones en desarrollo longitudinal de la lengua en niños y que, naturalmente, manejan evidencia contrastadora con datos fáctico-normativos. Datos de contrastación que, en muchas ocasiones, ni siquiera se acomodan al *conocimiento de agente*, al proceder de *normas* no previstas: ¿cómo, si no, se podrían incorporar las gramáticas infantiles?, ¿de qué otro modo incluir los puzzles (así catalogados por D. Crystal) que se hallan en las producciones en casos de disfunción verbal?

Por otro lado, a la par de los matices particulares sobre «lo empírico», según las evidencias a las que se recurra por la naturaleza misma de la investigación, es importante asimismo contemplar en el denominado *contexto de descubrimiento* la naturaleza de las producciones y muestras verbales, de modo que ni los filtros de conocimiento de agente ni los tamices de una carga excesiva de teoría incapaciten y hagan inoperativo el *método retroductivo*. La actividad investigadora consiste no tanto en aplicar y contrastar teorías con ánimo de verificarlas cuanto en diseñar marcos de comprensión y procedimientos de análisis para sistematizar datos y observaciones. Hanson (1958, p. 75) señala con contundencia que «El problema no es la utilización de teorías, sino el hallazgo de éstas». La naturaleza factual de la materia y del objeto de estudio en investigaciones lingüísticas no sujetas a sesgos y servidumbres inapelables de escuela revela datos inesperados, procesos insospechados, categorías novedosas, gracias a la observación continuada y al refinamiento consecuente de los datos en la teoría y de la teoría en los datos. En el contexto de descubrimiento se trazan investigaciones sustentadas en datos factuales constantemente revisitados, así que son susceptibles de contrastación no circular; si no en coordenadas espaciotemporales, como es lo común cuando se trata de *eventos* en las ciencias naturales, sí en marcos interpersonales de *acciones* definidas por la fortaleza de su regularidad. De modo que, en suma, tanto los inventarios de muestras de discurso como las categorías y el etiquetado de las unidades disfrutan de naturaleza factual siempre y cuando no estén constreñidas por algún bias previo y hayan sido establecidas sobre procedimientos retroductivos que garanticen su ontología real objetificada. Hanson (1958, p. 112) señala con perspicacia que

el paradigma de observador no es el de quien ve y comunica lo que todos los observadores comunes ven y comunican, sino quien ve en objetos habituales lo que nadie ha visto anteriormente.

Así se investiga en lingüística de corpus y en lingüística de construcciones: apenas hay garantías de antemano sobre unidades establecidas; es la observación cuidadosa y continuada la que induce cambios en la teoría que paulatinamente se va construyendo. Posteriormente, será la *evidencia convergente*, que procede de diferentes planteamientos y fuentes de pesquisa (Schönefeld, 2011), la que otorgará mayor o menor fuerza empírica a las propuestas teóricas sobre inventarios sistematizados y datos etiquetados.

Finalmente, resulta sobresaliente en el libro la cuestión del paralelismo y la forzada identificación de corte positivista entre las ciencias naturales (que se toman como *el* modelo de cientificidad) y ciencias humanas como la lingüística, lo cual afecta tanto a tentativas de atribuir propiedades ontológicas ficticias cuanto a metodología —que no es sino técnica probatoria o experimental— y hasta a pretensiones explicativas y predictivas, que en absoluto se compadecen en ninguna dimensión con la naturaleza propia de lo que son *acciones* humanas. Es lo que, con agudeza, cataloga la autora como «cientifismo por emulación» (López Serena, 2019, p. 196). En esta línea, se dedica el capítulo 6 al tema del cambio lingüístico y a la visión evolucionista que lo contempla desde el prisma positivista de la evolución neodarwinista (con metáforas no inocuas), frente a la consideración del cambio como *acción* explicable en su finalidad y en coordenadas sociales en las que las variables no son ni rígidas ni mensurables. Donde la visión positivista aduce consideraciones generales de *leyes de evolución* a las que se otorga validez para todos los casos, la concepción del cambio lingüístico que da cabida a la compleja diversidad de factores involucrados en los comportamientos humanos incorpora coordenadas que canalizan garantías de análisis detallado sobre elementos que inciden en la variación y que pueden promover mudanzas en la lengua. El abordaje epistemológico que con brillantez se expone en el capítulo 6 transita por tres derroteros, a saber: (a) que la variación en la lengua no es una cuestión de causalidad sino de finalidad, (b) que la ontología de los hechos lingüísticos descansa en su naturaleza de *acciones* interindividuales (frente a la ontología de los *eventos* de naturaleza exclusivamente espacio-temporal) y (c) que la metodología de investigación sobre la variación y el cambio requiere no tanto de modelos evolutivos generales cuanto de variables, no todas ellas regulares para que la variación se ancle como mudanza. Diferenciar entre *explicación racional*,

