



Revista Española de Lingüística

Órgano de la Sociedad Española de Lingüística

RSEL

53/2

julio-diciembre 2023

REVISTA ESPAÑOLA DE LINGÜÍSTICA

(RSEL)

53/2

Edita

SEL

<https://doi.org/10.31810/RSEL.53.2>

REVISTA ESPAÑOLA DE LINGÜÍSTICA (RSEL)

ISSN: 0210-1874 • eISSN: 2254-8769

Depósito Legal: M-24.769-1971

DIRECTORA: M. 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 Chincoa: produccionRSEL@gmail.com



FECYT-S2023/2023
Fecha de certificación: 20 de julio de 2023 (7ª convocatoria)
Válida hasta: 28 de julio de 2024

ÍNDICE 53/2 (2023)

ARTÍCULOS

Significados procedimentales, enriquecimiento contextual y variación geolectal: en torno al perfecto compuesto (y el perfecto simple) 9
Procedural meaning, contextual enrichment and geolectal variation: about the compound (and the simple) past
JOSÉ AMENÓS PONS

La interrelación entre evidencialidad y ergatividad (escindida): una comparación tipológica entre el español, las lenguas romances y otras familias lingüísticas 45
The interplay between evidentiality and (split) ergativity: a typological comparison of Spanish, the Romance languages and other linguistic branches
VÍCTOR LARA BERMEJO

Atomicity and countability in linguistically-informed instruction: Acquisition of L2 English noun types 89
Atomicidad y contabilidad en la enseñanza lingüísticamente informada: adquisición de los tipos de nombres en inglés como L2
DAKOTA J. THOMAS-WILHELM Y ELISABET PLADEVALL-BALLESTER

Consideraciones sobre la evaluación morfológica de estudiantes de inglés como L2119
Considerations towards morphological assessment of English language learners
BRENDA VARGAS VEGA Y PIERRE-LUC PAQUET

Enunciación y significado aproximativo en la construcción «estar por + infinitivo»141
Uttering and approximative meaning in the Spanish construction «estar por + infinitive»
ADRIÀ PARDO LLIBRER

SECCIÓN MONOGRÁFICA

FENÓMENOS DE LA DIGLOSIA ÁRABE: DE LA EDAD MEDIA A LA EDAD CONTEMPORÁNEA

Introducción 169
Introduction
MONTSERRAT BENÍTEZ FERNÁNDEZ Y MARÍA ÁNGELES GALLEGO

A/al- inicial como identificador de arabismos en la Recopilación de Diego de Guadix 179
Initial a/al- as an identifier of arabisms in the Recopilación by Diego de Guadix
MARÍA ÁNGELES GALLEGO Y PATRICIA GIMÉNEZ EGUÍBAR

<i>Variación lingüística en fuentes árabes de Al-Ándalus: el tratado de 'Abd al-Ra'ūf (Córdoba, s. x)</i>	201
<i>Linguistic Variation in Arabic Sources from al-Andalus: the Manual of 'Abd al-Ra'ūf (Cordova, 10th c.)</i>	
ESTEFANÍA VALENZUELA MOCHÓN	
<i>What is White Arabic? New labels in a changing Arab world</i>	229
<i>¿Qué es el árabe blanco? Etiquetas nuevas en un mundo árabe cambiante</i>	
ANA IRIARTE DÍEZ, CLAUDIA LAABE, NINA VAN KAMPEN Y MONTSERRAT BENÍTEZ FERNÁNDEZ	
<i>Variación lingüística o la emergencia de nuevas variedades altas de la lengua árabe: Marruecos como estudio de caso</i>	267
<i>Linguistic variation or the emergence of new high varieties of the Arabic language: Morocco as a case study</i>	
MONTSERRAT BENÍTEZ FERNÁNDEZ	
<i>From mixed arabic to «educated written dāriġa»: diglossic variation in moroccan written production</i>	293
<i>Desde el árabe mixto hacia la «dāriġa escrita culta»: La variación diglósica en la producción escrita marroquí</i>	
ROSA PENNISI	
NOTAS	
<i>La psicología cognitiva a examen: sobre el libro Conversaciones sobre la mente cognitiva del profesor José Eugenio García-Albea</i>	317
JOSÉ MANUEL IGOA	
RESEÑAS	329

ARTÍCULOS

SIGNIFICADOS PROCEDIMENTALES, ENRIQUECIMIENTO CONTEXTUAL Y VARIACIÓN GEOLECTAL: EN TORNO AL PERFECTO COMPUESTO (Y EL PERFECTO SIMPLE)


José AMENÓS PONS¹

Universidad Complutense de Madrid²

Resumen

En este trabajo argumentamos la utilidad de integrar factores semánticos y pragmáticos en la descripción del perfecto compuesto y simple de las lenguas románicas, a fin de dar cuenta de su diversidad funcional intra e interlingüística. Partiendo de la tesis de la infradeterminación semántica (Sperber y Wilson, 1995) y de la idea de que el significado de los tiempos verbales tiene naturaleza procedimental (Moeschler, 1994, 1998; Nicolle, 1997, 1998; Saussure, 2003), desarrollamos una descripción de los perfectos basada en un conjunto limitado de combinaciones de rasgos semánticos. Algunos de estos rasgos son invariables y comunes a todas las lenguas románicas, mientras que otros presentan opciones paramétricas binarias, cuyo valor varía de una lengua a otra e incluso, dentro de una misma lengua, de una variedad geolectal a otra. Nuestra propuesta es compatible con la idea de una semántica procedimental rígida pero infraespecificada, susceptible de enriquecimiento contextual. El conjunto de rasgos y opciones paramétricas que proponemos explica los tipos de interpretaciones posibles en cada variedad. Además, los rasgos semánticos reciben sistemáticamente enriquecimiento contextual, y este desarrolla en direcciones diferentes las potencialidades de la forma verbal, semánticamente infraespecificadas. La competición entre expresiones lingüísticas semánticamente próximas es asimismo un factor importante en la distribución de las rutas interpretativas accesibles para cada una.

Palabras clave: Semántica procedimental; enriquecimiento pragmático; perfecto compuesto; perfecto simple

1. jamenos@ucm.es.  <https://orcid.org/0000-0003-2980-7003>

2. La investigación que ha conducido a este artículo ha sido parcialmente financiada por el proyecto *Evidentiality, Perspectivisation and Subjectivisation at the Interfaces of Language (EPSILone)*, PGC2018-094223-B-C21, Ministerio de Economía y Competitividad (España). El autor agradece especialmente las sugerencias de Victoria Escandell-Vidal a las versiones previas del texto, así como los comentarios de dos revisores anónimos.

PROCEDURAL MEANING, CONTEXTUAL ENRICHMENT AND GEOLECTAL VARIATION: ABOUT THE COMPOUND (AND THE SIMPLE) PAST

Abstract

In this paper, we argue for the suitability of integrating semantic and pragmatic factors in the description of the compound and the simple past of Romance languages, to account for their intra- and interlinguistic functional diversity. Building on the semantic underdeterminacy thesis (Sperber and Wilson, 1995) and on the idea that tenses have procedural meaning (Moeschler, 1994, 1998; Nicolle, 1997, 1998; Saussure, 2003), we develop a description of the compound and the simple past in Romance, based on a limited set of combinations of semantic features. Some of these features are invariant and common to all Romance languages, while others present binary parametric options, whose value varies from one language to another and even, within the same language, from one geolectal variety to another. Our proposal is compatible with the idea of a rigid but underspecified procedural semantics, susceptible to contextual enrichment. The set of features and parametric options we propose account for the interpretations available in each variety. Additionally, semantic features systematically receive contextual enrichment, and this enrichment develops the underspecified semantics of the verb tense in different directions. Competition between semantically close linguistic expressions is also an important factor in the distribution of interpretative routes accessible to each tense.

Keywords: Procedural semantics; pragmatic enrichment; compound past; simple past

RECIBIDO: 18/01/2023

APROBADO: 02/05/2023

1. INTRODUCCIÓN

Es un hecho generalmente aceptado que en la determinación de la referencia temporal de los enunciados interactúan factores semánticos, sintácticos y pragmáticos. La información de los tiempos verbales se complementa con la que aportan los adverbios y otras expresiones temporales, además de la situación de comunicación y el conocimiento del mundo. En las últimas décadas, asimismo, se ha establecido con fuerza la concepción de las relaciones temporales del discurso no como un hecho propiamente semántico, sino como el producto de la combinación del contenido tempo-aspectual de las formas verbales, las relaciones conceptuales entre los eventos y el tipo de discurso (Smith, 2003; Unger, 2006; Bres, 2003, 2007, entre otros).

Sin embargo, en las descripciones lingüísticas escasean los análisis detallados sobre cómo tiene lugar la interacción entre semántica, sintaxis y pragmática. Esta preocupación está en el centro del presente trabajo. Nuestro objeto de estudio principal es el pretérito perfecto compuesto (en adelante, PC) románico en general y español en particular. La interacción de los rasgos semánticos, el entorno sintáctico-discursivo y las inferencias pragmáticas produce una variedad de interpretaciones como la que puede surgir de los siguientes enunciados:

- (1) El problema *ya ha sido resuelto* por los poderes del estado.
- (2) Paco *ha estado* muchas veces en París.
- (3) María *ha trabajado* 30 años en este hospital.
- (4) Luisa *ha llegado* hace un rato.
- (5) Este verano *han cambiado* el ascensor de la comunidad.

Fuera de contexto, el PC de casi todos estos enunciados tiene más de una interpretación posible: en varios de ellos (1, 4 y 5) podrían darse lecturas de antepresente (en las que lo relevante es la localización temporal del evento en sí)³, pero todos admitirían asimismo interpretaciones en las que se destaca un estado de presente (en adelante, EP⁴) de cierto tipo. Así, el predicado télico de (1), si se entiende como evento único y no repetido, recibe una lectura resultativa, pero si desaparecen esas dos condiciones podría entenderse como experiencial ('ha sido resuelto en otras ocasiones'); en (2), más claramente, la interpretación experiencial del predicado atélico se impone, gracias a la combinación de este con la expresión 'muchas veces'; (3) puede entenderse como un uso continuativo del PC, si el enunciado se relaciona con una situación en la que María sigue trabajando en el mismo hospital; sin embargo, la interpretación es experiencial si no se da la continuidad de la situación; el PC de (4) y el de (5) pueden entenderse como resultativos, dado que se trata de eventos télicos no repetidos.

3. Aunque el término «antepresente» procede directamente de Bello (1847), el sentido en el que se usa en nuestro trabajo es más específico y restringido que el de la propuesta original. Dicho término refiere, en nuestro caso, a las interpretaciones del PC en las que se destaca una relación de naturaleza temporal (típicamente, interpretaciones hodiernas y de pasado reciente) entre la situación y el tiempo del habla. Así pues, los casos en los que no se infiere relación temporal, sino aspectual, quedan fuera de esta denominación.

4. Con el término «estado de presente» (EP) nos referimos a una situación o estado de cosas que se considera válido en el tiempo del habla (o en el que se tome como referencia en la interpretación de la forma verbal). Un EP puede ser bien el resultado directo de un evento anterior, bien la experiencia causada por este, o bien la continuidad del mismo. Preferimos al término EP al término «resultado presente», que resulta resbaladizo al pensar en las interpretaciones experienciales y en las continuativas.

No todas esas interpretaciones serían posibles en las distintas áreas del mundo hispanohablante: las de antepresente, en especial, tienen fuertes restricciones geográficas, fuera y dentro del español europeo. Las resultativas, las experienciales y las continuativas están más extendidas, pero no son igual de naturales en todas las áreas geográficas (ver RAE-ASALE, 2009 y Azpiazu 2019, para una visión de conjunto).

Una referencia importante en el estudio de la interacción de semántica, sintaxis y pragmática en las interpretaciones del PC es el trabajo de Depraetere (1998). La autora se centra en el *present perfect* inglés⁵, y demuestra con claridad una serie de hipótesis:

1. La decisión del hablante de usar el PC indica su intención de focalizar una situación pasada desde la perspectiva del presente.
2. La interacción de ciertos rasgos lingüísticos (progresividad; telicidad; irreversibilidad) lleva a implicaciones semánticas de tipo resultativo, pero estas no dependen del tiempo verbal, ya que pueden surgir sin el PC. Lo específico de este tiempo verbal es que los resultados se entienden en relación con el presente.
3. Los predicados télicos introducen implicaciones semánticas de resultado, a diferencia de los atélicos. Aun así, en el caso de los predicados atélicos, en ciertos entornos discursivos puede ser relevante la interpretación de un resultado; este tendría carácter de implicatura conversacional.

Depraetere (1998) puntualiza asimismo que el resultado de un predicado télico en PC puede entenderse como una situación que permanece en el tiempo del habla, o bien como una experiencia; eso depende, ante todo, del tipo de evento y del lapso que lo separa del tiempo del habla: en el caso de eventos reversibles, cuanto mayor sea el lapso, más improbable es que el resultado permanezca en el presente⁶.

Un punto clave de la aproximación de Depraetere (1998) es la idea de que la interpretación de un resultado (de un EP, en nuestros propios términos) no está especificada en la semántica del PC, sino que se determina en el discurso. Interpretar el PC exige siempre establecer una relación entre el evento y el tiempo del habla, pero la naturaleza de esa relación está infraespecificada, esto es, no está codificada en la semántica de la forma verbal. El modo en que se entiende en cada caso la

5. Por comodidad, en adelante usaremos en este trabajo la abreviatura PC también para referirnos a la forma verbal del inglés que es morfológicamente equivalente al PC español; lo mismo haremos después con las formas verbales de otras lenguas distintas del español.

6. Una idea similar, aunque menos detallada, se expone en Sperber y Wilson (1993) y Smith (1990).

relación con el tiempo del habla depende del tipo de predicado, del lapso temporal, del tipo de discurso y del contexto de interpretación.

Las propuestas de Depraetere (1998) contienen análisis bien precisos de un conjunto de casos particulares en una sola lengua: el inglés. Al mismo tiempo, dejan cuestiones sin desarrollar. En especial, Depraetere no considera las interpretaciones de antepresente, dado que el PC del inglés carece de ellas. ¿Cómo extrapolar ese tipo de análisis a otras lenguas, como el español y otras lenguas románicas, en las que dichas lecturas sí son posibles? Por otra parte, ¿cómo dar cuenta de la variación en las interpretaciones disponibles en distintas variedades geolectales de una misma lengua, así como en lenguas distintas, genética y tipológicamente emparentadas? ¿Cómo formalizar adecuadamente el detalle de esa variación?

En un enfoque que aspira a integrar gramática y pragmática, no basta con establecer una lista de usos y valores del PC, asignándoles un área geográfica, sino que resulta necesario determinar la manera en que, en distintas variedades, interactúan un conjunto de rasgos semánticos y de posibilidades pragmáticas comunes en sus principios, aunque dispares en gran parte en cuanto a sus resultados. Esbozar una descripción de este tipo centrada en el español, pero atendiendo también a otras lenguas románicas como el portugués y el francés⁷, es el objetivo del presente artículo. Para ello recurriremos a la noción de «significados procedimentales» (Blakemore, 1987; Wilson y Sperber, 1993; Moeschler, 1994, 1998; Nicolle, 1997, 1998; Saussure, 2003; Escandell-Vidal y Leonetti, 2011, entre otros).

Los contenidos de nuestro artículo, más allá de esta introducción, se estructuran del siguiente modo. En el apartado 2 profundizaremos en la idea de una semántica procedimental de los tiempos verbales y justificaremos su interés. Posteriormente, en el apartado 3, contrastaremos dos desarrollos distintos (y en parte, alternativos) de la noción de «significados procedimentales». Tras sopesar las aportaciones de cada tipo de desarrollo, en el apartado 4 nos centraremos en uno de ellos al formular nuestra propuesta para el PC (y, complementariamente, para el pretérito perfecto simple); propondremos la aplicación de nuestros análisis no solo a lenguas románicas distintas (portugués, español y francés), sino también a distintas variedades del español.

7. La elección de estas dos lenguas no es casual: a pesar de su proximidad geográfica, el uso del PC en cada una de ellas es completamente distinto: en francés, el PC recibe no solo interpretaciones de EP y de antepresente, sino que funciona como tiempo perfectivo de uso general; en portugués, sin embargo, el PC solo admite interpretaciones acumulativas, que entendemos como un subtipo de las experienciales (Squartini y Bertinotto, 2000; Azpiazu, 2019).

Finalmente, en el apartado 5 estableceremos algunas conclusiones, subrayando las aportaciones de nuestro trabajo.

2. LOS SIGNIFICADOS PROCEDIMENTALES

La multiplicidad de matices interpretativos de cualquier forma verbal, y en concreto del PC español y románico, supone un reto para las descripciones que sostienen que el PC tiene una semántica única e invariable en todos sus usos. Así, el PC español puede recibir interpretaciones de aspecto perfectivo (pasado hodierno como en (4) o, más generalmente, antepresente como en (5)) y de aspecto perfecto⁸ (en sus distintas manifestaciones: resultativo, experiencial, continuativo, como se ejemplifican en (1), (2) y (3)). En cuanto al lapso temporal que separa el evento del tiempo del habla, puede tratarse de una (escasa) distancia objetiva o de una cercanía meramente subjetiva, sentida por el hablante (se trata en este último caso de las interpretaciones denominadas «de noticias recientes») (RAE-ASALE, 2009).

Así pues, un análisis monosémico (esto es, un análisis que defiende que la forma verbal tiene un contenido semántico único e invariable) debe dar cuenta de la relación entre unos rasgos semánticos estables y la multiplicidad de interpretaciones que deriva de ellos. Este empeño es compatible con un enfoque como el nuestro, que parte de una visión sistemática de las relaciones entre el código lingüístico y los aspectos pragmáticos que inciden en la interpretación⁹. Nos basamos para ello en dos concepciones básicas: la tesis de la infradeterminación semántica y la distinción entre las expresiones de significado conceptual y las de significado procedimental.

La tesis de la infradeterminación semántica (presente en distintos enfoques teóricos, pero adoptada con decisión por Sperber y Wilson, 1995) postula una diferencia entre el significado lingüísticamente codificado en un enunciado y la proposición que dicho enunciado expresa. Lo codificado constituye simplemente un indicio de lo que se quiere expresar. Así se explica que pueda haber diversidad de interpretaciones de un núcleo semántico común, en función del tipo de enriquecimiento contextual que reciba.

8. Definimos el «aspecto perfectivo» como aquel que focaliza las situaciones en su conjunto, de principio a fin, y las presenta como completas o acabadas. En cambio, el aspecto perfecto focaliza un estado de cosas que resulta de un proceso previo (Comrie, 1976; RAE-ASALE, 2009).

9. Una alternativa consistiría en renunciar de entrada a considerar que a una forma verbal tenga que corresponderle un significado único. Esta opción es la que toma, por ejemplo, Delgado Díaz (2021) desde la lingüística variacionista. Las limitaciones de espacio nos impiden considerar con detalle las ventajas e inconvenientes de esta posibilidad, que no adoptaremos.

En cuanto a la distinción entre expresiones de significado conceptual y expresiones de significado procedimental, se basa en la idea de que no todas las unidades lingüísticas contribuyen de la misma manera a los procesos de interpretación. Algunas clases de expresiones (nombres, adjetivos, verbos, la mayoría de los adverbios) codifican conceptos. Otras, en cambio, codifican instrucciones de procesamiento, es decir, indicaciones sobre cómo combinar las representaciones conceptuales para interpretarlas. Las expresiones lingüísticas pertenecientes a las categorías cerradas o funcionales (entre las que encuentran los tiempos verbales) pertenecen a estas últimas. Así pues, los tiempos verbales no contienen representaciones que puedan glosarse directamente como conceptos, sino restricciones sobre el procesamiento inferencial de las representaciones conceptuales de estados y eventos (Moeschler, 1994, 1998; Nicolle, 1997, 1998; Saussure, 2003).

En lo que al PC se refiere, el hecho de que cualquier situación expresada en PC deba ponerse en relación con el tiempo del habla (o, en su defecto, con otra coordenada temporal que pueda servir de referencia) no es propiamente una representación conceptual situada conscientemente en la mente de los hablantes, sino un condicionante básico y fundamental del procedimiento de interpretación de la forma, que en buena parte es automático, no consciente ni deliberado.

Más allá de la instrucción procedimental esquemática dictada por la semántica del PC, el proceso de interpretación continúa hasta que el oyente alcanza una interpretación que armoniza de forma intuitivamente adecuada el esfuerzo invertido en el trabajo de procesamiento y los resultados cognitivos de dicho esfuerzo, según unas expectativas de relevancia (Sperber y Wilson, 1995).

La breve caracterización que acabamos de presentar puede sin duda despertar algunos interrogantes. Para nosotros son, fundamentalmente, tres:

1. Dado que la morfología verbal entra sistemáticamente en contacto con el lexema verbal, y este pertenece a las clases léxicas, ¿cuánto hay finalmente de conceptual y cuánto de procedimental en los tiempos verbales, y cómo interactúan ambos componentes?
2. ¿Cuánto hay de base semántica y cuánto de enriquecimiento pragmático en cada tipo de interpretación de una forma verbal?
3. ¿Cómo entender la variación inter e intralingüística de una expresión procedimental?

Para los objetivos del presente trabajo son cruciales las dos últimas preguntas; a ellas dedicaremos los apartados 3 y 4. Para llegar a ellas es necesario, sin embargo,

responder a la pregunta 1. Esta respuesta implica un posicionamiento teórico, que pasamos a esbozar aquí.

Como categorías funcionales, el tiempo y el aspecto gramatical son categorías procedimentales que restringen la interpretación del contenido explícito. Estas restricciones pueden incidir en varios niveles: en primer lugar, en la construcción de la representación mental de la situación y en la determinación de la referencia temporal del enunciado (aspectos que, en términos de Sperber y Wilson, 1995, forman parte de las explicaturas básicas del enunciado). En segundo lugar, pueden incidir en las inferencias sobre la actitud del hablante (explicaturas de nivel superior, según Sperber y Wilson). Los enunciados (1) a (5), presentados en el apartado precedente, son ejemplos del primer nivel, mientras que (6) y (7) ejemplifican el segundo nivel:

- (6) Tranquilo, que mañana a estas horas *ya has terminado* (vs. *ya habrás terminado*).
- (7) - ¡Anda, qué escayola llevas!
+ Sí, es que *me he roto* la pierna esquiando (vs. *me rompí* la pierna).

En (6), la elección del PC, en lugar del futuro compuesto, sugiere una inmediatez que es ante todo subjetiva: al sustituir una forma verbal de anterioridad con respecto al futuro por una de anterioridad con respecto al presente, obtenemos un efecto de proyección psicológica del hablante, relacionable pragmáticamente con el deseo o la impaciencia. En cuanto a (7), el uso del PC, en lugar del perfecto simple (en adelante, PS), no implica necesariamente una mayor cercanía temporal del evento, sino una relevancia actual sentida por el hablante, en cuanto a dicho evento y a sus resultados.

Por otra parte, las categorías de tiempo y aspecto gramatical interactúan con el aspecto léxico, cuyo contenido es conceptual. En este trabajo asumimos que los significados procedimentales son rígidos e invariables, aunque semánticamente infradeterminados. En cambio, los significados conceptuales son flexibles; pueden modificarse en función de las exigencias contextuales, para asegurar la interpretación. Una argumentación detallada de dicha postura puede verse en Escandell-Vidal y Leonetti (2000, 2011).

Conviene recordar aquí que las categorías sintácticas funcionales poseen significado procedimental, y es precisamente esta naturaleza (de la que se deriva, en especial, su carácter abstracto, difícilmente verbalizable y no abierto a la introspección) la que está en el origen de su rigidez, necesaria para poder llevar a cabo su función. La estructura funcional debe ser capaz de encauzar la combinación de

los contenidos conceptuales de las categorías léxicas con la información contextual adecuada.

El aspecto léxico, por tanto, está sometido a ajustes y enriquecimientos en el proceso interpretativo por la acción del aspecto gramatical, pero no a la inversa. La amplísima bibliografía que (desde De Swart, 2000) ha estudiado los fenómenos de coacción aspectual da cuenta del modo en que tiempo, aspecto gramatical y aspecto léxico interactúan en la obtención de distintas interpretaciones de una forma verbal¹⁰. Los análisis propuestos por De Swart sugieren que el mayor alcance sintáctico de una categoría incide en su poder sobre las demás. En el caso de las formas verbales, la jerarquía puede representarse con el siguiente esquema (De Swart, 1998, p. 342):

[Tense [Aspect [Eventuality Description]]]

3. LA SEMÁNTICA PROCEDIMENTAL DE LOS TIEMPOS VERBALES EN CONTRASTE INTERLINGÜÍSTICO

En trabajos de distintos autores, la noción de «significados procedimentales» se ha desarrollado de maneras diferentes, no siempre compatibles entre sí¹¹. En este apartado nos centraremos en dos formas de entender la estructura de una descripción semántica procedimental, que se han utilizado en descripciones contrastivas en las que el español se compara con otras lenguas románicas. Nuestra intención, al presentar los dos enfoques, es reflexionar sobre la eficacia de cada uno al formalizar la semántica de las formas verbales. Una de las trampas potenciales en las que puede caer una descripción gramatical formulada desde la pragmática es, precisamente, su indefinición al tratar de armonizar aspectos codificados y no codificados. Así pues, ¿en qué medida evitan estos enfoques dicho problema?

Los dos tipos de enfoque a los que nos referimos son los siguientes:

1. TIPO 1. El *modus operandi* de este tipo de descripción consiste en introducir en la descripción semántica una variable infraespecificada, que debe saturarse pragmáticamente. La variación interlingüística se explica en función de formas y tendencias específicas de saturación (esta vía es la que se

10. En el apartado anterior, basándonos en Depraetere (1988), hemos evocado brevemente el papel que el tipo de predicado tiene en la interpretación de los distintos tipos de EP. Volveremos a mencionar esta cuestión en el apartado 3.

11. Una panorámica de las distintas direcciones que ha tomado esta noción puede verse en Escandell-Vidal, Leonetti y Ahern (2011). El caso específico de los tiempos verbales se aborda en Amenós Pons (2020).

sigue en Depraetere, 1998 y en Nishiyama y Koenig, 2010, para el PC inglés; se encuentra asimismo en Amenós Pons, 2010a, 2011 para el PC español y el PC francés).

- TIPO 2. En este tipo de descripción, la estructura de una expresión procedimental contiene diferentes fases o capas. Al contrastar lenguas emparentadas se postula una base semántica común a todas y un conjunto de rasgos semánticos con realizaciones paramétricas alternativas [\pm rasgo] propias de cada lengua (Escandell-Vidal, 2010, 2014, 2019a y 2021, para el futuro español y románico; Escandell-Vidal, 2022, para el imperfecto español).

El contraste entre el PC del español (peninsular estándar) y el francés, en una descripción procedimental de tipo 1, se abordó en Amenós Pons (2010a, 2011). Conviene recordar aquí que la diferencia esencial en el uso del PC en ambas lenguas se encuentra en el hecho de que, en francés no literario, el PC cubre no solo las posibilidades de un perfecto y de un antepresente, sino todas las de un tiempo perfectivo general (como es, en español, el PS). El contraste entre PS y PC en francés solo se mantiene claramente activo en los textos narrativos literarios (Corblin y De Swart, 2004; Abeillé y Godard, 2021).

Amenós Pons (2010a, 2011) argumenta que el PC español y el francés poseen una semántica común, que puede glosarse como una instrucción de procesamiento cuyo contenido es el siguiente: el oyente tiene que representarse mentalmente una situación de cualquier tipo como acotada, situarla en un punto anterior al tiempo de habla (H) y a la vez considerar su relación con H. Esta relación no se establece directamente, sino a través de una coordenada abstracta R (momento de referencia), cuya relación con H es de coincidencia $\langle R, H \rangle$.

En este enfoque, R es una coordenada abstracta que debe saturarse pragmáticamente. Esto puede hacerse bien a través de la inferencia de un EP, de una conexión temporal (proximidad) entre la situación y el tiempo del habla H o de una conexión discursiva (en este último caso no se asume una proximidad temporal entre la situación y H, sino entre el hablante y su discurso de manera general). La saturación de R en una de esas tres direcciones depende del tipo de predicado, así como de las hipótesis contextuales accesibles (entorno lingüístico, conocimiento del mundo, contexto previo).

Las diferencias entre el español y el francés se explican como rutinas pragmáticas predominantes en cada lengua: R tiene a entenderse como una conexión temporal

en español, y discursiva en francés. Las diferencias entre las variedades del español podrían (hipotéticamente) explicarse de la misma manera.

La semántica común del PC en español y en francés abre la puerta, en teoría, a las mismas posibilidades interpretativas en las dos lenguas. Dichas posibilidades se representan en la Figura 1.

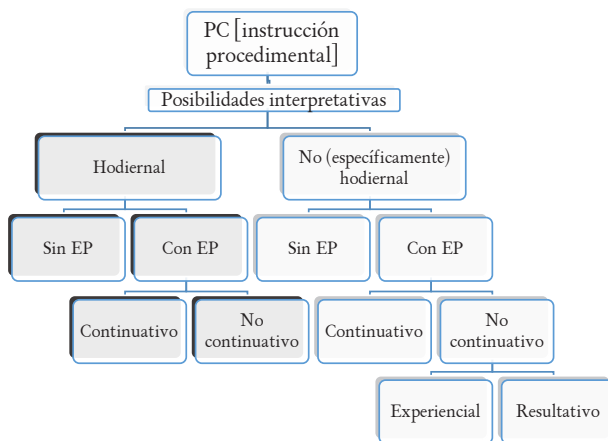


Figura 1. Posibilidades interpretativas del PC español peninsular estándar y en francés (Amenós Pons, 2010a, 2011)

Todos los usos del PC en español y en francés, tal como se reflejan en la figura 1, proceden de la instrucción procedimental común introducida en los párrafos precedentes. La aplicación automática de esa instrucción abre dos vías interpretativas posibles: el evento puede entenderse como un pasado hodiernal o no ser así. La etiqueta «hodiernal» admite dos grandes tipos de interpretación: de proximidad temporal (es decir, sin EP contextualmente relevante) o aspectual (con la representación contextualmente relevante de un EP, que puede entenderse bien como la continuación de la situación en el tiempo del habla, o bien como un resultado que sigue a su término).

En cuanto a la etiqueta «no (específicamente) hodiernal», incluye no solo aquellos usos en los que el PC actúa como un pasado no hodiernal (frecuentes en francés, pero mucho menos en español peninsular), sino los casos en los que el contexto no permite (o no requiere) establecer con precisión la localización temporal de la situación. Nuestro enunciado (7), más arriba, sería una muestra de tal indefinición.

Tanto en los usos hodiernales como en los no hodiernales del PC, la configuración aspectual (en la línea de lo descrito por Depraetere, 1998), el entorno discursivo y el conocimiento del mundo harán que tenga o no prominencia un EP; de aquí

que se diferencie entre interpretaciones sin EP o con EP. En los casos en los que se percibe un EP contextualmente relevante, este puede surgir ya sea como implicación semántica (particularmente, con los predicados télicos) o como fruto de una inferencia pragmática (en el caso de los predicados atélicos). En cualquiera de los dos casos, sin embargo, la interpretación de un EP no surge automáticamente del PC, sino su interacción con el contexto (lingüístico y/o pragmático) en que se procesa la forma.

Comparando los usos hodiernales con los no hodiernales, se observa en la figura 1 que la diferencia entre EP experiencial y resultativo solo se establece para los no hodiernales: eso se debe a que se requiere cierto intervalo de separación entre la situación de la que se habla y la coordenada H para que las interpretaciones experienciales sean posibles; asimismo se requiere una estimación del intervalo para determinar si un estado de cosas obtenido como resultado de una situación previa se mantiene aún en H. Nada de todo eso es posible en las interpretaciones hodiernales del PC¹².

En el análisis de Amenós Pons (2010a, 2011), que una instrucción procedimental común pueda satisfacerse de maneras contrastadas en lenguas diferentes se justifica por el hecho de que las instrucciones procedimentales se consideran rígidas, pero infraespecificadas y, por tanto, no aportan información sobre cómo deben satisfacerse. Los fenómenos de especialización pragmática pueden entenderse, por tanto, como el fruto de las necesidades de los sistemas lingüísticos específicos de los que forma parte una expresión. Así, en francés se ha producido una especialización del PS en discursos formales de carácter narrativo; el modo en que contrastan el PS y PC, y las tendencias interpretativas de este último, son claramente distintas dentro y fuera de los textos literarios narrativos. Algunos factores que permiten calificar esta evolución como una especialización pragmática son los siguientes: (1) los significados procedimentales no forman parte del conocimiento consciente de los hablantes, mientras que la propiedad o impropiedad del uso del PC y el PS en distintos tipos de texto sí es un conocimiento consciente de los hablantes de francés; (2) la asociación de cada tiempo a un tipo de texto, y su uso dentro de este, es un hecho subjetivo, ligado a una decisión (parcialmente) consciente; (3) la diferencia que existe entre el español y el francés en el uso del PS y el PC no se reproduce en ningún otro elemento del

12. Un revisor anónimo ha señalado la conveniencia de explicar por qué la unidad de tiempo «día» tiene tanta importancia en la estructura de la Figura 1. En la formulación original del autor, este hecho tiene que ver precisamente con que la condición hodiernal del PC impone restricciones interpretativas parcialmente distintas de las del PC no hodiernal: tal como acabamos de explicar, la distinción entre tipos de interpretaciones aspectuales no continuativas (es decir, experienciales y resultativas) resulta pertinente en los usos no hodiernales del PC, pero no en los usos hodiernales.

sistema verbal, por lo que no puede afirmarse que sea el fruto de una evolución de los sistemas verbales en su conjunto.

Con todo, a pesar de los argumentos aducidos, algunos contrastes entre el PC del español y el del francés sí parecen sistemáticos, es decir, gramaticales y no solo discursivos. Por ejemplo, contrariamente a lo que ocurre en español, en francés oral y escrito informal el uso del PC es indiferente al contraste hodiernal vs. no hodiernal. Eso no es algo que en francés dependa del enriquecimiento contextual ni de la subjetividad de los hablantes; se trata, por tanto, de un hecho que debería reflejarse en algún rasgo del tiempo verbal del francés, pero no en el del español.

Por otra parte, si bien en español peninsular las interpretaciones del PC pueden ser temporales, aspectuales o de proximidad psicológica, ninguna puede tener una fundamentación puramente discursiva o de registro (a diferencia de lo que ocurre en francés). Este contraste debería también asociarse a algún rasgo contrastado de la forma verbal de ambas lenguas. Al no hacerse así, la propuesta semántica de Amenós Pons (2010a, 2011) pierde capacidad explicativa y predictiva, ya que la clave de la diferencia entre el español y el francés queda fuera de la semántica verbal. Esta es, quizás, una insuficiencia inherente a las propuestas que más arriba hemos denominado «de tipo 1», al incorporar una noción poco definida (la de «rutinas pragmáticas») como único complemento de la caracterización de un núcleo semántico general e infraespecificado.

Por todo ello, en el próximo apartado nos orientaremos a la construcción de una propuesta descriptiva «de tipo 2» para el PC español que pueda extenderse a otras lenguas románicas y que pueda dar cuenta, en cada caso, no solo de la diversidad de interpretaciones del PC, sino de la variación intra e interlingüística.

4. HACIA UNA NUEVA DESCRIPCIÓN PROCEDIMENTAL DEL PC (Y EL PS) ESPAÑOL Y ROMÁNICO

4.1. *Más allá del tiempo y el aspecto*

Las descripciones presentadas en el apartado 3 incluyen, sin cuestionarlas, unas coordenadas de base temporalista, inspiradas claramente en las de Reichenbach (1948). En la bibliografía se ha debatido abundantemente de qué modo pueden esas coordenadas dar cabida a las nociones aspectuales, y no solo a las temporales (véase Borik, 2006 y las referencias que contiene). Con todo, algo que difícilmente cabe en un sistema de coordenadas temporales es un contraste como el del PC/PS francés contemporáneo, que, en lo fundamental, no implica diferencias de tiempo

ni de aspecto, sino que afecta a la relación del hablante con su propia producción discursiva. Esta observación no es nueva; se encuentra ya en Benveniste (1966), quien considera la elección del PC o la del PS como característica de una actitud enunciativa diferente del hablante ante el enunciado que produce¹³.

Así, postular una base semántica de carácter temporal no parece la mejor opción para una caracterización unificada del PC románico contemporáneo, en el que coexisten formas del PC de interpretación únicamente aspectual (como en portugués), otras que alternan tiempo y aspecto (como en español) y otras en las que lo determinante la perspectiva enunciativa (como en francés).

Una alternativa al manejo de coordenadas temporales es la que plantean trabajos como el de Speas (2008). Al estudiar lenguas que poseen, en su sistema verbal, morfemas gramaticales dedicados a la expresión de la evidencialidad (Aikhenvald, 2004), Speas se inspira en Reichenbach para proponer un sistema de coordenadas, pero las despoja de un carácter intrínsecamente temporal. Para Speas, los tiempos verbales codifican relaciones entre situaciones; esas relaciones permiten inferir información sobre la localización temporal, así como sobre la fuente de información o el compromiso del hablante. Las situaciones no son objetos temporales, sino conjuntos de representaciones mentales («*sets of assumptions*»). Esas representaciones pueden proceder de situaciones de mundo, de estados mentales propios y ajenos o de situaciones imaginarias; no tienen, por tanto, un carácter esencialmente cronológico, aunque esa posibilidad no se excluye, y de hecho puede llegar a ser la predominante en una lengua.

Escandell-Vidal (2010, 2014, 2019a, 2021, 2022, en prensa) se sirve de esa misma idea en sus descripciones procedimentales del futuro románico y del imperfecto español: ambos tiempos verbales se caracterizan, precisamente, por la facilidad con la que pueden expresar matices relacionados con posicionamientos del hablante¹⁴. Optar por un análisis de este tipo no implica sostener que los tiempos verbales

13. Benveniste (1966) defiende la existencia de dos planos de enunciación diferentes: el primero de ellos es el plano de la historia, caracterizado por la presentación de hechos (reales o ficticios) sin establecer vínculo alguno con el sujeto y las coordenadas de la enunciación; el segundo es el plano del discurso, en el que se establece una relación entre la enunciación y los hechos presentados. El PS en tercera persona es la piedra angular de un relato de hechos pasados en el plano de la historia, ya que desvincula los hechos relatados del presente en que se relatan; el PC es, en cambio, característico del relato de hechos pasados en el plano del discurso, puesto que enlaza los hechos con el presente de la enunciación.

14. El uso del futuro para expresar conjetura referida al presente (como en el enunciado «llaman a la puerta; será el cartero») es el más frecuente en la mayoría de las variedades geográficas de español, al menos en el uso espontáneo de la lengua (Escandell-Vidal, 2010, 2014). En cuanto al imperfecto, su uso se presta a una gran cantidad de matices no estrictamente temporales (RAE-ASALE, 2009; Amenós Pons, 2010a; Escandell-Vidal, 2022).

románicos codifican sistemáticamente la modalidad o la fuente de información, sino asumir como natural que puedan aportar información sobre estas: las interpretaciones evidenciales y las temporales se basan en idénticos sistemas de relaciones.

Speas (2008) y Escandell-Vidal (2021, 2022) construyen sus análisis sobre tres nociones primitivas:

- Situación evaluada (SE): la situación sobre la que se predica una proposición p .
- Situación de referencia (SR): la situación con respecto a la cual se evalúan las condiciones de verdad de p .
- Situación de discurso (SD): la situación en la que tiene lugar la enunciación de p .

Entre estas tres nociones pueden establecerse dos clases de relaciones binarias:

- Accesibilidad (\nearrow) vs. no accesibilidad (\downarrow) desde el entorno de percepción del hablante.
- Inclusión (\sqsubset) vs. ordenamiento ($<$) de una situación con respecto a otra.

El núcleo semántico de cada forma verbal se caracteriza por una configuración específica de relaciones binarias. Según cuál sea la configuración, la forma verbal será compatible con ciertas funciones, mientras que tendrá restricciones (más o menos estrictas) para otras. Así, por ejemplo, una relación de accesibilidad entre SE y SR favorece las interpretaciones temporales de presente y pasado (situaciones a las que un individuo tiene o ha tenido acceso perceptual), y es especialmente compatible con la evidencialidad directa (basada en la percepción). En cambio, una relación de no accesibilidad favorece las interpretaciones temporales de futuro y es especialmente compatible con la evidencialidad indirecta. Por otro lado, una relación de inclusión entre SR y SD permite inferir que el foco de la enunciación (y de la percepción) no es la SE en sí misma, sino el vínculo entre SE y SR. En contrapartida, una relación de ordenamiento entre SR y SD lleva a inferir que el foco de la enunciación es la SE y su posición con respecto a SD.

Todo esto, sin embargo, no agota el contenido semántico de la forma verbal, sino que este se encuentra subsecuentemente especificado en relación con una serie de opciones paramétricas, relacionadas con la disponibilidad de determinadas opciones interpretativas.

A continuación, esbozaremos esta idea para la descripción del PC románico, que desarrollaremos para el español peninsular (estándar), el portugués europeo y el francés. Nuestra exposición nos llevará a hablar no solo de este tiempo verbal, sino también del PS, ya que el uso que en cada lengua se hace de uno de los dos

tiempos está necesariamente condicionado por la utilización que se hace del otro¹⁵. Posteriormente plantearemos una reflexión sobre la medida en que una propuesta de este tipo puede dar cabida a la variación intralingüística del PC y el PS en español.

4.2. PC y PS románico: núcleo semántico común y rasgos paramétricos

Presentamos en la tabla 1 la formalización de la instrucción procedimental del PC, compuesta por un núcleo semántico común a las distintas lenguas y un parámetro fundamental de variación entre estas. Dicho parámetro está en la base de las diferencias interlingüísticas en el uso y la interpretación de la forma verbal.

NÚCLEO SEMÁNTICO COMÚN	SR $\not\subset$ SE SR \subset SD
PARÁMETRO DE VARIACIÓN	$\pm SE \cap SD \neq 0$
Portugués (europeo)	SR $\not\subset$ SE SR \subset SD + $ SE \cap SD \neq 0$
Español (peninsular estándar)	SR $\not\subset$ SE SR \subset SD + $ SE \cap SD \neq 0$
Francés	SR $\not\subset$ SE SR \subset SD - $ SE \cap SD \neq 0$

Tabla 1. Semántica procedimental del PC románico (ejemplificado en el portugués europeo, español peninsular y francés)

La tabla 1 muestra que el PC románico codifica siempre dos tipos de relaciones: por una parte, una relación de accesibilidad entre SE y SR (la situación evaluada es accesible desde la situación de referencia). Por otra parte, una relación de inclusión entre la SR y la SD (la situación de referencia está incluida en la situación de discurso).

La primera relación favorece que una situación expresada en PC se localice como anterior al tiempo del habla. En cuanto a la segunda, la inclusión de la situación de

15. Para un análisis interlingüístico del papel de la interdependencia de las interpretaciones de los dos perfectos, entendida en términos de competencia entre las formas, puede verse Schaden (2009a y b). En términos más generales, Leonetti (2022) destaca la importancia de los mecanismos de competencia entre formas lingüísticas próximas en el desarrollo de distintos fenómenos interpretativos. Según Leonetti, la existencia de estos fenómenos muestra las ventajas de las descripciones que prestan atención a los fenómenos inferenciales, frente a los modelos descriptivos basados únicamente en el código.

referencia en la situación de discurso lleva a que el enunciado focalice no la situación evaluada en sí, sino el vínculo de esta con la situación de referencia.

Las relaciones entre las nociones primitivas SR, SE y SD no tienen una naturaleza intrínsecamente temporal (no se codifica en ellas una relación necesariamente secuencial), sino que pueden entenderse contextualmente como aspectuales, temporales o discursivo-enunciativas. En el apartado 3 se ha postulado una variable infraespecificada R, susceptible de saturarse con un contenido aspectual, temporal o discursivo. Esa variable, como tal, no resulta necesaria en el enfoque que proponemos aquí. Sin embargo, el principio de la infraespecificación lingüística que subyace en la descripción del apartado 3 es fundamental para entender que un mismo tipo de relación pueda revestir una naturaleza variable en función de las informaciones (semánticas, sintácticas y pragmáticas) con las que interactúe la forma verbal.