explicación funcional y explicación evolutiva se vuelve esencial en el examen de la complejidad de la variación y del cambio (López Serena, 2019, p. 210 y ss.). No en vano, la profesora Araceli López Serena es docente de materias de historia de la lengua en la Universidad de Sevilla y ha cultivado de manera particular esta esfera bajo el paraguas de las teorías lingüísticas integradoras de, entre otros, E. Coseriu y R. Anttila, y al abrigo de la filosofía de una ciencia lingüística acorde con la naturaleza ontológica de las prácticas verbales como *acciones* intersubjetivas que se producen en contextos particulares.

La publicación de un libro de Epistemología de la Lingüística en tiempos supersónicos como los que vivimos constituye de entrada un reto de disonancia, ya que la reflexión meta- sobre la investigación que se desarrolla no se compadece bien con la premura, sino que requiere disposición detenida ante el presente metodológico diverso. Se incrementa el desafío cuando, ciertamente, los trabajos que se contienen en el volumen tienen visos descriptivos y han sido publicados de modo particular previamente en distintos medios. Ello prueba que el modo de investigar de la profesora Araceli López Serena se acoge siempre a principios de rigor de estimación epistemológica cuando se trata de definir el objeto de interés y edificar la metodología adecuada tanto a los propósitos de la indagación como, sobre todo, a la naturaleza de los datos. Los criterios de juicio estimativo figuran como recursos indispensables para dar calidad y garantía a la pesquisa.

Al margen del contexto de vorágine, y más allá del papel de tañido que tiene este libro, la esfera epistemológica resulta consustancial a la propia labor de investigación rigurosa que, como lingüistas, debemos cultivar. Las garantías del conocimiento que se alcance descansan en la valoración continuada de los enfoques y de los métodos que se seleccionan. Investigar no es de ninguna manera el simple manejo de recursos o la mera proyección de *un* modelo teórico; las pesquisas rutinarias sin ejecución de decisiones justificadas no casan con procesos de estimación ponderada en los que se fundamenta el saber científico objetivo y demostrado. Hacer investigación es elegir desde prismas meta- entre opciones, la selección sopesada es etapa y proceder *sine qua non* en la actividad científica. La historia de la lingüística contemplada en progresión temporal y cualitativa no es sino valoración epistemológica articulada en un marco transversal meta-teórico (Zamorano Aguilar, 2008, 2022). Sin duda, la dimensión formativa para la investigación escrupulosa orientada hacia la profundidad y la innovación ha de incluir por necesidad la vertiente epistemológica. Este libro de la profesora López Serena contiene importantes destellos de por dónde discurre ese decisivo aprendizaje.