El rasgo crucial de diferenciación entre las lenguas románicas se encuentra en el hecho de que, en cada lengua, deba o no existir algún solapamiento entre la situación evaluada y la situación de discurso. Este parámetro está expresado en el parámetro $\pm |SE \cap SD| \neq 0$. Si una lengua presenta la opción paramétrica $+ |SE \cap SD| \neq 0$, al interpretar el PC se entenderá necesariamente que la intersección entre la situación evaluada y la situación de discurso no puede estar vacía, es decir, que existe (cierto grado de) solapamiento entre la situación evaluada y la de discurso. Si presenta la opción contraria ($- |SE \cap SD| \neq 0$), se interpretará que no tiene por qué existir solapamiento, si bien este no es imposible. El solapamiento puede no darse propiamente entre el desarrollo temporal de la situación evaluada y el de la situación de discurso, sino entre el resultado de aquella y la situación de discurso.

En definitiva, para cada una de las tres lenguas estudiadas aquí, la semántica del PC está constituida por un conjunto de tres relaciones complementarias, que son las que se expresan en la tabla 1. Dos de estas relaciones son comunes a las tres lenguas y constituyen, por tanto, el núcleo semántico de la forma verbal románica. El tercer rasgo es invariable en una lengua dada, pero cambia de una lengua a otra, es decir, está sujeto a variación paramétrica.

El PC del español y el portugués tiene el valor paramétrico $+ |SE \cap SD| \neq 0$, lo que significa que se interpreta un solapamiento entre SE y SD; el PC francés, en cambio, tiene el valor $- |SE \cap SD| \neq 0$, y por tanto no exige ese solapamiento. Una consecuencia directa de esta diferencia es que las interpretaciones del PC tienen más restricciones en portugués y español que en francés.

Para explicar esta idea presentamos, primeramente, en la tabla 2, los distintos tipos de interpretación del PC posibles en las tres lenguas de nuestro estudio, que ejemplificaremos más abajo.

	EP	ANTEPRES	DISC
Portugués (europeo)	✓	✗	✗
Español (peninsular estándar) ¹⁶	✓	✓	✗
Francés	✓	✓	✓

Tabla 2. Rutas interpretativas del PC en portugués, español y francés

Las opciones de la tabla 2 sintetizan los tipos de desarrollo contextual que, en cada lengua, puede recibir la relación entre la situación evaluada y la situación de referencia. La opción EP se refiere a la (im)posibilidad de que, en la interpretación del PC, la relación se entienda como aspectual (un resultado o una experiencia). Por otra parte, la opción AntePRES hace referencia a la posibilidad o imposibilidad de que la relación entre SE y SR se interprete en una lengua como una conexión de naturaleza temporal (típicamente, un pasado hodiernal o reciente). Finalmente, la opción «discursivo-enunciativa» Disc capta la idea de que, en algunas lenguas (pero no en otras), el vínculo entre SE y SR puede entenderse como la pertinencia de inferir la inclusión del hablante y el oyente en el propio discurso, sin que exista (necesariamente) algún tipo de conexión aspectual o temporal. A continuación, mostramos un ejemplo de cada tipo para las tres lenguas analizadas. Los ejemplos permiten observar qué es posible y qué no para el PC de cada una:

- (8) Interpretaciones EP
 - (8a) Port. O Manuel *tem visitado* a avó. (PC)¹⁷
 - (8b) Esp. Manuel *ha visitado* (varias veces) a su abuela. (PC)
 - (8c) Fr. Manuel *a rendu visite* à sa grand-mère (plusieurs fois). (PC)
- (9) Interpretaciones AntePRES
 - (9a) Port. Esta manhã, Maria *foi* ao médico. (PS)

16. Las rutas interpretativas del PC en español peninsular estándar coexisten, dentro de la península ibérica, con sistemas aparentemente distintos, en los que las rutas interpretativas no coinciden. Este aspecto se tratará más adelante, en el apartado 4.3.

17. En portugués, el uso del PC lleva necesariamente a interpretar que la situación que se predica ha tenido lugar un número indeterminado de veces, en un intervalo no especificado y que se prolonga hasta la situación de discurso (Mateus *et al.*, 2003).

- (9b) Esp. Esta mañana, María *ha ido* al médico. (PC)
 (9b) Fr. Ce matin, Maria *est allée* chez le médecin. (PC)
- (10) Interpretaciones Disc
 (10a) Port. Há quarenta e quatro anos atrás, a 19 de Dezembro de 1978, exactamente às 8.27 da manhã, a França *foi mergulhada* na escuridão. (PS)
 (10b) Esp. Hace cuarenta y cuatro años, el 19 de diciembre de 1978, exactamente a las 8.27 horas, Francia *se sumió* en la oscuridad. (PS)
 (10c) Fr. Il y a quarante-quatre ans, le 19 décembre 1978 à 8 h 27 exactement, la France *est plongée* dans le noir. (PC)

Es importante precisar que, a diferencia de los parámetros presentados en la tabla 1, las opciones de la tabla 2 no son rasgos semánticos del PC, sino distintos tipos de interpretación. Así, los rasgos paramétricos del PC en una lengua (tabla 1) están presentes en todas las interpretaciones posibles de la forma verbal (de hecho, constituyen la base de estas). En cambio, cada tipo de interpretación implica una ruta interpretativa determinada, aunque sin excluir, en algunos casos, la ambigüedad (por ejemplo, si no existen suficientes informaciones contextuales para discriminar una ruta interpretativa u otra).

Como se ha dicho, la interpretación que recibe el PC está en función del tipo de predicado, de las expresiones localizadoras (si las hay), del género discursivo y de las expectativas de relevancia. Esto es, por tanto, lo que determina el enriquecimiento contextual que recibe en cada caso la semántica de la forma verbal. La importancia de dicho enriquecimiento contextual, vinculada a la armonización de los distintos tipos de informaciones lingüísticas y extralingüísticas, permite explicar, por ejemplo, por qué, dentro de una misma variedad lingüística del español, un uso no hodiernal del PC resulta aceptable en algunos casos, pero no en otros:

- (11) *Se han casado* hace unas semanas y se les ve muy enamorados.
 (12) #*Ayer se ha levantado* temprano y se le ve muy preocupado.

En (11), el localizador «hace unas semanas» resulta contradictorio con las restricciones del PC en español, al dificultar el establecimiento de un vínculo objetivo entre SE y SD. Sin embargo, la interpretación se salva si el interlocutor es capaz de percibir, en la relación entre la situación evaluada y la situación de discurso, unos efectos de sentido que no se derivarían si no se estableciera tal relación: la situación actual se explica por la enunciación del hecho pasado, causante del estado de cosas presente. En cambio, en (12), en ausencia de otras precisiones, la interpretación

fracasa porque la relación del pasado con el presente no produce efecto alguno, y por tanto no se puede justificar interpretativamente el uso del PC. Este tiempo verbal, en cambio, resultará aceptable en una situación en la que sea manifiesto que el personaje no se ha acostado la última noche, es decir, siempre que el PC se entienda como referido a la ocurrencia más reciente del evento «levantarse».

El término «rutas interpretativas», usado en el título de la tabla 2, requiere alguna aclaración suplementaria: es evidente que el tipo de EP no es siempre el mismo, dado que existen interpretaciones resultativas, experienciales y continuativas del PC. Asimismo, las interpretaciones de AntePRES pueden presentar distintos matices (pasado inmediato, pasado hodierno, pasado prehodierno sin referencia expresa). El tipo de relación entre el hablante y su discurso también puede revestir tintes variados. La noción de «rutas interpretativas», precisamente, trata de capturar esta pluralidad: la tabla 2 contiene un esbozo de la manera en que, en cada lengua, puede restringirse el espacio cognitivo de búsqueda de una interpretación satisfactoria. En la elección de una ruta u otra (entre aquellas disponibles en una lengua dada) incide la particular combinación de rasgos lingüísticos (tipo de predicado, expresiones localizadoras) y pragmático-discursivos (conocimiento del mundo, situación de enunciación, información derivada del procesamiento de enunciados previos en el discurso) con los que se relaciona la forma verbal en los enunciados.

De este modo, el PC del portugués europeo y el del español peninsular estándar comparten el valor paramétrico $+ [SE \cap SD] \neq 0$. Esto tiene como consecuencia que ninguna de las dos lenguas admite un uso del PC fundado únicamente en el tipo de discurso (oral vs. escrito) o en el registro (informal vs. formal). Según la tabla 2, tanto el portugués como el español aceptan que el vínculo entre la situación evaluada y la situación de referencia se interprete como aspectual (ruta interpretativa EP). Esto no excluye, sin embargo, la existencia de contrastes interlingüísticos en las restricciones de los tipos de EP accesibles en cada lengua, como veremos enseguida.

En las restricciones del PC en las distintas lenguas es fundamental el uso que cada una de ellas hace de la forma simple de pasado perfecto, es decir, el PS. Al igual que ocurre con el PC, el PS de las lenguas románicas tiene un núcleo semántico común a todas, complementado por un parámetro con dos opciones, causante de la variación interlingüística. Esta idea queda recogida en la tabla 3:

NÚCLEO SEMÁNTICO COMÚN	SR $\not\subset$ SE SR < SD
PARÁMETRO DE VARIACIÓN	$\pm SE \cap SD = 0$
Portugués (europeo)	SR $\not\subset$ SE SR < SD - $ SE \cap SD = 0$
Español (peninsular estándar)	SR $\not\subset$ SE SR < SD + $ SE \cap SD = 0$
Francés	SR $\not\subset$ SE SR < SD + $ SE \cap SD = 0$

Tabla 3. Semántica procedimental del PS románico (ejemplificado en el portugués, español y francés)

En sus rasgos nucleares, la semántica del PS se diferencia de la del PC en el hecho de que la situación de referencia precede siempre a la situación de discurso¹⁸. Esto se expresa con la fórmula $SR < SD$, que junto con la relación $SR \not\subset SE$ ('la situación evaluada es accesible desde la situación de discurso') constituye el núcleo semántico del PS. Así las cosas, la variación fundamental entre los PS románicos se encuentra en la posibilidad o imposibilidad de que, en cada lengua, la forma verbal exprese algún tipo de solapamiento de la situación evaluada con respecto a la situación de discurso. El PS del español peninsular estándar y el del francés contienen el valor paramétrico $+ |SE \cap SD| = 0$, es decir, la intersección entre SE y SD tiene que estar vacía. Esto significa que no puede haber solapamiento; por tanto, en estas lenguas, el PS localiza siempre situaciones desconectadas de la situación de discurso. En cambio, el PS portugués contiene el rasgo $- |SE \cap SD| = 0$, es decir, el PS portugués carece de restricciones en cuanto al solapamiento entre la situación evaluada y la situación de discurso (puede haber, por tanto, solapamiento entre SE y SD). Así, en esta lengua el PS puede operar contextualmente no solo como una forma aspectualmente perfecta, sino también como una forma de aspecto perfecto. En consecuencia, el PS portugués puede expresar sentidos resultativos que, en español peninsular estándar y en francés, están reservados al PC.

18. En cambio, tal como se representa en la tabla 1, en el PC la situación de referencia está incluida en la situación de discurso ($SR \subset SD$).

- (13) *Comprei* cervejas; estão no frigorífico. (PS)
- (14) *He comprado* cervezas; están en la nevera. (PC)
- (15) *J'ai acheté* des bières ; elles sont dans le réfrigérateur. (PC)

El PC portugués solo puede obtener interpretaciones de aspecto perfecto (no perfectivas), y aun en este terreno compite con el PS. En portugués, los distintos tipos de interpretaciones de aspecto perfecto se han repartido entre las dos formas verbales. El PC portugués se vincula siempre a la pluralidad eventiva; solo admite interpretaciones experienciales de tipo durativo o acumulativo, en las que la repetición del evento se entiende como una pluralidad indivisible o incontable (Mateus *et al.*, 2003):

- (16) O profesor *tem chegado* tarde [un número indeterminado de veces]
- (17) *O profesor *tem chegado* tarde quatro veces.
- (18) *O profesor *tem chegado* às quatro.

En cambio, como se observa en (13), las interpretaciones resultativas del PC no son posibles en portugués actual. Este hecho parece ser el resultado de un proceso de especialización (siglo XVI), fruto de la competencia con el PS y con *haver* + participio (Wigger, 2005; Azpiazu, 2019).

En contraste, el PC español peninsular estándar puede recibir interpretaciones temporales de antepresente y también aspectuales resultativas, experienciales y continuativas, como las que se han mostrado al principio de este trabajo, en (1), (2) y (3). Recordemos que el valor paramétrico $+ |SE \cap SD| = 0$ del PS orienta hacia la representación de una situación evaluada que no puede solaparse con la situación de discurso; esto hace del PC el candidato óptimo para la expresar los distintos matices de aspecto perfecto.

La semántica del PC y la de PS delimitan un conjunto de opciones interpretativas disponibles, pero en la distribución intervienen también factores relacionados con la optimización funcional del contraste entre dos formas potencialmente en competencia (Schaden, 2009a y b). Por tanto, la instrucción procedimental del PC no determina completamente la interpretación de la forma verbal, pero sí limita o excluye unas posibilidades, a la vez que favorece otras. Esto tiene relación no solo con la semántica del propio PC, sino con la del PS en cada lengua, y con el sistema del que ambos forman parte. En cada sistema, la competición entre dos formas verbales próximas (PC y PS) se resuelve de manera parcialmente diferente.

Que el PS portugués (a diferencia de su equivalente morfológico español) admita sin dificultad lecturas de perfecto resultativo no implica que se haya convertido en un tiempo de aspecto perfecto, sino que puede funcionar ocasionalmente como tal, si el tipo de predicado y el entorno sintáctico y discursivo lo favorecen: por ejemplo, en (13), la interpretación resultativa es posible por la combinación del PS con un predicado télico (un logro); a esto se añade el hecho de que la afirmación de que las cervezas están en la nevera resalta no el evento de la compra en sí, sino el resultado de esta en la situación de discurso: la disponibilidad de cervezas¹⁹.

En la tabla 2 (rutas interpretativas del PC), la ruta interpretativa Disc está señalada como no disponible para el español peninsular. Con todo, como se ha visto en el enunciado (11), es posible que, en esta lengua, la interpretación del PC esté ligada al relieve psicológico; este surge solo como resolución de un desajuste entre la necesidad de establecer algún tipo de vinculación entre SE y SD, en un entorno como el de (11), en el que el localizador *hace unas semanas* bloquea la interpretación de una proximidad de pasado reciente. El desajuste se salva con la obtención de un efecto de perspectiva: la proximidad no es algo objetivo, sino subjetivo, sentido por el hablante²⁰.

El efecto de perspectiva, por tanto, se produce sin que desaparezca el vínculo aspectual o temporal (uno u otro, en función del tipo de predicado y del entorno discursivo) con la situación de referencia²¹. Dicho de otro modo, el español peninsular

19. En el PS del español peninsular estándar, en un entorno similar, se derivaría igualmente la interpretación de un estado resultante. La diferencia con el portugués, sin embargo, es que el PS español transmitiría dos informaciones: que la compra se realizó en un momento del pasado no conectado con el presente, y que el resultado de esa compra sigue teniendo vigencia en la situación de discurso. En portugués, en contrapartida, solo la segunda información se pone de relieve. Esto es lo que permite al PS portugués funcionar como perfecto resultativo en muchos más entornos que el del español.

20. Efectos interpretativos relacionados con la perspectiva subjetiva del hablante, surgidos de la necesidad de resolver conflictos interpretativos, se han descrito asimismo para otros tiempos verbales, y en especial para determinadas interpretaciones del pretérito imperfecto de indicativo como las citativas («¿Pero tú no llegabas mañana?»). La inferencia de un sujeto de percepción como forma de resolver en la interpretación desajustes entre rasgos semántico-sintácticos se ha detectado asimismo en otros ámbitos de la gramática, como en las combinaciones de la cópula *estar* con adjetivos ligados a la predicación de individuo («La última película de Almodóvar está genial»). Con respecto al imperfecto citativo, véase Saussure, 2003; Amenós Pons, 2010a y 2020; Escandell-Vidal, 2022 y en prensa. Para un análisis más global, véase Escandell-Vidal y Leonetti (2011) y Escandell-Vidal (2019b).

21. Entendemos que las rutas interpretativas no son necesariamente excluyentes. Además, existen muchos casos de ambigüedad, en los que la interpretación del enunciado no exige determinar, por ejemplo, si la interpretación buscada es de EP o de AntePRES (Azpiazu, 2019). Por otra parte, el grado de prominencia contextual de un EP, con un predicado télico, puede ser variable (Amenós Pons, 2010a). Todo esto apoya la idea de que, en la interpretación de una conexión aspectual, temporal o psicológica entre SE y SR, es crucial el papel de la pragmática.

estándar admite sin dificultad el denominado PC de noticias recientes, que es una de las interpretaciones posibles del enunciado (7), repetido abajo por comodidad²². En cambio, no son posibles en español peninsular estándar usos de pasado perfectivo general como los del francés (basados en la inclusión enunciativa de hablante y oyente en el discurso), representados en (19).

- (7) - ¡Anda, qué escayola llevas!
+ Sí, es que *me he roto* la pierna esquiando.
- (19) *J'ai bien travaillé* toute la semaine, Raymond *est venu* et *m'a dit* qu'il avait envoyé la lettre. *Je suis allé* au cinéma deux fois avec Emmanuel qui ne comprend pas toujours ce qui se passe sur l'écran. Il faut alors lui donner des explications. Hier, c'était samedi et Marie *est venue*, comme nous en étions convenus. *J'ai eu très envie* d'elle parce qu'elle avait une belle robe à raies rouges et blanches et des sandales de cuir. (Albert Camus, *L'étranger*)²³

Gracias al PC, la relevancia actual del evento en PC de (7) se interpreta como una valoración subjetiva del hablante, fundamentada en su propia experiencia o percepción directa. El uso del PC en ese enunciado no implica solo proximidad temporal (cercanía al momento del habla), sino que expresa también (y ante todo) un juicio subjetivo del hablante, que se erige en protagonista. La noción que expresa el PC de noticias recientes, por tanto, está muy relacionada con la evidencialidad directa. Por supuesto, esto no significa que la fuente de información esté codificada como tal en el PC, sino que el entorno en el que este se interpreta favorece la inferencia de dicha fuente (para un análisis de las «estrategias evidenciales» en el PC español, véase Azpiazu, 2016).

En lo que atañe al PC francés, ciertamente permite las interpretaciones resultativas, experienciales y continuativas, y también las interpretaciones de antepresente; se detectan asimismo interpretaciones de proximidad psicológica, que combinan

22. Un revisor anónimo ha señalado que el PC del enunciado (7) no recibe necesariamente interpretación de noticias recientes, sino que en algunos casos puede entenderse como resultativo o de pasado reciente. Compartimos plenamente esta observación. Para nosotros, esto pone de manifiesto que el tipo de relación entre la situación evaluada y la situación de discurso está lingüísticamente infraespecificada, y que el hecho de inclinarse por una interpretación u otra es algo que ocurre durante el proceso de interpretación, cuando el oyente pone en relación sus expectativas y necesidades informativas con las informaciones lingüísticas que le brinda el enunciado y con las hipótesis contextuales accesibles.

23. El uso del PC como tiempo narrativo en esta novela de Camus se justifica, en gran parte, por la recreación de la lengua oral que lleva a cabo el autor: todo el relato está puesto en boca de su protagonista. Para un análisis lingüístico de este uso, véase De Swart y Molendijk (2002); Amenós Pons (2010b).

las rutas interpretativas AntePRES y Disc. Estos tipos de interpretaciones, y los mecanismos que las hacen posibles, son los mismos que en español. Allí donde, en cambio, el francés se diferencia del español es en el uso narrativo del PC, representado en (19). El vínculo entre SE y SR, en dicho uso narrativo, no es propiamente temporal ni aspectual, sino que se trata de un rasgo caracterizador de aquellos géneros que pertenecen al universo del discurso (frente al universo de la historia, según la terminología de Benveniste, 1966). Los géneros que favorecen el uso narrativo del PC son los vinculados a la oralidad, en los que se establece una relación (siempre implícita, y a veces explícita) entre los hechos narrados y la situación de enunciación. Ahora bien, al tratarse de un uso convencional no marcado (y no de una elección del hablante), el PC francés en uso narrativo no transmite la inferencia pragmática de una implicación directa del hablante; se difumina y se pierde, por tanto, el efecto de evidencialidad directa que (7) transmite en español. Esta es, sin duda, una diferencia importante entre el español y el francés.

4.3. *Apuntes en torno a la variación intralingüística del PC*

En el seno de una misma lengua pueden (y de hecho suelen) coexistir varios sistemas, organizados en función de distintas variables. Entre estas se encuentran las situacionales (informal/formal, oral/escrito) y las geolectales. Al comparar la gramática de dos o más subsistemas es esperable que alguna de las opciones paramétricas que definen la semántica de una expresión lingüística (por ejemplo, un tiempo verbal) pueda diferir de un subsistema a otro. Por supuesto, si existe variación paramétrica semántica, los tipos de enriquecimiento contextual que admitirá la expresión se verán afectados. Veamos cómo se relaciona todo esto con el PC (y por extensión, con el PS) de dos de las lenguas que consideramos aquí: el francés y el español.

En cuanto al contraste PC/PS, coexisten en francés actual un sistema espontáneo o natural, materializado en la lengua oral y en la escrita no literaria, y un sistema cultivado, arcaizante, característico de la escritura literaria²⁴. En el primer sistema,

24. Escandell-Vidal (2019a), apoyándose en Moreno Cabrera (2013), establece una serie de características que permite diferenciar cuándo nos encontramos ante un uso propio de un sistema natural y cuándo estamos ante un uso propio de un sistema cultivado, no natural. Un sistema natural se adquiere tempranamente, en la infancia; se usa activamente y no está restringido a un registro o situación específica; se interpreta siguiendo patrones cognitivos generales, que combinan información lingüística y extralingüística. En contrapartida, un sistema cultivado es de adquisición tardía y suele requerir instrucción formal; los hablantes pueden tener de él solo un conocimiento pasivo; su uso se

la semántica del PC es la que hemos descrito en el apartado precedente: $SR \not\subset SE$; $SR \subset SD$; $- |SE \cap SD| \neq 0$. Estos rasgos permiten que el PC esté abierto a las rutas interpretativas que más arriba, en la tabla 2, hemos denominado EP, AntePRES y Disc. Eso implica que el PC francés admite el uso narrativo. En cuanto al PS, está casi ausente del sistema espontáneo, con excepción de algunas unidades fraseológicas, como la que se muestra en (20), o de su aparición ocasional en textos periodísticos como (21), a caballo entre el sistema natural y el uso deliberado de un sistema arcaizante recuperado con voluntad estilística²⁵. En este último caso, el periodista comienza una crónica deportiva en PC, pero cambia al PS cuando la crónica se transforma en pura narración, para dar así relieve al encadenamiento de eventos, sin que se infiera precedencia temporal de las situaciones en PS con respecto a las que se expresan en PC.

(20) Fiston ! Il *fut* un temps où j'étais comme toi, il viendra un jour où tu seras comme moi.

(21) La dernière victoire de Lorient en Ligue 1 remontait à avant les fêtes, au 14 décembre 2008, au Havre (3-1). Hier, les Merlus ont renoué avec le succès après dix journées d'attente [...].

Lorient, désormais onzième avec 35 points, a fait la différence dans le premier quart d'heure, en inscrivant donc coup sur coup deux buts. Amalfiato *expédia* d'abord une frappe pure du droit dans la lucarne gauche de Janot, en ayant tout le temps de reprendre un centre de Gameiro (1-0, 12e). À peine l'engagement effectué par les Verts, le ballon *brûla* les pieds de Bayal pour arriver à nouveau dans les pieds de Gameiro, double passeur décisif, cette fois pour Vahirua (2-0, 13e), qui *dut céder* sa place à la mi-temps. [...] (*Le Monde*)

limita a un conjunto determinado de situaciones y registros; se interpreta de acuerdo con convenciones discursivas específica, que hay que aprender.

25. Un estudio detallado de los entornos y factores que favorecen la aparición del PS en francés actual se encuentra en Labeau (2022). Tras llevar a cabo un amplio estudio de corpus, la autora concluye que la división entre el uso narrativo del PS y del PC no se ajusta estrictamente al contraste escritura vs. oralidad, sino que existe un continuum que va desde los registros más formales de la lengua escrita hasta los más informales de la lengua oral. En la producción oral formal, e incluso en la lengua coloquial, se detectan usos ocasionales del PS de carácter formulaico y limitados a ciertos verbos (especialmente, la tercera persona del verbo *être*). En cuanto a los medios de comunicación escritos, el PC es el tiempo dominante, pero Labeau menciona el cultivo deliberado del PS en el diario *Le Monde*, posiblemente amparado en el prestigio lingüístico de esta forma verbal.

En el sistema cultivado, arcaizante, el PC presenta los mismos rasgos semánticos que los del PC en el español peninsular estándar: $SR \not\subset SE$; $SR \subset SD$; $+ \mid SE \cap SD \mid \neq 0$. En este sistema, el PC puede enriquecerse contextualmente como EP o como AntePRES, pero no como Disc. Esto excluye el uso narrativo, para el cual se recurre al PS (cuya semántica se ha descrito en el apartado anterior). En la medida en que esta última forma verbal es prototípica de los textos literarios narrativos, se ha convertido ante todo en una marca lingüística de narración, por encima de la expresión de matices temporales.

En lo que se refiere al español, es bien conocida la existencia de fuerte variación geolectal en el uso de los perfectos. La variación geográfica en el uso del PC español ha dado lugar a distintas propuestas descriptivas, entre las que se encuentran las de Veiga (2014) y Azpiazu (2019). Tomamos aquí como referencia la de Azpiazu, que a su vez se basa en la de Veiga, a la que matiza y completa. Reproducimos la propuesta de Azpiazu en la tabla 4, para después relacionarla con nuestros planteamientos.

	SUBSISTEMA A		SUBSISTEMA B			
	Variedad A1	Variedad A2	Variedad B1a	Variedad B1b	Variedad B2a	Variedad B2b
P1	+	+	-	-/+	+	+
P1 + P2	+	+	-	-/+	+	+
P1 + P2 + P3a	+	-	-	-	+	-/+
P1 + P2 + P3a + P3b	-	-	-	-	+	-/+
Regiones	Español peninsular, variedades cultas de Perú y Bolivia	México, Colombia, Venezuela, Canarias...	Noroeste peninsular	Río de la Plata	Habla popular de Bolivia	Centro peninsular, hablas populares de Perú, Ecuador

Tabla 4. Subsistemas del PC español, con sus variedades correspondientes (Azpiazu, 2019)

En el modelo de Azpiazu, el Perfecto 1 (P1) representa el uso experiencial y el continuativo (como en *He visitado muchas veces el Museo del Prado* o en *Ha trabajado veinte años en este hospital*). P2 representa el uso resultativo con indeterminación temporal (como en *Han reparado la puerta tres veces pero sigue estropeada*). P3a se refiere a hechos sucedidos en un pasado inmediato, hodiernal (como en *¿Has entendido lo que te he dicho?*) con o sin referencia precisa, mientras que P3b codifica eventos prehodiernales de referencia precisa (*El año pasado se han jubilado cinco empleados en la empresa*).

Azpiazu distingue dos grandes subsistemas geolectales: el subsistema A agrupa las variedades del español en las que existe oposición entre PC y PS; en la variedad A1, el PC puede obtener lecturas continuativas, experienciales, resultativas y de pasado hodiernal. En la variedad A2, solo puede obtener lecturas continuativas y experienciales. En cuanto al subsistema B, incluye los usos de las áreas en las que

no existe oposición entre el PC y el PS (es decir, aquellas zonas en las que uno solo de los tiempos tiende a concentrar todas las lecturas que, en el subsistema A, se reparten entre los dos tiempos verbales).

Recordemos que, en nuestros propios términos, las varias interpretaciones de un tiempo verbal no presuponen necesariamente una diferencia en la semántica de la forma verbal propiamente dicha. Por otra parte, la disponibilidad de unas rutas interpretativas u otras en cada variedad geolectal no se debe únicamente a la semántica del PC en cada una, sino al modo en que dicha semántica se combina con la del PS. En cada variedad, las combinaciones de los dos tiempos (con opciones paramétricas variadas) autorizan distintas resoluciones de la competición entre PC y PS.

Así, traduciendo el contenido de la tabla 4 a las rutas interpretativas que hemos descrito en el apartado anterior (tabla 2), diremos que, en la variedad A₁, el PC presenta las opciones de enriquecimiento contextual EP y AntePRES; la semántica del PC y del PS, en esta variedad, es la que hemos descrito para el español peninsular estándar²⁶. En contrapartida, en la variedad A₂, la única ruta interpretativa accesible es la de EP; que esto pueda ocurrir implica que el PS, en esta variedad, ha tomado el valor paramétrico $- |SE \cap SD| = 0$, lo que permite un reparto de tareas diferente entre el PC (cuya semántica es como la del español peninsular estándar) y el PS.

En síntesis, las diferencias entre las variedades A₁ y A₂ no se deben a una distinta semántica del PC, sino a una diferencia en el PS, cuyo valor paramétrico es $+ |SE \cap SD| = 0$ en la variedad A₁, y $- |SE \cap SD| = 0$ en la variedad A₂. Esto significa que, en México, Colombia, Venezuela y Canarias, el PS puede interpretarse como antepresente, cosa que no ocurre en español peninsular estándar.

Por otra parte, en la variedad B_{1a}, el PC tiende a estar ausente del sistema espontáneo del idioma. Así las cosas, el PS toma la opción paramétrica $- |SE \cap SD| = 0$. De este modo, el PS del sistema B_{1a} permite tanto interpretaciones aspectualmente perfectivas como de perfecto (al igual que ocurre en portugués).

En contrapartida, en la variedad B_{2a} es el PS el que tiende a estar ausente, con lo que en esta última variedad todas las rutas interpretativas del PC están abiertas: EP, AntePRES, Disc. Para que esto sea posible, el PC de la variedad B_{2a} adopta el valor paramétrico $- |SE \cap SD| \neq 0$. La semántica del PC, en la variedad B_{2a}, permite que el PC se utilice como tiempo narrativo principal en secuencias no hodiernas, que se ven vinculadas a la experiencia del hablante (Azpiazu, 2019, p. 190):

26. Tal como hemos señalado en el apartado 4.2, en la variedad A₁ existe el denominado PC de noticias recientes, pero su interpretación, típicamente, está vinculada a las interpretaciones de AntePRES en entornos en los que una proximidad temporal objetiva entre SE y SD queda descartada.

- (22) *He tenido* los estudios primarios y secundarios en el Colegio La Salle de La Paz, los doce años en el mismo colegio. Luego *he salido* a estudiar los estudios universitarios en la Universidad, primero Católica de Chile y luego en la Universidad del Estado de Chile. *He tenido* estudios de ciencias sociales y políticas, y administrativas en la escuela que se encuentra a cargo de la Universidad de Chile. *He terminado* esa etapa de estudio y regresé a mi país en la que... por motivos políticos, en los cuales yo no tenía cabida en ese momento (español de La Paz, en Soto, 2014, p. 141).

Las variedades B1b, propia del Río de la Plata, y B2b, característica de Perú y Ecuador, así como de algunas zonas del centro de la península ibérica, presentan situaciones lingüísticas especialmente complejas. En el Río de la Plata existe gran variación microlectal, por una parte y, por otra, un contraste entre oralidad y escritura relacionable con la coexistencia de un sistema espontáneo (que tiende a restringir el PC a los usos de P1) y un sistema cultivado (que puede ampliar sus usos a los del sistema P2) (Kempas, 2006; Rodríguez Louro, 2009; Parrinha, 2014; Azpiazu, 2019).

En la variedad B1b, el PS presenta el valor paramétrico - $|SE \cap SD| = 0$, mientras que la semántica del PC no difiere de la que hemos descrito para las variedades A1 y A2. Sin embargo, en la variedad B1b parecen coexistir dos sistemas (uno espontáneo y otro cultivado), y en cada uno de ellos la especialización funcional entre PC y PS se ha resuelto de manera algo diferente: el PS es en ambos sistemas el tiempo verbal con más rendimiento funcional, pero este es mayor en el sistema espontáneo que en el sistema cultivado.

En lo que se refiere al PC de Perú y Ecuador (dentro de la variedad B2b), admite las rutas interpretativas EP, AntePRES y Disc, lo que implica que (al igual que el PC de la variedad B2a) puede recibir interpretaciones no solo aspectuales y temporales, sino también discursivo-enunciativas. Estas últimas se relacionan con la subjetividad, con la evidencialidad directa: el hablante se convierte en el protagonista y experimentante. Ahora bien, lo que parece caracterizar el sistema B2b, en Perú y Ecuador, es que la ruta interpretativa Disc puede entenderse también como una expresión de evidencialidad indirecta; el PC puede referirse a un hecho no vivido, del que no se tiene certeza (Dumont, 2013), ejemplificado en (21):

- (23) A. Pero *me han dicho* que no es tan bueno, porque si se daña la filmadora, se daña todo, por ejemplo...
C. Hmm, *no he sabido* eso, pero ha de ser (Dumont, 2013)

Esta característica del español andino parece relacionada con la influencia del sistema evidencial del quechua, como lengua en contacto con el español (Azpiazu, 2019). Así, la coexistencia o la proximidad geográfica entre lenguas favorece la frecuencia de determinadas interpretaciones, mientras que excluye o restringe otras. También puede motivar caminos evolutivos específicos, como parece ocurrir aquí, dado que no existen otras variedades del español en las que el PC pueda expresar evidencialidad indirecta (es decir, casos en los que, a través del PC, el hablante se presenta a sí mismo como fuente de una información basada en su experiencia personal).

La frecuencia de las interpretaciones del PC vinculadas a la proximidad temporal subjetiva en el español del centro peninsular (en particular, en áreas de Salamanca y Madrid) es lo que lleva a Azpiazu a incluir el PC de esta zona dentro de la variedad B2b. Ahora bien, aun admitiendo que esto pueda ser así, hay que considerar que el PS de estas áreas de la península ibérica tiene una semántica parcialmente distinta de la de Perú y Ecuador: el PS del español peninsular posee los rasgos $SR \nearrow SE$; $SR < SD$; $+ |SE \cap SD| = 0$, mientras que el de Perú y Ecuador tiene los rasgos $SR \nearrow SE$; $SR < SD$; $- |SE \cap SD| = 0$. Así, el PS peninsular no admite solapamiento entre la situación evaluada y la situación de discurso, mientras que el PS andino sí lo admite. Una consecuencia de esta diferencia es que las interpretaciones evidenciales del español de Salamanca y Madrid surgen de la necesidad de ajustar contextualmente las posibilidades interpretativas de los dos tiempos verbales que compiten en el sistema: el PC admite cualquier ruta interpretativa, pero la semántica del PS no permite solapamiento entre SE y SD. Conciliar las posibilidades del sistema lleva a considerar que, con el PC, la proximidad entre la situación evaluada y la de discurso puede no ser un hecho objetivo, sino una percepción subjetiva del hablante. Tal cosa no ocurre en el español andino, donde la interpretación evidencial surge sin conexión temporal de SE y SD:

- (24) [...] estuve un mes no más [en mi tierra] después me regresé / me enfermé / mi garganta *se ha cerrado* y todo me ha pasado / no no se abrió mi garganta / todo enfermedad me agarró gripe todo y total amarilla *m'he vuelta* (español de Quito, Escobar 1997, citado por Azpiazu, 2016)
- (25) Salimos ayer mi hija y yo del Corte Inglés y nos hemos encontrado un monedero tirado. Por supuesto, sin dinero, pero tenía todas las tarjetas. Entonces lo *hemos entregado* allí en el Corte Inglés («Hoy por hoy Salamanca», *Cadena Ser*, Azpiazu, 2016)

Todo esto lleva a considerar que, en las hablas populares de Bolivia, Perú y Ecuador (variedades B2a y B2b), el PC posee una semántica específica, con un valor

paramétrico diferente del de las demás variedades. En esos países, el PC presenta la opción paramétrica - $|SE \cap SD| \neq 0$: no se requiere solapamiento entre la situación de enunciación y la de discurso. Esto permite que la conexión entre ambas situaciones se entienda únicamente como enunciativa, sin intervención directa del aspecto ni del tiempo²⁷. Esto favorece las interpretaciones evidenciales, sin necesidad de que estén ligadas a la proximidad temporal (a diferencia de lo que ocurre en el español de Salamanca o en el de Madrid, que Azpiazu incluye también dentro de la variedad B2b: en estas zonas, las interpretaciones evidenciales surgen de la necesidad de aproximar temporalmente un evento concluido a una situación de discurso objetivamente posterior y separada de aquel).

El conjunto de sistemas del español peninsular y americano, tal como los hemos descrito, se resume en la tabla 5. Para evitar reiteraciones y reducir la complejidad de la tabla, incluimos solamente el rasgo paramétrico que cambia de una variedad a otra (y no los rasgos semánticos invariables).

	SUBSISTEMA A		SUBSISTEMA B			
	Variedad A1	Variedad A2	Variedad B1a	Variedad B1b ²⁸	Variedad B2a	Variedad B2b ²⁹
Rasgo paramétrico del PC	+ $ SE \cap SD \neq 0$	+ $ SE \cap SD \neq 0$	--	+ $ SE \cap SD \neq 0$	- $ SE \cap SD \neq 0$	- $ SE \cap SD \neq 0$
Rasgo paramétrico del PS	+ $ SE \cap SD = 0$	- $ SE \cap SD = 0$	- $ SE \cap SD = 0$	- $ SE \cap SD = 0$	--	Centro peninsular: + $ SE \cap SD = 0$ Perú y Ecuador: - $ SE \cap SD = 0$
Regiones	Español peninsular, variedades cultas de Perú y Bolivia	México, Colombia, Venezuela, Canarias...	Noroeste peninsular	Río de la Plata	Habla popular de Bolivia	Centro peninsular, hablas populares de Perú, Ecuador

Tabla 5. Variación paramétrica del PC y PS en español

Pretender una descripción detallada de los valores paramétricos en las distintas variedades del español y de las restricciones que afectan a sus opciones de enriquecimiento contextual sería un empeño inabarcable para un artículo como este.

27. Escobar (1997) explica que los tres PC de (24) responden a «hechos que le sucedieron específicamente a la informante, pues los experimentó físicamente» (citado en Azpiazu, 2016, p. 314).

28. A diferencia de la variedad A2, la variedad B1b se caracteriza por un fuerte contraste entre un sistema espontáneo y un sistema cultivado; en el sistema espontáneo, dentro de las opciones que permite la semántica de los dos tiempos verbales, la competición entre PC y PS se resuelve claramente a favor del PS, con lo que las opciones de enriquecimiento pragmática del PC están altamente restringidas.

29. Por razones de coherencia expositiva, mantenemos las denominaciones de los subsistemas que propone Azpiazu (2019), si bien la diferencia paramétrica del PS en el centro peninsular frente a las hablas populares de Perú y Ecuador cuestionan la adecuación de postular un único sistema para todas esas zonas.

Nuestro objetivo, aquí, ha sido sugerir el modo en que un planteamiento como el nuestro, dotado de una arquitectura estable y flexible a la vez, puede dar cuenta de la variación lingüística. Por supuesto, describir la realidad del uso no consiste solo en etiquetar los rasgos semánticos y los valores paramétricos. Las interpretaciones de una expresión lingüística se construyen a partir de sus rasgos semánticos, y de ahí la utilidad de contar con un sistema descriptivo basado en un conjunto cerrado de rasgos determinados. Sin embargo, cada expresión forma parte de un sistema específico, establecido en un entorno lingüístico y social determinado; eso lleva a realizaciones tendencialmente diferentes del conjunto limitado de rasgos semánticos propios de la expresión, vinculadas a mecanismos de competición entre dos formas próximas como son el PC y el PS.

5. CONCLUSIONES

Al comenzar el artículo, hemos planteado la conveniencia de integrar factores semánticos y pragmáticos en la descripción del PC. Al mismo tiempo, hemos constatado la necesidad de buscar una forma sistemática de relacionar los distintos elementos para lograr descripciones lo suficientemente explícitas, adecuadas y coherentes. Por otra parte, al comparar lenguas próximas y directamente relacionadas, como es el caso de las románicas, se observa enseguida que coexisten rasgos comunes con una variabilidad evidente. Cualquier descripción lingüística debería incorporar esta constatación.

Apoyándonos en la tesis de la infradeterminación semántica (Sperber y Wilson, 1995) y en la idea de que el significado de los tiempos verbales tiene naturaleza procedimental (Moeschler, 1994, 1998; Nicolle, 1997, 1998; Saussure, 2003), hemos desarrollado una descripción del PC y (y por extensión, del PS) basada en la consideración de que la estructura de una expresión procedimental contiene diferentes fases o capas, y de un conjunto limitado de combinaciones de unos pocos rasgos semánticos está en el origen de todas las interpretaciones de un tiempo verbal, ya sean temporales, aspectuales, discursivas o evidenciales (Escandell-Vidal, 2010, 2014, 2021, 2022). Así, hemos sostenido que el PC y el PS de las lenguas románicas poseen, respectivamente, una base semántica común a todas, completada por un rasgo semántico con realizaciones paramétricas alternativas propias de cada variedad lingüística.

Nuestra propuesta es compatible con la idea de una semántica procedimental rígida pero infraespecificada, susceptible de enriquecimiento contextual, a la vez que evita

recurrir a nociones poco definidas (y no falsables) como la de «rutinas pragmáticas» que se encuentran en otras aproximaciones pragmáticas.

El conjunto de rasgos y opciones paramétricas que proponemos explica los tipos de interpretaciones posibles en cada variedad lingüística. Una idea que se desprende de la observación del uso de la lengua es que, desde un punto de vista sincrónico, no hay necesariamente una correlación evidente entre la morfología verbal del PC y los tipos de interpretaciones que permite en cada variedad. Ciertamente, la relación que el PC establece entre la situación evaluada y la situación de discurso tiene su origen en la presencia de un auxiliar en presente de indicativo. Sin embargo, la competencia de esta forma verbal con la del PS ha dado lugar a soluciones diferentes en cada variedad lingüística, hasta el punto de que en algunas variedades la intersección entre SE y SD puede estar vacía, lo que tiende a difuminar su diferenciación funcional con el PS.

Así, los significados lingüísticos reciben sistemáticamente enriquecimiento contextual, y este desarrolla en direcciones diferentes las potencialidades expresivas de la forma. Establecer el papel de la semántica sin dejar de lado el de la pragmática ha sido nuestro empeño. En los estudios sobre el PC y el PS, las perspectivas y los enfoques descriptivos han sido muy variados. Una aproximación procedimental como la presente puede, tal vez, arrojar nueva luz sobre la manera en que la semántica gramatical y léxica, el conocimiento del mundo y los procesos cognitivos interactúan en la interpretación de las formas, en una sola lengua y más allá de esta.

REFERENCIAS

- Abeillé, A. y Godard, D. (dir.) (2021). *La grande grammaire du français*. París: Actes Sud.
- Aikhenvald, A. (2004). *Evidentiality*. Oxford: Oxford University Press.
- Amenós Pons, J. (2010a). *Los tiempos de pasado del español y el francés: semántica, pragmática y aprendizaje de E/LE. Perspectivas desde la Teoría de la Relevancia*. Tesis doctoral. Madrid: UNED.
- Amenós Pons, J. (2010b). Telling a story in French and in Spanish: Past tenses and temporal relations revisited. En I. Witczak-Plisiecka (Ed.), *Pragmatic perspectives on language and linguistics 2009, vol. 1: Speech Actions*, pp. 395-424. Cambridge: Cambridge Scholars Publishing.
- Amenós Pons, J. (2011). Cross-linguistic variation in procedural expressions: semantics and pragmatics. En V. Escandell-Vidal, M. Leonetti y A. Ahern (Eds.), *Procedural Meaning: Problems and Perspectives*, pp. 233-266. Leiden: Brill.
DOI: https://doi.org/10.1163/9780857240941_011
- Amenós Pons, J. (2020). Eventos y situaciones. La referencia temporal. En V. Escandell-Vidal, J. Amenós Pons y A. Ahern (Eds.), *Pragmática*, pp. 166-189. Madrid: Akal.
- Azpiazu, S. (2016). Evidencialidad en el pretérito perfecto compuesto del español. Revisión y propuesta. En R. González Ruiz, D. Izquierdo Alegría y Ó. Loureda Llamas (Eds.),

- La evidencialidad en español: Teoría y descripción*, pp. 303-325. Madrid / Frankfurt del Meno: Iberoamericana / Vervuert.
- Azpiazu, S. (2019). *La composicionalidad temporal del pretérito compuesto en español*. Berlín: De Gruyter. DOI: <https://doi.org/10.1515/9783110633658>
- Bello, A. (1847). *Gramática de la lengua castellana destinada al uso de los americanos*. Santiago de Chile: Imprenta del Progreso. (Edición manejada: Madrid, Edaf 1984).
- Benveniste, E. (1966). *Problèmes de linguistique générale 1*. París: Gallimard.
- Blakemore, D. (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.
- Borik, O. (2006). *Aspect and Reference Time*. Oxford: Oxford University Press.
- Bres, J. (2003). Temps verbal, aspect et point de vue : de la langue au discours, *Cahiers de praxématique 41*, 55-84.
- Bres, J. (2007). Et plus si affinités... Des liaisons entre les instructions du plus-que-parfait et les relations d'ordre, *Cahiers Chronos 18*, 139-157.
- Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.
- Corblin, F.; de Swart, Henriette (Coord.) (2004). *Handbook of French semantics*. Stanford: CSLI Publications.
- De Swart, H. (2000). Tense, aspect and coercion in a cross-linguistic perspective. En M. Butt and T. H. King (Eds.), *Proceedings of the Berkeley Formal Grammar conference University of California, Berkeley*, pp. 1-20. Stanford: CSLI Publications.
- De Swart, H. y Molendijk, A. (2002). Le passé composé narratif: une analyse discursive de *L'étranger* de Camus. En B. Laca (Ed.), *Temps et aspect : de la morphologie à l'interprétation*, pp. 193-211. París: Presses Universitaires de Vincennes.
- Delgado-Díaz, G. (2021). *The Evolution of Spanish Past Forms*. Londres/Nueva York: Routledge.
- Depraetere, I. (1998). On the resultative character of present perfect sentences, *Journal of Pragmatics 29*, 597-613.
- Dumont, J. (2013). Another look in the present perfect in an Andean variety of Spanish: Grammaticalization and evidentiality in Quiteño Spanish. En J. A. Cabrelli (Ed.), *Selected Proceedings of the 16th Spanish Linguistics Symposium*, pp. 279-291. Somerville, MA: Cascadilla Proceedings Project.
- Escandell-Vidal, M.V. (2010). Futuro y evidencialidad, *Anuario de Lingüística Hispánica XXVI*, 9-34.
- Escandell-Vidal, M. V. (2014). Evidential futures: The case of Spanish. En P. de Brabanter, M. Kissine, y S. Sharifzadeh (Eds.), *Future times, future tenses*, pp. 219-247). Oxford: Oxford University Press.
- DOI: <https://doi.org/10.1093/acprof:oso/9780199679157.001.0001>
- Escandell-Vidal, M. V. (2019a). El futuro del español. Sistema natural frente a usos cultivados, *Verba Hispánica 26*, 15-33. DOI: <https://doi.org/10.4312/vh.26.1>
- Escandell-Vidal, M. V. (2019b). Evidential implicatures and mismatch resolution. En R. Carston, B. Clark y K. Scott (Eds.), *Relevance, Pragmatics and Interpretation*, pp. 66-79. Cambridge: Cambridge University Press.
- DOI: <https://doi.org/10.1017/9781108290593>
- Escandell-Vidal, M. V. (2021). The semantics of the simple future in Romance: Core meaning and parametric variation. In Baranzini, L. and De Saussure, L. (Ed.) *Aspects of Tenses, Modality and Evidentiality, Cahiers Chronos 31*, Leiden: Brill.

- Escandell-Vidal, M. V. (2022). Facetas del imperfecto: del significado a las interpretaciones. En M. Martínez-Atienza de Dios (Ed.), *En torno a la delimitación de determinadas categorías lingüísticas*, pp. 159-182. Berlín: De Gruyter.
DOI: <https://doi.org/10.1515/9783110767834>
- Escandell-Vidal, M. V. (en prensa). Understanding the Spanish imperfecto. Temporal dimensions and evidential effects. En J. Egetenmeyer, S. Dessì-Schmid y M. Becker (Eds.), *Time, Aspect and Discourse Structure*. Berlín: De Gruyter.
- Escandell-Vidal, M. V.; Leonetti, M. (2000). Categorías funcionales y semántica procedural. En M. Martínez, et al. (Coord.), *Cien años de investigación semántica: De Michel Bréal a la actualidad, vol. 1*, pp. 363-378. Madrid: Ediciones Clásicas.
- Escandell-Vidal, M. V. y Leonetti, M. (2011). On the rigidity of procedural meaning. En M.V. Escandell-Vidal, M. Leonetti y A. Ahern (Eds.), *Procedural meaning: Problems and perspectives*, pp. 81-103. Leiden: Brill.
DOI: https://doi.org/10.1163/9780857240941_005
- Escandell-Vidal, M. V.; Leonetti, M.; Ahern, A. (Eds.) (2011). *Procedural Meaning: Problems and Perspectives*. Bingley: Emerald.
- Escobar, A. M. (1997). Contrastive and innovative uses of the present perfect and the preterite in Spanish in contact with Quechua. *Hispania* 80, 859-870.
DOI: <https://doi.org/10.2307/345107>
- Kempas, I. (2006). *Estudio sobre el uso del pretérito perfecto prehodierno en el español peninsular y en comparación con la variedad del español argentino hablada en Santiago del Estero*. Tesis doctoral. Helsinki: Universidad de Helsinki.
- Labeau, E. (2022). *The decline of the passé simple*. Leiden: Brill.
- Leonetti, M. (2022). Sobre los mecanismos de competición. *Revista Española de Lingüística* 52/2, 47-84. DOI: <https://doi.org/10.31810/rsel.52.2.3>
- Mateus, M. H. et al. (2003). *Gramática da Língua Portuguesa*. Lisboa: Caminho. 6ª edición.
- Moeschler, J. (1994). Anaphore et déixis temporelles. Sémantique et pragmatique de la référence temporelle. En J. Moeschler, et al., *Langage et pertinence. Référence temporelle, anaphore, connecteurs et métaphore*, pp. 39-105. Nancy: Presses Universitaires de Nancy.
- Moeschler, J. (1998b). Pragmatique de la référence temporelle. En J. Moeschler (Ed.), *Le temps des événements*, pp. 157-180. París: Kimé.
- Moreno Cabrera, J. C. (2013). *Cuestiones clave de la lingüística*. Madrid: Síntesis.
- Nicolle, S. (1997). A Relevance-theoretic account of *be going to*. *Journal of Linguistics* 33, 355-377.
- Nicolle, S. (1998). A relevance theory perspective on grammaticalization. *Cognitive Linguistics* 9-1, 1-35.
- Nishiyama, A.; Koenig, J-P. (2010). What is a perfect state? *Language* 86, 611-646. DOI: <https://doi.org/10.1353/LAN.2010.0014>
- Parrinha, S. R. (2014). Valores pragmáticos y discursivos en el uso del pretérito perfecto compuesto en el español de Buenos Aires. En S. Azpiazu (Ed.), *Formas simples y compuestas del pasado en el verbo español*, pp. 103-115. Lugo: Axac.
- Real Academia Española y Asociación de Academias de la Lengua Española (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.

- Reichenbach, H. (1948). *Elements of Symbolic Logic*. Nueva York y Londres: The Free Press / Macmillan.
- Rodríguez Louro, C. (2009). La referencia indefinida y la expresión de pasado en el español rioplatense argentino. En L. Colantoni y C. Rodríguez Louro (Eds.), *Perspectivas teóricas y experimentales sobre el español de la Argentina*, pp. 283-297. Madrid / Frankfurt del Meno: Iberoamericana / Vervuert.
- Saussure, L. de (2003). *Temps et pertinence. Éléments de pragmatique cognitive du temps*. Bruselas: De Boeck / Duculot.
- Schaden, G. (2009a). Present perfects compete. *Linguistics and Philosophy* 32 (2), 115-141. DOI: <https://doi.org/10.1007/s10988-009-9056-3>
- Schaden, G. (2009b). *Composés et surcomposés. Le 'parfait' en français, allemand, anglais et espagnol*. París: L'Harmattan.
- Smith, C. (2003). *Modes of discourse*. Cambridge: Cambridge University Press.
DOI: <https://doi.org/10.1017/CBO9780511615108>
- Smith, N. (1990). Observations on the pragmatics of tense. *UCL Working Papers in Linguistics* 2: 82-94. <https://www.phon.ucl.ac.uk/publications/WPL/90papers/UCLWPL2%205%20Smith.pdf>
- Soto, G. (2014). El pretérito perfecto compuesto en el español estándar de nueve capitales americanas: frecuencia, subjetivización y deriva aorística. En S. Azpiazu (Ed.), *Formas simples y compuestas del pasado en el verbo español*, pp. 131-146. Lugo: Axac.
- Speas, P. (2008). On the syntax and semantics of evidentials. *Language and Linguistics Compass* 2(5), 940-965. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00069.x>
- Sperber, D. y Wilson, D. (1995). *Relevance. Communication and cognition*. Oxford: Blackwell (2a edición).
- Sperber, D. y Wilson, D. (1993). Pragmatics and time. *UCL Working Papers in Linguistics* 5, 277-298.
- Squartini, M.; Bertinetto, P. M. (2000). The simple and compound past in Romance languages. En Ö. Dhal (Ed), *Tense and aspect in the languages of Europe*, pp. 403-439. Berlín y Nueva York: De Gruyter. DOI: <https://doi.org/10.1515/9783110197099>
- Unger, C. (2006). *Genre, relevance and global coherence*. Nueva York: Palgrave.
- Veiga, A. (2014). Diacronía de *he cantado/canté* en el sistema verbal español. Subsistemas y variantes. En J.L. Ramírez Luengo, E.P. Velásquez Upegui (Eds.), *La historia del español hoy. Estudios y perspectivas*, pp. 151-179. Lugo: Axac.
- Wigger, L.-G. (2005). *Die Entwicklungsgeschichte der romanischen Vergangenheitstempora am Beispiel des Pretérito Perfeito Composto im Portugiesischen*. Tübinga: Universidad de Tübinga.
- Wilson, D. y Sperber, D. (1993). Linguistic form and relevance. *Lingua* 90, 1/2, 1-25.

José Amenós Pons

Dpto. Lengua Española y Teoría de la Literatura

Facultad de Filología, Edif. D

Despacho 363 (1.ª planta)

Universidad Complutense de Madrid

28040 Madrid

LA INTERRELACIÓN ENTRE EVIDENCIALIDAD Y ERGATIVIDAD (ESCINDIDA): UNA COMPARACIÓN TIPOLOGICA ENTRE EL ESPAÑOL, LAS LENGUAS ROMANCES Y OTRAS FAMILIAS LINGÜÍSTICAS

Víctor LARA BERMEJO¹
Universidad de Cádiz

Resumen

El español exhibe un sistema morfológico de evidencialidad a través de su futuro morfológico y condicional. Los trabajos más recientes apuntan al oeste de la península ibérica como epicentro de esta particularidad y un aumento exponencial en los últimos cien años de los usos evidenciales frente a los temporales o modales. En este artículo, queremos demostrar la correlación estrecha entre el nacimiento de la evidencialidad y los patrones ergativos que, no solo explican el origen geográfico del español, sino la evolución en otras variedades romances y en otras familias lingüísticas.

Palabras clave: evidencialidad; español; iberorromance; ergatividad; futuro; condicional

THE INTERPLAY BETWEEN EVIDENTIALITY AND (SPLIT) ERGATIVITY: A TYPOLOGICAL COMPARISON OF SPANISH, THE ROMANCE LANGUAGES AND OTHER LINGUISTIC BRANCHES

Abstract

Spanish exhibits a morphological system of evidentiality by means of its morphological future and conditional tense. The latest studies point out that the western part of the Iberian Peninsula is the epicentre of this phenomenon, which has increased exponentially in the last century to the detriment of temporal and modal values. In this article, I aim to demonstrate the narrow correlation between the birth of evidentiality and ergative patterns

1. victor.lara@uca.es.  <https://orcid.org/0000-0002-1068-8553>

that, not only justify the geographical origin for Spanish, but they also explain the evolution of evidentiality elsewhere in the Romance spectrum and in other linguistic families.

Keywords: evidentiality; Spanish; Ibero-Romance; ergativity; future, conditional

RECIBIDO: 28/11/2022

APROBADO: 11/06/2023

1. INTRODUCCIÓN²

La evidencialidad se ha convertido en los últimos tiempos en un tema prolífico en estudios que, en numerosas ocasiones, suscita controversias sobre su propia etiqueta. No todos los autores concuerdan en determinar qué es evidencial y no todos coinciden en la taxonomía que se puede aplicar a las lenguas romances a cuenta de sus valores evidenciales. Por un lado, Squartini (2001) establece que las lenguas románicas descritas en la tabla 1 cuentan con un sistema de evidencialidad, porque el futuro morfológico (FM) y el condicional (COND) pueden servir para indicar inferencia y reportatividad.

	Portugués	Español	Francés	Italiano
FM	I / R	I	I	I
COND	I / R	I / (R)	I / R	R

Tabla 1: Evidencialidad en los romances mediante el FM y el COND [I=inferencial / R=reportativo] (Squartini, 2001)

Por otro, Aikhenvald (2004) alega que, aunque toda lengua tiene estrategias de evidencialidad, ya que siempre hay maneras de expresar la fuente de información de un enunciado, la evidencialidad *per se* es un sistema flexivo que marca de manera primaria, hegemónica o frecuente fuente de información frente a otro tipo de lecturas que ese mismo paradigma pueda denotar. Escandell-Vidal (2014, 2021), por su parte, argumenta que el significado codificado por el FM, en español, se caracteriza mejor en términos evidenciales como presente anti experiencial, lo que admite interpretaciones tanto conjeturales como temporales. En última instancia, Lara Bermejo (2021a) indica que el español peninsular cuenta con un paradigma evidencial encarnado en el FM y el COND, el primero de los cuales hace referencia a una conjetura, mientras que el segundo, además de servir para una inferencia

2. Este artículo se enmarca en el proyecto *Evidencialidad, subjetivización y perspectivización en las interfaces de la lengua*, con referencia PID2019-104405GB-I00, financiado por el Ministerio de Ciencia e Innovación (España).

referida al pasado, surge para la reportatividad. A su vez, el portugués presenta la misma casuística, con el añadido de que su FM también puede emplearse para la reportatividad, mientras que el gallego tan solo exhibe un patrón de evidencialidad mediante el FM (pero no el COND) y el catalán oscila según su variedad, puesto que la parte valenciana sí ha desarrollado evidencialidad a través del FM, mientras que Cataluña y Baleares carecen de evidencialidad. La teoría en la que se fundamenta el autor es la propugnada por Aikhenvald (2004).

Además de los trabajos sobre variedades romances, la evidencialidad ha sido analizada pormenorizadamente en lenguas de la rama túrquica (véase Johanson, 2018), urálica y altaica (véase Skribnik, 1998; Skribnik y Kehayov, 2018), tibetana y mongoles (véase Brosig y Skribnik, 2018; DeLancey, 2018) o de la zona norteamericana (véase Thornes, 2018). En todos los estudios que se llevan a cabo, se da cuenta del tipo de evidencialidad que surge en las lenguas correspondientes, del origen de la estrategia que ha resultado en un paradigma flexivo evidencial o de hasta qué punto la formación de esta posibilidad se ha debido al contacto lingüístico. Sin embargo, no existen investigaciones que traten de esclarecer qué propicia que una variedad dada, en un momento determinado, resuelva crear un paradigma flexivo que connote fuente de información. En otras palabras, no sabemos si la aparición de un sistema evidencial (tal cual lo entiende Aikhenvald) responde a unos requisitos previos de índole lingüística, los cuales fomenten su formación en un lugar y tiempo concretos.

Los análisis de Lara Bermejo (2021a) indican que el sistema morfológico evidencial del español nace en el oeste de la península ibérica, el cual lo va contagiando al resto de dicha variedad y a sus lenguas romances vecinas, siempre dentro del entorno peninsular. En otro trabajo, Lara Bermejo (2021b) defiende que el italiano creó su paradigma evidencial a partir de las variedades meridionales. En lo que concierne al comportamiento universal, los estudios sobre evidencialidad se concentran en lenguas que muestran un patrón de alineamiento ergativo o de ergatividad escindida y, a tenor de lo que se sabe del español peninsular occidental y de las variedades italianas sureñas en su comportamiento dialectal, encontramos también numerosos modelos que se alinean con tendencias ergativas.

La ergatividad, al igual que la evidencialidad, es un concepto muy debatido en lingüística, ya que hay autores que la restringen a un rasgo meramente flexivo y morfológico (Dixon, 1994), mientras que otros amplían la definición a estrategias sintácticas e, incluso, discursivas (McGregor, 2009). En principio, la ergatividad es el alineamiento morfológico basado en diferenciar el sujeto de una transitiva

(el cual recibe caso ergativo) frente al sujeto de una intransitiva, cuyo caso (el absolutivo) coincide con el del objeto directo de una transitiva. Sin embargo, como explicaremos con más detalle en epígrafes posteriores, esta taxonomía presenta escisiones en función de las características semánticas del sujeto de una intransitiva, puesto que, si este es agentivo y volitivo, puede llegar a declinarse en ergativo, mientras que, si es no agentivo y no volitivo, opta por el absolutivo. En definitiva, las lenguas con ergatividad escindida suelen comportarse así al atender a la oposición inergativo versus inacusativo.

A la luz de esta simultaneidad (patrones ergativos y paradigma evidencial), en este artículo, pretendemos desvelar si dicha coincidencia es tan solo aparente o, efectivamente, hay una correlación clara entre ergatividad (escindida) y evidencialidad. Nuestra hipótesis parte de la afirmación de que la existencia de una es consecuencia de la otra y que, por tanto, no es casual que la evidencialidad en español haya nacido en el área geográfica donde lo ha hecho y en el momento en que se ha producido, y lo mismo podemos argüir para el italiano u otras zonas de otras variedades distintas a las romances o indoeuropeas. Con el fin de comprobar la veracidad de nuestra hipótesis, a continuación, desarrollaremos el comportamiento de ambas categorías para, más tarde, estudiar las tendencias lingüísticas del español, las lenguas romances y otras familias. Para ello, en 2, nos detendremos en analizar el fenómeno de la evidencialidad en términos generales; en 3, nos enfocaremos en estudiar la ergatividad o sus sistemas escindidos desde una perspectiva tipológica; en 4, profundizaremos en el interrogante que explique las motivaciones existentes para que concurren ambos factores (ergatividad y evidencialidad) y qué tres posibilidades se dan para la creación de este fenómeno; en 5, analizamos el caso del español y las demás lenguas romances, mientras que, en 6, culminaremos con las conclusiones.

2. EVIDENCIALIDAD

En la introducción, anunciábamos que la evidencialidad no suscita consenso, ya que, dependiendo del autor que consultemos, la definición es más abarcadora o más restrictiva. Mientras que, para Aikhenvald (2004), la evidencialidad es un sistema morfológico que primariamente, sistemáticamente o unívocamente marca fuente de información, Palmer (1986) denomina evidencialidad a cualquier recurso que configure fuente de información, ya sea mediante adverbios como *aparentemente*, circunloquios, verbos modales o el empleo de ciertos tiempos verbales que

puedan connotar inferencia o reportatividad, aunque dicho uso sea secundario o esporádico. En este artículo, seguiremos los postulados de Aikhenvald (2004, 2018a).

Si bien no toda lengua posee un sistema flexivo que haga referencia a la fuente de información, aquellas que sí lo han desarrollado no exhiben un patrón homogéneo: o su paradigma es defectivo, al surgir para cierta temporalidad o personas gramaticales, o es opcional, o priman la marcación de un tipo de fuente información frente a todo lo demás. A todo ello hay que agregar el hecho de que la marcación morfológica de la evidencialidad en una lengua dada puede clasificarse en diferentes sistemas, atendiendo al número de tipos de fuente de información que se configuren: podemos hallar sistemas que discernan dos opciones, tres, cuatro y, en menor medida, cinco o más. Sin embargo, en cada uno de dichos subsistemas encontramos cierta coherencia, ya que siempre hay un interés especial por marcar una información que no sea de primera mano, como la reportatividad, y, si las opciones se multiplican, el énfasis no solo se aplica a aquello que no es de primera mano, sino a la dicotomía entre evidencia directa e indirecta: lo que uno ve o percibe auditivamente frente a lo que infiere, asume o concluye por su conocimiento del mundo u otro proceso mental.

El hecho de que haya lenguas que carezcan de un sistema de evidencialidad no significa que no lo puedan constituir, ya que, como la propia Aikhenvald (2004, 2018a) demuestra, existen varios procesos que tienden a convertir una serie de constituyentes en morfemas evidenciales, a saber: gramaticalización de ciertos verbos, cópulas, tiempos y modos verbales específicos, marcadores locativos y deícticos e, incluso, estrategias evidenciales que se hubieran fosilizado. La autora afirma que el futuro y el condicional son propensos a ser reanalizados como marcas de evidencialidad e incluso ejemplifica el condicional en francés, aunque no etiqueta esta lengua como poseedora de evidencialidad; sobre esta particularidad, volveremos más adelante. En última instancia, Aikhenvald (2004, 2018b) también advierte de que la evidencialidad puede deberse al contacto lingüístico.

A pesar de que los estudios de la autora a la que recurrimos son bastante célebres, ningún autor que ha tratado sobre el tema aplicado a las lenguas romances ha profundizado en la taxonomía de dicho paradigma, indicando si las lenguas romances poseen un sistema que discerna entre dos, tres o más tipos de fuente. En cualquier caso, como defenderemos posteriormente, las afirmaciones que se han hecho sobre la existencia de la evidencialidad en el espectro romance, en especial el francés, no se ajustan a lo que aquí entendemos por tal.

En el plano geolingüístico, contamos con varios mapas que tratan de ejemplificar gráficamente la extensión de este fenómeno y su idiosincrasia. Observemos el mapa de la figura 1.

El mapa que se presenta en la figura 1 indica que la codificación de la evidencialidad en las lenguas del mundo se configura mediante cinco estrategias fundamentales: afijos verbales (lo más usual), la reutilización de un tiempo verbal (estrategia que se da sobre todo en los Balcanes, Oriente Medio y zonas del Pacífico), partículas separadas (la

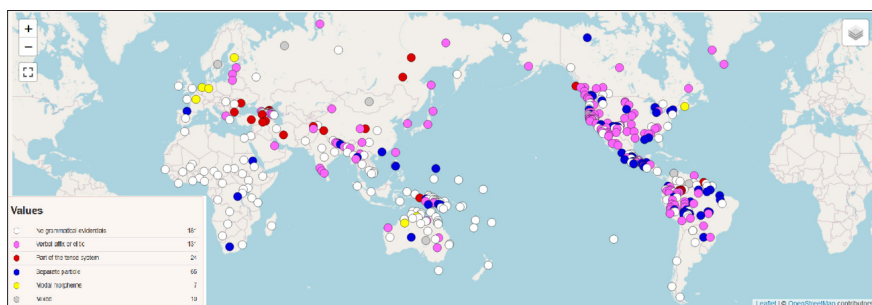


Figura 1. Evidencialidad en el mundo (De Haan, 2013)

alternativa más frecuente después de los morfemas), los morfemas modales (finés y lenguas germánicas, entre otras), y una mezcla de las anteriores. Asimismo, los puntos en blanco indican que dichas lenguas no poseen marcación morfológica evidencial.

El mapa de la figura 1 presenta varios problemas. El primero de ellos es englobar verbos modales como morfemas de modo. De acuerdo con De Haan (2013), oraciones como (1-3) del holandés, el alemán y el francés (sacadas de su trabajo), respectivamente, exhiben una estrategia morfológica de evidencialidad mediante el verbo modal, mientras que las oraciones (4-5) del inglés y el español, respectivamente, no se contemplan como estrategias morfológicas de evidencialidad, a pesar de presentar exactamente la misma alternativa y el mismo verbo.

- (1) *Het moet een goede film zijn*
Esta deber.3SG.PRS. una buena película ser.INF.
(‘Debe de ser una buena película’)
- (2) *Er muss beim Arzt sein*
Él deber.3SG.PRS. en+el médico ser.INF.
(‘Debe de estar en el médico’)
- (3) *Il doit être à Paris maintenant*
Él deber.3SG.PRS. estar.INF. en París ahora
(‘Debe de estar en París ahora’)

- (4) *It must be 8 o'clock in Tokyo*
 Ello deber.3SG.PRS. ser.1NF. 8 en punto en Tokio
 ('Deben de ser las ocho en Tokio')
- (5) Debe de haberse olvidado lo que le he pedido

El segundo de ellos es precisamente la contradicción que acabamos de soslayar, ya que los mismos verbos modales constituyen una prueba de la existencia de evidencialidad en ciertas lenguas, pero en otras no. Además, si bien De Haan (2013) aplica una perspectiva mucho más abarcadora en su concepción de la evidencialidad, no apuntala lo defendido por Squartini (2001) para las lenguas romances, pues el primero no entiende que el español tenga evidencialidad (el italiano y el portugués ni siquiera están incluidos en la muestra), pero el segundo defiende justo lo contrario. En último lugar, la misma distribución de las estrategias contraviene lo analizado por otros autores: De Haan (2013) indica que el francés expresa evidencialidad mediante verbos modales, mientras que Squartini (2001) insiste en que son los tiempos verbales (en concreto, el futuro y el condicional) los que se arrojan dicha facultad.

A pesar de los problemas que, en nuestra opinión, plantea el mapa 1, la distribución cartográfica es prácticamente coincidente con la que proporciona la célebre monografía de Aikhenvald (2004), reproducida en el mapa de la figura 2.

El mapa de la figura 2 muestra que los sistemas de evidencialidad, reiteramos, tal cual los concibe la autora, es decir, paradigmas morfológicos que marcan fuente de información de manera frecuente (independientemente de que también puedan

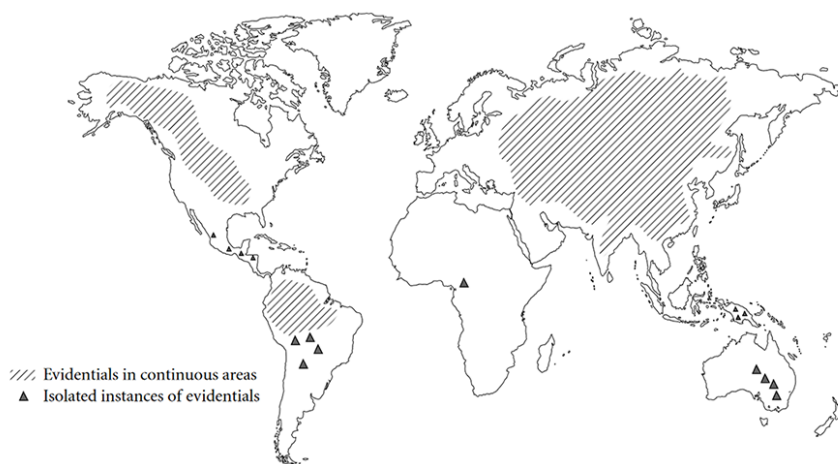


Figura 2. Evidencialidad morfológica en el mundo (Aikhenvald, 2004)

expresar otros valores), se difunden por el Cáucaso, Asia Central, parte de Asia Pacífico, el interior de Norteamérica y la zona amazónica de Sudamérica. De manera residual, encontramos islas de evidencialidad en regiones de Australia, así como en las inmediaciones del golfo de Guinea y el espacio que media entre la Amazonía y el Cono Sur. El cotejo de los mapas que aparecen en las figuras 1 y 2 refleja que la evidencialidad no se da en todas las lenguas, aunque es susceptible de hacerlo. Igualmente, las incongruencias entre ambos vuelven a poner de relieve la falta de unanimidad a la hora de etiquetar qué es evidencialidad, pero estas disensiones no son óbice para comprobar que este fenómeno se disemina por las lenguas túrquicas, fino-urálicas, altaicas, tibetanas y de otras familias no indoeuropeas que se concentran en Norteamérica, la Amazonía, el Sudeste asiático, Australia, así como en el euskera. El mayor punto de fricción, en realidad, tiene que ver con las lenguas romances.

3. ERGATIVIDAD

La ergatividad es entendida como la marcación diferencial del sujeto de una transitiva (A) en contraposición a la del sujeto de una intransitiva (S) y el objeto de una transitiva (O), los cuales reciben la misma marca (Dixon, 1994). En otras palabras, la ergatividad promociona el rasgo de agentividad (casi siempre encarnado en un ente humano), de volición o de control, frente a otros parámetros semánticos. Sin embargo, las lenguas ergativas suelen transitar hacia modelos nominativo-acusativos, presentando una fase intermedia en la que exhiben comportamientos de ambas opciones. Asimismo, Dixon (1994) afirma la escasa probabilidad de hallar lenguas puramente ergativas o puramente nominativas y la relativa frecuencia de encontrar idiomas con un patrón escindido.

El euskera, sin ir más lejos, se suele catalogar como una lengua ergativo-absolutiva, pero la configuración de sus argumentos promociona la escisión en función de los rasgos de control, volición, animacidad o agentividad, si bien esta posibilidad no está extendida en todo el idioma y depende, en gran parte, de la variedad dialectal (Aldai, 2009) (véase 6-7, de Hualde y Ortiz de Urbina, 2003, pp. 389-390).

- (6) *Jonek korritu du*
Jon.ERG. correr AUX.
(‘Jon corrió’)
- (7) *Xabier atera da*
Xabier.ABS. salir AUX.
(‘Xabier salió’)

Las ocurrencias (6-7) muestran el trato distinguidor del sujeto de un verbo intransitivo, poniendo de relieve la escisión basada en los rasgos semánticos mencionados. En (6), el sujeto se construye en ergativo porque el evento de *correr* suele ser volitivo, controlado y agentivo, mientras que (7) opta por un sujeto en absolutivo porque el evento de *salir* no necesariamente implica los factores anteriores. En otras palabras, el vasco diferencia en la marcación de caso los inergativos de los inacusativos.

La ergatividad ha estado tradicionalmente asociada con un comportamiento puramente morfológico, pero la bibliografía ha ampliado la definición de esta alternativa poniendo de relieve el hecho de que la sintaxis también puede atender a las etiquetas de volitivo, controlable o agentivo. Uno de los defensores de esta visión más abarcadora es McGregor (2009), quien asegura que la ergatividad puede materializarse no solo mediante la morfología flexiva en el verbo, el nombre o el pronombre, sino a través del orden de constituyentes. Para el autor, un patrón ergativo es aquel que aúna el paciente y el autor bajo un mismo comportamiento, a diferencia del agente, que recibe otra clase de trato. En esta clasificación hallaríamos el orden de palabras de los argumentos en los verbos intransitivos y transitivos, ya que el sujeto gramatical del inacusativo (paciente) suele comportarse como el objeto de una transitiva; en cambio, el agente de una transitiva suele ser igual que el agente de una intransitiva (8-11).

- (8) Viene un coche
- (9) Ha nacido mi hijo
- (10) El atleta corre (la maratón)
- (11) El vigilante observa a los visitantes

En (8-9), los sujetos sintácticos de *venir* y *nacer* suelen colocarse de manera no marcada en una posición prototípica del objeto, debido a su carencia de volición, agentividad o control. No obstante, aunque (10) también sea intransitiva, independientemente de su versión inergativa o con objeto directo, el sujeto sintáctico coincide con el semántico, por lo que su posición no marcada es siempre la anteposición al verbo, como ocurre con el sujeto de la transitiva en (11).

McGregor (2009) va más allá y postula lo que él denomina ergatividad discursiva. Su configuración atiende a la presentación sistemática del autor o el paciente al principio de la oración como sinónimo de información nueva, frente al agente, que suele interpretarse como información conocida (12-14).

- (12) (A mí) me gusta el chocolate
- (13) Me aburre esa serie

(14) Se me ha vertido el agua

Las oraciones (12-14) reflejan lo que McGregor (2009) ha denominado ergatividad discursiva, ya que en las tres ocasiones el paciente, el experimentante o el autor (que no el agente) de la acción se ha colocado en primera posición en el discurso por ser las entidades humanas. En todos los casos, el sujeto sintáctico se halla en posición de objeto y, aunque (12-13) muestra un experimentante e incluso un paciente, (14) promociona *me* al ser el autor del evento *verter*: papel semántico diferente del agente, al ser el autor el causante de un evento de manera involuntaria (Givón, 2001).

Los ejemplos anteriores certifican además el empleo del dativo para todos esos papeles temáticos, lo cual no es baladí. Este caso suele presentar una alta topicalidad, pues frecuentemente encarna un ente humano, y esa misma característica está detrás de otro patrón ergativo (en su acepción más abarcadora) con el denominado sujeto dativo. Nos detendremos en esta circunstancia en epígrafes posteriores, pero en este momento queremos resaltar la existencia de elementos declinados en dativo, pero con claros comportamientos de sujeto en situaciones de falta de volición, control o agentividad.

En cualquier caso, como hemos advertido previamente, la sistematicidad en la configuración morfológica de la ergatividad es divergente según la lengua, como bien ilustran los mapas 3, 4 y 5 (obsérvese la difusión *grosso modo* coincidente de los patrones ergativos con la difusión de los paradigmas evidenciales del mapa ofrecido en la figura 2).

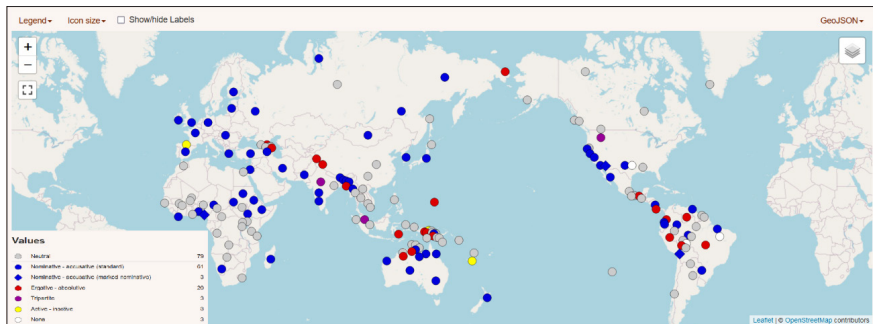


Figura 3. Alineamiento morfológico en pronombres (Comrie, 2013a)

Los mapas ofrecidos en las figuras 3, 4 y 5 revelan los patrones de alineamiento morfológico en función de tres parámetros: la flexión pronominal, la marcación de los sintagmas nominales y la flexión verbal. Aunque no siempre coinciden los

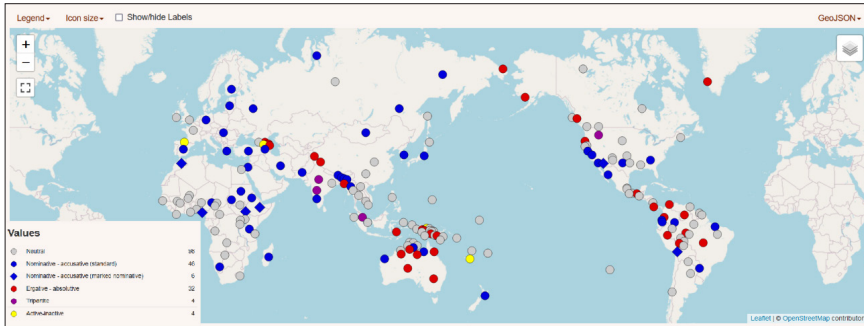


Figura 4. Alineamiento morfológico en sintagmas nominales (Comrie, 2013b)

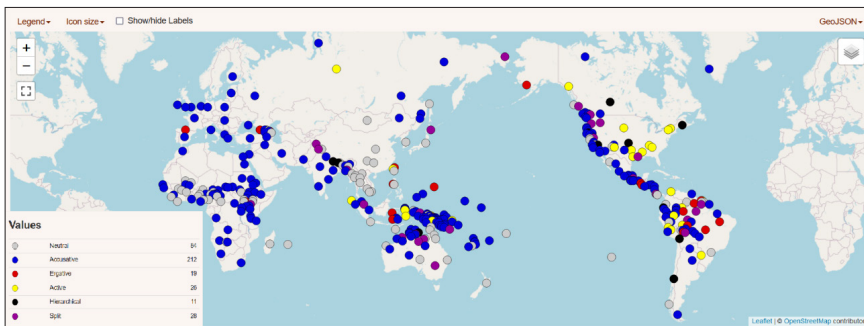


Figura 5. Alineamiento morfológico en la flexión verbal de persona (Siewierska, 2013)

tres factores en una misma lengua, por regla general, aquella que no exhibe un patrón nominativo-acusativo es coherente con esta medida en la mayor parte de contextos lingüísticos cartografiados. Aunque, como hemos avanzado, las ilustraciones sobre la ergatividad dibujan una extensión de dicho fenómeno similar a la de la evidencialidad, estas presentan varios problemas, entre los cuales destaca la no contemplación de particularidades dialectales que puedan contradecir las normas del estándar. En cualquier caso, la ergatividad (escindida) se reparte por gran parte de las variedades indígenas de Norteamérica y la Amazonía, las lenguas tibetanas, el Sudeste asiático, variedades autóctonas de Australia o algunas lenguas túrquicas.

No nos explayaremos más sobre la idiosincrasia de la ergatividad, ya que el objetivo era dar un panorama sucinto de lo que se entiende como tal, pero queremos subrayar la importancia también de la evolución diacrónica, al igual que indicábamos para la evidencialidad, puesto que la (falta de) existencia de patrones ergativos suele deberse a procesos paulatinos en el tiempo que, veremos, han sido cruciales para el fomento de la evidencialidad.

4. EVIDENCIALIDAD Y ERGATIVIDAD

La tesis que sostenemos se fundamenta en el hecho de que la evidencialidad y la ergatividad están estrechamente relacionadas y que la existencia de la primera es, por regla general, consecuencia de la segunda (pero no al revés), salvo en casos de contacto lingüístico. Para determinar esta conexión, a continuación, pasamos a tratar la simultaneidad de ambos fenómenos a nivel tipológico, para más tarde ocuparnos del español y del italiano en particular y, por último, de las demás variedades romances. Asimismo, prestaremos especial atención a la importancia de la diacronía en la evolución de los dos rasgos y veremos cómo los aparentes contraejemplos son, en realidad, una fotografía del desarrollo de dicha diacronía o el resultado del contacto lingüístico.

La conexión entre ergatividad y evidencialidad se basa en el hecho de que ambas características promocionan los rasgos de volición, control, agentividad y animacidad, frente a sus antónimos. Este comportamiento queda claro para el funcionamiento de la ergatividad, ya sea pura o escindida, pero se antoja mucho más sutil para la evidencialidad. Si analizamos los sistemas evidenciales a nivel universal, comprobamos que coinciden con lenguas con alineamientos ergativo-absolutivos; de la misma manera, aquellas lenguas que no entran dentro de esta categoría, pero que sí han desarrollado sistemas evidenciales, con independencia de su complejidad, presentan patrones claramente ergativos que son anteriores a la creación del paradigma evidencial. Por último, hallamos un grupo de lenguas que difieren de las anteriores, porque la existencia de la evidencialidad en ellas se explica por contacto lingüístico. En cualquiera de estas posibilidades, la misma evolución del sistema evidencial tiende a favorecer los rasgos de (falta de) volición, control, humano o agentivo, por lo que su configuración tampoco es el resultado de un proceso arbitrario. Analicemos las tres opciones.

4.1. *Lenguas con ergatividad y evidencialidad morfológicas*

En el primer caso, los mapas reproducidos anteriormente son una buena muestra de esta coincidencia que, reiteramos, no es gratuita. No vamos a repetir aquí todo lo que hemos esbozado en el apartado sobre evidencialidad, ya que hemos hecho referencia a familias ergativas con sistemas evidenciales, pero sí queremos centrarnos en su configuración y su desarrollo. La propia tipología estudiada por Aikhenvald (2004) incide en una serie de jerarquías y estadios implicativos como los siguientes: si una lengua tiene un sistema binario, este se organiza de acuerdo con la dicotomía

de evidencia de primera mano frente a la que no lo es o, en una subcategoría, evidencia directa frente a indirecta. Aquellas variedades que solo marcan un único tipo de fuente de información lo hacen para la que no resulta de primera mano o no es directa. También es común la marcación de información reportativa frente a una marca cero para lo demás. No obstante, si una lengua dada desarrolla tres distinciones evidenciales, estas comúnmente se atienen a los parámetros de directa, inferida y reportativa, o visual frente a inferida y no visual, o incluso visual frente a inferida y reportativa. En cualquiera de los casos, hay una especial inclinación para promocionar aquella evidencia que se ha obtenido de manera menos controlable o volitiva en comparación con otros mecanismos de obtención de dicha información. Además, en estos sistemas, es incluso probable la distinción meramente de fuentes de información de tercera mano (reportativa o citativa), incidiendo en la carencia de control sobre la fuente.

Lo mismo puede decirse de los sistemas con cuatro a más distinciones, ya que en todos ellos existe una tendencia a discriminar la fuente directa de la indirecta, o la fuente visual frente a la incontrolable, como la inferencia. En otras palabras, surge un interés especial en establecer de manera clara si aquello que el hablante expresa lo ha obtenido de una forma controlable, volitiva y, hasta cierto punto, agentiva. La mayor o menor nitidez en las distinciones evidenciales se concentran en esta misma casuística, puesto que, eventualmente, se empieza a discernir entre grados de volición (no es lo mismo lo visual, que lo auditivo o lo olfativo) o grados de control (no es lo mismo la inferencia de uno mismo que lo que otra persona dice).

Como es sabido, toda lengua tiene estrategias de evidencialidad, es decir, todo idioma es capaz de matizar de dónde procede la información si el hablante considera oportuno explicitar esta posibilidad. Igualmente, toda lengua tiene estrategias de corte ergativo, en el sentido de que puede clarificar el nivel de control, volición o agentividad de una acción determinada. Sin embargo, a nuestro juicio, la gramaticalización de dichas estrategias y su eventual obligatoriedad en el discurso es un salto cualitativo, porque presupone una cada vez mayor importancia para establecer por defecto los valores semánticos que están detrás de su funcionamiento. En el caso de la ergatividad, para marcar la volición, el control y la agentividad de cualquier acción; en el caso de la evidencialidad, para marcar el grado de volición, control y agentividad en la manera de obtener una información. La distribución evidencial sugiere esto mismo en las lenguas puramente o parcialmente ergativas, pero en aquellas que no lo son, pero han creado mecanismos ergativos, la dicotomía se establece de forma inexorable bajo los mismos parámetros: es el caso del alemán o el español peninsular. Idéntica

elección hacen las variedades que, sin contar con estrategias ergativas, adoptan un paradigma evidencial por contacto lingüístico: es el caso de los iberorromances (salvo el español) o el *Sprachbund* de los Balcanes.

En la clase que nos ocupa, las lenguas que son morfológicamente ergativas o tienen un alineamiento escindido, el factor volitivo, de control o agentivo es tan relevante que incluso podemos hallar especializaciones egofóricas. Estas se refieren a la fuente de información que tiene que ver con los procesos del hablante, en comparación con la (no) marcación de evidenciales para la segunda y tercera personas. En takuu, la marca egofórica surge para actos volitivos, mientras que los actos incontrolables y sobrevenidos se consignan mediante otro morfema distinto (Sun, 2018). La egoforidad se da ampliamente en lenguas tibetanas y birmanas, de la zona del Daguestán, ciertas lenguas mongoles y en áreas de Papúa Nueva Guinea (Aikhenvald, 2018a), es decir, en variedades ergativas. En tártaro, la especialización evidencial no solo se opone entre evidencia directa e indirecta, sino si esta ha sido asumida de manera volitiva o no (Forcker, 2018). La direccionalidad que defendemos en este artículo, al admitir que los patrones o alineamientos ergativos promocionan la creación de la evidencialidad se ve también en los trabajos históricos que se han hecho de las lenguas habladas en el Cáucaso. Friedman (2018), por ejemplo, data la implantación del paradigma evidencial en el sur de dicha cadena montañosa en el período medieval, es decir, mucho después de su nacimiento. El recorrido que hace de otras familias turcas, urálicas o indoaryanas corroboran esto mismo, ya que determina el origen temporal de la creación del paradigma evidencial: en todos estos casos, el alineamiento ergativo es idiosincrático de dichas familias y no ha sido un desarrollo a lo largo de los años.

La importancia de los factores de volición, control o agentividad se da hasta en la adquisición de los evidenciales. Fitneva (2018) expone que los hablantes de turco empiezan a reproducir primero los evidenciales directos a los dieciocho meses y, a los veintiuno, los indirectos. Dentro de estos últimos, primero aprenden los referidos a la inferencia y, en última instancia, los reportativos. Lo mismo sucede en búlgaro, donde los nuevos hablantes adoptan primero los evidenciales que marcan inferencia y más tarde los reportativos. El mismo camino se repite en coreano, donde el reportativo en el último en aprenderse, o el japonés, donde se replica este comportamiento.