REFERENCIAS

- Bartsch, R. (1982). The concepts 'rule' and 'norm' in linguistics. *Lingua*, 58, 51-83.
- Bartsch, R. (1983). *Norms of language. Theoretical and Practical Aspects*. Londres: Longman.
- Bernardo Paniagua, J. M.^a (1995). *La construcción de la lingüística. Un debate epistemológico*. Valencia: Publicaciones de la Universidad de Valencia.
- Botha, R. (1973). *The Justification of Linguistic Hypothesis. A Study of Nondemonstrative Inference in Transformational Grammar*. La Haya: Mouton.
- Botha, R. (1981). *The Conduct of Linguistic Inquiry. A Systematic Introduction to the Methodology of Generative Grammar*. La Haya: Mouton.
- Bunge, M. (1969). *Scientific research. Strategy and Philosophy*. Nueva York: Springer. Traducción al español de M. Sacristán. *La investigación científica. Su estrategia y su filosofía*. Esplugues de Llobregat: Ariel, 1983 (2^a ed.).
- Bunge, M. (1983). *Lingüística y filosofía*. Barcelona: Ariel.
- Caravedo, R. (1999). *Lingüística del Corpus. Cuestiones teórico-metodológicas aplicadas al español*. Salamanca: Ediciones de la Universidad de Salamanca.
- Carr, P. (1990). *Linguistic Realities. An Autonomist Metatheory for the Generative Enterprise*. Cambridge: Cambridge University Press.
- Chafe, W. (1992). The importance of corpus linguistics to understanding the nature of language. En J. Svartvik (ed.). *Directions in Corpus Linguistics* (pp. 98-103). Berlín: Walter de Gruyter.
- Elffers, E. (1978). Description and prescription in linguistics. En W. Zonneveld y F. Weerman. *Linguistics in The Netherlands (1977-1979)* (pp. 124-135). Dordrecht: Foris.
- Fernández Pérez, M. (1984). El carácter de la ciencia lingüística. *Verba*, 11, 129-156.
- Fernández Pérez, M. (1986). *La investigación lingüística desde la filosofía de la ciencia (A propósito de la lingüística chomskiana)*. Santiago de Compostela: Publicaciones de la Universidad de Santiago de Compostela.
- Fernández Pérez, M. (1993). Sociolingüística y Lingüística. *Lingüística Española Actual*, 15 (2), 149-248.
- Hanson, N. R. (1958). *Patterns of Discovery. An Inquiry into the Conceptual Foundations of Science*. Cambridge: Cambridge University Press. Traducción al español de E. García Camarero y A. Montesinos. *Patrones de descubrimiento: investigación sobre las bases conceptuales de la ciencia. Observación y explicación: guía de la filosofía de la ciencia*. Madrid: Alianza, 1977.
- Hanson, N. R. (1971). *Observation and Explanation: A Guide to Philosophy of Science*. Londres: Harper and Row. Traducción al español de E. García Camarero y A. Montesinos. *Patrones de descubrimiento: investigación sobre las bases conceptuales de la ciencia. Observación y explicación: guía de la filosofía de la ciencia*. Madrid: Alianza, 1977.
- Hopper, P. J. (1998). Emergent Grammar. En M. Tomasello, (ed.). *New Psychology of Language. Cognitive and Functional Approaches to Language Structure* (pp. 155-175). Mahwah (New Jersey): Lawrence Erlbaum.

- Itkonen, E. (1978). *Grammatical Theory and Metascience. A Critical Investigation into the Methodological and Philosophical Foundations of 'Autonomous' Linguistics*. Amsterdam: John Benjamins.
- Itkonen, E. (1979). Qualitative and Quantitative Analysis in Linguistics. En T. Perry (ed.). *Evidence and Argumentation in Linguistics* (pp. 334-366). Berlín: W. de Gruyter.
- Itkonen, E. (1983). *Causality in linguistic theory*. Nueva York: Croom Helm.
- Itkonen, E. (2020). Concerning the scope of normativity. En A. Mäkilähde; V. Leppänen y E. Itkonen. *Normativity in Language and Linguistics* (pp. 29-67). Amsterdam: John Benjamins.
- Jiménez Ruiz, J. L. (2007). *Metodología de la investigación lingüística*. Alicante: Publicaciones de la Universidad de Alicante.
- Kemmer, S. y Barlow, M. (1999). Introduction: A Usage-Based Conception of Language. En M. Barlow y S. Kemmer (eds.). *Usage Based Models of Language* (pp. vii-xxviii). Stanford: Center for the Study of Language and Information.
- Langacker, R. (1987). *Foundations of Cognitive Grammar*. Stanford: Stanford University Press.
- Martí Sánchez, M. (1998). *En torno a la cientificidad de la Lingüística. Aspectos diacrónicos y sincrónicos*. Alcalá de Henares: Publicaciones de la Universidad de Alcalá de Henares.
- Martín Miguel, F. (1998). *La gramática de Halliday desde la filosofía de la ciencia*. Almería: Publicaciones de la Universidad de Almería.
- Moure, T. (1996). *La alternativa no discreta en lingüística: una perspectiva histórica y metodológica*. Santiago de Compostela: Publicaciones de la Universidad de Santiago de Compostela.
- Perry, T. (1979). *Evidence and Argumentation in Linguistics*. Berlín: W. de Gruyter.
- Saussure, F. de (1916). *Cours de Linguistique générale*. París: Payot. Traducción al español de A. Alonso, notas críticas de T. de Mauro, *Curso de Lingüística General*. Madrid: Alianza, 1983.
- Schönefeld, D. (2011). *Converging evidence. Methodological and Theoretical Issues for Linguistic Research*. Amsterdam/Filadelfia: John Benjamins.
- Wunderlich, D. (1974). *Grundlagen der Linguistik*. Hamburgo: Rowohlt. Traducción al inglés de R. Lass. *Foundations of Linguistics*, Cambridge: Cambridge University Press.
- Zamorano Aguilar, A. (2008). En torno a la historia y la historiografía de la lingüística. Algunos aspectos teóricos y metateóricos. En F. Carriscondo y C. Sinner (eds.). *Lingüística española contemporánea. Enfoques y soluciones* (pp. 244-277). Múnich: Peniöpe.
- Zamorano Aguilar, A. (2022). *La gramatización del español en el Perú del siglo XIX. Contribución a la historia de las ideas lingüísticas en América Latina*. Berlín: Peter Lang, cap. 1, 17-55.