4.2. *Lenguas con patrones ergativos y evidencialidad morfológica*

Por otro lado, hallamos lenguas o variedades que, sin poder catalogarse como ergativo-absolutivas, presentan evidencialidad por contar con ciertos patrones con un cariz ergativo: es el caso del alemán, ciertos idiomas del Báltico o, de manera incipiente, el griego moderno, entre otros. El alemán presenta verbos modales que, a juicio de De Haan (2013), son prueba de evidencialidad. Ya hemos defendido aquí por qué no contemplamos esta posibilidad para etiquetar una variedad dada como poseedora de evidencialidad. Sin embargo, este mismo idioma sí exhibe una estrategia morfológica que marca fuente de información mediante el estilo indirecto, lo cual, según Aikhenvald (2004), suele ser también una fuente para desarrollar evidencialidad morfológica. A pesar de que el alemán está considerado como lengua nominativo-acusativa e incluso mantiene flexión pronominal y nominal casual coherente con dicha clasificación, presenta comportamientos propios de un alineamiento que incide en el rasgo humano, volitivo o agectivo frente a la función sintáctica. Nos referimos al sujeto dativo.

El dativo como marca de caso suele consignar papeles temáticos relacionados con el rasgo humano (Givón, 2001). Así, surge para materializar el experimentante, el receptor, el poseedor, el beneficiario e incluso el autor de un evento. En consecuencia, el dativo suele promocionarse al inicio del discurso al ser el rasgo humano un factor de máxima topicalidad. Observemos los siguientes ejemplos del español y el italiano (15-18).

- (15) Me gusta la pizza
 (16) Me he roto la pierna
 (17) *Mi piace la pizza*
 Me gusta.3SG.PRS.IND. la pizza
 ('Me gusta la pizza')
 (18) *Mi sono rotto la gamba*
 Me ser.1SG.PRS.IND. romper.PCP. la pierna
 ('Me he roto la pierna')

Las frases (15-18) proporcionan el orden no marcado de verbos con dativo experimentante y tema nominativo o con dativo poseedor y tema nominativo. El elemento que encarna los rasgos de humano se expresa en primer lugar, a pesar de no ser el sujeto sintáctico, sin mayores repercusiones morfológicas en estos idiomas, pero hay lenguas germánicas que sí sufren otro tipo de reajustes en su paradigma casual.

El concepto de sujeto dativo o no canónico alude a un elemento semánticamente humano que, morfológicamente, se marca en dativo, por lo que sintácticamente no es el sujeto, pero en el orden de palabras sí funciona como tal. Las lenguas germánicas son un buen ejemplo de este comportamiento. Eythórsson y Barðdal (2005) arguyen que esta familia lingüística ha exhibido siempre sujetos oblicuos (en especial, dativos), aunque en la actualidad no todas ellas recurran a esta posibilidad con la misma frecuencia.

- (19) *Him ofhreow þaes mannes*
 3SG.DAT. apenar.PST. el.GEN. hombre.GEN.
 ('Sentía pena por el hombre')

El ejemplo (19) (extraído de Eythórsson y Barðdal, 2005), del inglés medieval, certifica que no existía ningún elemento nominativo con el que supuestamente se estableciera la concordancia. En su lugar, teníamos un dativo experimentante y un argumento que podía ser genitivo o incluso estar introducido por una preposición, constituyendo, así, un sintagma preposicional. La inexistencia de un sujeto nominativo todavía se atestigua en danés y alemán (20-21, sacados de Eythórsson y Barðdal, 2005).

- (20) *Mér er kalt*
 1SG.DAT. ser.PRS. frío
 ('Tengo frío')
- (21) *Mir ist kalt*
 1SG.DAT. ser.3SG.PRS. frío
 ('Tengo frío')

Mientras que el sujeto expletivo *es* es opcional, la forma sin expletivo es la más usual. La anteposición del dativo sin una marca de nominativo en el resto del enunciado se cataloga como prueba de que el verdadero sujeto es dicho dativo, pero, para Cole *et al.* (1980), las lenguas germánicas, sobre todo el alemán, presentan más evidencias de sujeto dativo y de promoción del rasgo humano por encima de la función sintáctica (22-24).

- (22) *Mir grauet von mir selbst*
 1SG.DAT. horrorizar.3SG.PRS. de 1SG.DAT. REFL.
 ('Me horrorizo [a mí mismo]')

- (23) *Mir wird's schlecht und*
 1SG.DAT. ser.3SG.PRS.+NEUT. malo y
 – *graut's vor der Zukunft*
 PRODAT. horrorizar.3SG.PRS.+NEUT. de el futuro
 ('Me pone malo y me horroriza el futuro')
- (24) *Mitglied bei uns zu sein bedeutet aber auch,*
 Miembro con 1PL.DAT. a ser.INF. significar.3SG.PRS. pero también,
das Gefühl zu haben, – geholfen zu werden
 el sentimiento a tener.INF., PRODAT. ayudar.PCP. a ser.INF.
 ('Ser nuestro miembro significa también tener la sensación de ser ayudado')

Los ejemplos (22-24) (reproducidos de Cole *et al.* 1980) evidencian que los oblicuos expresados desarrollan características que aprobarían el test para ser sujetos (Cole *et al.*, 1980). En primer lugar, (22) indica la posibilidad de tener un reflexivo correferente con un dativo. En las oraciones (23-24), vemos la capacidad de omisión del referente oblicuo (dativo, en este caso) no solo en una subordinada de infinitivo, sino en una coordinada. Y, aunque (23) exhiba el expletivo *es* como sujeto sintáctico, la elisión a la que hacemos referencia es una prueba irrefutable del carácter de sujeto que el dativo posee en esos ejemplos. Según los autores, esta posibilidad surge desde siempre en las lenguas germánicas, independientemente de que cada lengua, por su lado, se haya ido deshaciendo de esta posibilidad o la haya mantenido. Es más, el comportamiento del alemán se observa en el inglés medieval y las lenguas escandinavas en su etapa antigua (actualmente, estas presentan neutro de materia, al igual que en las zonas romances con evidencialidad, como veremos más adelante, Baunmüller, 2000). El inglés y las lenguas escandinavas, sin embargo, se han ido orientando a la reconversión del experimentante al caso nominativo, pero el alemán se ha mantenido igual y el islandés parece estar en una fase intermedia, aunque con mucha producción de sujeto oblicuo. No es casual que sea el alemán el que presenta evidencialidad, en su caso por medio del estilo indirecto.

La misma direccionalidad emerge igualmente en griego; por ejemplo, Verhoeven (2008) indica que los verbos con objeto experimentante tienden a topicalizarse y a anteceder al verbo cuando el estímulo es no agentivo. Es decir, las características semánticas del sujeto sintáctico son determinantes para la ubicación del objeto experimentante, trasladándose a una posición prototípica del sujeto si el constituyente declinado en nominativo no es volitivo o agentivo. Lo mismo podemos argüir para el orden de valencias ante un verbo inacusativo, donde el sujeto sintáctico se posiciona en una ubicación prototípica del objeto (véase 25, de Karantzola y Lavidas, 2014).

- (25) *Espase to vazó*
 romper.PST.3SG neut.NOM. jarrón
 ('El jarrón se rompió')

Karantzola y Lavidas (2014) muestran que el griego se caracteriza por un fenómeno poco común interlingüísticamente y que se documenta sobre todo en lenguas ergativas y, de manera extraordinaria, en inglés. Se trata de la labilidad, consistente en el empleo de un mismo lexema para expresar tanto causa como efecto, sin ninguna marca de ningún tipo que modifique el verbo en función del número de argumentos. Esta opción causativa se ha desarrollado en griego con algunos verbos inacusativos o de una baja transitividad prototípica. No es de extrañar que, en lo que respecta a la evidencialidad, Giannakidou y Mari (2012) aseguren que el morfema $\theta\alpha$, que sirve para crear valores de temporalidad futura, ha ido evolucionando a un morfema de evidencialidad que expresa inferencia. Aunque ambas lecturas (la temporal y la evidencial) han convivido, actualmente es el valor de fuente de información el que prima como lectura no marcada, haciendo que efectivamente el griego moderno haya gramaticalizado en su futuro la evidencia inferencial.

No obstante, las lenguas indoeuropeas no son las únicas en mostrar esta correlación. En chukchi, en la península de Kamchatka, se marca un mismo sujeto humano como ergativo o absoluto en función del control que dicho constituyente tiene sobre la acción (Polinskaja y Nedjalkov, 1987). En cingalés, Dixon (1994) argumenta que un mismo sujeto sintáctico puede declinarse en nominativo o dativo, según la agentividad, volición o control de este. Lo mismo podemos apuntar para otras lenguas con patrones ergativos o de ergatividad escindida, como el guaraní, el quechua o variedades de la zona del Pacífico austral (Dixon, 1994), puesto que todas ellas tienen herramientas frecuentes para discernir entre un sujeto agentivo, volitivo y con control frente al que carece de dichos matices. En definitiva, su comportamiento sintáctico y morfológico atiende a las características semánticas del sujeto, siempre en base a los rasgos de agentividad, animacidad, control o volición. En todas estas lenguas aducidas, observamos un paradigma morfológico de evidencialidad y la correlación es tan frecuente que Chirikba (2003) expone para la familia abjasia la posibilidad del futuro como recurso especializado de conjetura. Lo mismo ocurre en coreano, donde la existencia también de patrones lábiles en los causativos (Kim, 2012) ha permitido el desarrollo de un sistema de evidencialidad (Aikhenvald, 2018a).

La simultaneidad de los patrones ergativos y la evidencialidad, la volvemos a ver en variedades portuguesas de Brasil y África. En el primer caso, Carvalho (2016) explora la caída de los reflexivos en ciertos dialectos del portugués brasileño, provocando el nacimiento de la labilidad también en dicha zona. Y es precisamente en esa zona amazónica donde advertimos un arraigo de la evidencialidad, tanto en sus lenguas autóctonas como en las variedades lusófonas vecinas (Aikhenvald y Dixon, 1998). El caso del portugués africano revela la misma dinámica: ya sea en sus formas criollas o por influencia de lenguas bantúes, las variedades caboverdianas, angoleñas y mozambiqueñas también presentan labilidad y patrones ergativos y, en muchas de ellas, evidencialidad (Duarte de Oliveira y De Araújo, 2019).

La tesis que propugna que son los patrones ergativos los que favorecen la creación de la evidencialidad y no al contrario se ve con el caso del inglés y, como defenderemos *a posteriori*, del francés. La lengua inglesa no solo presenta una alta proporción de labilidad, sino que también encierra en alguna de sus variedades modificaciones en los objetos en función de la semántica. En cuanto a la primera característica, es a partir de finales de la Edad Media cuando dicho idioma empezó a incrementar la nómina de verbos lábiles (McMillion, 2006). La razón, en este caso, de acuerdo con el autor, es el fuerte rechazo que los hablantes de inglés tuvieron ante el recurso de reflexivos, asociados con la migración escandinava de la época. La inclinación a eliminar el pronombre reflexivo en aquellos verbos que siempre lo habían presentado provocó la paulatina transitivización de los mismos. Igualmente, este idioma ha evolucionado en cuanto a la marcación de sus objetos en oraciones ditransitivas a un modelo como el ejemplificado en (26-27). Veámos en (19) la posibilidad del inglés de tener oraciones sin sujeto nominativo en la misma época en la que se incrementó la labilidad, pero actualmente, esta lengua ha tendido a lo que Dunn et al. (2017) denominan *nominative sickness* (en contraposición al *dative sickness* del alemán), pero aún observamos comportamientos que priman el rasgo humano frente al estatus sintáctico del constituyente, en este caso, en la relación de los objetos (véase 26-27).

- (26) *I gave the book to John*
 Yo dar.1SG.PST. el libro a John
 ('Le di el libro a John')
- (27) *I gave John the book*
 Yo dar.1SG.PST. John el libro
 ('Le di el libro a John')

La posibilidad de (27), existente también en dialectos meridionales italianos, responde a la alta topicalidad que un ente humano recibe en el discurso, independientemente de la función sintáctica que desempeñe. El trato específico que incluso en el entorno del objeto se da a los referentes humanos refleja de nuevo un comportamiento basado en las características semánticas de los argumentos, más que en su estatus sintáctico, ya que la anteposición que veíamos en el par anterior es imposible con referentes no humanos (véase 28-29).

- (28) *He sent the book to the store*
 Él enviar.3SG.PST. el libro a la tienda
 ('Envió el libro a la tienda')
- (29) **He sent the store the book*
 Él enviar.3SG.PST. la tienda el libro
 ('Envió el libro a la tienda')

Asimismo, la variedad estándar opta por distinguir el género gramatical en función de las características biológicas de los referentes, contraponiendo de nuevo animado versus inanimado. En el caso de los animales, si se sabe de antemano el sexo, el género del pronombre atiende a dicha característica, pero si no, emerge el neutro *it*. Siguiendo con la norma de este idioma, la misma elección de los pronombres de objeto obedece a patrones de animacidad o rasgo humano frente al caso sintáctico o el rol semántico que puedan desempeñar. Además, según Wagner (2004), también posee dialectalmente neutro de materia y a ello hay que añadir el intercambio de papeles temáticos para oraciones con un sujeto no agentivo (30).

- (30) *There came three men*
 Ahí venir.PST. tres hombres
 ('Vinieron tres hombres')

El posicionamiento del sujeto en una ubicación prototípica del objeto provoca el surgimiento del deíctico *there* y, aunque el ejemplo no permita ver la concordancia del verbo al conjugarse en pretérito, el presente adopta una concordancia sintáctica, aunque el sujeto se halle pospuesto. Las características ergativas del inglés son ahora menores que las que se daban en estadios anteriores, ya que Dixon (1994) augura una tendencia hacia la universalización de los modelos nominativo-acusativos en todo el comportamiento gramatical de una lengua dada, pero, aunque sus verbos modales puedan emplearse con valor evidencial, no podemos concluir que su

significación primaria, hegemónica o unívoca sea esa. El inglés es, por tanto, un buen ejemplo de que, a pesar de contar con comportamientos claramente ergativos, no ha desarrollado un paradigma morfológico de evidencialidad. Es más, expone los mismos componentes que las lenguas bálticas, el griego moderno e incluso el español peninsular dialectal, como demostraremos más adelante, pero, de momento, los elementos que propician la creación de evidencialidad (todos ellos de carácter ergativo) permanecen sin la existencia de esta, mientras que el surgimiento de los paradigmas evidenciales descritos antes se han visto motivados por los mismos parámetros que existen en inglés.

4.3. *Evidencialidad por contacto lingüístico*

En última instancia, contamos con lenguas que sí han desarrollado un sistema de evidencialidad, tal cual lo entendemos aquí, pero las cuales, a su vez, no exhiben comportamientos ergativos. La razón de esta contradicción se halla en el hecho de que dicho paradigma es contagio directo de una lengua vecina. Esta contingencia está ampliamente documentada. Sin ir más lejos, Aikhenvald (2018b) demuestra la presencia de dicho fenómeno en el judeoespañol de Turquía, precisamente como contagio del turco; lo mismo se aduce para el *Sprachbund* de los Balcanes o el armenio. Pero incluso bajo esta posibilidad, la lengua que desarrolla evidencialidad por contacto con otra puede adoptar el morfema de la que ha establecido la influencia. Soper (1996) defiende que el tayiko (de la familia irania) ha adoptado el sufijo *-mis* del uzbeko (de la rama túrquica) para fuente que no es de primera mano. Lo mismo ocurrió con este sufijo en su adopción por parte del kyrz por contacto con el azerí, siempre designando inferencia o reportatividad, es decir, evidencia indirecta. El propio turco mantiene una distinción morfológica entre la evidencia directa frente a la indirecta, ya sea esta última si proviene de una tercera fuente o de procesos inferenciales del propio hablante (Ünal, 2018).

Aikhenvald (2004) asegura que precisamente el turco es el mayor irradiador de evidencialidad de su zona circundante, por lo que las lenguas que lo rodean han ido moldeando un sistema flexivo evidencial como consecuencia del contacto lingüístico con el turco. Lo mismo arguye la autora para el *Sprachbund* báltico, ya que el letón y el lituano aceptan la evidencialidad como paradigma morfológico por contagio del estonio. Igual comportamiento se atestigua en el *Sprachbund* de los Balcanes (Aikhenvald, 2004) y en todo el Cáucaso. Esta influencia ha saltado más allá de la cadena montañosa euroasiática, incluso en la aceptación de la labilidad, pues, aunque este rasgo es una estrategia causativa que se atestigua fundamentalmente

en lenguas ergativas o con algún tipo de patrón escindido (Haspelmath, 1993; Letuchiy, 2004, 2009), el ruso se ha contagiado de la misma por la cercanía de este idioma con las ergativas caucásicas (Letuchiy, 2015).

La fuerte interrelación de las familias lingüísticas que comparten frontera en Asia Central es de una gran magnitud. Aunque la bibliografía dice que el turco es el mayor difusor de evidencialidad a su alrededor, esto no quiere decir que, previamente, no haya adoptado esta posibilidad por contacto lingüístico con el Cáucaso. Así lo dictamina Chirikba (2003), al ver que las etapas más primigenias de variedades caucásicas (y, por ende, ergativas) ya disponían de sistemas evidenciales y que estas no entraron estrechamente en contacto con las lenguas túrquicas hasta finales de la Edad Media. En cualquier caso, la influencia recíproca que las familias túrquica, caucásica, altaica, fino-urálica e incluso indoaryana han tenido es innegable, demostrándose sobre todo en el asentamiento de paradigmas evidenciales desde los Balcanes hasta Siberia, pasando por todo Asia Menor, el Cáucaso y Asia Central. La direccionalidad de este hecho sigue estando sin resolver, pero la interconexión entre todas ellas es tan grande que Johanson (2003) afirma que la evidencialidad se va perdiendo en los dialectos turcos que limitan con lenguas indoeuropeas que escapan a este fuerte vínculo. Las relaciones estrechas también afloran en persa, donde, de acuerdo con Perry (2000), su futuro ha terminado connotando inferencia de manera primaria.

La península ibérica es otro ejemplo de este comportamiento, ya que la evidencialidad que se atestigua en portugués, gallego y catalán parece estar propiciada por el contagio directo del español, si bien sobre este asunto nos detendremos a continuación. La correlación de patrones ergativos y evidencialidad vuelve a aparecer en otros dialectos del español que no son los europeos. Y de nuevo el contacto lingüístico también juega un papel en el contagio tanto de patrones ergativos como de la creación de evidencialidad morfológica. Así, el español en contacto con el quechua presenta leísmo y omisión del pronombre átono en contextos en los que el estándar lo requiere por someterse este último idioma al modelo animado – no animado, en lugar de la función sintáctica de los constituyentes (véase 31-32, de Fernández-Ordóñez, 1999).

(31) Las elecciones, yo nunca entendí

(32) Le quiero, Gloria

El quechua posee igualmente evidencialidad morfológica, la cual ha provocado incluso que el español de dicha zona no solo se sirva del FM o el COND para

expresar fuente de información, sino que emplee los tiempos compuestos también para dicha finalidad (Pfander y Palacios, 2013).

5. LAS LENGUAS ROMANCES

La argumentación teórica que hemos desarrollado en el epígrafe anterior puede demostrarse además con la situación de las variedades romances. El hecho de que la marcación morfológica para la evidencialidad sea una rama más de un comportamiento coherente para distinguir entre los conceptos de volición, control, agentividad o animacidad se corrobora con los ejemplos actuales e históricos. Hemos analizado las tres posibilidades que propician la creación morfológica de un paradigma evidencial a nivel universal, y estas mismas tres opciones pueden extrapolarse también a las lenguas romances. Como aludíamos al principio del artículo, la mayor controversia es la que rodea a los idiomas de esta familia, seguramente por la misma falta de consenso en establecer qué es evidencialidad. Sin embargo, a continuación, veremos que han sido los patrones ergativos los que han causado el desarrollo de un sistema para marcar fuente de información en el español peninsular occidental, así como en las variedades meridionales del italiano, mientras que el resto de sus dialectos y las demás lenguas iberorromances han difundido este mismo paradigma como consecuencia del contacto lingüístico, si bien cada una se halla en una fase distinta. Por último, sostendremos que el francés es el único idioma que no ha sucumbido a este fenómeno y que se comporta como el inglés, ya que presenta claros patrones ergativos que todavía no han desencadenado en un paradigma morfológico evidencial.

5.1. *El español peninsular y el italiano*

La evidencialidad en español ha sido estudiada por numerosos académicos, entre los que destacan los trabajos de Escandell-Vidal (2014, 2021) o Squartini (2001). En todos ellos, los datos manejados proceden de la propia introspección o de cuestionarios *ad hoc*. En todos los casos, se afirma la existencia de evidencialidad encarnada en el empleo del FM para la inferencia o conjetura y el COND para la reportatividad. Tal es así, que Escandell Vidal (2019) aduce que, al menos en el caso del FM, los valores temporales se restringen al lenguaje culto, ya que en el plano relajado, coloquial, no culto y como primer valor semántico que todo nativo aprende antes de su escolarización, la lectura inferencial se impone absolutamente como la no marcada y frecuente. Los datos de esta autora proceden igualmente de

corpus y ejemplos encontrados en la bibliografía que maneja, cuya metodología no engloba registros dialectales.

El caso de Aaron (2014) es muy ilustrativo, ya que investiga la evolución semántica del FM español y llega a la conclusión de que tanto el significado temporal como el evidencial han coexistido desde la aparición del FM, pero la significación conjetural ha sido históricamente escasa y restringida al habla culta. Aún a principios del siglo xx, la autora encuentra más muestras de FM conjeturales en contextos elevados que en contextos no cultos, pero en todos los casos, el FM como conjetura es minoritario. No obstante, señala que a medida que avanza el siglo xx, el FM conjetural aumenta en el habla no culta, pero nunca supera en número al valor temporal.

El matiz acerca de la tipología textual sobre la que se basa la mayor parte de los análisis aplicados al español es relevante, puesto que restringe la amplitud del funcionamiento de la evidencialidad a la variedad estándar con algunas líneas hacia realidades vernáculas. Es, *grosso modo*, el mismo caso para el resto de variedades romances, como veremos más adelante. Además del hecho de que los trabajos mencionados se fundamentan en el estándar o en el plano culto, ninguno de ellos establece estadísticamente el valor no marcado de todas las lecturas posibles de ambos tiempos. Esto provoca que los autores concluyan que prácticamente todos los idiomas romances posean evidencialidad a través del FM y/o el COND, ya que en todos ellos existe la lectura evidencial en convivencia con la temporal o la modal. Sin embargo, cuando cotejamos los resultados que arrojan las herramientas dialectales, observamos un comportamiento específico que se repite en todas las lenguas.

Los primeros datos referidos al plano no culto y dialectal, los encontramos en los atlas realizados en la primera mitad del siglo pasado, los cuales abarcan más de 1.500 localidades en total. Estos primeros trabajos se basaban en la repetición de palabras y frases preestablecidas por parte de informantes de entorno rural, mayores y de escasa alfabetización (NORM). La metodología de encuesta es por definición controvertida, porque en principio no garantiza la espontaneidad del hablante, pero el perfil del encuestado asegura la producción lingüística sin condicionamientos del estándar o de usos cultos. Los atlas que se llevaron a cabo fueron el *Atlas linguistique de France* (ALF), que cubre Francia (salvo Bretaña), la zona francófona de Bélgica y su homóloga en Suiza; el *Atlante linguistico ed etnografico dell'Italia e della Svizzera meridionale* (AIS), que cubre Italia, la zona italiana de Suiza e Istria; el *Atlasul lingvistic român* (ALR), que cubre Rumanía y Moldavia; y el *Atlas lingüístico de la Península Ibérica* (ALPI), que cubre toda la península más las islas Baleares, por lo que tenemos una representación geográfica y lingüística de los

romances europeos de hace un siglo en lo que respecta a numerosos fenómenos de todo tipo y no solo léxicos o fonéticos. El diseño de los cuestionarios preveía una gran cantidad de palabras y frases que tenían como objetivo registrar una particularidad gramatical, léxica o fonética específica; es más, el repertorio auguraba la aparición del mismo fenómeno en diversas frases para, precisamente, contrastar su frecuencia o el posible comportamiento gradual en función del contexto sintáctico, semántico, fonético, etc. En lo que respecta al FM y el COND, sus resultados han sido cartografiados por Lara Bermejo (2021b), como reproduciremos seguidamente, pero hemos de hacer dos observaciones: la primera es que no hay datos sobre el rumano, ya que dicha lengua no desarrolló un FM y COND homologables al resto de romances, puesto que su estrategia se basa en un auxiliar de volición; la segunda hace referencia a la falta de oraciones que contuvieran un FM o un COND con matiz evidencial.

Si observamos el ALF, sus preguntas con FM eran *Si nous ne mangeons pas nos prunes, elles se moisiront bientôt; Finiras-tu?; Ils feront ce qu'ils voudront; Quand mon fils sera grand, je l'enverrai à Paris; Je saurai ça mieux que lui; Je verrai; Des pommes, nous n'en aurons guère; Nous nous reverrons; Pourquoi ne vous mariez-vous pas? Vous trouverez bien quelqu'un qui vous ira; Toi, tu iras là*. En el AIS, estas se correspondían con *Lo manderò; Non dormirò stanotte; Quando sarà grande mio figlio; Te lo darò se lo vuoi; Le venderò domani; Marciranno ben presto; Faranno ciò che vorranno*. Solo el ALPI previó una única oración con un FM inferencial: *Aún no habrá venido* (el resto de las oraciones tenían una lectura temporal y una única modal). En lo que respecta al COND, ninguno de estos atlas previó lecturas temporales o evidenciales, sino únicamente modales.

En los mapas de las figuras 6 y 7, ilustramos el empleo no evidencial del FM y COND en los romances europeos de hace un siglo, sacados de Lara Bermejo (2021b).

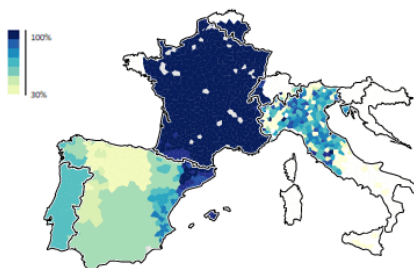


Figura 6. FM no evidencial a principios del siglo xx (Lara Bermejo, 2021b)

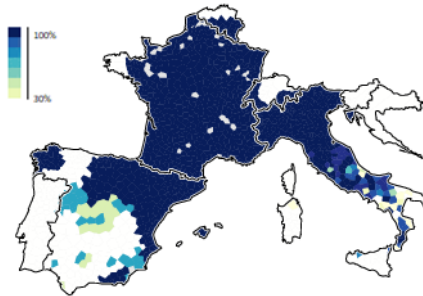


Figura 7. CONd no evidencial a principios del siglo XX (Lara Bermejo, 2021b)

Los mapas ofrecidos en las figuras 6 y 7, del FM y el CONd, respectivamente, dan cuenta de la extensión de ambos tiempos como recursos no evidenciales hace cien años. Si bien el rumano está excluido de este estudio, debido a que sus estrategias de futuro y condicional no siguieron el mismo camino que en el resto de variedades, el estudio de Lara Bermejo (2021b) demuestra que ambos tiempos eran, ante todo, alternativas no evidenciales en la lengua oral y en contextos no cultos en casi todo el espectro romance, salvo en zonas muy acotadas. En primer lugar, el FM era escaso o no existente en la mitad sur de Italia, así como sus islas; igualmente, este tiempo no era tampoco frecuente en el oeste del español peninsular para dichos valores, si bien surgía en mayor proporción. El CONd, por su parte, ocupa un espacio mayor en la península itálica, salvo el extremo meridional y las islas, mientras que era inexistente en portugués y gran parte del español.

La fotografía que muestran los datos de hace cien años da pistas acerca del nacimiento de la evidencialidad en los distintos romances (aspecto sobre el cual nos detendremos a continuación), pero antes es necesario comentar una particularidad que viene a colación de las dos observaciones que hacíamos antes: la frase que el ALPI introdujo con un FM inferencial (pero no un CONd reportativo) estaba concordada en perfecto (*Aún no habrá llegado*). Este detalle es esencial, porque los tiempos perfectos se prestan en mayor medida a aceptar los valores evidenciales antes que los simples. Stage (2003) afirma que el FM conjetural en francés solo es posible cuando está flexionado en perfecto, mientras que no es aceptado por la variedad estándar en su forma simple.

Los porcentajes reflejados en los mapas 6 y 7 llevan a postular que, hace cien años, el FM era evidencial en el oeste del español (ya que sus lecturas no evidenciales son secundarias) y, en principio, en todo el sur de Italia y, paulatinamente en el resto del país, ya que la zona austral no presenta ni una sola ocurrencia de FM no evidencial.

Ya Ledgeway (2009) avisa de que el FM en la mitad sur de Italia es únicamente conjetural desde la época decimonónica, cuando dejó de expresar temporalidad, matiz que empezó a ser consignado por el presente. Lo mismo podemos argüir para el caso del COND, pero no hubo ninguna pregunta en los cuestionarios de ningún atlas con valor evidencial para el COND, lo cual vuelve a poner de manifiesto la escasa probabilidad de que este sirviera para lecturas evidenciales en el plano no culto.

Para poder afirmar esta última idea, es necesario comparar la evolución del FM y el COND a nivel dialectal y no solo describir el estado de hace cien años. No obstante, el análisis exhaustivo es solo posible en los iberorromances gracias a los corpus actuales que, con una metodología que sí favorece la espontaneidad, llevan realizándose desde la década de 1990. Nos referimos al *Corpus Oral y Sonoro del Español Rural* (COSER) para el español; *Corpus Dialectal para o Estudo da Sintaxe* (CORDIALSIN) para el portugués; *Corpus Oral Informatizado de la Lengua Gallega* (CORILGA) para el gallego; y *Corpus Oral Dialectal* (COD) y *Corpus Dialectal del Català* (DIALCAT) para el catalán. Todos ellos se basan en la entrevista semidirigida con el mismo perfil de hablante que el de los atlas de principios del siglo xx. En este sentido, los resultados garantizan la evolución dialectal del FM y del COND sin condicionamientos estándares durante los últimos cien años. Hay que añadir que el número total de municipios encuestados en estos corpus alcanza las 300 localidades.

El mapa 6 sugiere que el FM empezaba a ser evidencial en el español occidental de hace un siglo, ya que los valores no evidenciales eran minoritarios, aunque podían surgir hasta un 30% de las ocasiones en las que se expresara dicho tiempo. Es decir, las ocurrencias de FM con valor conjetural (*Aún no habrá venido*) no superaban las de matiz temporal, por lo que dicho tiempo verbal no se había convertido aún en una marca morfológica de evidencialidad propiamente dicha, aunque sí parecía estar orientándose hacia esa conversión. Para averiguar si los resultados cartográficos eran el preludio de la especialización del FM (y el COND) como evidenciales, es necesario compararlos con la casuística actual (mapas de las figuras 8 y 9).

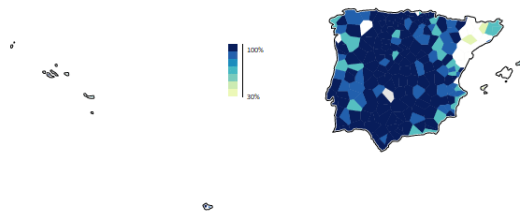


Figura 8. FM evidencial en la actualidad (Lara Bermejo, 2021a)



Figura 9. CONd evidencial en la actualidad (Lara Bermejo, 2021a)

Los mapas presentados en las figuras 8 y 9, extraídos de Lara Bermejo (2021a), reflejan que el empleo evidencial llega a ser el único en gran parte del español peninsular, en especial el oeste, en consonancia con los datos del mapa ofrecido en la figura 6. De nuevo, el FM da mayores incidencias de evidencialidad que el CONd, ya que este va a rebufo del primero. Por tanto, la evidencialidad en los iberorromances nace en el oeste del español peninsular y, posteriormente, se va extendiendo al resto del territorio. Las conclusiones de los mapas de las figuras 8 y 9 se sintetizan en la tabla 2.

	Portugués	Gallego	Español	Catalán
Futuro	Evidencial	Evidencial	Evidencial	Temporal (Cataluña y Baleares) Evidencial (Valencia)
Condicional	Evidencial	Modal	Evidencial	Modal

Tabla 2. Valores no marcados del FM y el CONd en los iberorromances (Lara Bermejo, 2021a)

La tabla 2 indica que el FM es evidencial en todas las variedades iberorromances, incluida el catalán, si bien de Valencia, pero no de Cataluña ni Baleares, mientras que el CONd lo es tan solo en español y portugués, pero no en gallego y catalán. Los números indican que el valor primario, hegemónico y, en muchos casos, único del FM en español es el inferencial. Lo es también primariamente en portugués, gallego y el catalán de Valencia. En el caso del CONd, podemos aducir lo mismo en español y portugués. Empero, la cartografía insiste en que es el español peninsular occidental el foco de la evidencialidad en dicho idioma, ya que es la zona donde, hace un siglo, el FM daba menos proporción de valor temporal y el área donde actualmente el FM es únicamente evidencial y nunca temporal o modal. El CONd también es más fuerte como evidencial en la misma región frente a otros territorios.

De acuerdo con lo defendido a lo largo de este artículo, la aparición de un paradigma morfológico de evidencialidad ha tenido que ser consecuencia de una serie

de patrones ergativos previos que han motivado su surgimiento. Esta asunción implica que, si la evidencialidad del español ha nacido en el oeste peninsular, dicha área geográfica ha de poseer una serie de rasgos de corte ergativo y que, además, tenían que estar ya presentes hace más de un siglo, puesto que la evidencialidad se configura en esa época. Las particularidades dialectales coincidentes con dicha zona apuntalan nuestra teoría.

Por un lado, el español peninsular occidental presenta labilidad. Los verbos lábiles registrados en el español peninsular dialectal son *entrar*, *quedar* y *caer* (Lara Bermejo, 2020).

- (33) El coche entró en el garaje – Entré el coche en el garaje
- (34) Los libros se quedaron en la mesa – Me quedé los libros en la mesa
- (35) El vaso (se) ha caído – He caído el vaso

Los enunciados (33-35) (extraídos de Lara Bermejo, 2020) muestran la transitivización de los tres lexemas inacusativos a costa de la oposición léxica *meter-entrar*, *dejar-quedar* y *tirar-caer*. El surgimiento de este fenómeno se acota al oeste peninsular, si bien su incidencia geográfica ha ido descendiendo con el paso de los años, a tenor de los registros en la bibliografía. Mientras que las monografías dialectales de 1960 y 1970 circunscribían la labilidad a casi toda Castilla y León, Extremadura y parte de Andalucía occidental, así como zonas del oeste de Castilla-La Mancha (Zamora Vicente, 1970; Alvar, 1996; Montero Curiel, 2006; García Mouton, 1994; Ariza, 2008), el trabajo de campo llevado a cabo por Lara Bermejo (2020) localiza el fenómeno en Extremadura y la zona más occidental de Castilla y León, sin llegar a León, con incidencia desigual en el oeste de Castilla-La Mancha. Sin embargo, la mayor o menor proporción de la labilidad se supedita a una serie de factores de índole semántico-sintáctica que reproducimos en (i-iv).

- i. *Entrar* > *quedar* > *caer*
- ii. Sujeto humano involuntario > sujeto no humano > sujeto humano voluntario
- iii. Objeto no afectado > objeto afectado
- iv. Atético > tético

Las jerarquías de (i-iv) establecen que, si un hablante transitiviza el verbo *caer*, lo hará necesariamente con los de su izquierda; asimismo, si transitiviza cualquiera de dichos verbos con un sujeto no humano, lo hará con un sujeto humano involuntario y, así, sucesivamente. Como se puede observar, la probabilidad de que un

verbo sea lábil es inversamente proporcional al grado de transitividad del sujeto, el objeto y la acción.

El comportamiento causativo que acabamos de describir no es el único trazo de ergatividad, ya que podemos hacer referencia a, al menos, tres rasgos lingüísticos que se atestiguan en (parte de) la zona donde se esparce la labilidad o en aquella área donde hace un siglo se daba y que responden a patrones no nominativo-acusativos. El primero de ellos es el léismo, laísmo y loísmo, así como el desarrollo de una concordancia propia para entidades no contables (neutro de materia). Su incidencia responde a patrones de animacidad, agentividad o volición, expandiéndose por gran parte de Castilla y León, Cantabria, Asturias, País Vasco, Navarra, La Rioja, Madrid y Castilla-La Mancha, con conatos en el este de Extremadura (Fernández-Ordóñez, 1999). Su extensión, por tanto, coincide con la región más oriental de la labilidad hace cien años, así como con los municipios del este de Extremadura, el oeste castellanomanchego y del centro-oeste de Castilla y León donde la labilidad sigue apareciendo actualmente.

El segundo de ellos es el empleo de pronombres reflexivos en verbos con un sujeto no volitivo o no humano, precisamente para marcar esa falta de agentividad. Lara Bermejo (2020) hace constar la existencia de frases como en (36-37).

- (36) Ayer me soñé que me tocaba la lotería
 (37) Hay que esperar a que la fruta se madure

Los enunciados de (36-37) (sacados de Lara Bermejo, 2020) suponen una estrategia para marcar la semántica de los argumentos del verbo, ya que el sujeto del mismo no exhibe las mismas propiedades en todas las oraciones; el reflexivo emerge precisamente en aquellos que carecen de agentividad, volición o control.

El tercer rasgo tiene que ver con un tipo de léismo, denominado como falso léismo por Fernández-Ordóñez (1999), consistente en el uso sistemático del dativo *le(s)* independientemente de la función sintáctica de objeto y el sexo del referente para marcar cortesía. Gómez Seibane (2021) demuestra que su utilización escapa por completo a la influencia del verdadero léismo, puesto que responde a la marcación de un referente humano mediante el caso gramatical que suele expresarlo: el dativo. Es el mismo tipo de estrategias que se da en lenguas germánicas, como veíamos antes.

Los ejemplos aducidos anteriormente son una muestra de la insistencia de dicha zona geográfica de marcar de manera relevante los rasgos que tienen que ver con humano, agentivo o volitivo frente a lo contrario. Se dan de manera dialectal,

aunque hay particularidades que han trascendido dicha región o se han establecido incluso en el estándar (véase el léismo de persona singular), pero reflejan claros patrones ergativos en todas sus acepciones (morfológica, sintáctica y discursiva), los cuales han dado pie a la creación de evidencialidad, precisamente en tipos de información que escapan al control o la volición del hablante. En primer lugar, el FM es el que se especializa para la inferencia, es decir, para un tipo de fuente de información que escapa al control del sujeto, pero que sigue partiendo de él mismo y no de un tercero. Las conjeturas que consignan los FM se enfocan en el presente o el futuro, pero, en un segundo estadio, el COND se convierte en evidencial para marcar también inferencia, aunque referida a un evento pasado (véase 38-39, sacados de Lara Bermejo, 2021a y 2023). En última instancia, este último tiempo verbal empieza a darse para expresar reportatividad; de nuevo, una fuente que escapa a la volición del hablante y que suma un nivel más en dicha falta de control, puesto que la información sale de un tercero.

- (38) En este pueblo habrá unas cuatrocientas personas o por ahí, pero hay muchos ingleses
- (39) No sé cómo irían los pobrecitos, porque todo lo que dicen no se ponían zapatos cuando se iban a casar algunos, pobrecillos

Las demás variedades romances, ya sean de la península ibérica o del resto de Europa, funcionan de manera idéntica, con una salvedad: algunas de estas han adoptado un paradigma evidencial por contacto lingüístico. El único idioma que ha desarrollado de manera autónoma un sistema morfológico de evidencialidad ha sido el italiano, gracias a los derroteros de sus dialectos meridionales. Si volvemos a los mapas 6 y 7, el inicio de la centuria de 1900 indica que el FM era inexistente como valor temporal y, en menor medida, el COND tampoco lo era ni en dicho matiz ni en el modal. La deducción de que, en esa época, el FM ya era evidencial en esa región viene dada por la propia bibliografía que trata el tema, la comparación con el español y la misma concatenación de elementos de índole ergativa que se acumulan ahí.

A falta de corpus dialectales parejos a los de las lenguas iberorromances, la única constancia de la que disponemos es la que se atribuye a trabajos como los de Ledgeway (2009) o Rohlf's (1968). En estos, se afirma que el FM es, ante todo, una estrategia de conjetura, si bien su valor temporal sigue siendo abundante. Berretta (1994) afirma que, en la adquisición del italiano como L1, los aprendientes asumen el uso del FM primero en su versión inferencial para, más tarde, incorporar la

temporal. Este camino no es exclusivo de las variedades del sur, sino que puede circunscribirse a toda la geografía italiana, aunque es cierto que el sur ha ido más allá al no recurrir al FM para valores temporales y sustituirlo por otras alternativas. Reiteramos aquí las investigaciones de corte histórico, en las que se apunta al hecho de que el sur abandonó el FM como matiz temporal para especializarlo para la inferencia. Como acontecía con el español, ambas lecturas han convivido siempre, pero en un momento dado, una región determinada ha resuelto gramaticalizar este tiempo para la fuente de información.

La bibliografía da a entender que, lo que era común para la zona más austral de Italia, ha ido en aumento en el resto del país y que el sur ha sido el foco de este fenómeno. Sin embargo, como ocurre con el español, el lugar de nacimiento de la evidencialidad en italiano no es algo casual y responde, a nuestro juicio, a la conjunción que se produce entre patrones ergativos y creación de un sistema evidencial, siendo siempre este consecuencia de los anteriores. El italiano meridional presenta una casuística similar a la descrita para el español occidental. De acuerdo con Cerullo (2021), existe una serie de verbos inacusativos que se emplean de forma transitiva a expensas de las construcciones perifrásticas, formadas por verbo con bajo contenido léxico (*fare, portare...*) más preposición o adverbio (*su, giù, fuori...*) (40-43, extraídos de Cerullo, 2021).

- (40) *Esco il cane*
 Salir.1SG.PRS.IND. el perro
 ('Saco el perro')
- (41) *Avete rimasto qualcosa nello spogliatoio?*
 Haber.2PL.PRS.IND. quedar.PCP. algo en+el vestuario
 ('¿Habéis dejado algo en el vestuario?')
- (42) *Arrivami il sale*
 Llegar.IMP.+me el sal
 ('Pásame la sal')
- (43) *Tornami la penna*
 Volver.IMP.+me la bolígrafo
 ('Devuélveme el bolígrafo')

El inventario de verbos lábiles en el italiano meridional dialectal sigue siendo, a día de hoy, una incógnita y, prácticamente, tan solo disponemos de las referencias de Cerullo (2021), sin que haya todavía un estudio exhaustivo a este respecto.

Además de ella, Trumper (1997) aduce que la transitivización se produce con el verbo *entrare*, mientras que Sornicola (1997) lo señala para el caso de *rimanere*. Además, de acuerdo con la autora, esta transitivización de inacusativos ha derivado en la reorganización de clíticos de tercera persona en función del rasgo humano del referente. Así, ciertos verbos intransitivos permiten la elección de un clítico acusativo si la referencia es humana a costa del canónico de objeto indirecto. Los factores semánticos que inciden en la labilidad italiana replican los referidos para el español: surge en contextos de baja transitividad, con sujetos no volitivos u objetos sin afectación.

Asimismo, aparte de presentar labilidad en el extremo sur, el entorno meridional posee otras estrategias que se focalizan sobre el factor humano, agentivo, volitivo o de control. Las variedades sureñas cuentan con el marcador diferencial de objeto, como el español estándar. De acuerdo con Ledgeway et al. (2019), los distintos dialectos del sur encumbran las entidades humanas que funcionan como objeto directo mediante la preposición *a*, a diferencia de otros referentes inanimados o no humanos. Aunque no todo ente humano necesariamente viene precedido por la preposición, es la mayor o menor referencialidad de este la que promueve el marcador diferencial, tratando, por tanto, de una manera específica las referencias que tienen que ver con los rasgos de animacidad o de humano.

Igualmente, a tenor de lo descubierto por Sornicola (1997), la zona austral de la península itálica puede llegar a universalizar los clíticos de acusativo para cualquier contexto sintáctico si se refieren a una entidad humana, pero, aunque el favorecimiento del acusativo contravenga la tendencia de aunar en el dativo el rasgo humano, la razón subyace en la transitivización de verbos intransitivos. Y, al igual que en el español dialectal, la misma área meridional ha desarrollado neutro de materia en distintos contextos sintácticos (Maiden, 1997). Por último, Loporcaro (1997) incide en la existencia de objetos indirectos humanos reconvertidos en objetos directos incluso en oraciones ditransitivas, al estilo del inglés, con el referente humano antepuesto al paciente, siempre con referencia a los dialectos meridionales, es decir, en la zona donde había evidencialidad ya hace un siglo.

5.2. *Los iberorromances y el francés*

Mientras que el español occidental y el italiano meridional han creado un sistema morfológico evidencial a través del FM y/o el COND precisamente por los factores ergativos dialectales que propiciaban este desarrollo, el resto del español peninsular y las variedades italianas centro-norteñas lo han incorporado como

contagio de sus respectivos epicentros. No han sido las únicas, ya que las demás lenguas iberorromances han seguido el mismo proceso. El cotejo de los mapas 6 a 9 es claro a este respecto: hace cien años, ni el portugués, ni el gallego ni el catalán presentaban porcentajes bajos de FM temporal, por lo que su significación primaria era esa misma. Los datos actuales reflejan un cambio drástico de tendencia en las lenguas occidentales y un proceso de cambio en marcha en el catalán, en distintas fases según la variedad catalanoparlante. Así, la zona valenciana ha terminado desarrollando evidencialidad en el FM (pero no en el COND), ya que la fuente de información de tipo inferencial es la primaria, pero la catalana y balear aún no han llegado a ese estadio, si bien admiten dialectalmente esta posibilidad; es más, Cataluña la acepta en mayor grado ahora que hace cien años.

Cabe señalar que ninguna de dichas áreas lingüísticas presenta una secuencia de patrones ergativos como la descrita para el español peninsular occidental, por lo que la única posibilidad de que hayan imitado el modelo español es mediante contacto. Aunque Bazenga y Rodrigues (2019) y Segura da Cruz (1991) avisan de un incipiente *lheísmo*, al estilo del léismo de persona en español, en ciertas zonas meridionales del país, no hallamos ni labilidad, ni confusión pronominal en función de la semántica de los actantes, ni siquiera una divergencia en el auxiliar de los tiempos perfectos (incluso en los casos tan restringidos en los que se acepta en gallego y portugués). El máximo exponente de configuración sintáctica atendiendo a los parámetros de agentividad y volición es el orden no marcado de constituyentes con verbos inacusativos, ya que también es frecuente la posición posverbal, y un incipiente *lheísmo* de cortesía para diferenciar el referente humano del no humano (Bazenga *et al.*, 2016).

La propia geolingüística de los mapas 8 y 9 apoya la tesis sobre el contacto lingüístico, pues no solo ha penetrado la evidencialidad antes en las variedades más al oeste, en el límite con su foco, sino que, para el catalán, ha sido la valenciana la que ha sucumbido a esta particularidad, repitiendo el comportamiento secular que ha tenido con respecto a otros fenómenos lingüísticos que aparecen en el catalán y que ahora están incluidos gracias a su incorporación a través de Valencia. A tenor de lo analizado por Fernández-Ordóñez (2011), la península ibérica se ha ido contagiando recurrentemente fenómenos, con independencia de la lengua hablada, y estas influencias han sido tanto léxicas, como fonéticas y morfosintácticas, si bien la direccionalidad geográfica no ha sido pareja. Así, los cambios fonéticos suelen escindirse entre el norte y el sur (y viceversa), mientras que la morfosintaxis ha recorrido un camino horizontal: de este a oeste y viceversa. Por ejemplo, los

indefinidos *alguien* y *nadie* en detrimento de *alguno* y *ninguno* son un contagio del occidente peninsular, mientras que los pronombres *nosotros* y *vosotros* son una influencia del este. La entrada del FM y COND también es originaria del este peninsular, hasta tal punto de que, según Cunha y Cintra (1992), no fue nunca popular en portugués, donde se acotó al lenguaje elevado.

El caso del catalán es primordial para corroborar lo que sostenemos: su zona colindante con el francés carece también de evidencialidad. Ya hemos comentado que, a nuestro juicio, el francés no es todavía evidencial y este hecho resulta relevante, porque los rasgos galorrománicos en el catalán son sobradamente conocidos y la evidencialidad del catalán se da cuanto más al sur de su zona nos encontremos, siendo el español el claro contagiador de esta particularidad, como lo ha sido del alomorfo *-ra* por *-se* para el imperfecto de subjuntivo (Lara Bermejo, 2019). El hecho de que el catalán ha ido incorporando muy tardíamente el FM y el COND para lecturas evidenciales se ve enfatizado por las mismas afirmaciones de la bibliografía al respecto, ya que Badia i Margarit (1962), Wheeler *et al.* (1999) y Solà y Rigau (2002) insisten en que en catalán la inferencia se consigna mediante el modal *deure*, si bien todos ellos soslayan muy de pasada la posibilidad de que, coloquialmente, se pueda dar el FM, aunque remotamente. Hay que tener en cuenta la fecha de publicación, ya que todavía a mediados del siglo pasado se aseguraba la escasa aparición del FM con valor inferencial y, desde luego, la inexistencia del COND a este efecto.

La realidad del francés es una prueba adicional a toda la argumentación: carece de evidencialidad, incluso a pesar de presentar conatos de labilidad. La labilidad en francés es un tema relativamente recurrente en la bibliografía académica, pero no todos los autores emplean el término de la misma manera. Larjavaara (2000) opta por una definición muy abarcadora, ya que admite ciertos reflexivos en la misma categoría, pero hace notar el comportamiento lábil de verbos como *sortir*, *monter*, *descendre* y *apprendre*, sin que ello suponga un choque con el estándar. La aceptabilidad por parte de la variedad normativa no es *per se* un criterio para nuestra argumentación, aunque tenemos que reconocer que su aceptación también en dicho plano es un salto cualitativo en el prestigio de la labilidad, pero el francés actual cuenta también con usos lábiles de verbos que el estándar rechaza por completo, en especial, *tomber*. Resaltamos además la circunstancia de que la labilidad siempre suele atañer a los mismos verbos interlingüísticamente.

La labilidad no es el único elemento que se fundamenta en los rasgos semánticos del sujeto en esta lengua, sino que Bilous (2011) subraya la plausibilidad de sujetos

no volitivos pospuestos precisamente para marcar dicha carencia de agentividad, control o volición (44), convirtiendo cualquier verbo en transitivo.

- (44) *Il est venu trois hommes*
 3SG.MASC. ser.3SG.PRS. venir.PCP.MASC.SG. tres hombres
 ('Vinieron tres hombres')

El ejemplo (44) (de Bilous, 2011) muestra la posibilidad de enviar a una posición de aparente objeto los argumentos que, en una posición no marcada, son el sujeto, como ilustrábamos para el inglés. La reorganización argumental provoca la transitivización del verbo inacusativo *venir*, haciendo que el francés recurra al expletivo por tener que producir obligatoriamente un sujeto. A todo ello, se le suma la distinción del auxiliar en la configuración de los tiempos perfectos, en función de la inacusatividad del verbo: *être* para los inacusativos y *avoir* para los inergativos y transitivos.

El problema, a nuestro juicio, de los análisis sobre el francés en lo que respecta a la evidencialidad es que se basan sobre todo en el lenguaje culto. No solo eso, sino que ciertos autores determinan que el francés posee evidencialidad por el mero hecho de que su FM y COND puedan connotar inferencia y reportatividad, sin que haya una reflexión acerca de la frecuencia de ese matiz semántico en confrontación con el valor temporal o modal. Recordemos que el español siempre se caracterizó por un FM que podía servir para la temporalidad y la inferencia, pero esta última posibilidad se limitaba al plano culto (así lo constatan los mapas y la bibliografía) y representaba un porcentaje mínimo frente al valor temporal de dicho tiempo. Es probable que más adelante el FM y el COND se transformen realmente en evidenciales al exhibir hegemoníicamente esta lectura semántica, sobre todo teniendo en cuenta los rasgos ergativos que se advierten en dicha lengua, pero no podemos afirmar que sea el caso en la actualidad.

La importancia del género discursivo es esencial, puesto que el mismo español prueba que, aunque el COND valga como valor inferencial referido a un evento pasado tanto a nivel coloquial como en contextos elevados, su significado reportativo se limita al registro periodístico y, por ende, a un plano culto. La probabilidad de que el COND se comporte como un reportativo ha aumentado con el tiempo, pero aún hoy, esta posibilidad está condicionada por el género (Romero Gualda, 1994). De vuelta al francés, Van de Weerd (2018) muestra ejemplos de COND reportativo en documentos jurídicos del siglo XVIII y Vatrican (2010) establece el uso del COND evidencial en contextos periodísticos. Es más, a pesar de que parte de la bibliografía

ha aceptado su condición de lengua con evidencialidad por permitir el FM para conjeturas en circunstancias muy concretas, aquellos autores que han llevado a cabo una investigación sobre lengua oral a este respecto coinciden en determinar que el valor conjetural es mínimo, no se suele producir en el lenguaje no culto y que, por tanto, está constreñido a un registro muy específico, por lo que el FM en francés es actualmente un recurso temporal por defecto (Poplack y Turpin, 1999, Lyons, 1968, Tomaszkiwicz, 1988). Incluso Aaron (2007) llega a afirmar que los ejemplos con FM conjetural que salen en obras tan célebres como en Palmer (1986) o Fleischman (1982) son inventados y no reflejan la casuística real.

La situación del francés (e históricamente del español) a este respecto se corrobora con la evolución de otros romances: Rohlf (1968) también restringe el uso del COND reportativo en italiano al periodismo; y, finalmente, Oliveira (1985) también se refiere al COND evidencial portugués como una estrategia típica de los textos periodísticos para marcar la reportatividad. De nuevo, sobre el COND reportativo en francés, Aikhenvald (2004), siguiendo los datos de Dendale (1993), confirma que el recurso es meramente periodístico o surge en textos elevados, sin que nadie aporte un solo dato de COND evidencial oral o no culto. Otros autores, como Bermúdez (2016), para el caso del español, o Kronning (2014) para el francés, afirman que el COND reportativo aparece en otra tipología textual, como el lenguaje científico y literario, pero siguen siendo enunciados cultos. Pero incluso la posibilidad de que el COND reportativo emerja en el español no culto, como sugiere Bermúdez (2016), se constriñe a una cuestión de frecuencia, siendo aún poco común.

El trabajo de Martines (2017) demuestra el empleo temprano del FM como conjetura en el catalán, pero circunscrito al siglo XIII. La muestra replica el caso del francés y de los demás romances, donde la convivencia de la lectura temporal y evidencial ha existido siempre, con clara preponderancia del primero sobre el segundo y con este último acotado al lenguaje culto. Por tanto, es en el sur de Italia y en casi toda la península ibérica, donde el valor evidencial ha saltado al plano coloquial o no culto, revirtiendo la complementariedad, ya que la lectura temporal es actualmente en dichos casos una cuestión del registro elevado. En consecuencia, la existencia de la evidencialidad en portugués, gallego y parte del catalán responde al hecho de que el español dialectal lo ha esparcido al resto de la península ibérica, mientras que el italiano estándar y las variedades septentrionales han incorporado la evidencialidad morfológica por influencia de los dialectos del sur.

Sin embargo, ya sea en el español o en las demás lenguas romances, la direccionalidad en la creación de la evidencialidad o en el uso cada vez más frecuente

de ciertas estrategias evidenciales se repite insistentemente. El primer tiempo en sucumbir es el FM, seguido del COND; la primera fuente en consignarse es la inferencia, seguida de la reportatividad. Ya nos hemos explayado a este respecto, sobre todo para el español y el italiano, pero el portugués ofrece una prueba más: tanto su FM como su COND sirven para la inferencia y la reportatividad. Mientras que la versión reportativa de ambos tiempos es una cuestión de registro y de lenguaje culto, la elección inferencial es la primaria en el lenguaje oral y no elevado, con la misma complementariedad que en español. Una vez más, vemos la especialización de una marca incontrolable (inferencia) que sale de uno mismo para, más tarde, crear otra que carece de volición, pero que posee menos control todavía por parte del hablante (reportatividad).

6. CONCLUSIONES

El sistema de evidencialidad en español, encarnado en el FM y el COND, para connotar inferencia y reportatividad, nace en el oeste de España como consecuencia de los rasgos ergativos que previamente se han ido imponiendo en esta zona. Con el tiempo, ha ido contagiando esta particularidad al resto de la región castellanoparlante, además de a la zona catalanohablante de Valencia, al portugués europeo y al gallego. Su nacimiento y evolución coincide con el desarrollo de la evidencialidad en italiano, cuyo FM se especializa como inferencial hace más de un siglo en el sur del país, precisamente gracias a la concatenación de rasgos ergativos que también se daban en ese territorio.

El resto de idiomas romances carecen de un sistema flexivo evidencial, a pesar de contar con las mismas estrategias para consignar inferencia y reportatividad, pero la frecuencia del matiz evidencial es secundaria y ha estado siempre restringida al registro culto. Este hecho es especialmente relevante para el francés y, aunque el propio idioma posea patrones ergativos que favorezcan el nacimiento de la evidencialidad, la frecuencia de uso todavía no permite postular tal situación. Sin embargo, aunque la tendencia a la ergatividad parece ser condición *sine qua non* para el desarrollo de la evidencialidad (véase el alemán, las lenguas amerindias o del Cáucaso), la existencia de esta característica se puede dar como consecuencia del contacto lingüístico. Es precisamente lo que ha sucedido en los iberorromances (salvo el español), las lenguas que rodean el turco, las variedades hispanas de los Andes o en los *Sprachbünde* de los Balcanes y el Báltico.

La razón que permite explicar la correlación entre ergatividad y evidencialidad reside en el hecho de que ambas características inciden en los rasgos semánticos que tienen que ver con el grado de volición, agentividad y control por parte del sujeto. Es por ello que son la conjetura (evidencia indirecta) y reportatividad (evidencia indirecta y de una tercera fuente) las primeras en conformarse en el nacimiento y desarrollo de la evidencialidad, frente a otro tipo de fuente directa, que permanece sin marcar. Lo vemos no solo en los romances, sino de manera general en la tipología lingüística, a tenor de los datos expuestos, y permite entender la sistematicidad tanto en las fuentes de información que suelen marcarse como en la probabilidad de existencia de la evidencialidad en una lengua o variedad dada.

Agradecimientos

Quisiera expresar mi gratitud a los revisores que evaluaron el manuscrito, por sus comentarios y sugerencias constructivas, así como por su lectura pormenorizada.

REFERENCIAS

- Aaron, J. (2007). El futuro epistémico y la variación: gramaticalización y expresión de la futuridad desde 1600. *Moenia*, 13, 253-274.
- Aaron, J. (2014). A certain future: epistemicity, prediction, and assertion in Iberian Spanish future expression. *Studies in Hispanic and Lusophone Linguistics*, 7(2), 215-240.
- Aikhenvald, A. (2004). *Evidentiality*. Oxford: Oxford University Press.
- Aikhenvald, A. (2018a). *The Oxford handbook of evidentiality*. Oxford: Oxford University Press.
- Aikhenvald, A. (2018b). Evidentiality and language contact. En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 148-172. Oxford: Oxford University Press.
- Aikhenvald, A. y Dixon, R. M. W. (1998). Evidentials and areal typology: a case study from Amazonia. *Language Sciences*, 20(3), 241-257.
- Aldai, G. (2009). Is Basque morphologically ergative?: Western Basque versus Eastern Basque. *Studies in Language*, 33, 783-831.
- Alvar, M. (1996). *Manual de dialectología hispánica*. Barcelona: Ariel.
- Ariza, M. (2008). *Estudios sobre el extremeño*. Cáceres: Universidad de Extremadura.
- Badia i Margarit, A. M. (1962). *Gramàtica catalana*. Barcelona: Enciclopedia Catalana.
- Baumüller, K. (2000). Gender in North Germanic: a diasystematic and functional approach. En B. Unterbeck et al. (Eds.). *Gender in grammar and cognition*, pp. 35-53. Berlín: Mouton de Gruyter.
- Bazenga, A. et al. (2016). Variantes sintáticas (padrão e não padrão) em português: representações sociais e atitudes linguísticas de falantes madeirenses. En A. Razky et al. (Eds.). *Atas do IV CIDS*, pp. 1-15. París: Université Paris-Sorbonne.

- Bazenga, A. y Rodrigues, L. (2019). O uso do clítico *lhe* em variedades do português. En A. M. Ferreira *et al.* (Eds.). *Pelos mares da língua portuguesa*, pp. 17-33. Aveiro: Universidad de Aveiro.
- Bermúdez, F. (2016). Rumores y otros malos hábitos. El condicional evidencial en español. *Cuadernos de Lingüística de El Colegio de México*, 3(2), 35-69.
- Berretta, M. (1994). Il futuro italiano nella varietà nativa colloquiale e nella varietà di apprendimento. *Zeitschrift für romanische Philologie*, 110(1-2), 1-36.
- Bilous, R. (2011). *Transitivité et marquage d'objet différentiel*. Toronto: University of Toronto.
- Brosig, B. y Skribnik, E. K. (2018). Evidentiality in Mongolic. En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 554-578. Oxford: Oxford University Press.
- Carvalho, J. (2016). What causes the alternation of agentive verbs in Brazilian Portuguese? *U. Penn Working Papers in Linguistics*, 22(1), 61-70.
- Cerullo, G. (2021). The transitivisation of unaccusatives in Italian (and Romance). Comunicación dada en el *Forschungskolloquium*, 9 de febrero de 2021, Universität Hamburg.
- Chirikba, V. (2003). Evidential category and evidential strategy in Abkhaz. En A. Aikhenvald y R. M. W. Dixon (Eds.). *Studies in evidentiality*, pp. 243-272. Amsterdam: John Benjamins.
- Cole, P. *et al.* (1980). The acquisition of subjecthood. *Language*, 56(4), 719-743.
- Comrie, B. (2013a). Alignment of case marking of pronouns. En M. Dryer y M. Haspelmath (Eds.). *The world atlas of language structures (WALS)* (Chapter 99). Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Comrie, B. (2013b). Alignment of case marking of full noun phrases. En M. Dryer y M. Haspelmath (Eds.). *The world atlas of language structures (WALS)* (Chapter 98). Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Cunha, C. y Cintra, L. (1992). *Nova gramática do português contemporâneo*. Lisboa: João Sá de Costa.
- De Haan, F. (2013). Coding of evidentiality. In M. Dryer y M. Haspelmath (Eds.). *The world atlas of language structures (WALS)* (Chapter 78). Leipzig: Max Planck Institute for Evolutionary Anthropology.
- DeLancey, S. (2018). Evidentiality in Tibetic. En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 580-594. Oxford: Oxford University Press.
- Dendale, P. (1993). Le conditionnel de l'information incertaine: marqueur modal ou marqueur évidentiel? En G. Hilty (Ed.). *XXe Congrès International de Linguistique et Philologie Romanes*, pp. 165-176. Tübingen: Francke.
- Dixon, R.M.W. (1994). *Ergativity*. Cambridge: Cambridge University Press.
- Dryer, M. S. y Haspelmath, M. (2013). *The world atlas of language structures (WALS)*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Duarte de Oliveira, M. S. y De Araújo, G. A. (2019). *O português na África atlântica*. São Paulo: FFLCH.
- Dunn, M. *et al.* (2017). Dative sickness: a phylogenetic analysis of argument structure evolution in Germanic. *Language*, 93(1), 1-22.
- Escandell-Vidal, V. (2014). Evidential futures: the case of Spanish. En P. De Brabanter *et al.* (Eds.). *Future times, future tenses*, pp. 221-246. Oxford: Oxford University Press.

- Escandell-Vidal, V. (2021). The semantics of the simple future in Romance: core meaning and parametric variation. En L. Baranzini y L. de Saussure (Eds.). *Aspects of tenses, modality, and evidentiality*, pp. 9-31. Leiden: Brill.
- Eythórsson, T. y Barðdal, J. (2005). Oblique subjects: a common Germanic inheritance. *Language*, 81(4), 824-881.
- Fernández-Ordóñez, I. (1999). Leísmo, laísmo y loísmo. En I. Bosque y V. Demonte (Dirs.). *Gramática descriptiva de la lengua española*, pp. 1317-1397. Madrid: Espasa.
- Fernández-Ordóñez, I. (2011). *La lengua de Castilla y la formación del español*. Madrid: Espasa.
- Fitneva, S. A. (2018). The acquisition of evidentiality. En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 186-201. Oxford: Oxford University Press.
- Fleischman, S. (1982). *The future in thought and language. Diachronic evidence in Romance*. Cambridge: Cambridge University Press.
- Forker, D. (2018). Evidentiality in Nakh-Daghestanian languages. En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 490-509. Oxford: Oxford University Press.
- Friedman, V. (2018). Where do evidentials come from? En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 124-147. Oxford: Oxford University Press.
- García Mouton, P. (1994). *Lenguas y dialectos de España*. Madrid: Arco Libros.
- Giannakidou, A. y Mari, A. (2012). An evidential analysis of Greek and Italian future morphemes. Comunicación dada en el congreso *The Nature of Evidentiality*, 14-16 de junio de 2012, Universiteit Leiden.
- Givón, T. (2001). *Syntax*. Ámsterdam / Filadelfia: John Benjamins.
- Gómez Seibane, S. (2021). Conectando las formas de tratamiento y el leísmo en la correspondencia de los siglos XVIII y XIX: el leísmo de cortesía. *Revista Internacional de Lingüística Iberoamericana*, 38, 129-150.
- Haspelmath, M. (1993). More on the typology of inchoative/causative verb alternations. En B. Comrie y M. Polinsky (Eds.). *Causatives and transitivity*, pp. 87-121. Ámsterdam / Filadelfia: John Benjamins.
- Hualde, J. I. y Ortiz de Urbina, J. (2003). *A grammar of Basque*. Berlín: Mouton de Gruyter.
- Johanson, L. (2003). Evidentiality in Turkic. En A. Aikhenvald y R. M. W. Dixon (Eds.). *Studies in evidentiality*, pp. 273-291. Ámsterdam / Filadelfia: John Benjamins.
- Johanson, L. (2018). Turkic indirectivity. En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality* (pp. 510-524). Oxford: Oxford University Press.
- Karantzola, E. y Lavidas, N. (2014). On the relation between labilisations and neuter gender: evidence from Greek diachrony. *Linguistics*, 52(4), 1025-1059.
- Kehayov, P. y Vihman, V. (2014). The lure of lability: a synchronic and diachronic investigation of the labile pattern in Estonian. *Linguistics*, 52(4), 1061-1105.
- Kim, H. S. (2012). Transitives, causatives and passives in Korean and Japanese. En M. Shibatani et al. (Eds.). *Argument structure and grammatical relations*, pp. 241-255. Ámsterdam / Filadelfia: John Benjamins.
- Kronning, H. (2014). Pour une linguistique contrastive variationnelle: le conditionnel épistémique d' «emprunt» en français, en italien et en espagnol. En H. Helland y C. Meklenborg (Eds.). *Affaire(s) de grammaire: Mélanges offerts à Marianne Hobaek Haff à l'occasion de ses soixante-cinq ans*, pp. 67-90. Oslo: Novus Forlag.

- Lara Bermejo, V. (2019). El pretérito imperfecto de subjuntivo en la Península Ibérica del siglo XX. *Verba*, 46, 313-338.
- Lara Bermejo, V. (2020). *Construcciones causativas y labilidad en español*. Berlín: De Gruyter.
- Lara Bermejo, V. (2021a). Futuro y condicional en las lenguas iberorromances: de marca de tiempo a marca de evidencialidad. *Borealis*, 10(1), 145-162.
- Lara Bermejo, V. (2021b). Evolution of the Romance evidentiality along the 20th century: a Coserian approach. *Concordia Discors v. Discordia Concors*, 16, 105-131.
- Lara Bermejo, V. (2023). El condicional en las lenguas romances de la Península Ibérica. *RESLA*, 36(1), 30-59.
- Larjavaara, M. (2000). *Présence ou absence de l'objet: limites du possible en français*. Helsinki: Academia Scientirum Fennica.
- Ledgeway, A. (2009). *Grammatica diacronica del napoletano*. Tübingen: Max Niemeyer.
- Ledgeway, A. et al. (2019). Differential object marking and the properties of D in the dialects of the extreme south of Italy. *Glossa: a journal of general linguistics*, 4(1), 1-25.
- Letuchiy, A. B. (2004). Lability of verbs and its relations to verb meaning and argument structure (based on data from Indo-European, Arabic, Turkish and other languages). *Vortrag beim LENCA-2-Symposium an der Universität Kasan*, 11, 1-5.
- Letuchiy, A. B. (2009). Towards a typology of labile verbs: lability versus derivation. En P. Epps y A. Arkhipov (Eds.). *New challenges in typology: transcending the borders and refining the distinctions*, pp. 247-268. Berlín: Mouton de Gruyter.
- Letuchiy, A. B. (2015). Historical development of labile verbs in modern Russian. *Linguistics*, 53(3), 611-647.
- Loporcaro, M. (1997). Puglia and Salento. In M. Maiden y M. Parry (Eds.). *The dialects of Italy*, pp. 338-348. Londres: Routledge.
- Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
- Maiden, M. (1997). Inflectional morphology of the noun and adjective. In M. Maiden y M. Parry (Eds.). *The dialects of Italy*, pp. 68-74. Londres: Routledge.
- Martines, J. (2017). L'émergence des futurs épistémiques romans. L'exemple du catalan médiéval du XIIIème siècle. En L. Baranzini (Ed.). *Le futur dans les langues romanes*, pp. 133-168. Berna: Peter Lang.
- McGregor, W. B. (2009). Typology of ergativity. *Language and Linguistic Compass*, 3(1), 480-508.
- McMillion, A. (2006). *Labile verbs in English: their meaning, behaviour and structure*. Estocolmo: Universidad de Estocolmo.
- Montero Curiel, P. (2006). *El extremeño*. Madrid: Arco Libros.
- Oliveira, F. (1985). O futuro em português: alguns aspectos temporais e/ou modais. In *Actas do 1º Encontro da Associação Portuguesa de Linguística*, pp. 353-373. Lisboa: Associação Portuguesa de Linguística.
- Palmer, F. R. (1986). *Mood and modality*. Oxford: Oxford University Press.

- Perry, J. R. (2000). Epistemic verb forms in Persian of Iran, Afghanistan and Tajikistan. En L. Johanson y B. Utas (Eds.). *Evidentials. Turkic, Iranian and neighbouring languages*, pp. 229-257. Berlín: Mouton de Gruyter.
- Pfander, S. y Palacios, A. (2013). Evidencialidad y validación en los pretéritos del español andino ecuatoriano. *Círculo de Lingüística Aplicada a la Comunicación*, 54, 65-99.
- Polinskaja, M. S. y Nedjalkov, V. P. (1987). Contrasting the absolutive in Chuckchee : syntax, semantics and pragmatics. *Lingua*, 71(1-4), 239-269.
- Poplack, S. y Turpin, D. (1999). Does the futur have a future in (Canadian) French? *Probus*, 11, 133-164.
- Rohlf, G. (1968). *Grammatica storica della lingua italiana e dei suoi dialetti*. Turín: Einaudi.
- Romero Gualda, M. V. (1994). *El español hablado en los medios de comunicación*. Madrid: Arco Libros.
- Segura da Cruz, M. L. (1991). *O falar de Odeleite*. Lisboa: Universidad de Lisboa.
- Siewierska, A. (2013). Alignment of verbal person marking. En M. Dryer y M. Haspelmath (Eds.). *The world atlas of language structures (WALS)* (Chapter 100). Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Skribnik, E. K. (1998). K voprosu o neochevidnom naklonenii v mansijskom jazyke (struktura i semantika). *Yazyki korennyh narodov Sibiri: sbornik nauchnyh trudov*. Vyp. 4, 197-215.
- Skribnik, E. K. y Kehayov, P. (2018). Evidentials in Uralic languages. En A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 525-553. Oxford: Oxford University Press.
- Solà, J. y Rigau, G. (2002). *Gramàtica del català contemporani*. Barcelona: Empúries.
- Soper, J. (1996). *Loan syntax in Turkic and Iranian*. Bloomington: Eurolingua.
- Sornicola, R. (1997). Campania. In M. Maiden y M. Parry (Eds.). *The dialects of Italy*, pp. 330-337. Londres: Routledge.
- Squartini, M. (2001). The internal structure of evidentiality in Romance. *Studies in Language*, 25(2), 297-334.
- Stage, L. (2003). Les valeurs modales du futur et du present. In M. Birkelund et al. (Eds.). *Aspects de la modalité* (pp. 203-216). Tübingen: Max Niemeyer Verlag.
- Sun, J. T-S. (2018). Evidentials and person. In A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality* (pp. 47-63). Oxford: Oxford University Press.
- Thrones, T. (2018). Evidentiality in the Uto-Aztecan languages. In A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality* (pp. 409-430). Oxford: Oxford University Press.
- Tomaszkiewicz, T. (1988). L'interprétation modale du futur polonais et français en corrélation avec la valeur sémantique du lexème verbal. *Studia Romanica Posnaniensia*, 13, 167-174.
- Trumper, J. (1997). Calabria and southern Basilicata. In M. Maiden y M. Parry (Eds.). *The dialects of Italy* (pp. 355-375). Londres: Routledge.
- Ünal, E. (2018). Evidentials, information sources and cognition. In A. Aikhenvald (Ed.). *The Oxford handbook of evidentiality*, pp. 175-185. Oxford: Oxford University Press.

- Van de Weerd, J. (2018). Vers les origines sémantiques du conditionnel épistémique. Étude d'un genre juridique en français classique (XVIe-XVIIIe siècles). *Langue Française*, 4, 77–89.
- Vatrican, A. (2010). La modalité et le conditionnel de rumeur en français et en espagnol. *Modèles Linguistiques*, 31(62), 83–94.
- Verhoeven, E. (2008). (Non-)canonical marking of experiencer objects: a typological comparison of Chinese, Korean, Turkish and Modern Greek. *Language Typology and Universals*, 61(1), 81–92.
- Wagner, S. (2004). Gendered pronouns in English dialects – a typological perspective. En B. Kortmann (Ed.). *Dialectology meets typology. Dialect grammar from a cross-linguistic perspective*, pp. 479–496. Berlín: Mouton de Gruyter.
- WALS: véase Dryer, M. y Haspelmath, M. (2013).
- Wheeler, M. et al. (1999). *Catalan: a comprehensive grammar*. Londres: Routledge.
- Zamora Vicente, A. (1970). *Dialectología española*. Madrid: Gredos.

Víctor Lara Bermejo
Universidad de Cádiz,
Facultad de Filosofía y Letras,
Avenida Doctor Gómez Ulla, s/n,
11003 Cádiz (España)

ATOMICITY AND COUNTABILITY IN LINGUISTICALLY-INFORMED INSTRUCTION: ACQUISITION OF L2 ENGLISH NOUN TYPES

Dakota J. THOMAS-WILHELM¹
University of Illinois System

Elisabet PLADEVALL-BALLESTER²
Universitat Autònoma de Barcelona

Abstract

While the literature shows that there is strong support for the effects of explicit instruction on explicit language knowledge, there has been little research on the effects of different types of explicit instruction on language acquisition. This study attempts to bridge the gap between Instructed Second Language Acquisition and Generative Second Language Acquisition by creating an instructional context that employs Generative theories and findings in order to teach L2-English articles and noun types. The study investigates L1-Mandarin learners in three different instructional contexts (linguistically-informed instruction, traditional instruction, and no extra instruction) acquiring countable and uncountable nouns in an indefinite-singular article context in L2 English. As measured by an elicited-sentence imitation task and an acceptability judgment task at pre-test, immediate posttest, and three-week-delayed post-test, the results find that learners receiving explicit instruction on the semantic features needing to be reassembled make the greatest gains over the duration of the study. These findings suggest that if linguistically-informed instruction were implemented in a systematic way throughout an entire grammar course, it may lead to greater linguistic gains in a shorter amount of time.

Keywords: atomicity; countability; Generative Second Language Acquisition; Instructed; Second Language Acquisition; Linguistically-Informed Instruction; Language Pedagogy

1. dakotat@uillinois.edu.  <https://orcid.org/0000-0002-4845-0161>
2. elisabet.pladevall@uab.cat.  <https://orcid.org/0000-0002-2793-9179>

ATOMICIDAD Y CONTABILIDAD EN LA ENSEÑANZA LINGÜÍSTICAMENTE INFORMADA: ADQUISICIÓN DE LOS TIPOS DE NOMBRES EN INGLÉS COMO L2

Resumen

Si bien la bibliografía muestra evidencias de los efectos de la instrucción explícita en el conocimiento explícito de la lengua, ha habido poca investigación sobre los efectos de los diferentes tipos de instrucción explícita sobre la adquisición de segundas lenguas. Este estudio pretende acercar dos enfoques (la adquisición de segundas lenguas en contextos de instrucción y la adquisición de segundas lenguas desde una perspectiva generativista) mediante la creación de un contexto de instrucción que emplea la teoría generativista para la enseñanza del uso de los artículos con los diferentes tipos de nombres en inglés L2. El estudio analiza tres contextos diferentes (instrucción lingüísticamente informada, instrucción tradicional y no instrucción) en la adquisición de sustantivos contables y no contables precedidos del artículo indefinido singular en inglés L2 por parte de hablantes de chino mandarín. Mediante una tarea de imitación de oraciones y una tarea de juicios de aceptabilidad desarrolladas en tres momentos (antes de la intervención pedagógica, inmediatamente después y al cabo de tres semanas), los resultados muestran que los estudiantes que recibieron instrucción explícita sobre las características semánticas que necesitan ser reestructuradas en la adquisición de los tipos de sustantivos mejoran de manera más significativa. Estos hallazgos sugieren que la instrucción lingüísticamente informada implementada de manera sistemática a lo largo de todo un curso de gramática podría conducir a un mayor aprendizaje en un período de tiempo más corto.

Palabras clave: atomicidad; contabilidad; adquisición generativa de segundas lenguas; adquisición de segundas lenguas en contextos de instrucción; instrucción lingüísticamente informada; pedagogía lingüística

RECIBIDO: 10/09/2022

APROBADO: 01/03/2023

1. INTRODUCTION

The general consensus within second language (L2) grammar instruction is that explicit instruction is more effective than implicit instruction (Norris & Ortega, 2000; Spada & Tomita, 2010). What remains less clear is what type of knowledge is developed from explicit instruction (Norris & Ortega, 2000) and why certain linguistic forms are more difficult to acquire than others (Slabakova, 2009b, 2013). Furthermore, functional morphology which often lacks saliency and can be syntactically and semantically complex (e.g., English noun types) can prove difficult for

L2 learners (Slabakova, 2013). Within L2 acquisition research, there exists a gap between Generative Second Language Acquisition (GenSLA) theoretical findings and Instructed Second Language Acquisition (ISLA) pedagogy. To date, only a handful of studies have been conducted and informed by GenSLA research (Lopez, 2017; Lopez & Sabir, 2017; Sabir, 2018; Umeda, Snape, Yusa, & Wilstshier, 2017), and there still exists a need to link theoretical GenSLA research and ISLA pedagogy (Marsden & Slabakova, 2019; Whong, Gil, & Marsden, 2013). The current study was born out of this need for further research that encourages links between GenSLA and ISLA paradigms. In our study, linguistically-informed instruction refers to a method of instruction that directly applies GenSLA theoretical findings in an ISLA pedagogical environment. The study reports on the development of this instructional context based on GenSLA theories of Feature Reassembly (Lardiere, 2008, 2009a, 2009b), the Bottleneck Hypothesis (Slabakova, 2008, 2009a, 2009b, 2019), and the Cline of Difficulty (Cho, 2012; Slabakova, 2009a, 2013).

In our study, we create an instructional intervention for L1-Mandarin, L2-English learners in a classroom setting, which specifically teaches the semantic features of English noun types. Previous research (Cho & Slabakova, 2014; Choi & Ionin, 2017; Choi, Ionin, & Zhu, 2018; Choi, Zhu, & Ionin, 2019; Hua & Lee, 2005; Slabakova, 2013) has found that the underlying semantic feature settings of noun types, that is [\pm count] and [\pm atomic], prove to be particularly problematic in the acquisition of English noun types for learners from generalized classifier languages. The specific aim of this study is to investigate the effects of different types of instruction on the non-native English acquisition of noun type distinctions in different article contexts by Mandarin-speaking students.

The article starts with a brief overview of atomicity and countability research in noun type acquisition and instruction, and an examination of recent GenSLA theories on the reassembly of L1 features for the L2 and the acquisition of functional morphology. We then present the details of the study, our results, and the primary findings. This paper concludes with considerations of limitations and implications for both further research and future language pedagogy.

2. LITERATURE REVIEW

When the English noun type distinctions are examined, their complexity is highlighted and demonstrates why errors are widely reported, even amongst advanced L2 learners. Within the theoretical literature, there is wide agreement surrounding atomicity and countability in terms of what these features mean and

how they are represented in languages that do not have a countable/uncountable distinction with their nouns. The definition of countability adopted in the current study comes from Gillon (1992), which places the denotations of both countable and uncountable nouns in the same domain with the features [+count] and [-count]. While the [\pm count] feature can easily separate countable nouns from uncountable nouns, it is not sufficient to distinguish between the two different uncountable noun types in English. In order to differentiate between substance- and object-uncountable nouns, which are both [-count], in English, we also need to consider atomicity. The definition of atomicity adopted in the current study is from Choi and Ionin (2017, p. 1): «A noun is atomic iff it cannot be divided into smaller parts which still bear the property denoted by the NP.» In other words, a noun is [+atomic] if you can divide it into smaller units which would still denote the same properties as the original noun. In English, these semantic features [\pm atomic] and [\pm count] are combined to make three different noun types, as in (1)–(3).

- (1) [+count, +atomic] Countable Nouns: *dog, cat*
- (2) [-count, +atomic] Uncountable-Object Nouns: *furniture, money,*
- (3) [-count, -atomic] Uncountable-Substance Nouns: *wine, cheese*

2.1. Noun Type Acquisition

English noun type distinctions are notoriously difficult to acquire by L2 learners whose L1s are Generalized Classifier Languages (GCLs). Previous research (Choi & Ionin, 2017; Choi *et al.*, 2018; Choi *et al.*, 2019) shows that GCLs, such as Mandarin, do not have a fully grammaticized distinction between countable and uncountable nouns. According to Cheng & Sybesma (1998, 1999, 2014) the link between atomicity and morphosyntax of nouns in GCLs is more direct than the countability distinction. While noun types in their bare form (i.e., without a classifier) provide a distinction between [+atomic] and [-atomic], the classifier system is grammaticized for a distinction between [+count] and [-count] features. In addition, while countability is encoded directly on the morphosyntax of the Mandarin classifier system, plural marking in Mandarin is restricted to [+human] nouns and can also be morphologically realized in the demonstrative determiners modifying nouns (Cheng & Sybesma, 1998, 1999, 2014). See examples (4)–(6) below from Choi *et al.* (2018, p. 154).

- (4) *shu*(*-men)
 book(-PL)
 «book(s)»
- (5) *jiaju*(*-men)
 furniture(-PL)
 «furniture»
- (6) *shui*(*-men)
 water(-PL)
 «water»

Only a handful of studies have investigated the acquisition of English noun types (Choi & Ionin, 2017; Choi *et al.*, 2018; Choi *et al.*, 2019; Hua & Lee, 2005; Inagaki, 2013). Hua and Lee (2005) conducted a study with L1-Mandarin, L2-English participants investigating their sensitivity to [\pm count] feature settings with articles and quantifiers in count-selective and mass-selective contexts. Using a grammaticality judgment task (GJT) and forced-choice elicitation task (FCET), they found that L1-Mandarin, L2-English learners demonstrated sensitivity to different nouns presented in count-selective and mass-selective contexts but did not make the same distinctions solely on a word-by-word basis. A study by Inagaki (2013) investigated quantity judgments by L1-Japanese, L2-English speakers of countable and uncountable nouns using a picture-matching task (PMT) based on Barner and Snedeker (2005). Inagaki (2013) found that L2-English learners performed similarly to NSs in basing their quantity judgments on number for [+count] nouns and volume for [-count] nouns. Furthermore, they found that the interpretation of flexible nouns was not altered by changes in syntax.

In a series of studies conducted by Choi and colleagues (Choi & Ionin, 2017; Choi *et al.*, 2018; Choi *et al.*, 2019), they investigated both countability and atomicity in L1-Korean and L1-Mandarin, L2-English speakers. Using a self-paced reading task (SPRT) and a GJT, Choi and Ionin (2017) found that L2-English learners displayed sensitivity to plural marking on [-count, -atomic] nouns but not [-count, +atomic] nouns as measured by the GJT and SPRT. This led them to conclude that L2-English learners from GCLs struggle with the incompatibility between plural marking and atomicity in English. With a cloze task, Choi *et al.* (2018) found that L2-English learners correctly used -s with [+count] and optionally overuse it with [-count, +atomic] nouns but do not use it with [-count, -atomic] nouns. Finally, Choi *et al.* (2019) used a PMT to investigate [\pm count] and object-/kind-reference. In their results, they found that L1-Korean and L1-Mandarin learners

of English allow plural marking on [-count, +atomic] nouns to denote objects to a greater extent than English native speakers.

2.2. *Generative Second Language Acquisition*

In this study, the acquisition of noun-types will be approached through the Feature Reassembly Hypothesis (Lardiere, 2008, p. 4), which states that, «the ways in which grammatical features are morphologically combined and conditioned may well affect their acquirability and overt realization in SLA.» One of the primary assumptions of this proposal is that learners look for morphological correspondences between their L1 and L2. Under her hypothesis, the assembly of features in the L2 causes greater difficulty when they differ from the assemble in the L1. Thus, depending on how features are assembled (either overt or covert realization is possible) and conditioned, some features may be just as difficult to acquire as others.

Following on the proposal for the Feature Reassembly Hypothesis (Lardiere, 2008, 2009a, 2009b), Slabakova (2008, 2009a, 2009b) makes predictions for learnability based on what is 'easy' and what is 'hard' to acquire. She predicts that learning situations where re-assembly of grammatical features is required are more challenging than situations in which simple remapping of L1 to L2 morphemes are required. She calls this the Bottleneck Hypothesis (Slabakova, 2008, 2009a, 2009b) and the Cline of Difficulty (Slabakova, 2009a). According to the Bottleneck Hypothesis, language learnability lies within language variation. Slabakova (2016) goes on to say that the motivation for her proposal is practical: areas of grammar that are more difficult should get more instructional effort and attention. The 'Bottleneck' of language learning, therefore, lies in functional morphology.

Within the Bottleneck Hypothesis (Slabakova, 2008, 2009a, 2009b), the Cline of Difficulty (Cho & Slabakova, 2014; Slabakova, 2009a) makes specific predictions on the degrees of difficulty in feature mapping, where different language backgrounds and target languages pose different levels of difficulty in their learning situations. According to the Cline of Difficulty (Cho & Slabakova, 2014; Slabakova, 2009a), morpheme to morpheme mapping with no reassembly is the easiest to acquire, while morpheme to morpheme with reassemble is expected to be more difficult. In Mandarin, the countability and atomic features are morphologically realized, but they are assembled differently than English. Therefore, L1-Mandarin/L2-English learners must map their L1 features to English and then reassembly them strictly onto nouns.

The Feature Re-Assembly Hypothesis by Lardiere (2008, 2009a, 2009b) assumes that the learners will assess their interlanguage grammars and rebuild the features to become more native-like. While it may be possible to create a steady-state grammar that is similar to a NS, their ability to produce structures that are similar to those of the of a NS may take longer to develop. Additionally, the Bottleneck Hypothesis (Slabakova, 2008, 2009a, 2009b) predicts that functional morphology is especially difficult for L2 learners and that this will considerably slow down their acquisition process. Therefore, it is predicted that learning the functional morphology of English noun types will be particularly difficult for these L2 learners, as the L2 grammar is much different from the L1 grammar. This highlights the importance of what requires more focus in language instruction.

Having established the GenSLA theoretical background, we now turn our attention to ISLA and the previous research on linguistically-informed instruction before describing the study and presenting our results.

2.3. *Instructed Second Language Acquisition*

In the SLA literature, ISLA research regularly tackles questions such as *Which type of instruction is of most use?* and *Which language features most benefit from classroom input?* It is generally agreed upon in the literature that L2 grammar instruction can be effective (Norris & Ortega, 2000; Spada & Tomita, 2010) although some structures, such as English articles and noun types, are notoriously problematic. In the literature, many studies have investigated the overall effects of implicit or explicit grammar instruction. Implicit instruction generally involves exposing the learners to a target structure without any attention to rules or negative feedback, whereas explicit instruction focuses on the rules. Norris and Ortega (2000) and Spada and Tomita (2010) found there is widespread support for providing some form of explicit instruction during language lessons for both simple and complex grammar forms, although they were unable to conclude whether instruction can improve implicit knowledge or just explicit. According to Ellis (2009), there are a number of differences between implicit and explicit knowledge, including awareness, access, and use of the linguistic knowledge. Furthermore, Loewen (2015) states that recent research in ISLA has been informing the question of «whether explicit instruction leads to the acquisition of explicit knowledge in L2 learners» (p. 31). While it is generally assumed that explicit knowledge is a result of explicit instruction (Doughty, 2003), the effect of different types of explicit instruction on explicit knowledge is still unknown. While there continue to be disagreements in

the literature as to whether explicit knowledge can become implicit knowledge, it has yet to be determined if explicit instruction that targets explicit knowledge may or may not have long-term effects on implicit knowledge of forms.