MILAGROS FERNÁNDEZ PÉREZ

Universidad de Santiago de Compostela

ROMERO DÍAZ, J. *Categorías léxicas en español y japonés. Un estudio comparativo*. Madrid: Ediciones Universidad Autónoma de Madrid (Colección Japón, 8), 2023, 356 pp., ISBN 978-84-8344-917-2.

El presente libro es un estudio comparativo entre las principales categorías léxicas del español y del japonés. Consta de siete capítulos y cinco anexos. El capítulo primero sirve de introducción teórica al libro. En él se expone el marco analítico que se utiliza a lo largo de los restantes capítulos de la obra. El capítulo segundo trata sobre el nombre en japonés y en español. El capítulo tercero está completamente dedicado al estudio del nombre adjetival en japonés y su correspondencia en español. El capítulo cuarto se ocupa del adjetivo en español y en japonés. El capítulo quinto compara el verbo en las dos lenguas. El capítulo sexto está dedicado al nombre verbal japonés y sus correspondencias en la lengua española. Por último, el capítulo séptimo enuncia las conclusiones y las implicaciones pedagógicas de las descripciones anteriores. Después de las pertinentes referencias bibliográficas, se incluyen cinco anexos. El primero de ellos da una lista de las abreviaturas utilizadas a lo largo del libro. El segundo es una útil tabla de correspondencias de los nombres adjetivales japoneses en español. En el anexo tercero se incluye la encuesta realizada para el análisis de los nombres adjetivales. El cuarto anexo incluye una tabla de correspondencias entre los adjetivos japoneses y españoles. El quinto anexo presenta la encuesta en japonés para el análisis de los nombres verbales.

En la introducción teórica del capítulo primero, se expone, en primer lugar, la definición de las categorías léxicas utilizando rasgos binarios distintivos, tal como se propuso en las primeras versiones de la gramática generativa. Se detallan algunas de las propuestas en este marco para analizar las categorías léxicas del japonés. A continuación, se proporciona el marco teórico de los análisis que se ofrecen en el libro, que es el *Lexicón Generativo* de Pustejovsky. En este modelo, las entradas léxicas aparecen enriquecidas con una serie de rasgos de diversa naturaleza, que sirven para explicar tanto las posibles combinaciones léxicas permisibles como la creación de nuevos significados mediante la combinación sintagmática de esas unidades (la coacción de tipo, la co-composición). En una serie de subsecciones, se analiza la estructura argumental, la estructura eventiva y la estructura de *qualia* de las entradas léxicas enriquecidas de este modelo de análisis.

El estudio de la categoría léxica de nombre en español y japonés se aborda en el capítulo segundo. Se analizan los errores en que incurren las personas japonesas que aprenden español y que están determinados por la diferente estructura léxica de los sustantivos en las dos lenguas.

El capítulo tercero está dedicado al estudio de una categoría léxica característica de la lengua japonesa: el denominado *nombre adjetival*, conocido en la gramática japonesa mediante la expresión *keiyō dōshi* ‘verbo adjetival’. Esta categoría plantea cuestiones de análisis lingüístico bastante complejas en relación con la aparición de la cópula, la asignación de funciones sintácticas, su comportamiento sintáctico y su interpretación semántica. En todos estos casos, el modelo de Lexicón Generativo ofrece, tal como muestra el texto, una buena forma de plantear y analizar estas cuestiones. El capítulo termina con un estudio empírico de la influencia de esta categoría en los errores de estudiantes japoneses de español, sobre la base de un estudio empírico mediante una encuesta realizada a cien personas.