Previous research on the effects of instruction have only looked at traditional implicit and explicit instruction. Bruhn de Garavito (2013) further examined the effect of formal instruction and argues that the methodological decisions surrounding grammar instruction often depend on the theoretical position about the role of input. She relates L2 Spanish object pronoun instruction with GenSLA empirical research finding about the acquisition of object pronouns. She investigates how they are currently taught in language textbooks and how input that learners receive might be improved. Bruhn de Garavito identifies differences between English and Spanish object pronouns that cause confusion to L2 Spanish learners. She considers these differences in how they relate to the findings in GenSLA research on clitic position, and further suggests how these findings might be applied to language teaching. The author concludes that the most important takeaway message should be that «research carried out within the generative tradition has something to say regarding language teaching» (p. 32). She identifies this as a two-way street, though. While GenSLA empirical findings can inform language pedagogy, language pedagogy should also inform the research that is conducted within GenSLA, and that all the work in both fields needs to be more comprehensible to those in the other field.

In this study, the type of explicit instruction used, linguistically-informed instruction, is relatively new and there are no instructional intervention studies, to date, that have investigated the effects of instruction with regard to L2-English noun types. The majority of previous research on English noun types has investigated quantity judgments of countable and uncountable nouns with children and L2 learners. With mixed results on the effects of linguistically-informed instruction on the acquisition of L2-English features in the literature review, and no studies to date on the effects of linguistically-informed instruction of noun types, the present study adds to this limited field by looking at the development and reassembly of English noun type semantic features in an indefinite-singular context. We now turn out attention to the study, which will explain the participants, the methods of data collection, and the data analysis.

3. THE STUDY

The aim of this study is to test previous proposals (Lopez, 2017; Sabir, 2018; Snape & Yusa, 2013) that make recommendations for innovating teaching material (Marsden & Slabakova, 2019) by investigating the acquisition of L2-English noun types in an indefinite-singular context (e.g., *a dog*, **a furniture*, **a toothpaste*). Using a pre-/post-test design, this study investigates the effect of instruction on L1-Mandarin speakers in three different instructional contexts—linguistically-informed instruction (LING), traditional instruction (TRAD), and no extra instruction (NOEX). We attempt to do this and explore the overarching question of whether or not it is beneficial for L2 acquisition to create a teaching and learning environment that is informed by both GenSLA research and instructed second language acquisition (ISLA) pedagogy. Informed by the Feature Reassembly Hypothesis (Lardiere, 2008, 2009a, 2009b), the Bottleneck Hypothesis (Slabakova, 2008, 2009a, 2009b), and the Cline of Difficulty (Slabakova, 2009a), this new instructional pedagogy teaches L2-English noun types through semantic universals: [\pm atomic] and [\pm count]. We investigate the effects of instruction through two primary research questions:

1. What is the effect of each type of instruction on the reassembly of L1 noun type features for the L2?
2. What type of instruction will lead to greater gains in the reassembly of L1 noun type features for the L2 at both immediate and delayed post-test?

In the literature, there is positive evidence in support of the effects of explicit instruction (Bowles, 2011; Bruhn de Garavito, 2013; Loewen, 2015; Long, 1983, 2009; Lopez, 2017; Lopez & Sabir, 2017; Master, 1994, 2002; Norris & Ortega, 2000; Snape & Yusa, 2013; Sonbul & Schmitt, 2013; Spada & Tomita, 2010); therefore, it is predicted overall that explicit instruction will facilitate the reassembly of the L1 [count] and [atomic] features for the L2. In other words, following explicit instruction, learners should realize that the indefinite article context is count-noun-selective, therefore only permitting singular countable nouns, that is nouns with the feature settings [+count, +atomic]. Following the Cline of Difficulty (Slabakova, 2009a) and the Bottleneck Hypothesis (Slabakova, 2008, 2009a, 2009b), it is predicted that the reassembly of the [count] and [atomic] features will pose moderate difficulty because of the differences in their morphological realizations in the L1 and L2. It is predicted that instruction will mediate this reassembly process. Furthermore, it is predicted that linguistically-informed instruction (Lopez, 2017;

Lopez & Sabir, 2017; Sabir, 2018), by the development of metalinguistic knowledge via teaching abstract linguistic structures, will foster an environment for more successful reassembly of the [count] and [atomic] features than traditional and no-extra-instruction learners.

3.1. Participants

The participants consisted of 65 L1-Mandarin, L2-English learners who, at the time of data collection, were enrolled in a 16-week credit-based ESL course at a university in the Midwestern United States. While all of them were enrolled in a grammar course, many of the students were also enrolled in one or more of the other skill courses (reading, writing, listening, or speaking). It is important to note here that the grammar structures covered in this study are not explicitly instructed in these classes, but rather assigned as «self-study» units to the student with no assessments or knowledge checks given over the content/structures. These L2 learners were divided into three participant groups based on the type of instructional intervention they received. Prior to participation in the study, all participants completed a bio-data and language use questionnaire (BLUQ) and the Oxford Quick Placement Test (OQPT) (U. C. L. E. Syndicate, 2001). Table 1 below shows the descriptive statistics of the participants.

	LINGUISTICALLY-INFORMED INSTRUCTION	TRADITIONAL INSTRUCTION	NO EXTRA INSTRUCTION
NUMBER	30 (11 male, 19 female)	18 (9 male, 9 female)	17 (5 male, 12 female)
AGE	17-30 (M = 20.38)	17-22 (M = 19.39)	18-25 (M = 19.47)
OQPT SCORE	20-48 (M = 35.9)	19-52 (M = 33.94)	13-45 (M = 31.94)

Table 1. Participant descriptive statistics

As can be seen in Table 1, there was an imbalance of the number of participants in each group. While some might view this as problematic, the majority of analysis is conducted within group, and any between group analysis in the results is conducted as a gain score analysis, which significantly decreases any effect of number of participants between groups.

The LING group received one-hour of instructional intervention that was informed by GenSLA theory and used the semantic features of the noun types in teaching the differences in feature settings for English articles and noun types, as well as crosslinguistic differences between Mandarin and English. The TRAD group received one-hour of instructional intervention on English articles and noun

types using their assigned grammar textbook, *Grammar & Beyond 4* (Bunting, Diniz, & Reppen, 2009). The third group, NOEX, did not receive any instructional intervention as part of the research study. It must be acknowledged that the decision to only include one-hour of intervention was to maintain an instructional environment which is similar to that of an actual classroom. In the grammar courses from which the participants were recruited, any curricular item is only even given approximately one to two hours of explicit instruction and practice in the classroom. All instructional intervention took place in the hour immediately preceding the immediate post-test data collection session.

3.1. *Instruments and Data Collection*

This study used a pre-test–intervention–post-test design, with a delayed post-test; see Table 2 below. The same tasks were completed at all three data collection times. The pre-test (T₀) was conducted one week before the instructional intervention, and the immediate post-test (T₁) was conducted immediately following the instructional intervention. As aforementioned, the delayed post-test (T₂) was conducted three weeks following T₁. Two tasks were used at each point: an elicited-sentence imitation task and an acceptability judgment task.

WEEK	LING GROUP	TRAD GROUP	NOEX GROUP
0	Pre-participation questionnaires (BLUQ & OQPT)		
1	To data collection (ESIT & AJT)		
2	1-hour instruction using linguistically-informed materials + T ₁ data collection (ESIT & AJT)	1-hour instruction using textbook + T ₁ data collection (ESIT & AJT)	No extra instruction + T ₁ data collection (ESIT & AJT)
3			
4			
5	T ₂ data collection (ESIT & AJT)		

Table 2. Data collection procedure for all groups

The elicited-sentence imitation task (ESIT) and acceptability judgment task (AJT) each consisted of 36 items: 3 nouns for each noun type (countable, uncountable-object, and uncountable-substance) with each noun type in the countable-noun-selective indefinite context *a* __. The tasks also included 24 items that tested the same noun types in definite-singular and definite-plural contexts, but those results are not reported here. All data collection was done in an on-campus computer lab, and both tasks were administered in an online form using IBEX Farm (<https://spellout.net/ibexfarm/>). Although ESITs have been greatly debated they have been

proven to be a valid task in collecting data on a speaker's implicit knowledge (Ellis *et al.*, 2009; Spada, Shiu, & Tomita, 2015). As Ellis (2005b, p. 45) states, participants' «failure to imitate a sentence at all or to reproduce it in such a form that they did not create an obligatory context for the target structure of a sentence [is] coded as 'avoidance'.» If a participant appropriately imitates a sentence where the target structure is correctly supplied, then it is accepted that the form or structure has become part of the learner's implicit knowledge.

In the ESIT, participants heard a sentence and were asked to repeat the sentence using correct English following a 3000ms delay. Each sentence was opinion-oriented. During the 3000ms delay, the participants were asked to mark whether they agreed or disagreed with the content of the sentence. This was included to encourage the participants to focus on the meaning of each statement and prevent rote repetition. Example items can be seen in examples (7)–(9) below.

- (7) [+count, +atomic] Countable Nouns
Everyone likes a boy as their first child.
- (8) [–count, +atomic] Uncountable-Object Nouns
*Americans do not have a furniture in their house.
- (9) [–count, –atomic] Uncountable-Substance Nouns
*Mexicans add a garlic to their tacos.

Participant imitations were audio recorded and transcribed by the researcher. Following a similar rating scale to Ellis *et al.* (2009), each imitated sentence was scored depending on their ability to correct the sentence. An imitated statement was given a 2 if the statement was grammatically imitated with an appropriate article and noun; a 1 was given if the statement was imitated with either an appropriate article or noun; and a 0 was given if the sentence was not imitated grammatically.

The AJT in this study was untimed and targeted the same article and noun type combinations as the ESIT for a total of 36 items (different sentences than those included in the ESIT); see examples (10)–(12). Participants were asked to rate the acceptability of each item on a 7-point Likert scale ranging from *very acceptable* (7) to *not acceptable at all* (1).

- (10) [+count, +atomic] Countable Nouns
There is a boy sitting on the ground with his hands on his head.
- (11) [–count, +atomic] Uncountable-Object Nouns
*We have a furniture in that room

(12) [-count, -atomic] Uncountable-Substance Nouns

*Although they think he is crazy, he puts a salt on everything.

3.3. *Instructional Intervention*

The instructional intervention consisted of a single 60-minute lesson for each the LING and TRAD groups. As explained above, the single one-hour intervention was chosen to mirror the amount of time curricular items are given in this particular skill-based program. For both the LING and TRAD learners, a form-focused explicit instruction approach was used to match what is normally done in their regular grammar courses. All instruction took place in a university classroom adjacent to the computer lab used for data collection. The instructional intervention was conducted in the second week of the 5-week study.

The NOEX group did not receive any instructional intervention on English articles or noun types as part of the study. In addition, this unit of the textbook is assigned as self-study, so these participants did not receive explicit instruction on these topics in their regularly-scheduled grammar courses. The TRAD group were explicitly taught about article choice with countable and uncountable nouns using a form-focused approach and their course-assigned textbook, *Grammar & Beyond 4* (Bunting *et al.*, 2009). Grammar structures and explanations presented in this book take a corpus-based approach, meaning that the grammar presentations are based on an analysis of an English corpus and should represent real-world use. During the instructional intervention, the grammar structures and explanations were presented, followed by dedicated time for practice using the exercises in both the textbook and workbook. Examples of the textbook definitions from *Grammar & Beyond 4* (Bunting *et al.*, 2009, p. 105) can be found in (13)–(15).

- (13) Use *a/an* with a singular count noun when the noun is not specifically identified or when it is first mentioned and new to the reader.
- (14) Do not use *a/an* with noncount or plural nouns. Use *some* or \emptyset .
- (15) Use *a/an* when introducing a count noun.

The LING group received instruction on the semantics of English articles and noun types using newly-created teaching materials. Motivated by previous research (Lopez, 2017; Lopez & Sabir, 2017; Sabir, 2018; Snape & Yusa, 2013), definitions of articles and noun types were presented by first defining the individual feature settings and then giving the explicit lexical forms. To facilitate

with feature reassembly and the bottleneck of language learning, learners were taught how these features are combined to make the various grammatical article and noun type combinations in English. Examples of the pedagogical definitions for atomicity and countability can be found in (16). After the presentation of the definitions of the semantic features, examples of their settings and corresponding noun types were presented as in (17).

- (16) Nouns: a word that refers to a person place, thing, or idea
- a. [\pm count]: the feature of a noun that determines whether or not it can have a plural form
 - b. [\pm atomic]: the feature of a noun that determines if it is made up of individuals
- (17) If a noun is...
- c. [+count, +atomic]:
 - i. In English, there is one type of nouns that is [+count, +atomic]:
 1. Countable Nouns: it has separate singular and plural forms, and the plural form takes a morphological -s; these nouns can be counted individually (e.g., *dog, dogs*)
 - d. [-count, \pm atomic]:
 - i. In English, there are two types of nouns that are [-count]:
 1. Uncountable-Substance Nouns [-atomic]: it refers to a substance that cannot be individuated or counted (e.g., *toothpaste*)
 2. Uncountable-Object Nouns [+atomic]: it refers to a set of objects that can be individuated into countable items (e.g., *furniture*)

The final part of the instructional intervention materials for LING learners, the participants received explicit instruction on notable differences between Mandarin and English grammar and mistakes commonly made by L1-Mandarin, L2-English language learners. Throughout the lesson, learners were encouraged to ask questions, and there were no apparent issues with their understanding of the concepts presented.

4. RESULTS

In order to explore the first research question, that is the effects of type of instruction on L2 acquisition, a two-way (3 noun types x 3 testing times) Repeated Measures Analysis of Variance (RM ANOVA) and Tukey Honest Significant Difference (HSD) post-hoc analysis were conducted individually for each group on the ESIT accuracy scores and AJT acceptability ratings. To investigate which method

of instruction led to greater gains, we conducted a series of two-way (3 noun types x 3 learner groups) factorial RM ANOVAs on mean gain scores for each of the tasks and testing periods.

4.1. *Elicited-Sentence Imitation Task*

Pre-test (T₀), immediate post-test (T₁), and delayed post-test (T₂) results for the three learner groups and each noun type examined in the indefinite singular context can be found in Tables 3, 4, and 5, showing the mean percentage accuracy in their imitations.

In the majority of the data, there is an increase in imitation scores. Descriptive statistics show that LING learners improved in all noun types from T₀ to T₁ and T₀ to T₂. The RM ANOVA showed a significant effect of time ($F_{2,801} = 30.50, p < .001, \eta_p^2 = .071$) between pre- and post-test scores. There was also a significant effect of noun type ($F_{2,801} = 56.88, p < .001, \eta_p^2 = .124$) meaning that there were differences in how each noun type was treated irrespective of time. Finally, there was also a significant interaction of time and noun type ($F_{4,801} = 2.69, p < .05, \eta_p^2 = .013$), meaning that time affected the relationship between knowledge of noun types. A Tukey post-hoc analysis revealed a significant increase in overall imitation scores from T₀ to T₁ (mean difference = 13.52, $p < .001$) and from T₀ to T₂ (mean difference = 14.81, $p < .001$). The post-hoc analysis revealed significant increase in imitations scores from T₀ to T₁ for object-uncountable (mean difference = 12.22, $p < .05$) and substance-uncountable nouns (mean difference = 18.33, $p < .001$). Similar results were also found T₀ to T₂ for both noun types (object-uncountable: mean difference = 21.11, $p < .001$); substance-uncountable: mean difference = 15.56, $p < .001$).

NOUN TYPE	LING (n = 30)		
	T ₀	T ₁	T ₂
[+count, +atomic] a dog	87.22 (sd = 25.49)	97.22 (sd = 11.52)	95.00 (sd = 16.84)
[-count, +atomic] *a furniture	61.11 (sd = 24.61)	73.33 (sd = 28.24)	82.22 (sd = 25.21)
[-count, -atomic] *a toothpaste	64.44 (sd = 28.29)	82.78 (sd = 27.19)	80.00 (sd = 26.82)

Table 3. Mean percentage of accuracy of imitation for LING learners in the elicited-sentence imitation task

In looking at the post-hoc analysis of the interactions for LING learners, the analysis revealed that prior to instruction, LING learners displayed significantly higher imitation scores for countable nouns than object- (mean difference = -26.11, $p < .001$) and substance-uncountable (mean difference = -22.78, $p < .001$) nouns. At T₁, the post-hoc analysis found that LING learners still performed significantly better with countable nouns than object- (mean difference = -23.89, $p < .001$) and substance-uncountable (mean difference = 14.44, $p < .01$) nouns.

NOUN TYPE	TRAD (n = 18)		
	T ₀	T ₁	T ₂
[+count, +atomic] a dog	80.56 (sd = 31.35)	89.81 (sd = 26.39)	90.91 (sd = 26.38)
[-count, +atomic] *a furniture	67.59 (sd = 27.74)	67.59 (sd = 29.39)	73.15 (sd = 28.67)
[-count, -atomic] *a toothpaste	63.89 (sd = 26.45)	69.44 (sd = 35.58)	65.74 (sd = 36.11)

Table 4. Mean percentage of accuracy of imitation for TRAD learners in the elicited-sentence imitation task

Table 4 shows the descriptive statistics for TRAD learners. In the descriptive statistics for the TRAD learners' ESIT data, we can see learners' improvement in two noun types (countable and substance) from T₀ to T₁ and all noun types from T₀ to T₂. The RM ANOVA revealed a significant effect of noun type ($F_{2,456} = 19.61, p < .001, \eta_p^2 = .079$) but not for time ($F_{2,456} = 1.18, p = .301, \eta_p^2 = .007$) nor the interaction of noun type and time ($F_{4,456} = .59, p = .67, \eta_p^2 = .005$).

The post-hoc analysis revealed that TRAD learners treated object-uncountable nouns differently than countable nouns at T₁ (mean difference = -22.22, $p < .01$), and substance-uncountable nouns different than countable nouns at T₁ (mean difference = -20.37, $p < .05$) and T₂ (mean difference = -25.17, $p < .01$).

NOUN TYPE	NOEX (n = 17)		
	T ₀	T ₁	T ₂
[+count, +atomic] a dog	88.24 (sd = 21.42)	95.10 (sd = 18.04)	84.31 (sd = 30.81)
[-count, +atomic] *a furniture	64.71 (sd = 25.09)	67.65 (sd = 27.97)	61.76 (sd = 32.54)
[-count, -atomic] *a toothpaste	62.75 (sd = 29.74)	80.39 (sd = 28.42)	67.65 (sd = 35.81)

Table 5. Mean percentage of accuracy of imitation for NOEX learners in the elicited-sentence imitation task

Table 5 shows the descriptive statistics for NOEX learners. Unexpectedly, NOEX learners improved with all noun types from T₀ to T₁ but displayed little change from T₀ to T₂. For NOEX learners, the RM ANOVA found a significant effect of time ($F_{2,450} = 5.76, p < .01, \eta_p^2 = .0249$) and noun type ($F_{2,450} = 31.68, p < .001, \eta_p^2 = .123$) but not the interaction of time and noun type ($F_{4,450} = 1.00, p = .407, \eta_p^2 = .009$) meaning that the difference between pre- and post-test scores was significant as well as the differences between noun types but not the interaction of these variables. In other words, there was overall improvement irrespective of noun type and there are differences in noun types irrespective of time, but the analysis found no noun type that was better than another at a specific time.

In looking at the post-hoc analysis for noun types for NOEX learners, the analysis revealed that, prior to instruction, these learners performed significantly better with countable nouns than object- (mean difference -23.53, $p < .01$) and substance-uncountable (mean difference -25.49, $p < .001$) nouns. For countable nouns, NOEX learners continued to perform significantly better than object-uncountable nouns at T₁ (mean difference = -27.45, $p < .001$) and T₂ (mean difference = -22.55, $p < .01$). Furthermore, from T₀ to T₁, NOEX learners performed significantly worse with regard to substance-uncountable nouns (mean difference = 17.65, $p < .05$).

In order to investigate which learner group made the greatest gains, the mean percentage gains from T₀ - T₁, T₀ - T₂, and T₁ - T₂ were compared across groups. The descriptive statistics are presented in Tables 6–8.

NOUN TYPE	LING (n = 30)		
	T ₀ - T ₁	T ₀ - T ₂	T ₁ - T ₂
[+count, +atomic]	10.00	7.78	-2.22
a dog	(sd = 16.14)	(sd = 17.36)	(sd = 10.48)
[-count, +atomic]	12.22	21.11	8.89
a furniture	(sd = 20.96)	(sd = 21.86)	(sd = 23.05)
[-count, -atomic]	18.33	15.56	-2.78
*a toothpaste	(sd = 20.22)	(sd = 26.60)	(sd = 23.60)

Table 6. Mean gain scores for LING learners in the elicited-sentence imitation task

NOUN TYPE	TRAD (n = 18)		
	T ₀ - T ₁	T ₀ - T ₂	T ₁ - T ₂
[+count, +atomic]	9.26	9.09	1.52
a dog	(sd = 25.71)	(sd = 25.13)	(sd = 22.92)
[-count, +atomic]	0.00	5.56	5.56
*a furniture	(sd = 21.39)	(sd = 20.61)	(sd = 20.61)
[-count, -atomic]	5.56	1.85	-3.70
*a toothpaste	(sd = 21.39)	(sd = 22.06)	(sd = 21.81)

Table 7. Mean gain scores for TRAD learners in the elicited-sentence imitation task

NOUN TYPE	NOEX (n = 17)		
	To - T1	To - T2	T1 - T2
[+count, +atomic] a dog	6.86 (sd = 16.73)	-3.92 (sd = 28.58)	-10.78 (sd = 37.60)
[-count, +atomic] *a furniture	2.94 (sd = 19.75)	-2.94 (sd = 26.51)	-5.88 (sd = 24.25)
[-count, -atomic] *a toothpaste	17.65 (sd = 19.96)	4.90 (sd = 24.84)	-12.75 (sd = 24.67)

Table 8. Mean gain scores for NOEX learners in the elicited-sentence imitation task

The RM ANOVA for To to T1 found a significant effect of learner type ($F_{2,186} = 3.10, p < .05, \eta_p^2 = .032$) but not for noun type ($F_{2,186} = 2.8, p = .063, \eta_p^2 = .029$) nor the interaction of learner type and noun type ($F_{4,186} = .966, p = .427, \eta_p^2 = .020$), meaning that there were significant differences between learner groups but not noun types. The Tukey post-hoc analysis revealed that LING learners made significantly greater gains on the ESIT than TRAD learners (mean difference = $-8.58, p < .05$). The post-hoc analysis did not reveal any other significant differences. Therefore, it appears that LING learners made greater gains than TRAD learners, regardless of noun type, from To to T1.

For To to T2, the RM ANOVA found a significant effect of learner type ($F_{2,179} = 7.61, p < .001, \eta_p^2 = .080$) but not for noun type ($F_{2,179} = 1.18, p = .309, \eta_p^2 = .013$) nor the interaction of these two variables ($F_{4,179} = 1.15, p = .335, \eta_p^2 = .025$). The post-hoc analysis revealed that LING learners made significantly greater gains than NOEX learners (mean difference = $-15.47, p < .001$). The mean difference in gains between LING and TRAD learners was only marginally significant (mean difference = $-9.85, p = .055$). In other words, from To to T2, LING learners made greater gains than both TRAD and NOEX learners regardless of noun type.

In the analysis of ESIT gains from T1 to T2, the RM ANOVA found a significant effect of learner type ($F_{2,179} = 4.66, p < .05, \eta_p^2 = .049$) and noun type ($F_{2,179} = 3.54, p < .05, \eta_p^2 = .038$) but not the interaction of these two factors ($F_{4,179} = .189, p = .944, \eta_p^2 = .004$). The Tukey post-hoc analysis found that irrespective of learner type, there was a significant difference between substance-uncountable and object-uncountable nouns (mean difference = $-9.74, p < .05$). Furthermore, LING learners made significantly greater gains than NOEX learners (mean difference = $-11.10, p < .05$). Unexpectedly, the post-hoc analysis also found that NOEX learners made significantly greater gains than TRAD learners (mean difference = $-10.87, p < .05$).

4.2. Acceptability Judgment Task

In the acceptability judgment task, participants were asked to rate the acceptability of sentences on a 7-point Likert scale. The raw acceptability ratings are used in the data analysis³. Pre-test (T₀), immediate post-test (T₁), and delayed post-test (T₂) results for the three learner groups and each noun type examined can be found in Tables 9–11, showing the acceptability ratings for each context.

NOUN TYPE	LING (n = 30)		
	T ₀	T ₁	T ₂
[+count, +atomic] a dog	4.79 (sd = 2.05)	5.16 (sd = 2.00)	5.27 (sd = 1.56)
[-count, +atomic] *a furniture	4.82 (sd = 2.00)	4.41 (sd = 2.14)	4.18 (sd = 2.16)
[-count, -atomic] *a toothpaste	4.49 (sd = 2.13)	4.26 (sd = 2.30)	4.63 (sd = 1.89)

Table 9. Mean acceptability ratings for LING learners in the acceptability judgment task

In the descriptive statistics, we see the data trending in the expected directions—acceptability ratings increase over time for countable nouns while they decrease for uncountable nouns. The descriptive statistics show that LING learners increase acceptability of countable nouns from T₀ to T₁ and T₂. For uncountable nouns, they decrease from T₀ to T₁ and T₂ for object-uncountable nouns and only decrease from T₀ to T₁ for substance-uncountable nouns. The RM ANOVA for LING learners revealed a significant effect of noun type ($F_{2,801} = 7.98, p < .001, \eta_p^2 = .020$) but not for time ($F_{2,801} = .017, p = .841, \eta_p^2 = .0004$) nor the interaction of noun type and time ($F_{4,801} = 2.15, p = .073, \eta_p^2 = .011$). The post-hoc analysis revealed that, at T₂, LING learners appropriately displayed significantly higher acceptability ratings for countable nouns than object-uncountable nouns (mean difference = -1.09, $p < .05$).

NOUN TYPE	TRAD (n = 18)		
	T ₀	T ₁	T ₂
[+count, +atomic] a dog	5.11 (sd = 1.60)	4.85 (sd = 1.70)	5.19 (sd = 1.43)
[-count, +atomic] *a furniture	5.2 (sd = 1.77)	4.76 (sd = 2.00)	4.20 (sd = 1.64)
[-count, -atomic] *a toothpaste	4.15 (sd = 2.12)	4.91 (sd = 1.72)	4.57 (sd = 1.60)

Table 10. Mean acceptability ratings for TRAD learners in the acceptability judgment task

3. With acceptability ratings, a lower rating is a sign of rejection of a grammatical or ungrammatical structure while a higher rating is a sign of acceptance.

In the descriptive statistics, we see that TRAD learners decreased their acceptability ratings for countable and object-uncountable noun types from To to T₁ and increased for substance-uncountable nouns, even though they should have increased only for countable nouns. From To to T₂, we see an increase in acceptability ratings for countable nouns and substance-uncountable nouns and a decrease for object-uncountable nouns. The RM ANOVA did not find a significant effect of time ($F_{2,477} = .555, p = .574, \eta_p^2 = .002$), meaning there were no differences between pre- and post-test scores. On the other hand, there was a significant effect of noun type ($F_{2,477} = 3.52, p < .05, \eta_p^2 = .015$), meaning that there were differences in how each noun type was treated. The Tukey post-hoc analysis only revealed a significant difference between object- and substance-uncountable nouns prior to instruction (mean difference = $-1.06, p < .05$), with object-uncountable nouns having a significantly higher mean acceptability rating. There were no other significant pairwise comparisons.

NOUN TYPE	NOEX (n = 17)		
	To	T ₁	T ₂
[+count, +atomic] a dog	4.84 (sd = 2.04)	4.90 (sd = 1.73)	5.12 (sd = 1.69)
[-count, +atomic] *a furniture	5.51 (sd = 1.83)	5.14 (sd = 1.64)	5.00 (sd = 1.92)
[-count, -atomic] *a toothpaste	4.98 (sd = 1.97)	4.65 (sd = 1.97)	4.75 (sd = 1.83)

Table 11. Mean acceptability ratings for NOEX learners in the acceptability judgment task

For NOEX learners, the descriptive statistics show a slight increase in acceptability scores from To to T₁ and T₂ for countable nouns and a decrease in scores from To to T₁ and T₂ for uncountable nouns (both object and substance). The RM ANOVA failed to reveal a significant effect of time ($F_{2,450} = .554, p = .575, \eta_p^2 = .003$), noun type ($F_{2,450} = 2.05, p = .130, \eta_p^2 = .009$), or their interaction ($F_{4,450} = .614, p = .653, \eta_p^2 = .005$). Being as there were no significant effects or interactions, there were also no significant post-hoc pairwise comparisons.

In order to investigate which learner group made the greatest gains, the mean accuracy rating differences from To to T₁, To to T₂, and T₁ to T₂ were compared across groups. The descriptive statistics are presented in Tables 12–14.

NOUN TYPE	LING (n = 30)		
	To - T1	To - T2	T1 - T2
[+count, +atomic]	0.37	0.48	0.11
<i>a dog</i>	(sd = 1.69)	(sd = 1.49)	(sd = 1.35)
[-count, +atomic]	-0.41	-0.64	-0.23
<i>*a furniture</i>	(sd = 1.64)	(sd = 1.28)	(sd = 1.39)
[-count, -atomic]	-0.23	0.14	0.38
<i>*a toothpaste</i>	(sd = 2.13)	(sd = 1.63)	(sd = 1.99)

Table 12. Mean gain scores for LING learners in the acceptability judgment task

NOUN TYPE	TRAD (n = 18)		
	To - T1	To - T2	T1 - T2
[+count, +atomic]	-0.26	0.07	0.33
<i>a dog</i>	(sd = 1.30)	(sd = 1.26)	(sd = 0.88)
[-count, +atomic]	-0.44	-1.00	-0.56
<i>*a furniture</i>	(sd = 1.20)	(sd = 1.41)	(sd = 1.80)
[-count, -atomic]	0.76	0.43	-0.33
<i>*a toothpaste</i>	(sd = 1.66)	(sd = 1.18)	(sd = 1.36)

Table 13. Mean gain scores for TRAD learners in the acceptability judgment task

NOUN TYPE	NOEX (n = 17)		
	To - T1	To - T2	T1 - T2
[+count, +atomic]	0.06	0.27	0.22
<i>a dog</i>	(sd = 1.39)	(sd = 1.07)	(sd = 1.30)
[-count, +atomic]	-0.37	-0.51	-0.14
<i>*a furniture</i>	(sd = 1.09)	(sd = 1.84)	(sd = 1.38)
[-count, -atomic]	-0.33	-0.24	0.10
<i>*a toothpaste</i>	(sd = 1.62)	(sd = 2.14)	(sd = 1.55)

Table 14. Mean gain scores for NOEX learners in the acceptability judgment task

In looking at the descriptive statistics, it can be seen that LING learners made the greatest gains when compared to the other learner groups. In other words, acceptability ratings of countable nouns increased at all testing times and they decreased, as predicted, for uncountable-object nouns. While the other learner groups have made similar gains, they are not to the same extent as LING learners. The RM ANOVA for gains from To to T1 failed to produce any significant main effects for learner type ($F_{2,186} = .276, p = .759, \eta_p^2 = .003$), noun type ($F_{2,186} = 1.93, p = .148, \eta_p^2 = .020$), or their interaction ($F_{4,186} = 1.63, p = .170, \eta_p^2 = .034$). Since the ANOVA did not find any significant main effects, the Tukey post-hoc analysis also failed to reveal any significant pairwise comparisons. For the To to T2 data, the RM ANOVA revealed a significant main effect of noun type ($F_{2,186} = 8.50, p <$

.001, $\eta_p^2 = .084$) but not for learner type ($F_{2,186} = .257, p = .773, \eta_p^2 = .003$) nor the interaction of learner type and noun type ($F_{4,1867} = .764, p = .550, \eta_p^2 = .016$). The post-hoc analysis failed to reveal any meaningful significant pairwise comparisons. The analysis of AJT gains from T1 to T2 failed to reveal any significant effects of learner type ($F_{2,186} = .594, p = .553, \eta_p^2 = .006$) and noun type ($F_{2,186} = 2.02, p = .136, \eta_p^2 = .021$). It also failed to reveal any significant interaction of learner type and noun type ($F_{4,186} = .591, p = .670, \eta_p^2 = .013$). Furthermore, the post-hoc analysis failed to reveal any significant pairwise comparisons.

5. DISCUSSION AND CONCLUSIONS

To determine the effect of each type of instruction on the reassembly of L1 noun type features for the L2, we will look at the results for each of the groups separately. For the linguistically-informed instruction group, we found that, prior to instruction, the imitation scores were significantly higher for countable nouns than object- or substance-uncountable nouns. The low scores at pre-test suggest that the learners had not yet properly reassembled their L1 features for the L2 since they treated object-uncountable and substance-uncountable nouns like countable nouns. If they had begun the reassembly process, we would have expected that the pre-test results show some difference between noun types within each learner group. Furthermore, the results also found significant gains in the imitation scores of both object- and substance-uncountable nouns from pre-test to immediate post-test. Furthermore, while not always significant, these improvements in imitation scores were maintained into the delayed post-test. For the acceptability judgment task, there was an overall lack of significant differences in the analysis, but the descriptive statistics trended in the expected direction —increase in acceptability of countable nouns and decrease in acceptability of uncountable nouns. These findings are very suggestive that this form of instruction, that is, instruction that explicit teaches the semantic features of English noun types, may be more beneficial in the immediate development of implicit knowledge than explicit knowledge. This is contrary to previous research which found that explicit instruction was most beneficial for explicit knowledge (Bowles, 2011; Bruhn de Garavito, 2013; Loewen, 2015; Long, 1983, 2009; Lopez, 2017; Lopez & Sabir, 2017; Master, 1994, 2002; Norris & Ortega, 2000; Snape & Yusa, 2013; Sonbul & Schmitt, 2013; Spada & Tomita, 2010).

In other words, in a meaning-focused imitation task, learners showed greater improvement following explicit instruction on the semantic features of noun types in

an indefinite-singular context than when they were asked to make explicit judgements on the acceptability of sentences. This is in line with previous research by Ionin, Choi & Liu (2020) who found that L2 learners of English from GCLs showed slow-down effects in a self-paced reading task but did not find similar learner sensitivity as measured by a grammaticality judgment task. One explanation for this might be that the learners began to overthink during the acceptability judgment task, which led them to overanalyze what they had learned. Being as the pre-test results showed clear indication that learners were interpreting uncountable-object nouns on the basis of atomicity instead of countability, it might be the case that when prompted to make acceptability judgments, they were unable to overcome their L1 influence. Our results found that explicit instruction on noun type semantics proved more beneficial for building up implicit knowledge as measured by an elicited-sentence imitation task. The lack of significant interactions between noun types and testing time in the acceptability judgment task shows that the learners were unable to completely employ their new knowledge in an explicit manner.

The results of the traditional instruction group, on the other hand, only revealed significant differences between noun types on the elicited-sentence imitation task, but these were differences regardless of time. In other words, while the descriptive statistics showed that these learners increased their imitations scores for noun types across the course of the study, none of these increases were significant, neither were the differences between noun types. In the acceptability task, on the other hand, these learners were unable to recognize, following instruction, that uncountable nouns were unacceptable and countable nouns were acceptable, resulting in no significant differences. The lack of significance in the results for the traditional instruction learners suggests that the current method of instruction employed in English language classrooms is not effective in facilitating the reassembly of L1 features for the L2, which is contrary to the existing literature that suggests explicit instruction does lead to gains in linguistic knowledge (Bowles, 2011; Long, 1983, 2009; Master, 1994, 2002; Norris & Ortega, 2000; Spada & Tomita, 2010). This is likely due to the fact that these materials do not explicitly teach the differences between atomicity and countability, and how these features combine to make various types of nouns in English.

The results of the no-extra-instruction group presented an unexpected picture. In the imitation task, we found that no-extra-instruction learners improved on all noun types from pre-test to immediate post-test although the statistical analysis did not reveal this improvement to be significant. More specifically, the results showed an

increase in imitation scores from pre-test to immediate post-test but then a decrease in scores from immediate to delayed post-test. In other words, any improvement was not sustained. For the acceptability judgment task, while the results did not show any significant changes, the descriptive statistics did show an increase in the acceptability of countable nouns and a decrease in the acceptability of uncountable nouns. When taken together, the results for the no-extra-instruction group do show that no instruction did not facilitate the reassembly process at all. It is important to remind the readers here that all participants enrolled in the research project were also enrolled in their regularly-schedule ESL grammar courses. To that end, the chapter in *Grammar & Beyond 4* covered in this research project is not explicitly taught to those students but listed as a «self-study» chapter with no assessments provided over the content. Therefore, any improvement in imitation scores or acceptability ratings was likely due to factors external of the research project.

In summary, these findings are in-line with the theoretical assumption, based on the Cline of Difficulty (Slabakova, 2009b; Cho & Slabakova, 2014), that the learners have a greater difficulty with the definiteness feature than other semantic features because it is both context- and morpheme-bound in the L1, while it is only morpheme-bound in the L2. Therefore, the results did show that explicit teaching in a linguistically-informed manner proved to be more successful in helping the learners overcome the Cline of Difficulty (Slabakova, 2009b; Cho & Slabakova, 2014) and the bottleneck of second language acquisition (Slabakova, 2008, 2009a, 2009b, 2013).

The second research question investigated which type of instruction would lead to the greatest gains in the reassembly of L1 noun type features of the L2. In terms of teaching intervention, the overall results found the greatest number of significant gains with the learners that received explicit linguistically-informed instruction on the semantic features of English noun types. Before we discuss the results, we would like to once again reiterate that the decision to offer only one hour of instructional intervention was to maintain fidelity and ecological validity with the length of instruction these participants were previously exposed to. In our analysis of gains over the course of the study, we only found significant differences between learner groups in the elicited-sentence imitation task. The analysis of the imitation score gains found that linguistically-informed learners made greater gains than traditional learners from pre-test to immediate and delayed post-test. In the same analysis, we also found that linguistically-informed learners made greater gains in their imitation scores from pre-test to delayed post-test and immediate post-test to delayed post-test than no-extra-instruction learners. Unexpectedly, we also found

that no-extra-instruction learners made greater gains than traditional instruction learners from immediate to delayed post-test. Although we found no significant differences between learner groups, LING learners did perform descriptively better than the other two learner groups with regard to the acceptability judgment task.

As stated, the results for the group who received instruction on countability and atomicity differed for the two tasks, but also differed from both the traditional instruction and no extra instruction groups, suggesting that instruction on countability and atomicity was beneficial. This finding supports our prediction that explicit instruction will facilitate the reassembly of the L1 [count] and [atomic] features for the L2 as posited by the Feature Reassembly Hypothesis (Lardiere, 2008, 2009a, 2009b). In other words, following explicit instruction, LING learners were able to recognize the indefinite article context is count-noun-selective, and therefore, only permits singular countable nouns, that is nouns with the feature settings [+count, +atomic]. Being as the majority of significant improvement was found in the elicited-sentence imitation task and not the acceptability judgment task, it may be the case that the functional morphology of the indefinite article *a* is providing difficulty for the learners when they are employing their explicit knowledge. This is in support of the Bottleneck Hypothesis (Slabakova, 2008, 2009a, 2009b) which suggests that functional morphology is the *real slowdown* of the language acquisition process.

The lack of significant performance of the traditional instruction group is a clear indication that the current method of instruction employed in English language classrooms across the world is not as effective as instruction informed by GenSLA research. While the lack of findings in the traditional instruction group contradict previous research (Akakura, 2012; Master, 1994, 2002), it is important to note that the materials (Bunting *et al.*, 2009) used for this group's instruction taught differences in noun types based on the articles they can combine with and gave a short description of countable and uncountable nouns. These materials did not address the differences in atomicity and types of uncountable nouns. The linguistically-informed instruction, on the other hand, explicitly taught the differences of English noun types using the semantic features that distinguish them, atomicity and countability, as well as the articles they can combine with. This explicit instruction on feature assembly and presentation of articles that produce count-selective and uncountable-selective contexts seems to have led to significant gains in linguistic knowledge.

The results and findings of this study do not come without limitations. The first and most obvious limitation is the small number of participants included in the traditional and no extra instruction groups in comparison to the linguistically-informed

instruction group. While we originally collected data from 100 participants overall, quite a bit of the data was incomplete and had to be excluded from the analysis. In addition, it remains to be seen what the longer-term effects are of only one hour of instructional intervention. While it is quite unfortunate that we only administered a single one-hour instructional intervention, this was our attempt to adhere true to our «teacher-self». Further research would benefit from considering an extension of the instructional intervention period or conducting a «review lesson» as is often done in courses before a quiz, test, or exam. Furthermore, more research into the effects of noun type instruction in different article and singular/plurality contexts would be interesting to see if similar effects are observed with nouns presented in contexts with the definite article or the zero article. As a final remark, we hope that these findings will be accessible to researchers and instructors from both GenSLA and ISLA backgrounds. In an attempt to bridge GenSLA theoretical research and ISLA pedagogy, we must remind ourselves that GenSLA is not meant to be a pedagogy, but the results from GenSLA research should inform ISLA and current teaching pedagogy.

Acknowledgements

We are grateful to the audiences of the 2019 meeting of the American Association of Applied Linguistics and the 2019 meeting of Teaching English to Speakers of Other Languages International Conference and Language Expo, where parts of this paper were presented, for their comments.

REFERENCES

- Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research*, 16(1), 9-37.
- Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition*, 97(1), 41-66. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0010027704001787?via%3Dihub>
- Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, 33, 247-271.
- Bruhn de Garavito, J. (2013). What research can tell us about teaching: The case of pronouns and clitics. In M. Whong, K.-H. Gil, & H. Marsden (Eds.), *Universal Grammar and the Second Language Classroom* (pp. 17-34). Dordrecht: Springer.
- Bunting, J. D., Diniz, L., & Reppen, R. (2009). *Grammar and beyond 4*. New York: Cambridge University Press.

- Cheng, L. L. S., & Sybesma, R. (1998). *Yi-wang tang, yi-ge tang*: classifiers and massifiers. *Tsing Hua Journal of Chinese Studies*, 28(3), 385-412.
- Cheng, L. L. S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, 30(4), 509-542. DOI: <https://doi.org/10.1162/002438999554192>
- Cheng, L. L. S., & Sybesma, R. (2014). The syntactic structure of noun phrases. In C. T. J. Huang, Y. H. A. Li, & A. Simpson (Eds.), *The Handbook of Chinese Linguistics* (pp. 248-274). Hoboken: John Wiley & Sons, Inc.
- Cho, J. (2012). *Remapping nominal features in the second language*. (PhD Dissertation). University of Iowa.
- Cho, J., & Slabakova, R. (2014). Interpreting definiteness in a second language without articles: The case of L2 Russian. *Second Language Research*, 30(2), 159-190.
- Choi, S. H., & Ionin, T. (2017). Acquisition and processing of mass nouns in L2-English by L2 learners from generalized classifier languages: Evidence for the role of atomicity. In M. LeMendola & J. Scott (Eds.), *Proceedings of the 41st Annual Boston University Conference on Language Development (BUCLD 2016)* (Vol. 1, pp. 154-167). Somerville: Cascadilla Press.
- Choi, S. H., Ionin, T., & Zhu, Y. (2018). L1 Korean and L1 Mandarin L2 English learners' acquisition of the count/mass distinction in English. *Second Language Research*, 34(2), 147-177.
- Choi, S. H., Zhu, Y., & Ionin, T. (2019). Interpretation of count and mass NPs by L2 learners from generalized classifier L1s. In T. Ionin & M. Rispoli (Eds.), *Three Streams of Generative Language Acquisition Research: Selected Papers from the 7th Meeting of Generative Approaches to Language Acquisition – North America, University of Illinois at Urbana-Champaign* (pp. 253-270). Amsterdam: John Benjamins.
- Doughty, C. (2003). Instructed CLA: Constraints, compensation, and enhancement. In C. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256-310). Malden: Blackwell Publishing.
- Ellis, R. (2005b). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27(2), 141-172. DOI: <https://doi.org/10.1017/S0272263105050096>
- Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-25). Bristol: Multilingual Matters.
- Ellis, R., Loewen, S., Elder, C., Erlam, R., Philip, J., & Reinders, H. (Eds.). (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol, UK ; Buffalo [N.Y.]: Multilingual Matters.
- Gillon, B. (1992). Towards a common semantics for English count and mass nouns. *Linguistics and Philosophy*, 15(6), 597-639. DOI: <https://doi.org/10.1007/BF00628112>
- Hua, D., & Lee, T. H.-t. (2005). Chinese ESL learners' understanding of the English count-mass distinction. In L. Dekydtspotter, R. A. Sprouse, & A. Liljestr and (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference (GASLA 2004)* (pp. 138-149). Somerville: Cascadilla Proceedings Project.

- Inagaki, S. (2013). Syntax-semantics mappings as a source of difficulty in Japanese speakers' acquisition of the mass-count distinction in English. *Bilingualism: Language and Cognition*, 17(3), 464-477.
- Lardiere, D. (2008). Feature-assembly in second language acquisition. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 106-140). New York: Lawrence Erlbaum Associates.
- Lardiere, D. (2009a). Further thoughts on parameters and features in second language acquisition: A reply to peer comments on Lardiere's 'Some thoughts on the contrastive analysis of features in second language acquisition' in SLR 25(2). *Second Language Research*, 25(3), 409-422.
- Lardiere, D. (2009b). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, 25(2), 173-227.
- Loewen, S. (2015). *Introduction to Instructed Second Language Acquisition* (1 ed.). New York: Routledge.
- Long, M. H. (1983). Does second language instruction make a difference? A review of research. *TESOL Quarterly*, 17(3), 359-382.
- Long, M. H. (2009). Language teaching. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 3-5). Malden: Blackwell Publishing Ltd.
- Lopez, E. (2017). Teaching the English article system: Definiteness and specificity in linguistically-informed instruction. *Language Teaching Research*, 23(2), 200-217.
- Lopez, E., & Sabir, M. (2017). Article pedagogy: Encouraging links between linguistic theory and teaching practice. *RELC Journal*, 50(1), 188-201.
DOI: <https://doi.org/10.1177/0033688217730143>
- Marsden, H., & Slabakova, R. (2019). Grammatical meaning and the second language classroom: Introduction. *Language Teaching Research*, 23(2), 147-157.
DOI: <https://doi.org/10.1177/1362168817752718>
- Master, P. (1994). The effect of systematic instruction on learning the English article system. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 229-252). Cambridge, UK: Cambridge University Press.
- Master, P. (2002). Information structure and English article pedagogy. *System*, 30, 331-348.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
DOI: <https://doi.org/10.1111/0023-8333.00136>
- Sabir, M. (2018). Specificity and article use: a theoretically informed classroom intervention. *Instructed Second Language Acquisition*, 2(2), 137-163.
- Slabakova, R. (2008). *Meaning in the Second Language*. New York: Mouton de Gruyter.
- Slabakova, R. (2009a). Features or parameters: which one makes second language acquisition easier, and more interesting to study? *Second Language Research*, 25(2), 313-324.
DOI: <https://doi.org/10.1177/0267658308100291>
- Slabakova, R. (2009b). What is easy and what is hard to acquire in a second language? In M. Bowles, T. Ionin, S. Montrul, & A. Tremblay (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)* (pp. 280-294). Somerville: Cascadilla Proceedings Project.

- Slabakova, R. (2013). What is easy and what is hard to acquire in a second language: A generative perspective. In M. d. P. García Mayo, M. J. Gutierrez Mangado, & M. Martínez Adrián (Eds.), *AILA Applied Linguistics Series* (Vol. 9, pp. 5-28). Amsterdam: John Benjamins Publishing Company.
- Slabakova, R. (2016). *Second Language Acquisition* (First edition ed.). Oxford: Oxford University Press.
- Slabakova, R. (2019). The Bottleneck Hypothesis updated. In T. Ionin & M. Rispoli (Eds.), *Three Streams of Generative Language Acquisition Research* (pp. 319-345). Amsterdam: John Benjamins.
- Snape, N., & Yusa, N. (2013). Explicit article instruction in definiteness, specificity, genericity and perception. In M. Whong, K.-H. Gil, & H. Marsden (Eds.), *Universal Grammar and the Second Language Classroom* (pp. 161-183). Dordrecht: Springer.
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 60, 263-308.
- Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, 65(3), 723-751. DOI: <https://doi.org/10.1111/lang.12129>
- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, 60(2), 263-308. DOI: <https://doi.org/10.1111/j.1467-9922.2010.00562.x>
- U. C. L. E. Syndicate. (2001). Oxford Quick Placement Test: Paper and Pen Pack. In. Oxford: Oxford University Press.
- Umeda, M., Snape, N., Yusa, N., & Wilstshier, J. (2017). The long-term effect of explicit instruction on learners' knowledge on English articles. *Language Teaching Research*, 00(0), 1-21.
- Whong, M., Gil, K.-H., & Marsden, H. (2013). Introduction: Generative Second Language Acquisition and Language Pedagogy. In M. Whong, K.-H. Gil, & H. Marsden (Eds.), *Universal Grammar and the Second Language Classroom* (pp. 1-13). Dordrecht: Springer.

Dakota J. Thomas-Wilhelm

University of Illinois System

510 Devonshire Drive, Suite H

Champaign, IL 61820, United States

Elisabet Pladevall-Ballester

Departament de Filologia Anglesa i de Germanística

Carrer de la Fortuna. Edifici B. Facultat de Filosofia i Lletres

Universitat Autònoma de Barcelona

08193 Bellaterra (Cerdanyola del Vallès), Spain

CONSIDERACIONES SOBRE LA EVALUACIÓN MORFOLÓGICA DE ESTUDIANTES DE INGLÉS COMO L2

Brenda VARGAS VEGA¹

Universidad Autónoma de Querétaro

Pierre-Luc PAQUET²

Université de Montréal

Resumen

Este trabajo intenta ofrecer un panorama más completo de la adquisición morfológica del inglés como segunda lengua al analizar las respuestas obtenidas de una prueba de derivación tanto estadísticamente como con un análisis de errores. Dicha prueba se aplicó a 130 universitarios divididos en tres grupos de competencia y los análisis se centraron en palabras creadas con los sufijos *-able* y *-er*. En lugar de calificar las respuestas como «correctas/incorrectas», éstas se analizaron en un continuo desde «derivadas» hasta «inventadas». El objetivo general del trabajo es contemplar la manera en que se desarrolla el conocimiento morfológico considerando los sufijos evaluados, los tipos de errores y el dominio de la lengua. El análisis estadístico muestra diferencias significativas tanto en la adquisición de los sufijos como entre grupos de competencia. El análisis de errores nos dejó ver que existen diferencias en los tipos de errores que cometen los estudiantes de distintos niveles de competencia; p.ej., el grupo de menor competencia es el más propenso a añadir sufijos flexivos en lugar de derivativos. La relevancia del trabajo es demostrar que existe conocimiento parcial, haciéndonos reflexionar sobre temas de evaluación.

Palabras clave: morfología derivativa; inglés como segunda lengua; español como primera lengua; análisis de errores

1. brenda.vargas@uaq.mx.  <https://orcid.org/0000-0001-9867-0114>

2. ppaquet@uttyler.edu.  <https://orcid.org/0000-0002-7325-0807>

CONSIDERATIONS TOWARDS MORPHOLOGICAL ASSESSMENT OF ENGLISH LANGUAGE LEARNERS

Abstract

This paper attempts to give a more complete picture of the morphological acquisition of English as a second language by analyzing the responses obtained from a derivational test, both statistically and with error analysis. The test was administered to 130 university students divided into three proficiency groups and the analyses focused on words created with the suffixes *-able* and *-er*. Instead of scoring the answers as «correct/incorrect», they were analyzed on a continuum from «derived» to «invented». The overall aim of the paper is to identify the way in which morphological knowledge is developed considering the suffixes evaluated, the types of errors and language proficiency. The statistical analysis shows significant differences both in the acquisition of suffixes and between proficiency groups. The error analysis suggests that there are differences in the types of errors made considering students' proficiency levels; e.g., the lowest proficiency group is more likely to add inflectional suffixes instead of derivational suffixes. The relevance of the work is to demonstrate that partial knowledge exists, making us reflect on evaluation issues.

Keywords: derivational morphology; English as a second language; Spanish as a first language, error analysis

RECIBIDO: 26/08/2022

APROBADO: 01/03/2023

1. INTRODUCCIÓN

El presente estudio es resultado de un análisis complementario al presentado en una tesis doctoral (Vargas-Vega, 2018) que evalúa el conocimiento distribucional (descrito más adelante) sobre cuatro sufijos nominalizadores del inglés. Los datos aquí analizados surgen de los reactivos distractores de la prueba experimental de dicha tesis. En este trabajo se presenta un análisis diferente al realizado con los reactivos experimentales, ya que en lugar de tomar las respuestas como «correctas» o «incorrectas», éstas se analizan en tres niveles: i) modificación o no de la base; ii) derivación correcta o incorrecta; y iii) invención de palabras con sufijos que cumplen o no con las condiciones sintácticas y semánticas del contexto, con el propósito de identificar el nivel de conocimiento morfológico de los participantes. Como lo mencionan Donoso y Gómez (2018), analizar de los errores cometidos por aprendientes de una L2 es de gran importancia dado que permite identificar el estado de su interlengua y arrojar implicaciones pedagógicas sustanciales.

Aunque existen algunas investigaciones donde se analizan los errores, éstas suelen basarse en taxonomías prefabricadas y/o dichos estudios no suelen centrarse en sufijos particulares, sino en producciones escritas como narraciones o traducciones (Marín Serrano, 2013; Hoyle, 2018; Adamczyk, 2019). En este estudio se analiza la producción de palabras derivadas con dos sufijos que comparten varias propiedades con la intención de descubrir el nivel de interlengua específicamente en el área morfológica de estudiantes hispanohablantes de inglés.

Las preguntas de investigación que guían este trabajo son:

- i) ¿Existe una diferencia significativa en la adquisición de los sufijos *-able* y *-er*?
- ii) ¿Existe una diferencia significativa en la producción de derivados entre grupos de competencia?
- iii) ¿Qué tipos de errores cometen los estudiantes de diferente nivel de competencia cuando producen palabras derivadas en inglés?

2. ANTECEDENTES

Según Bauer y Nation (1993), en etapas muy tempranas de adquisición, los estudiantes de una L2 no suelen reconocer los afijos de las palabras, p.ej. reconocen *teacher* pero no *teach + er*. Con el tiempo, comienzan a desarrollar los sufijos flexivos (p. ej., *-s*, *-ing*) y finalmente los derivativos (p. ej., *-ation*, *-ful*). No obstante, este proceso de adquisición de afijos se ve afectado por la frecuencia, regularidad, productividad y predictibilidad de los mismos. En el mismo sentido, Archibald y Libben (2018) sugieren que el conocimiento morfológico cambia con el tiempo y que no todas las palabras se procesan de la misma manera. Ellos apuntan que los hablantes nativos de inglés pueden descomponer la palabra *refill* en *re + fill* por ser considerado un derivado semánticamente transparente, pero no sucede lo mismo con la palabra *research*, a la que la mayoría de los hablantes nativos accederían de manera léxica (completa) por considerarla un derivado no transparente a pesar de que diacrónicamente contenga el prefijo *re-*.

Por otro lado, de acuerdo con Tyler y Nagy (1989), el conocimiento de la morfología derivativa se compone de tres partes: el conocimiento relacional, el sintáctico y el distribucional. El primero se refiere al conocimiento o percepción de que dos palabras están relacionadas (p. ej., *argue-argument*); el segundo consiste en saber que los sufijos derivativos marcan la categoría sintáctica de las palabras en inglés (p.ej., que el sufijo *-ize* genera verbos); y el tercero es el conocimiento de las

restricciones en la concatenación de bases y afijos. Por ejemplo, se puede añadir el sufijo *-ness* a adjetivos, pero no a verbos (*quietness* frente a **playness*). Vargas-Vega y Velásquez Upegui (2019) encontraron que este aspecto es el último en la adquisición del inglés como L2, tal como Tyler y Nagy (1989) lo habían sugerido para la L1. Libben (2012, *apud* Archibald & Libben, 2018), complementa la propuesta de Tyler & Nagy (1989) al incorporar el aspecto semántico al conocimiento de la morfología derivativa. Él sugiere que este conocimiento consiste, entre otras cosas, en la habilidad de repetir, comprender y producir palabras multi-morfémicas con propiedades semánticas y sintácticas apropiadas.

La derivación puede dividirse en dos tipos: aquella que mantiene su clase gramatical (derivación homocategorial) y la que la cambia (heterocategorial) (Bauer, 1983). En el primer tipo, la derivación produce lexemas que corresponden a la misma clase gramatical que su base (p.ej. *hero* (N) -> *heroism* (N)) mientras que, en el segundo tipo, el proceso de derivación produce un lexema con una clase gramatical diferente al de la base (p.ej. *happy* (N) -> *happily* (ADV)). Se dice entonces que, en inglés, la prefijación comúnmente mantiene la clase gramatical mientras que la sufijación la cambia.

Respecto a la adquisición del conocimiento morfológico, Hayashi y Murphy (2011) descubrieron que primero se adquieren los sufijos flexivos, seguidos por los sufijos derivativos homocategoriales y por último se adquieren los sufijos derivativos heterocategoriales. El estudio de Ali, Hamid y Rahman (2021) apoya este último hallazgo mostrando que sus participantes obtuvieron mejores resultados en los derivativos homocategoriales que en los heterocategoriales.

En cuanto a la categoría gramatical de los sufijos, Schmitt y Zimmerman (2002) presentan mejores resultados en la producción de derivados nominales (63%) que adjetivales (54%), mientras que Al-Homoud (2017) y Vargas-Vega y Márquez Valencia (2022) no muestran una diferencia importante en la producción de palabras derivadas que generan estas categorías. Vargas-Vega y Márquez Valencia (2022), sin embargo, reportan un efecto de facilitación por cognados.

El desarrollo de la morfología derivativa es un proceso lento y progresivo (Karlsson, 2015; Schmitt y Zimmerman, 2002). Schmitt y Zimmerman (2002) proponen que incluso los estudiantes avanzados crean palabras inventadas al añadir sufijos correctos a bases que no los aceptan, por ejemplo **releasement* (en lugar de *release*); **minimizement* (en lugar de *minimization*); **persistment* (en lugar de *persistence*) y **survivation* (en lugar de *survival*). También señalan

que la derivación para formar adjetivos y adverbios parece ser más tardía que la de sustantivos y verbos.

En este estudio, se evalúa la producción controlada de dos sufijos deverbales elegidos de la escala de afijos propuesta por Bauer y Nation (1993). Dicha escala contiene siete niveles de afijos ubicados por su dificultad y ha sido usada por varios investigadores de inglés como L2 (Mochizuki y Aizawa, 2000; Schmitt y Zimmerman, 2002; Medellín y Auza, 2008; Ward y Chuenjundaeng, 2009; Danilović, Savić y Dimitrijević, 2013; Leontjev, Huhta y Mäntylä, 2016; Vargas-Vega, 2018; Vargas-Vega y Velásquez Upegui, 2019; entre otros). Sin embargo, la escala no considera ni la L1 de los aprendientes ni la frecuencia de las bases a las que se añaden los afijos (Gardner, citado en Karlsson, 2015); y hace falta validarla con trabajos experimentales, como se menciona en Leontjev y otros (2016, p. 76):

Debe enfatizarse que la clasificación de dificultad de afijos de Bauer y Nation (1993) es arbitraria y que es prematuro considerarla como el/un orden de adquisición. Además, hasta donde sabemos, aún hace falta que este orden de dificultad se corrobore empíricamente. (Traducción propia)

Este trabajo ayuda a resolver algunas de las inquietudes que surgen de la revisión de los antecedentes con respecto al impacto de la competencia en la L2 sobre el proceso de adquisición morfológica. Por ejemplo, ¿Qué características tienen las respuestas de los grupos de diferente competencia?, ¿Qué resulta más importante para sugerir un proceso de adquisición de sufijos: la competencia de los estudiantes en la L2 o las características de los sufijos?

3. EL PRESENTE ESTUDIO

El objetivo general del estudio es explorar la manera en la que se desarrolla el conocimiento morfológico considerando las siguientes variables: los sufijos evaluados, los tipos de errores y el dominio de la lengua. Nuestros objetivos particulares son:

1. Contemplar si los participantes muestran mejor desempeño derivando palabras con el sufijo nominal *-er* que con el sufijo adjetival *-able* (ambos cognados³ con el español y ubicados en el mismo nivel de dificultad de la escala de Bauer y Nation (1993)).

3. En lingüística, se reconocen principalmente dos tipos de cognados a nivel léxico: aquellos cuya escritura, significado y probablemente pronunciación son similares entre la L1 y la L2, conocidos como cognados verdaderos (p.e. doctor); y aquellos cuya escritura y pronunciación son similares, pero difieren en

no es posible porque la nota es muy larga

2. Observar si las producciones de palabras derivadas con sufijos distintos a *-er* y *-able* cumplen con los requisitos sintácticos y semánticos del contexto oracional.
3. Explorar las diferencias según niveles de competencia en la producción de palabras derivadas con los sufijos *-er* y *-able*.

Comenzaremos por exponer las similitudes y diferencias de los sufijos en cuestión y posteriormente, en la sección de metodología, explicaremos la selección de palabras meta y la creación de la prueba, entre otras cosas.

Tanto el nominalizador *-er* como el adjetivador *-able*, corresponden al nivel 3 según la escala de Bauer y Nation (1993); es decir, un nivel de baja dificultad. Sasao y Webb (2017) también ubican estos dos sufijos en nivel de principiantes basados en su estudio de conocimiento receptivo de 118 afijos derivativos⁴. Otras semejanzas que comparten estos sufijos son que ambos se añaden a bases verbales; ambos tienen alomorfos (*-er* frente a *-or*; *-able* frente a *-ible*) pero las formas elegidas son las más productivas de su clase (Bauer, Lieber, y Plag, 2013); es decir, *-er* es el sufijo más productivo para crear sustantivos con rol de agentes y *-able* es el sufijo más productivo para crear adjetivos; ambos tienen cognados con el español (p. ej. *painter* ‘pintor’, *recognizable* ‘reconocible’) y rara vez modifican ortográfica o fonológicamente la base a la que se añaden.

Algunas de las diferencias entre los sufijos son: 1) el sufijo *-er* es un nominalizador que crea el significado de agente (p. ej. *driver*) o instrumento (p. ej. *blender*) y se refiere al sujeto, mientras que el sufijo *-able* es un adjetivador que crea el significado de modalidad/posibilidad (p. ej. *breakable*) o juicio evaluativo (p. ej. *deplorable*) (Da Costa Moreira, 2014); 2) la productividad histórica de los nominalizadores deverbales es mayor que la productividad histórica de los adjetivadores deverbales (Bilynsky, 2015); y 3) las familias léxicas que denotan conceptos más concretos parecen ser de adquisición más temprana que las familias léxicas que denotan conceptos más abstractos. Por ejemplo, la familia de *abrir* (*abridor*, *abrelatas*)

significado, conocidos como falsos cognados (p.e. *embarazada/embarassed*). La propiedad de cognado se puede ver también a nivel morfológico como en los sufijos que aquí se estudian (p.e. *agradable/enjoyable*). En esta investigación se estudian palabras cognadas y no cognadas considerando la base, ya que el sufijo siempre será cognado (p.e. *compositores/composers frente a fumadores/smokers*).

4. El estudio de Sasao y Webb (2017) toma en cuenta 30 diferentes lenguas maternas, incluyendo el español (1348 participantes de 100 países diferentes con diferentes lenguas maternas). Una semejanza con nuestro estudio es que sólo analizan el conocimiento de afijos derivativos; y una diferencia es que lo hacen sin contexto.

y *pan* (*panadero*, *panadería*) se adquieren antes que la de *guapo* (*guapería*, *guapura*), ya que es más probable que los estudiantes las encuentren en sus libros de texto (Sánchez-Gutiérrez y Hernández Muñoz, 2018). Es importante señalar, sin embargo, que según Schmitt y Meara (1997), las asociaciones con palabras derivadas tienden a darse con base en los conceptos que evocan los derivados (p. ej. *employ* → *employer*) pero no necesariamente por su forma⁵.

4. METODOLOGÍA

4.1. Elección y distribución de los reactivos experimentales

La elección de los sufijos estudiados se basó en tres criterios: i) que tuvieran el mismo nivel de dificultad (Bauer y Nation, 1993); ii) que se añadieran a bases verbales; y iii) que tuvieran cognados en español. Además, como ya se mencionó, estos sufijos regularmente no modifican la base a la que se añaden (p. ej. *import* → *importer*, *do* → *doable*⁶) o en su defecto, ocasionan modificaciones mínimas como la duplicación de una consonante (p. ej. *jog* → *joggers*) o el cambio de una letra (p. ej. *compare* → *comparable*). Las palabras meta se eligieron tomando en cuenta las características de cognado y frecuencia.

Se eligieron 8 palabras meta para cada sufijo: dos cognadas de alta frecuencia (C/AF), dos cognadas de baja frecuencia (C/BF), dos no cognadas de alta frecuencia (NC/AF) y dos no cognadas de baja frecuencia (NC/BF)⁷ (ver tabla 1).

	<i>-er</i>	C	NC	<i>-able</i>	C	NC
AF		composers	dreamers		comparable	washable
		promoter	smokers	AF	applicable	regrettable
BF		planner(s)	joggers		observable	doable
		importer	punisher	BF	deplorable	bearable

Tabla 1. Características de las palabras meta⁸

5. Estos autores evalúan dos componentes del vocabulario: el conocimiento de los afijos y el conocimiento de la asociación de palabras. Exploran cómo estos aspectos se relacionan entre sí, con el tamaño de vocabulario y con la competencia general de la lengua. Advierten, basados en respuestas de hablantes nativos, que las asociaciones no siempre se dan con miembros de la misma familia de palabras, sino con conceptos que se relacionan (p.ej. *agreeable time*).

6. La variante de *-ible* suele generar más cambios en la base (p. ej. *admit* → *admissible*).

7. La frecuencia de las palabras derivadas se consultó en el British National Corpus, sección escrita, no académica.

8. C= cognado, NC= no cognado, AF= alta frecuencia, BF= baja frecuencia. Recordemos que nos referimos a cognados cuando vemos semejanzas entre los derivados del español y el inglés (no solo en sus terminaciones).

El diseño de la prueba usada en este estudio se basó en la prueba de estructura morfológica (TMS, por sus siglas en inglés) de Carlisle (2000), que ha sido usada exitosamente en numerosos estudios morfológicos tanto de la L1 (Muse, 2005; Stanfa, 2010) como de la L2 (Curinga, 2013; Kieffer y Lesaux, 2008, 2012; Kraut, 2015; Ramírez, Chen, Geva y Kieffer, 2010; Ramírez, Chen, Geva y Luo, 2011; Varatharajoo, Asmawi, Abdallah y Abedalaziz, 2015). La prueba de Carlisle (2000) fue diseñada para evaluar la conciencia de las relaciones entre las bases y las formas derivadas. Se divide en dos partes: una de derivación, donde hay que producir la forma derivada de la palabra para completar las oraciones como en el ejemplo (1); y otra de descomposición, donde hay que descomponer la forma derivada para completar las oraciones (ejemplo 2).

- (1) *Farm. My uncle is a ... [farmer]*
 (2) *Driver. Children are too young to... [drive]⁹*

Para este estudio se construyeron 16 oraciones que los participantes tenían que completar modificando la palabra entre paréntesis que aparecía al inicio de la oración (8 oraciones para cada sufijo¹⁰). Todas las oraciones tenían una longitud de 10 palabras como máximo para evitar el cansancio de los participantes y una mitad de las palabras meta de cada sufijo se colocó en posición pre-verbal mientras que la otra mitad se colocó en posición pos-verbal para intentar que los participantes leyeran las oraciones completas¹¹ (consultar el anexo 1 para ver los reactivos de la prueba). Los reactivos no contenían pseudopalabras porque éstas no son equivalentes a palabras reales. Las pseudopalabras funcionan bien para detectar el conocimiento sintáctico y semántico de los afijos, pero debido a que nosotros estábamos interesados en explorar también el conocimiento distribucional de los sufijos en diferentes grupos de competencia, decidimos no incluirlas¹².

La modalidad de aplicación de la prueba fue en línea por medio de una plataforma llamada Classmarker. Cada participante recibía los reactivos (oraciones) en un orden aleatorio distinto y no tenían límite de tiempo para su resolución. Aparecía una oración a la vez en la pantalla y no podían pasar a la siguiente si no completaban la

9. Ejemplos tomados de Carlisle (2000).

10. El limitado número de oraciones se debe a que, como ya se mencionó, se trata de los reactivos distractores de una tesis doctoral.

11. Esta prueba fue resultado de modificaciones derivadas de dos pruebas piloto y sugerencias de revisores expertos en inglés.

12. Mochizuki y Aizawa (2000), así como Lardiere (2006) han usado pseudopalabras, en caso de que el lector quiera saber cómo funcionan.

que estaban viendo en ese momento; es decir, no podían dejar reactivos sin contestar. Asimismo, no podían regresar a un reactivo anterior para evitar que corrigieran sus respuestas. A pesar de que la plataforma envía los resultados cuantitativos de manera inmediata, para el presente análisis, se capturó en excel cada una de las respuestas de los participantes para poder analizarlas (n=2080). No se penalizaron errores ortográficos o de concordancia de número (esto último en el caso del sufijo *-er*)¹³. La aplicación de la prueba fue en el horario de la clase de inglés en el salón de los estudiantes o en una sala de cómputo de sus facultades, y los profesores podían estar presentes. Se les aclaró a los estudiantes que sus resultados en esta prueba no afectarían su calificación en la asignatura.

4.2. *Participantes*

El nivel de los participantes fue previamente establecido por medio de pruebas simuladas de Cambridge aplicadas por la institución educativa a la que pertenecían (una universidad pública en Querétaro, México): En total participaron 55 estudiantes ubicados en el nivel A2, 46 estudiantes del nivel B1 y 29 estudiantes del nivel B2 de acuerdo con el Marco Común Europeo de Referencia para las Lenguas, dando un total de 130 participantes (52 hombres y 78 mujeres) en un rango de edad de entre 20 y 23 años y pertenecientes a dos facultades distintas. La diferencia en el número de participantes de los diferentes niveles de competencia se debe a que se tomaron grupos intactos y había más estudiantes en niveles básicos o intermedios que en niveles altos.

5. ANÁLISIS

En este estudio se realizan dos procedimientos: un análisis de errores y un análisis estadístico de varianza (ANOVA). El primero se llevó a cabo en tres niveles, que a continuación se describen.

En el primer nivel se identifica si los participantes realizan la tarea de modificar la base o no. El no hacerlo puede señalar falta de conocimiento del sufijo derivativo (error de omisión) o derivación cero¹⁴. Como no se puede saber la razón por la que no hubo derivación, este tipo de respuesta se cuantifica, pero no se intenta explicar.

13. La decisión de no penalizar errores ortográficos se ha visto en trabajos anteriores como los de Schmitt y Zimmerman (2002) y Karlsson (2015).

14. La derivación cero sucede cuando no es necesario añadir un sufijo para cambiar la categoría gramatical de la palabra. Por ejemplo, *cook* (v) o *cook* (n) se distinguen sólo por su contexto oracional.

En el nivel 2 se ubican todas las palabras modificadas tanto por derivación como por flexión que producen los estudiantes y que existen en la lengua inglesa. Si las palabras derivadas son adecuadas al contexto oracional por la información semántica y sintáctica que proveen, se consideran «adecuadas» (nivel 2.1.1); si existen pero no encajan en el contexto oracional, se consideran «inadecuadas» (2.1.2). Las respuestas existentes pero inadecuadas sugieren conocimiento léxico¹⁵.

El nivel 3 se compone de todas las palabras que inventan los estudiantes al añadir un sufijo que no corresponde a la base; es decir, los participantes sí realizan la tarea de derivar, pero lo hacen con un sufijo erróneo. Si el sufijo corresponde a la categoría gramatical impuesta por el contexto, se considera «aproximada» (nivel 3.1.1) pero si el sufijo que eligen cambia la categoría gramatical que se adecúa al contexto oracional, la respuesta se considera «no aproximada» (3.1.2). Las palabras inventadas aproximadas son de interés por el tipo de información que proveen. Por un lado, se puede considerar que los participantes ya cuentan con el conocimiento relacional y sintáctico de los sufijos aunque todavía no demuestran dominar su aspecto distribucional; por el otro, nos hacen indagar sobre las causas de dichos errores, ya que pueden deberse a un proceso de sobregeneralización del sufijo (como cuando un niño dice *rompido* en lugar de *roto*¹⁶), a la frecuencia o productividad del sufijo; o a un efecto de transferencia de la L1; por mencionar sólo algunos posibles factores. La tabla 2 presenta la clasificación de las respuestas.

	NIVEL 1 (modificación de bases)		NIVEL 2 (palabras reales)		NIVEL 3 (palabras inventadas)	
1. Derivación	1.1 Sí	2.1	2.1.1 respuesta adecuada	3.1	3.1.1 respuesta aproximada	
	<i>observable</i>	Derivados	<i>observable</i> (adj)	Derivados	<i>observative</i> (adj)	
	1.2 No	existentes	2.1.2 respuesta	inexistentes	3.1.2 respuesta no	
	<i>observe</i>		inadecuada		aproximada	

Tabla 2. Clasificación de las respuestas¹⁷

15. En este trabajo se entenderá por conocimiento léxico algo similar a un acceso directo al derivado dentro del lexicón mental en oposición a una descomposición morfológica para llegar a la formación y significado de la palabra derivada. Es decir, estamos usando «conocimiento léxico» como el conocimiento de palabras individuales memorizadas.

16. Cabe señalar que alguien que dice *rompido* en lugar de *roto* se da a entender perfectamente bien con su interlocutor, y que además este proceso de sobregeneralización es pasajero en los niños, como también puede serlo en el desarrollo de la L2.

17. Para facilitar la lectura, nos referimos a (1) como «derivación» aunque hubieron algunos pocos casos de adición de un sufijo flexivo que funcionaba en el contexto oracional. P. ej. *His decision to quit is regretted* (en lugar de *regrettable*). Las palabras derivadas que existen y son adecuadas al contexto

El segundo procedimiento fue un análisis estadístico que se realizó para explorar la diferencia de aciertos que obtuvieron los grupos de participantes, según el nivel de competencia, en la producción de palabras con los sufijos *-er* y *-able*. Es decir, el análisis estadístico se concentra en las respuestas del nivel 2.1.1 (palabras existentes y adecuadas) y se realizó con el paquete estadístico SPSS v.25.

6. RESULTADOS

Esta sección comienza con la descripción del primer nivel de análisis de errores; es decir, mostramos el porcentaje de respuestas en el que los participantes modificaron las bases en la tarea de derivación. Una vez descartado el porcentaje de respuestas donde no se modificó la base, se expone la distribución de tipo de errores (niveles 2 y 3). Finalmente, se muestra el análisis estadístico basado en las respuestas tradicionalmente correctas para observar si existe una relación creciente entre respuestas correctas y nivel de competencia.

6.1. *Análisis de errores*

En una tarea de producción morfológica, lo primero que nos ocupa es saber si los participantes fueron capaces de entender la tarea (nivel 1 de análisis). Como lo muestra la tabla 3, los participantes llevaron a cabo la tarea en 92.8% de los casos (respuesta 1.1); es decir, modificaron la palabra base añadiendo un sufijo (aunque el sufijo no siempre fue el esperado). Esto sugiere que entendieron la tarea morfológica. Complementariamente, los errores de omisión (nivel 1.2) representan el 7.2% y estos reactivos no se incluirán en los análisis siguientes.

Nivel de análisis	Tipo de respuestas	Núm. de respuestas	% respuestas	Núm. de respuestas	% respuestas	% Total de respuestas
		sufijo <i>-er</i>	sufijo <i>-er</i>	sufijo <i>-able</i>	sufijo <i>-able</i>	
NIVEL 1	1.1 palabras derivadas	989	47.5	942	45.3	92.8
	1.2 palabras no derivadas	51	2.5	98	4.7	7.2
	Total	1040	50%	1040	50%	100%

Tabla 3. Desempeño en la tarea (número de respuestas y porcentaje) 2080 ítems

(2.1.1), son las respuestas consideradas tradicionalmente correctas; mientras que las demás son consideradas tradicionalmente incorrectas. Fuente: Elaboración propia.

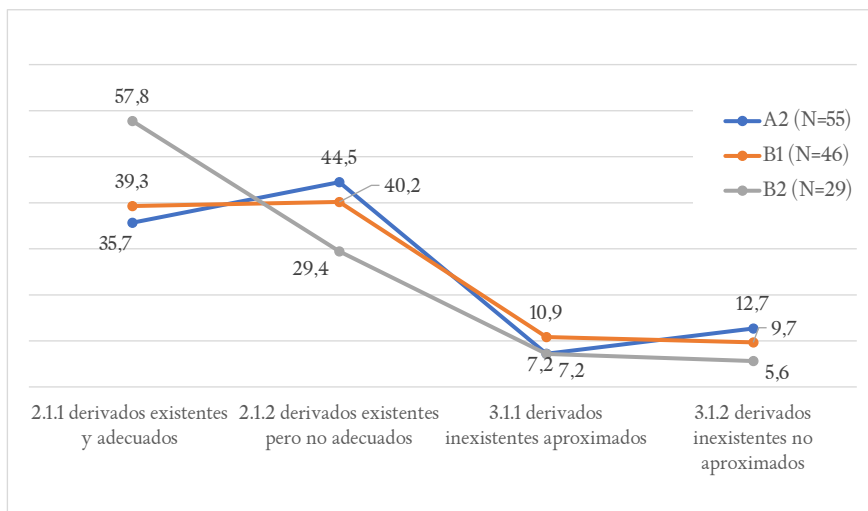
Como se observa en la tabla 3, el porcentaje de palabras derivadas para el sufijo *-er* tiende a ser mayor que para el sufijo *-able* (47.5% frente a 45.3% respectivamente), sumando un 92.8% de respuestas correspondientes a la categoría 1.1. Con base en esta categoría, podemos realizar el análisis de cómo se modificaron las bases, enfocándonos en los niveles 2 y 3. La tabla 4 nos permite identificar el porcentaje de palabras reales *frente a* palabras inventadas que crean los participantes, así como detectar si las palabras que inventan se acercan a los requisitos del contexto oracional. Asimismo, la tabla 4 compara el desempeño de los participantes en cada uno de los sufijos para identificar si alguno de los sufijos estudiados causa más problemas en la derivación.

NIVEL DE ANÁLISIS	TIPO DE RESPUESTAS	NÚMERO DE PALABRAS META CON EL SUFIJO <i>-er</i>	% sufijo <i>-er</i>	NÚMERO DE PALABRAS META CON EL SUFIJO <i>-able</i>	% SUFIJO <i>-able</i>	PORCENTAJE TOTAL
NIVEL 1	1.2 palabras no derivadas	51	2.5%	98	4.7%	7.2%
NIVEL 2 (palabras reales)	2.1.1 derivados existentes y adecuados (correctos)	567	27.3%	244	11.7%	39%
	2.1.2 derivados existentes pero no adecuados	275	13.2%	488	23.5%	36.7%
NIVEL 3 (palabras inventadas)	3.1.1 derivados inexistentes aproximados	105	5%	59	2.8%	7.9%
	3.1.2 derivados inexistentes no aproximados	42	2%	151	7.3%	9.3%
	Total	1040	50%	1040	50%	100%

Tabla 4. Distribución de tipo de respuestas (número de palabras meta y porcentaje) 2080 ítems

La tabla 4 ofrece dos evidencias de que los estudiantes muestran más problemas para derivar con el sufijo *-able* que con el sufijo *-er*. La primera es que producen un menor porcentaje de derivados existentes y adecuados con *-able* (11.7%) que con *-er* (27.3%) y la segunda evidencia es que producen un mayor porcentaje de derivados inexistentes y no aproximados con *-able* (7.3%) que con *-er* (2%). La primera evidencia coincide con los resultados de Schmitt y Zimmerman (2002), quienes encuentran una mejor producción de sustantivos que de adjetivos. Respecto a la formación de palabras inventadas según sufijo, la tabla 4 también muestra que los derivados inexistentes aproximados (respuesta 3.1.1) se usan en mayor porcentaje que los no aproximados en el caso de *-er* (5% frente a 2%), pero con el sufijo *-able*

sucede lo inverso (2.8% frente a 7.3%), sugiriendo un mejor conocimiento sintáctico y semántico del sufijo *-er*. Adicionalmente, el porcentaje de derivados inexistentes no aproximados (respuesta 3.1.2) es mayor para el sufijo *-able* que para el sufijo *-er* (7.3% frente a 2%), lo cual indica una generación de palabras más azarosa. Para finalizar el análisis de errores, se muestra en la figura 1 el porcentaje de cada tipo de respuesta por nivel de competencia.



En la figura 1, respuesta 2.1.1, se muestra que los participantes de mayor nivel de competencia (B2) producen un mayor porcentaje de respuestas correctas (57.8%) que los otros dos grupos. Aunque los participantes B2 crean palabras reales que no corresponden al contexto oracional (respuesta 2.1.2, 29.4%), éstas son casi la mitad de las que sí pertenecen al contexto (2.1.1). Por el contrario, en el caso de los grupos A2 y B1, la producción de palabras reales adecuadas y no adecuadas es casi igual. De hecho, en el caso del grupo A2, el porcentaje de las respuestas de derivados existentes no adecuados es incluso mayor que el de las respuestas existentes adecuadas, lo que sugiere un conocimiento léxico más que morfológico.

Para los tres grupos, la creación de palabras inventadas (nivel 3) es menor que de palabras reales (nivel 2). El grupo B1 parece experimentar con la L2 al producir un mayor porcentaje de palabras inexistentes aproximadas (3.1.1). Al ser intermedios, es esperable que se arriesguen a intentar formas lingüísticas basadas en su conocimiento parcial. Si interpretamos globalmente los resultados de esta figura, vemos un continuo de más correcto a más incorrecto (2.1.1 a 3.1.2). La tendencia, bien

dibujada por el grupo B2, es que las respuestas correctas incrementan con un mayor nivel de competencia al tiempo que disminuye la creación de palabras inventadas.

6.2. Análisis estadístico

Esta sección se concentra en un total de 811 aciertos de los participantes (respuesta 2.1.1). Comienza mostrando el análisis de medias y después se presentan los análisis de varianza para comparar el desempeño de los tres grupos de competencia en la producción de palabras con el sufijo *-er* y el sufijo *-able*. Es importante recordar que hubo ocho reactivos para evaluar la producción de palabras derivadas con cada sufijo. La tabla 5 muestra que la media de respuestas correctas es mayor en los reactivos con el sufijo *-er* (3.45 de ocho posibles) que con el sufijo *-able* (1.30 de ocho), y esta tendencia se mantiene en los tres grupos de competencia.

SUFIJO	GRUPO	N	MEDIA	95% DEL INTERVALO DE CONFIANZA PARA LA MEDIA	
				LÍMITE INFERIOR	LÍMITE SUPERIOR
aciertos <i>-er</i>	A2	55	3.33	2.81	3.85
	B1	46	3.07	2.67	3.46
	B2	29	4.28	3.59	4.96
	Total	130	3.45	3.14	3.75
aciertos <i>-able</i>	A2	55	.65	.38	.93
	B1	46	1.22	.82	1.62
	B2	29	2.66	2.02	3.29
	Total	130	1.30	1.04	1.56

Tabla 5. Estadística descriptiva de los aciertos con los sufijos *-er* y *-able* por nivel de competencia

Después de conocer las medias de la tabla 5, se llevó a cabo un análisis de varianza (ANOVA) unidireccional para comparar el desempeño de los tres grupos de competencia en la producción de palabras correctas con los dos sufijos estudiados. Los resultados entre grupos señalan que existen diferencias estadísticamente significativas entre las medias de los tres grupos de competencia, con el sufijo *-er*, $F(2, 127)=4.696$, $p<0.01$ y con el sufijo *-able*, $F(2, 127)=22.610$, $p<0.000$. Esta evidencia apuntaría a que el grupo B2 es estadísticamente diferente. Este hallazgo se explora en la tabla 6.

SUFIJOS	COMPARACIONES		DIFERENCIA DE MEDIAS	DESV. ERROR	SIG.	INTERVALO DE CONFIANZA AL 50%	
	ENTRE GRUPOS					LÍMITE INFERIOR	LÍMITE SUPERIOR
Aciertos en <i>-er</i>	A2	B1	.262	.341	1.000	-.21	.74
	A2	B2	-.949*	.392	.051	-1.49	-.40
	B1	B2	-1.211*	.405	.010	-1.77	-.65
	A2	B1	-.563*	.260	.097	-.92	-.20
Aciertos en <i>-able</i>	A2	B2	-2.001*	.298	.000	-2.42	-1.59
	B1	B2	-1.438*	.308	.000	-1.87	-1.01

Tabla 6. Comparaciones múltiples realizadas con la prueba post hoc Bonferroni¹⁸

El análisis *post hoc* de Bonferroni (ver tabla 6) nos permitió observar que en el caso del sufijo *-er*, las diferencias significativas de medias se encuentran entre los grupos de competencia A2 frente a B2, y B1 frente a B2. A partir de estas observaciones, se puede notar que el grupo de estudiantes avanzados (nivel B2) se distingue de los otros dos grupos en la producción de palabras con el sufijo *-er*. En cuanto al sufijo *-able*, las diferencias significativas de medias se encuentran en la comparación de todos los grupos; es decir, en A2 frente a B1, A2 frente a B2 y B1 frente a B2. Esto sugiere que la producción de palabras con este sufijo mejora en cada nivel de competencia.

7. DISCUSIÓN Y CONCLUSIONES

Los objetivos particulares de este trabajo fueron: i) identificar si los participantes muestran mejor desempeño derivando palabras con el sufijo *-er* que con el sufijo *-able*; ii) en caso de encontrar respuestas con otros sufijos, observar si éstas cumplen con los requisitos sintácticos y semánticos del contexto oracional; y iii) explorar las diferencias entre grupos de competencia.

En cuanto al primer objetivo, sugerimos que la derivación con el sufijo *-able* parece ser un desafío más grande que la derivación con el sufijo *-er* para los estudiantes de inglés. Al igual que Schmitt y Zimmerman (2002), notamos que los estudiantes presentan más dificultad para crear adjetivos derivados que sustantivos, aun cuando en nuestro caso, estos sufijos pertenecen al mismo nivel de dificultad según Bauer y Nation (1993). En este sentido, Karlsson (2015) sugiere que la habilidad de los estudiantes de producir palabras derivadas correctas es independiente al nivel de

18. La diferencia de medias es significativa en el nivel 0.5.

dificultad de los afijos propuesto por estos autores. El presente estudio sugiere que el nivel de competencia de los estudiantes es un mejor indicador para predecir errores que el nivel de dificultad de los sufijos.

La escala de Bauer y Nation (1993) presenta algunos conflictos como la ubicación de un sufijo en más de un nivel de dificultad debido a sus diferentes funciones, o que en ocasiones no son claros los criterios de inclusión a los niveles. A pesar de estos y otros problemas que se han detectado en la escala de dificultad, es importante subrayar su utilidad y recordar que sus mismos autores están conscientes de algunas limitaciones al reconocer que la escala es arbitraria, que no da cuenta de las palabras de muy alta frecuencia y que los niveles o sus componentes pueden variar dependiendo de los propósitos de investigación, aclarando además que esta escala se propone para el conocimiento receptivo de los afijos (Bauer y Nation, 1993). Los autores de este artículo reconocemos que las medidas de comprensión pueden ser muy distintas a las de producción tanto en el desempeño de los participantes como en el diseño de las tareas. En este estudio nos concentramos en evaluar la parte productiva (controlada), por lo que queda pendiente realizar más estudios que comparen la comprensión y la producción de sufijos, esta última con tareas controladas *frente a* producciones libres.

El efecto de frecuencia de palabra del que nos advierten Bauer y Nation (1993) se pudo distinguir en este estudio, ya que la palabra *smokers* fue producida correctamente por la mayoría de los participantes sin importar su nivel de competencia (121 repuestas correctas de un total de 130), lo que sugiere un conocimiento léxico. Estos autores discuten también una problemática que nos atañe directamente al señalar que el afijo *-ing* puede funcionar tanto como un sufijo flexivo como derivativo. Nosotros encontramos repuestas como *A more bearing life requires good attitude*, en lugar de *A more bearable life requires good attitude*, donde ambas opciones son posibles. El análisis fino de los datos nos permitió descubrir adicionalmente que los participantes de menor nivel de competencia son más propensos a usar sufijos flexivos que sufijos derivativos (*applied* en lugar de *applicable*, *compared* en lugar de *comparable*, *bearing* en lugar de *bearable*), tal como lo sugieren Bauer y Nation (1993), Hayashi y Murphy (2011) y Sasao y Webb (2017).