El capítulo cuarto es un detallado análisis de la categoría léxica de adjetivo en las lenguas española y japonesa, con un breve resumen final de las correspondencias entre ambas. El capítulo quinto lleva a cabo una comparación entre el verbo español y el japonés. Se analizan todos los aspectos relevantes: el tiempo verbal, el aspecto verbal y gramatical y el modo y la modalidad. Hay una amplia sección dedicada a la clasificación de los verbos japoneses, de gran utilidad para las personas estudiantes y estudiosas del japonés y de la tipología lingüística. El capítulo concluye con un análisis sintáctico-semántico de algunos verbos en español y en japonés.

El capítulo más interesante del libro es, sin duda, el sexto, dedicado al nombre verbal, una categoría fundamental de la sintaxis japonesa. La segunda sección de ese capítulo se centra en las construcciones de nombre verbal junto con el verbo ligero *する suru* ‘hacer’, como, por ejemplo, *勉強 する benkyō suru* ‘estudiar’ (literalmente, ‘hacer estudio’). Esta interesante construcción ha sido objeto de atención en la lingüística japonesa. El autor presenta los análisis de los lingüistas nipones Kageyama y Miyagawa respecto del argumento que es objeto directo del nombre verbal. El primero propone una regla de incorporación sintáctica del objeto al nombre verbal en casos como en *数学勉強 する sūgaku benkyō suru* ‘estudia matemáticas’ y el autor esgrime algunas objeciones a esta propuesta. El segundo combina una regla de incorporación sintáctica con una regla de composición léxica del objeto directo respecto del nombre verbal para casos como *成功 する seikō suru* ‘tiene éxito’, que es sometida también a una crítica lingüística. El autor realiza una propuesta en los términos del modelo del Lexicón Generativo y es en él donde se manifiesta de forma palmaria la utilidad de este modelo teórico, ya que es capaz de integrar y combinar los hallazgos de los dos autores referidos anteriormente salvando en gran medida las objeciones que se habían formulado anteriormente. La última parte de este interesante capítulo está dedicada al estudio de la correspondencia de

los nombres verbales japoneses en español. El estudio es de carácter empírico y se basa en los errores más comunes de los estudiantes japoneses de español al intentar encontrar el equivalente de estas construcciones japonesas con nombre verbal. El autor hizo una encuesta a más de cien estudiantes del Departamento de español de la Universidad de estudios extranjeros de Kobe durante el curso académico 2019-2020. El formulario en japonés de la encuesta aparece en el anexo quinto del libro (p. 356). Los resultados de la encuesta se clasifican por temas: hay una sección dedicada a deportes y juegos, otra a redes sociales, otra a los extranjerismos, además de una, muy interesante, dedicada a las onomatopeyas y a la mimesis y de una última dedicada a los compuestos adverbiales y oracionales.

El último capítulo está dedicado a enunciar las conclusiones e implicaciones pedagógicas del estudio. El libro es novedoso en tanto que aplica el modelo del Lexicón Generativo al estudio de la lingüística contrastiva entre el japonés y el español. En este sentido, resultará, sin duda, muy útil tanto para el aprendizaje del español partiendo de la lengua japonesa como para el aprendizaje del japonés desde el español; también, casi no hace falta decirlo, para la enseñanza del español desde el japonés y del japonés desde el español.

Todos los ejemplos del japonés, además de ofrecerse en su escritura original, están claramente glosados y traducidos, de manera que pueden ser perfectamente seguidos y entendidos de modo eficiente por aquellas personas que no tengan ningún conocimiento previo de esta lengua. Este extremo es particularmente importante porque significa que este libro puede ser aprovechado de modo muy eficiente por todas aquellas personas que, desde la lingüística, deseen obtener datos adecuados sobre una lengua tipológicamente muy diferente de las lenguas indoeuropeas. Todos los ejemplos japoneses son de primera mano y, por tanto, completamente fiables. Este no es un dato menor porque, en muchas ocasiones, cuando se citan ejemplos del japonés en estudios de lingüística tipológica o general, se toman de fuentes de segunda, tercera o incluso cuarta mano, con traducciones y glosas que a veces son inexactas o inadecuadas.

JUAN CARLOS MORENO CABRERA

Catedrático jubilado de Lingüística General de la Universidad Autónoma de Madrid

Edita
SEL

<http://revista.sel.edu.es>