En cuanto al objetivo (ii) (en caso de encontrar repuestas con otros sufijos, identificar si éstas cumplen con los requisitos sintácticos y semánticos del contexto oracional), los resultados sugieren que los participantes del nivel B1 son los que tienen mayor tendencia a añadir sufijos con estas características. Este tipo de repuestas es el que nosotros denominamos *derivados inexistentes aproximados*, ya que el sufijo cumple con la función sintáctica del contexto, pero no es posible añadirlo a la

base por restricciones de concatenación (el *conocimiento distribucional* de Tyler y Nagy, 1989). Tanto Karlsson, (2015) como Schmitt y Zimmerman (2002) reportan la creación de palabras inventadas en sus resultados como **arriviation* o **arrivement* en lugar de *arrival* y **couragable* o **courageful* en lugar de *courageous*). Schmitt y Zimmerman (2002) las llaman *formas creativas* y plantean que es posible que inclusive hablantes proficientes las produzcan basados en un conocimiento parcial de la morfología derivativa. Libben (2012, *apud* Archibald & Libben, 2018) las llama *construcciones multi-morfémicas novedosas* y considera que parte del conocimiento de la morfología derivativa es la habilidad de entenderlas y producirlas. Para nosotros es importante estudiar estas formas porque nos dejan ver tres cosas: 1) que los estudiantes tienen el conocimiento receptivo del sufijo (sin necesidad de evaluarlos con pseudopalabras); 2) que realmente están realizando una tarea de derivación (es decir, son conscientes de que las palabras se pueden descomponer); y 3) que sólo les falta terminar de desarrollar el conocimiento distribucional del sufijo.

Basados en los ejemplos de formas creativas citados por los autores ya mencionados, junto con los ejemplos de Vargas-Vega y Velásquez Upegui (2019), quienes reportan un sobreuso de *-ation* (p. ej. **refination* por *refinement*) y la creación de palabras como **dispersement* en lugar de *dispersal*, parece que la productividad del sufijo tiene incidencia en la creación de palabras derivadas (*-ation* y *-ment* son muy productivos). Una aportación del presente trabajo es que sus resultados apuntan a que la alta frecuencia de un miembro de la familia de palabras o de una colocación también puede influir en las respuestas de una tarea morfológica. Nosotros encontramos repuestas como *punishment* en lugar de *punisher* o *day-dreaming* en lugar de *day-dreamer*, donde el primer miembro de los pares es más frecuente que el segundo. Por otro lado, llama la atención que los derivados cognados no hubieran tenido un efecto facilitador claro. A pesar de que la mitad de las palabras meta eran cognadas y las que terminan con *-able* son idénticas o casi idénticas en español e inglés (*deplorable*, *observable*, *comparable*, *applicable*), el porcentaje de respuestas correctas en los reactivos con *-able* fueron bajos.

Por último, considerando las diferencias entre grupos de competencia del objetivo (iii), descubrimos que el grupo más avanzado (B2) se distingue de los otros grupos al producir un mayor porcentaje de respuestas correctas y un menor porcentaje de palabras inventadas. Cabe señalar que el hecho de que un estudiante no provea la respuesta esperada no significa que no sabe inglés. Por ejemplo, en el reactivo de *planner*, hubo un estudiante de nivel B2 que respondió *organizer*, lo que evidencia su comprensión lectora, su conocimiento del sufijo y su habilidad para comunicarse,

aunque en este estudio no se le consideró como respuesta correcta debido a que no surge de la derivación de la base proporcionada. Creemos que para fines de evaluación en el aula o en actividades comunicativas, los docentes pueden usar casos como éste o casos en los que se demuestra un conocimiento parcial para motivar a los estudiantes e impulsar su conocimiento.

Una limitación de este trabajo es que, a pesar de que se controló la frecuencia de las palabras meta, no se controló la frecuencia del resto de las palabras de las oraciones. Por ejemplo, en un reactivo como *The punished was reprehended by the (punisher)*, es posible que los participantes no conocieran la palabra *reprehended* y eso influyera en sus respuestas. Otra limitación es que, al evaluar sólo un sufijo de cada categoría gramatical, nuestros resultados aplican únicamente a estos dos sufijos, no la adquisición de sufijos nominales o adjetivales en general.

REFERENCIAS

- Adamczyk, M. (2019). El análisis de errores en la enseñanza polaca del inglés frente a los estudios de errores de español lengua extranjera. *Vistula University Working Papers*, 64(1), 28-39.
- Al-Homoud, F. (2017). EFL learners' receptive and productive knowledge of word derivatives. *International Journal of English Language and Linguistics Research*, 5(2), 43-56.
- Ali, M., Hamid, A., y Rahman, G. (2021). Acquisition of English derivational morphemes students of selected universities in Khyber Pakhtunkhwa, Pakistan. *Liberal Arts and Social Sciences International Journal (LASSIJ)*, 5(1), 338-355.
DOI: <https://doi.org/10.47264/idea.lassij/5.1.22>
- Archibald, J., y Libben, G. (2018). Morphological theory and second language acquisition. En J. Audring y F. Massini (eds.), *The Oxford handbook of morphological theory* (pp. 522-540). Oxford University Press.
- Bauer, L. (1983). *English word-formation*. Cambridge: Cambridge University Press.
- Bauer, L. y Nation, P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253-279. DOI: <https://doi.org/10.1093/ijl/6.4.253>
- Bauer, L., Lieber, R. y Plag, I. (2013). *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Bilynsky, M. (2015). The productivity of deverbal categories and suffixal models within shared-root deverbal paradigms as reflected in the Oxford English Dictionary. *SKASE Journal of Theoretical Linguistics*, 12(2), 32-45.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, 12(3), 169-190.
DOI: <https://doi.org/10.1023/A:1008131926604>
- Curinga, R. (2013). Direct and indirect effects of morphological awareness on reading comprehension for adolescent Spanish-English emergent bilinguals. *BUCLD 37 Online Proceedings Supplement*, 1-12.

- <https://www.bu.edu/buclcd/files/2013/06/Curinga.pdf>
- Da Costa Moreira, B. (2014). Two types of dispositional adjectives. *Revista Virtual de Estudos da Linguagem, número especial 8*, 186-196.
<http://www.revel.inf.br/files/7a70ca1e233064a8e187be5ff025c037.pdf>
- Danilović, J., Savić, J. D., y Dimitrijević, M. (2013). Affix Acquisition Order in Serbian EFL Learners. *Romanian Journal of English Studies, 10(1)*, 77-88.
DOI: <https://doi.org/10.2478/rjes-2013-0006>
- Donoso, E., y Gómez, R. (2018). Errores en discurso escrito en inglés por futuros profesores de inglés como lengua extranjera. *Literatura y Lingüística, 38*, 175-193. DOI: <http://dx.doi.org/10.29344/0717621X.38.1632>
- Hayashi, Y., y Murphy, V. (2011). An investigation of morphological awareness in Japanese learners of English. *Language Learning Journal, 39(1)*, 105-120.
DOI: <https://doi.org/10.1080/09571731003663614>
- Hoyle, V. (2018). *Informe final de investigación: Errores morfológicos, sintácticos y ortotográficos en la traducción del español al inglés en los alumnos de los cursos «Inversa 1» e «Inversa 2» de la Facultad de Humanidades y Lenguas Modernas*. Universidad Ricardo Palma. https://repositorio.urp.edu.pe/bitstream/handle/20.500.14138/2235/Informe_final_investigacion_2018_VHoyle.docx?sequence=1&isAllowed=y
- Karlsson, M. (2015). Advanced students' L1 (Swedish) and L2 (English) mastery of suffixation. *International Journal of English Studies, 15(1)*, 23-49.
DOI: <https://doi.org/10.6018/ijes/2015/1/196731>
- Kieffer, M. J. y Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing, 21(8)*, 783-804. DOI: <https://doi.org/10.1007/s11145-007-9092-8>
- Kieffer, M. J. y Lesaux, N. K. (2012). Direct and indirect roles of morphological awareness in the English reading comprehension of native English, Spanish, Filipino, and Vietnamese speakers. *Language Learning, 62(4)*, 1170-1204.
DOI: <https://doi.org/10.1111/j.1467-9922.2012.00722.x>
- Kraut, R. (2015). The relationship between morphological awareness and morphological decomposition among English language learners. *Reading and Writing, 28(6)*, 873-890.
DOI: <https://doi.org/10.1007/s11145-015-9553-4>
- Lardiere, D. (2006). Knowledge of derivational morphology in a second language idiolect. En M. Grantham, C. Shea y J. Archibald (Eds.) *Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006)*, 72-79. Cascadilla Proceedings Project.
- Leontjev, D., Huhta, A., y Mäntylä, K. (2016). Word derivational knowledge and writing proficiency: How do they link?. *System, 59*, 73-89.
DOI: <https://doi.org/10.1016/j.system.2016.03.013>
- Marín Serrano, F. (2013). Análisis y diagnóstico de errores en estudiantes de inglés como lengua extranjera. *Exedra: Revista Científica, (8)*, 182-198.
- Medellín Gómez, A. y Auza Benavides, A. (2008). Influencia del tamaño del vocabulario en el conocimiento morfológico de afijos. *Estudios de Lingüística Aplicada, 47*, 97-108.
DOI: <https://doi.org/10.22201/enallt.01852647p.2008.47.570>

- Mochizuki, M., y Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291-304.
DOI: [https://doi.org/10.1016/S0346-251X\(00\)00013-0](https://doi.org/10.1016/S0346-251X(00)00013-0)
- Muse, A. E. (2005). *The nature of morphological knowledge*. (Tesis de doctorado). Florida State University, Florida.
<https://diginole.lib.fsu.edu/islandora/object/fsu%03A180396>
- Ramírez, G., Chen, X., Geva, E. y Kieffer, H. (2010). Morphological awareness in Spanish-speaking English language learners: Within and cross-language effects on word reading. *Reading and Writing*, 23(3-4), 337-358.
DOI: <https://doi.org/10.1007/s11145-009-9203-9>
- Ramírez, G., Chen, X., Geva, E. y Luo, Y. (2011). Morphological awareness and word reading in English language learners: Evidence from Spanish and Chinese-speaking children. *Applied Psycholinguistics*, 32(3), 601-618. doi:
DOI: <https://doi.org/10.1017/S0142716411000233>
- Sánchez-Gutiérrez, C. H., y Hernández Muñoz, N. (2018). Development of derivational morphological awareness in Anglophone learners of Spanish: A relational knowledge study. *Foreign Language Annals*, 51, 369-388.
DOI: <https://doi.org/10.1111/flan.12344>
- Sasao, Y., y Webb, S. (2017). The word part levels test. *Language Teaching Research*, 21(1), 12-30. DOI: <https://doi.org/10.1177/1362168815586083>
- Schmitt, N., y Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19(1), 17-36. doi: <https://doi.org/10.1017/S0272263197001022>
- Schmitt, N. y Zimmerman, C. B. (2002). Derivative word forms: What do learners know?. *Tesol Quarterly*, 36(2), 145-171. DOI: <https://doi.org/10.2307/3588328>
- Stanfa, K. (2010). *Differentiating among students: The value added of a dynamic assessment of morphological problem-solving* (Tesis de doctorado). University of Pittsburgh, Pensilvania. <https://d-scholarship.pitt.edu/10259/>
- Tyler, A. y Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of Memory and Language*, 28(6), 649-667.
DOI: [https://doi.org/10.1016/0749-596X\(89\)90002-8](https://doi.org/10.1016/0749-596X(89)90002-8)
- Varatharajoo, C., Asmawi, A. B., Abdallah, N. y Abedalaziz, M. (2015). The awareness of morphemic knowledge for young adults' vocabulary learning. *The Malaysian Online Journal of Educational Science*, 3(2), 45-56. <https://eric.ed.gov/?id=EJ1085913>
- Vargas-Vega, B. (2018). *Morfología derivativa nominal en el proceso de adquisición del inglés como L2* (Tesis de doctorado). Universidad Autónoma de Querétaro.
- Vargas-Vega, B. y Velásquez Upegui, E. P. (2019). Adquisición de sustantivos deverbales del inglés como L2. *Cuadernos de Lingüística Hispánica*, (33), 189-206.
DOI: <https://doi.org/10.19053/0121053X.n33.2019.7992>
- Vargas-Vega, B. y Márquez Valencia, A.S. (2022). Producción de sufijos heterocategoriales por estudiantes mexicanos de inglés como segunda lengua. *Semas*, 3(5), 97-119.
- Ward, J., y Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37(3), 461-469. DOI: <https://doi.org/10.1016/j.system.2009.01.004>

ANEXO 1

1	(punish) The punished was reprimanded by the _____.
2	(jog) The _____ are out for their morning exercise.
3	(smoke) People who smoke are called _____.
4	(dream) Day-_____ live their life in a different way.
5	(import) Nissan is an _____ of cars.
6	(plan) The _____ of the project calculated the costs.
7	(promote) The _____ organizes the boxing matches.
8	(compose) Mozart and Beethoven are classic _____.
9	(bear) A more _____ life requires good attitude.
10	(do) I know this is _____ because it has been done.
11	(regret) His decision to quit is _____.
12	(wash) _____ paints are used by moms with little children.
13	(deplore) Repugnant and _____ are the words to describe his attitude.
14	(observe) Some circumstances make comets bright and _____.
15	(apply) Some _____ laws negatively affect people.
16	(compare) This one contains more words than any _____ dictionary.

Dra. Brenda Vargas Vega
 Universidad Autónoma de Querétaro Campus Aeropuerto
 Anillo Vial Fray Junípero Serra S/N.
 CP 76040 Querétaro, Qro. México

Pierre-Luc Paquet
 Université de Montréal
 Faculté des sciences de l'éducation Pavillon Marie-Victorin
 90, avenue Vincent-d'Indy
 Montréal (Québec) H2V 2S9

ENUNCIACIÓN Y SIGNIFICADO APROXIMATIVO EN LA CONSTRUCCIÓN «ESTAR POR + INFINITIVO»

Adrià PARDO LLIBRER¹
Universidad de Salamanca

Resumen

El presente estudio investiga los usos aproximativos de la construcción «*estar por + infinitivo*» en español peninsular. El significado aproximativo se define a partir de dos componentes fundamentales: proximidad y polaridad, identificados tradicionalmente en ciertos adverbios y locuciones como *casi* o *por poco*, que indican cercanía al evento denotado por el predicado sobre el que tienen ámbito, al tiempo que niegan su realización. Nuestro planteamiento analiza el funcionamiento próximo-polar de la construcción «*estar por + infinitivo*»: gramaticalmente, se relaciona con las perífrasis de inminencia, pero adquiere un valor aproximativo bajo ciertas restricciones enunciativas.

Palabras clave: *estar por + infinitivo*; componente próximo; componente polar; significado aproximativo; enunciación

UTTERING AND APPROXIMATIVE MEANING IN THE SPANISH CONSTRUCTION «ESTAR POR + INFINITIVE»

Abstract

In this article, I explore the approximative properties of the Spanish construction «*estar por + infinitive*» (roughly similar to English 'to be about to' + infinitive).

1. pardolibrer@usal.es.  <https://orcid.org/0000-0003-0372-9422>

Este trabajo ha sido posible gracias al proyecto CIPROM/2021/038 *Hacia la caracterización diacrónica del siglo XX (DIA20)*, de la Generalitat Valenciana, y al proyecto de I+D+I PID2021-125222NB-I00 *Aportaciones para una caracterización diacrónica del siglo XX*, financiado por MCIN/AEI/10.13039/501100011033/ y por FEDER ('Una manera de hacer Europa'). Agradezco a los revisores de este artículo sus comentarios y correcciones, que he tratado de incorporar o hacer constar en la medida de mis limitaciones; todos los errores corresponden exclusivamente al autor.

Approximative meaning, traditionally identified in adverbials like *almost*, is defined as the juncture of two main components: proximity, expressing closeness to the achievement of a predication, and polarity, triggering a negative reading. Syntactically defined as an imminence construction, I attempt to show how «*estar por + infinitive*» also acquires an approximative meaning under certain uttering constraints.

Keywords: *estar por + infinitive*; proximity; polarity; approximative meaning; uttering

RECIBIDO: 15/11/2022

APROBADO: 14/10/2023

1. INTRODUCCIÓN

El presente estudio analiza los usos con significado aproximativo de la construcción *estar por + infinitivo* (en adelante, *estar por + INF*). Generalmente entendida como una perífrasis de inminencia (con el sentido de ‘estar a punto de’; Álvarez Martínez, 1992; Kovacci, 1999; Pavón Lucero, 1999; *NGLE*: § 40.9q²), esta construcción puede desempeñar otras funciones, relacionadas con la negación (del tipo *eso está por ver*) o la modalidad (*estar por* como ‘estar a favor de’). Nuestro planteamiento se interesa por el significado aproximativo que esta construcción puede adquirir en ciertos contextos discursivos: en algunos casos, la enunciación por parte de un hablante de *estar por + INF* sugiere una lectura a caballo entre la inmediatez en la realización de un estado de cosas designado y su negación (*estoy por darme de baja de mi compañía telefónica* ≈ ‘estoy cerca de darme de baja, pero aún no lo he hecho’).

Para abordar el significado aproximativo de esta construcción (§ 2), exponemos la similitud entre *estar por + INF* y los denominados adverbios aproximativos, es decir, las palabras del tipo *casi* o *apenas*: estas formas presentan un componente próximo o aspectual y otro polar o negativo (§ 2.1), que permiten describir algunos casos de *estar por + INF* propios de la oralidad (§ 2.2). Para acotar estos casos, se deslindan los tres usos de la construcción descritos en las principales gramáticas

2. En adelante, usamos la abreviatura *NGLE* para referirnos a la *Nueva Gramática de la Lengua Española*: nos servimos de estas siglas para las citas en las que se hace referencia al conjunto del compendio gramatical de la Real Academia Española (2009), en oposición a la postura de otras gramáticas o autores. Asimismo, nos referimos con las siglas *GDLE* a la *Gramática Descriptiva de la Lengua Española* (Bosque y Demonte, 1999) y, para ambos compendios, remitimos al apartado, no a la página (únicamente cuando se trate de cuestiones concretas de algún apartado de la *GDLE* se remite a su autor, p. ej.: Kovacci, 1999).

del español³ (§ 3): en primer lugar (§ 3.1), se distinguen los usos de *estar por* + INF, como fórmula negativa, de los usos propiamente perifrásticos, con un sentido de inminencia; en segundo lugar (§ 3.2), se revisa el tratamiento que han recibido los usos modales de esta construcción, que categorizamos como sentido doxático. Hecha esta criba, nuestro análisis se interesa por el funcionamiento aproximativo de *estar por* + INF (§ 4): a partir de la observación de esta construcción en contextos orales –o que reproducen la oralidad– (§ 4.1), es posible estipular una serie de restricciones discursivas bajo las cuales la enunciación *yo estoy por* adquiere un sentido aproximativo (§ 4.2). En definitiva (§ 5), el sentido aproximativo, aunque relacionado formalmente con los sentidos inminente y doxático, se revela como un uso particular de *estar por* + INF, cuya lectura negativa responde a la situación enunciativa (lo que sitúa los diferentes usos en un mismo continuum funcional).

2. EL SIGNIFICADO APROXIMATIVO Y LA CONSTRUCCIÓN «ESTAR POR + INFINITIVO»

El término *adverbio aproximativo* es hasta cierto punto vago (Alcina y Blecua, 1975; Vera Luján, 1979; Carbonero Cano, 1978; Álvarez Martínez, 1992; Kovacci, 1999; Pavón Lucero, 1999; NGLE: § 40.9q), pues abarca diferentes formas que precisan de una clasificación más detallada. A grandes rasgos, por adverbios aproximativos se entienden los «adverbios que indican un valor próximo al denotado por la cláusula sobre la que inciden» (NGLE 40.9q). Ahora bien, es frecuente que, bajo la etiqueta de *adverbio aproximativo*, se agrupen por igual adverbios léxicos acabados en *-mente* (como, p. ej., *aproximadamente*, García-Page, 1991: 201) y otros que, más que aproximativos, vendrían a denotar cantidad (*escasamente*, García-Page, 1995:171) o tiempo (*recientemente*, García-Medall, 1993:159). De igual manera, se recogen en la bibliografía tanto locuciones adverbiales (p. ej., *más o menos*, García-Page, 2007) como unidades fraseológicas (p. ej., *ni fu ni fa*, Ruiz Gurillo, 1997) con un valor en principio aproximativo, cuyo uso, por otra parte, difiere del de formas como puedan ser *aproximadamente* o *prácticamente* y que, a su vez, presentan diferencias sintácticas con otros adverbios léxicos no acabados en *-mente* como son *casi* o *apenas* (Moreno Cabrera, 1984; González Rodríguez, 2008). La denominada investigación sobre adverbios aproximativos se desarrolla en torno a este último grupo de voces que, frente a otros adverbios igualmente tenidos por aproximativos, han recibido un tratamiento especializado

3. Compendios referidos en la nota anterior.

(Sadock, 1981; Horn, 2002; Amaral, 2007; Pons Bordería y Schweter, 2011; Ziegeler, 2000, 2016; Pardo Llibrer, 2018):

- (1) *Casi* me mato
- (2) *Apenas* hablo inglés
- (3) *Por poco* suspendo el examen

El interés de estos trabajos suele limitarse a *casi*, *apenas* y otros elementos afines (como la partícula mandarina *chā-yīdiar*, Li, 1976; o la construcción *por poco*, Pons Bordería y Schweter, 2005a, 2005b). Estas formas constituyen el principal objeto de estudio de la investigación sobre aproximativos (4-5), en oposición a otras formas con las que, sin ser idénticas, tienen un aire de familia (6):

- (4) Luis cobra casi 1000 € al mes → $\{ \dots_{28, 29} \text{ pero} \} \text{NO-30}$
(= Luis cobra no más de 1000 €)
- (5) Luis cobra apenas 1000 € al mes → $\text{SÍ-30}_{\{6, 31, 32, \dots\}}$
(= Luis cobra no menos de 1000 €)
- (6) Luis cobra aproximadamente 1000 € al mes → $\{ \dots_{28, 29} \text{ pero} \} \text{NO-30} / \text{SÍ-30}_{\{6, 31, 32, \dots\}}$
(= Luis puede cobrar más o menos de 1000 €)

El uso de *casi* (4) o *apenas* (5) determinan, respectivamente, una lectura inferior o superior a la cantidad modificada (Jayez, 1987; Jayez y Tovená, 2007, 2008), mientras que las formas del tipo *aproximadamente* (o *más o menos*, o *alrededor de...*) permiten una interpretación en ambas direcciones y, más que propiamente aproximativas, constituyen elementos que introducen vaguedad en el discurso (vaguificadores, Eklund, 2001, 2005; o, en sentido amplio, los llamados *hedges*, Wierzbicka, 1986; Wachtel, 1980, 1981; Channell 1980, 1985; Sauerland y Stateva 2007; Sauerland 2011). En contraposición a las formas vagas del tipo *aproximadamente*, el subconjunto formado por *casi*, *apenas* o *por poco* presenta un significado específicamente aproximativo (§ 2.1), cuya descripción parece ser replicable en algunos usos de la construcción *estar por* + INF (§ 2.2).

2.1. *Lo próximo y lo polar*

De acuerdo con los estudios sobre adverbios del tipo *casi* o *apenas*, el significado aproximativo se puede describir según la combinación de dos componentes (Sevi, 1998; Horn, 2002, 2009, 2011; Pons Bordería y Schweter, 2011; Greenberg y Ronen, 2013): un componente próximo, que denota cercanía a la realización del evento referido por

predicado sobre el que el adverbio tiene ámbito, y un componente polar, que implica una lectura negativa. Según estos dos componentes fundamentales de significado:

- (7) *Casi* me muero
 a. He estado cerca de morir (= CERCANÍA a P)
 b. No me he muerto ($\rightarrow \neg P$)

El uso de *casi* en (7) denota proximidad a la realización de ‘morir’ (P), al tiempo que entraña la negación del enunciado ($\neg P$). Se da la situación inversa con *apenas* (8):

- (8) *Apenas* hablo inglés
 a. Hablo inglés, pero poco (= CERCANÍA a $\neg P$)
 b. Hablo inglés ($\rightarrow P$)

El uso de *apenas* denota proximidad a la negación del predicado ($\neg P$), con lo que se connota un distanciamiento con respecto a ‘hablar inglés’ pero, en cualquier caso, el enunciado es formalmente afirmativo. Se da así una imagen especular entre el funcionamiento aproximativo de ambas formas:

	COMPONENTE PRÓXIMO	COMPONENTE POLAR
<i>Casi</i> me manché	Proximidad a P (= he estado cerca de mancharme)	$\neg P$ (= NO me he manchado)
<i>Apenas</i> me manché	Proximidad a $\neg P$ (= he estado cerca de no mancharme)	P (= sí me he manchado)

Figura 1. Componentes próximo y polar

En conclusión, el significado aproximativo se caracteriza por un componente próximo, que denota cercanía (*casi*) o alejamiento (*apenas*) con respecto al evento referido por el predicado, y un componente polar, que puede comportar la inversión de la polaridad del enunciado.

2.2. El caso de *estoy por*

En línea con el tratamiento aplicado a las formas del tipo *casi*, el presente estudio plantea la posibilidad de explicar ciertas realizaciones de la construcción *estar por* + INF en función de los compontes propios del significado aproximativo. Un ejemplo *ad hoc* sería el siguiente:

- (9) [un amigo a otro en un bar, esperando a un camarero que tarda en atenderlos]
 A: Madre mía, ¡cuánto tarda el camarero!
 B: (Oye/mira) (yo) estoy por irme...

De acuerdo con este ejemplo, el hablante B se vale de *estoy por irme* para manifestar un posicionamiento cercano al hecho de irse, si bien A puede inferir que la marcha de su interlocutor no se efectuará, al menos no de manera inmediata (= ‘estoy a punto de irme, aunque aún no me voy’).

A diferencia de los adverbios descritos en la sección anterior, el componente próximo no tendría aquí ámbito sobre un predicado determinado, sino que involucra todo el enunciado e invierte el punto de vista expresado por el hablante (García Negroni, 2009, 2016), como pueda ser la ‘intención de marcharse’. Este punto de vista puede inferirse tanto a través de un enunciado previo (la intervención de A) como a través del contexto (el hecho de que el camarero tarde en atenderlos); lo relevante aquí es que la lectura negativa (‘no me voy aún’) contrasta con el hecho de que *yo estoy por irme* se da en un marco discursivo previo (Anscombe, 1989) y orientado argumentativamente a la conclusión positiva ‘marcharse’. Luego, *estar por* + INF presenta usos discursivamente equiparables a los componentes del significado aproximativo, pero estos usos se identifican no tanto a partir de restricciones morfosintácticas, sino más bien a partir de restricciones enunciativas (García Negroni, 2017; García Negroni y Libenson, 2020): la formulación de (9), en primera persona del singular del presente de indicativo⁴ y dentro de contornos dialogales, hace de *estoy por* una construcción marcada en oposición a otras realizaciones de *estar por* + INF, que revisamos a continuación.

3. LAS CONSTRUCCIONES DEL TIPO «ESTAR POR + INFINITIVO» EN LA GRAMÁTICA

No todas las estructuras morfológicamente coincidentes con *estar por* + INF presentan el mismo significado. Independientemente del uso aproximativo arriba apuntado, pueden señalarse, a partir de las descripciones que ofrece la gramática, algunas estructuras sintácticas en las que:

Muestran sentido pasivo los infinitivos transitivos introducidos por la preposición *sin* cuando no llevan complemento directo [...] Un comportamiento similar muestra los infinitivos contruidos con la locución adverbial *a medio* (*un libro a medio leer*), así como los que se construyen como término de la preposición *por* en el contexto «(*estar*) *por* + infinitivo», con un significado próximo al de *estar* «(*estar*) *sin* + infinitivo» (NGLE: § 26.3.2d).

4. Un revisor anónimo señala acertadamente que esta formulación en primera persona del singular del presente de indicativo, desde el punto de vista pragmático, debe considerarse un acto de habla asertivo.

Así, *estar por* + INF se puede caracterizar, de entrada, como una locución emparentada con otras locuciones de tipo adverbial (*medio*) o preposicional (*sin*) y con sentido negativo. No obstante lo común de este uso de *estar por* + INF, la *GDLE* considera que este tipo de expresiones «[n]o debe[n] entenderse como perífrasis», por lo que «[c]abe también la posibilidad de considerar estas construcciones como ‘locuciones verbales’» (§ 51.1.5.8). Aunque en otras referencias gramaticales se vuelve sobre la denominación de perífrasis (*NGLE*: §28.3.2a), la consideración de locución para *estar por* + INF vendría a indicar un significado del verbo conjugado (*estar*) que difícilmente se corresponde con una forma propiamente auxiliar. Como señalan los principales compendios gramaticales de nuestra lengua (*GDLE*: § 52.1.1.4, *NGLE*: § 28.1.2), las perífrasis verbales se hallan en un continuum de mayor a menor prototipicidad. Este continuum va desde un mayor vaciamiento semántico del verbo auxiliar hasta otras combinaciones en que la misma forma auxiliar mantiene un significado más bien cercano al del verbo original (Fernández Marín, 2014; Garachana Camarero, 2017). La *NGLE* (§ 28.1.2: 530) lo ejemplifica como sigue:

Una parte de ellas [de las perífrasis] se explica mejor si se supone que *Pudo pagar sus deudas* se segmenta en la forma [*Pudo pagar*] [*sus deudas*], donde la perífrasis forma un grupo verbal que excluye el complemento directo de *pagar*, pero otras se analizan más adecuadamente si la segmentación se hace en la forma [*Pudo*] [*pagar sus deudas*], es decir, como un grupo verbal que contiene a su vez otro grupo en su interior.

Como en estos casos, la estructura *estar por* + INF se aleja del criterio de prototipicidad en lo que respecta a su categorización como perífrasis, por lo que es necesario desgranar todas sus posibles funciones.

En definitiva, la gramática oscila entre considerar *estar por* + INF o bien como una perífrasis no prototípica (tempoaspectual y «de fase preparatoria», *NGLE*: § 28.1.4c), o bien como una locución verbal relacionada con otras estructuras negativas como «el grupo preposicional ⟨*por* + infinitivo⟩ cuando modifica a *estar*, como en *una tesis que aún está por demostrar*» (*NGLE*: § 28.3.2a). Las diferentes denominaciones sugieren que *estar por* + INF constituye un único esquema formal con diferentes lecturas discursivas, de entre las que destacan un uso temporal con sentido de inminencia (§ 3.1) y un uso que llamamos doxático (§ 3.2), del cual se sirve el hablante para formular una opinión.

3.1. *Sentido inminente y sentido tempoaspectual*

La construcción *estar por* + INF puede presentar un sentido de inminencia (o de inmediatez⁵). Como señala Bravo Martín (2011: 2-3), el sentido de inminencia entronca con el significado tempoaspectual: sin ser exactamente equivalentes, dentro de las perífrasis tempoaspectuales existe la subcategoría de las perífrasis fasales, que «aluden a los diversos estadios que se pueden considerar en un proceso» (NGLE: § 28.1.4c). Estos estadios pueden remitir a las fases final (perífrasis terminativas: *acabar de comer*) e inicial (perífrasis incoativas: *empiezo a comer*), pero también a las fases durativa (perífrasis cursiva: *estoy comiendo*) y previa (esto es, las perífrasis de inminencia). Las perífrasis de inminencia remiten a la fase previa y se distinguen de las iniciales en que la acción que denotan todavía no ha comenzado; dentro de estas, se incluye ocasionalmente la construcción *estar por* + INF (cuyo «valor aspectual de inminencia o inmediatez en la realización de una acción es bastante frecuente»; GDLE: § 51.3:2.7). La inclusión de *estar por* + INF dentro de la heterogénea categoría de las perífrasis de inminencia es anecdótica, en comparación con otras estructuras que se recogen recurrentemente bajo este rasgo semántico:

- (10) La sopa *está para* comer
- (11) La tormenta *está al* caer
- (12) Fulanito *está a punto de* jubilarse
- (13) Me *voy a* duchar

De entre todos estos tipos, únicamente (13) se corresponde con una perífrasis. Los casos de (10) a (12) se pueden descartar como estructuras perifrásticas por los motivos que a continuación se aducen.

En primer lugar, todos estos casos presentan dificultades a la hora de categorizarlos como perífrasis: sin duda, denotan en cierto modo un sentido de inminencia, pero están más unidos por su parecido en torno a este rasgo semántico que por criterios formales definitorios. Por ejemplo, (10) es considerado, en su sentido de inminencia, como una forma más propia de la lengua clásica, cuya distribución sintáctica está más cerca de una estructura consecutiva con valor final (NGLE: § 28.3.2a). Por su parte, (11) presenta una nominalización del infinitivo mediante el

5. En cuanto a los términos de inmediatez e inminencia, cabe precisar que lo inmediato se entiende como un hecho que se sigue de otro (sin mediación de tiempo entre ambos), mientras que la inminencia remite más bien a la propiedad con la que definir tal hecho. Morfológicamente distintos, en adelante, se remite a este sentido como *inminente* o *de inminencia*, ya que pensamos que terminológicamente es el más adecuado (*inmediatez* proviene del participio pasado –ya realizado– *inmediato*).

artículo *el*, por lo que *estar* no funciona como verbo auxiliar. Ambos casos, (10) y (11), no pueden considerarse formalmente perifrásticos, aunque semánticamente (por las preposiciones finales *a* y *para*, que «imprime[n] a las oraciones [...] un acusado valor prospectivo»; *GDLE*: § 36.4.2.3) puedan asemejarse a la familia de perífrasis temporales.

Por su parte, la cuestión de la nominalización es determinante para distinguir (10) de (11) y (12), dado que en estos últimos ejemplos *estar a punto de* «no admite nominalizaciones: **El tren está a punto de la llegada*/*...*de eso*/* ¿*De qué está a punto el tren?*» (*GDLE*: § 51.3.2.7). Otras pruebas contribuyen a la caracterización de ‘*estar a punto de + INF*’ como perífrasis: los argumentos están seleccionados por el infinitivo y no admiten estructuras de relativo (**De llegar es de lo que está a punto el tren*). Ahora bien, «el carácter perifrástico no es completo ya que, cuando el sujeto es de persona, la secuencia es sustituible por una oración completiva ([*El coche estuvo a punto de arrollarme* →] *Estuve a punto de ser arrollado por un coche*)», además de que la anteposición de clíticos «no parece del todo agramatical: ??*Se están a punto de celebrar las elecciones* [frente a *Están a punto de celebrarse*]]» (*GDLE*: § 51.3.2.7)⁶. De esta manera, «el carácter perifrástico de esta construcción ha sido puesto en tela de juicio porque «*a punto de + infinitivo*» puede usarse como expresión predicativa sin el verbo *estar*, como en *una especie a punto de extinguirse*» (*GDLE*: § 28.3.2a). El cuestionamiento de dicho carácter perifrástico, en suma, nos indica que *estar a punto de* se trata de una perífrasis no prototípica, en la que el sentido de inminencia persiste sobre las modificaciones morfosintácticas, como muestra la combinación con ciertos sintagmas nominales télicos («*Estuve a punto de la quiebra*, *Estuve a punto del infarto*»; *GDLE*: § 51.3:2.7), sin necesidad de seleccionar un infinitivo.

Por último, (13) sí que se reconoce como perífrasis plena; sin embargo, un caso como *ir a + INF* no se incluye dentro de las perífrasis fasales, sino dentro de las temporales. Dentro de los valores identificados para *ir a + INF*, la gramática resalta que «[e]s obvio que este valor [de inminencia] está unido al temporal de futuridad» (*GDLE*: § 51.3.2.7, ej. 136):

- (14) a. ¿Qué hora es? Van a ser las tres (= «están a punto de ser las tres»)
- b. Señores, va a comenzar el partido (= «está a punto de comenzar el partido»)
- c. Vámonos, que va a llover (= «que está a punto de llover»)

6. Un amable revisor nos señala que la oración *Se están a punto de celebrar las elecciones* es perfectamente gramatical en su variedad (a efectos de este trabajo, nuestro punto de partida es el español peninsular que, en este caso, coincide con el de la Gramática).

Así, la relación entre expectativa de futuro e inminencia es indisociable, hasta tal punto que lo primero incluye lo segundo: «[e]se futuro suele ser inmediato en la realidad [...] Con la perífrasis [*ir a + INF*], la futuridad se ve más próxima e, incluso, más segura en la apreciación del hablante» (GDLE: § 51.3.2.1)⁷. Prueba de ello es que estos usos de *ir a + INF* se pueden sustituir por un futuro simple y el sentido de inmediatez queda subsumido en el temporal, dado que deriva del sentido temporal (más marcado mediante la forma sintética) (GDLE: § 51.3:2.7, ej. 138⁸):

- (15) - Todos sabemos que, antes o después, {vamos a morir/moriremos}
 - Si te lo digo, ¿me {vas a querer/querrás} más?
 - Aunque sea dentro de veinte años, ese dinero me lo {vas a tener/tendrás} que devolver
 - Cuando acabe la carrera, me {voy a poner/pondré} a trabajar

Ahora bien, en algunos contornos sintácticos los usos tempoaspectuales de *ir a + INF* no presentan un sentido de inminencia (aquellos en los que «no siempre es posible sustituir la perífrasis de futuro por un futuro simple»; GDLE: § 51.3:2.7, ej. 139):

- (16) a. Si se {van a molestar/*molestarán} tus padres, no vamos
 b. Como no {voy a acertar/*acertaré}, no juego
 c. Aunque {me voy a separar/*me separaré}, estoy tranquila⁹

Esta no conmutabilidad entre formas de futuro dentro de las cláusulas adverbiales impropias dispone, entre otras (vid. Sánchez. 2011), una prueba formal clave para deslindar, dentro de los usos tempoaspectuales, aquellos de los que se sigue un sentido de inminencia de aquellos de los que no.

El mismo criterio utilizado para *ir a + INF* puede aplicarse a los casos de *estar por + INF* cuando esta construcción tiene un sentido de inminencia (NGLE: § 28.3.2a). Ello se puede observar mediante los siguientes ejemplos:

- (17) El vídeo que estás por ver a continuación es muy divertido

7. Un revisor anónimo señala su disconformidad esta explicación. Pensamos que la consideración de, además de *más próxima*, *más segura* ha de entenderse como un matiz ilocutivo que incorpora la Gramática para este caso, en el que se sobrentiende que esta *apreciación del hablante* se da un contexto prototípicamente inmediato.

8. Ejemplo modificado para nuestra argumentación.

9. A propósito de este ejemplo, un revisor señala que los casos de (16) tienen un matiz dubitativo. Esto es cierto, pero pensamos que este matiz dubitativo se deriva de la contraposición entre las cláusulas condicionales y concesivas y la negación –explícita en (16a) y (16b) e implícita (‘tranquila’ = ‘no nerviosa’) de (16c)– presente en la cláusula siguiente.

- (18) Fulanito está por cumplir 18 años
 (19) Tus hijos están por civilizar

En un ejemplo como (17), la construcción *estar por* + INF tendría un sentido de inminencia que se ajustaría a la prueba de la conmutabilidad con las formas temporales (17'); sin embargo, casos como (18') o (19') complican este análisis:

- (17') El vídeo que {vas a ver/verás} a continuación es muy divertido
 (18') Fulanito {va a cumplir/?cumplirá} 18 años
 (19') Tus hijos {?van a civilizarse/se civilizarán}

Así, mientras que (17') permite la sustitución tanto por *ir a* + INF como por el futuro simple, la forma sintética resulta extraña en (18'); por su parte, ambas conmutaciones son dudosas en (19'). Ello se puede perfilar mediante la inclusión de estos predicados en cláusulas subordinadas impropias:

- (17'') Aunque el vídeo que {vas a ver/verás} es muy divertido, tiene su parte de seriedad
 (18'') Si Fulanito {va a cumplir/?cumplirá} 18 años, podrá votar en las próximas elecciones¹⁰
 (19'') *Si/?aunque/?como tus hijos van a civilizarse/se civilizarán, siguen siendo unos salvajes

De este modo, es posible distinguir un uso de *estar por* + INF con un sentido de inminencia en (17''), mientras que (18'') y (19'') se ajustan a otra caracterización semántica. El caso más evidente es el de (19) como fórmula negativa (o sea, como «grupo preposicional <por + infinitivo> cuando modifica a *estar*» (NGLE: § 28.3.2a):

- (19''') Tus hijos están por civilizar = Tus hijos no están civilizados

De acuerdo con esto, (19) recibe una interpretación negativa porque «no denota inminencia sino ausencia o carencia [...] El infinitivo que contiene presenta forma activa e interpretación pasiva» (NGLE: § 28.3.2a). Tal y como indica la GDLE (§ 51.3.2.7, ej. 70b¹¹), esta predicación en forma activa e interpretación pasiva puede incluir una cláusula subordinada completiva en función de sujeto:

- (20) Está por ver ₁que se acerque a casa de su hermano₁
 (≈ No se ha visto ₁que se haya acercado a casa de su hermano₁)

10. A propósito de este ejemplo, véase nota 6.

11. Ejemplo ampliado para nuestra argumentación.

Por su parte, (18^o) es ambiguo; tiene un valor claramente temporal, pero parece estar relacionado con las estructuras nominales de interpretación pasiva arriba aludidas, con la consiguiente lectura negativa (o sentido de carencia¹²), lo que enlaza con su difícil interpretación en términos de inminencia:

(18^o) Fulanito está por cumplir 18 años = Fulanito no ha cumplido 18 años

Por consiguiente, hasta donde interesa a nuestro estudio, casos de *estar por* + INF como los de (18) y (19) pueden descartarse como perífrasis con un significado temporal.

3.2. Sentido doxático

Además del sentido inminente contenido en el significado temporal de *estar por* + INF, algunos autores han insistido en el carácter marcadamente modal de esta construcción (Gómez Torrego, 1988; Olbertz, 2007; Giammatteo y Marcovecchio, 2010). Este carácter modal queda indicado en la gramática como sigue (GDLE: § 44.3.1.2):

Al valor de futuridad se añade, en ocasiones, un valor modal intencional. Ello ocurre, por ejemplo, en oraciones condicionales *en cuya prótasis solo cabe la perífrasis y no el futuro* o condicional simples, precisamente porque *lo dominante*, en estos casos, *es la modalidad* y no el tiempo. De hecho, lo ‘condicional’ es un tipo de modalidad¹³.

Este valor modal se reconoce, entonces, como *dominante* y se le atribuye un sentido intencional que dista mucho del sentido inminente. Asimismo, se indica la no conmutabilidad con el futuro sintético cuando *estar por* + INF se ubica en una cláusula adverbial impropia. A modo de ejemplo:

- (21) Yo estoy por legalizar las drogas
- a. Si yo {estoy por legalizar / *legalizaré} las drogas, es solo mi opinión
 - b. Yo no estoy por legalizar las drogas (sino [por legalizar] la prostitución)
 - c. Yo estoy por la legalización de las drogas

12. Cuestión aparte son las lecturas negativas que pueden derivarse de los usos de inminencia. Con respecto a la forma *estar a punto de*, la GDLE señala que «[c]on el auxiliar en pasado se indica una acción que no se realizó pero que estuvo muy cerca de realizarse. De ahí que sea equivalente semánticamente a las formas adverbiales *casi, por poco (no), etc.*» (§ 51.3.2.7). Tal paralelismo con estos adverbios, sin embargo, es una paráfrasis de las inferencias negativas que pueden seguirse de *estar a punto de* + INF conjugado en pasado; inferencias que, en cualquier caso, son cancelables (se trata de una clase de inferencias descrita por Kuteva (1998) como *action narrowly averted*).

13. Resaltado nuestro.

Un caso como (21) no puede conmutarse gramaticalmente mediante formas con valor de futuridad evidente (21a), sino –en todo caso y de acuerdo con la cita de arriba– por formas condicionales (?*Si yo legalizaría (o no) las drogas, es solo mi opinión*¹⁴). Además, la negación puede focalizarse sobre el argumento seleccionado por el verbo pleno (21b), así como este puede nominalizarse (21c) (como ocurre con los usos no perifrásticos con matices finales; cf. *supra* ejs. 10–11). Pudiendo parafrasearse por ‘estar a favor de’, denominamos este uso como sentido doxático, puesto que sirve al hablante para expresar una posición subjetiva con respecto a un estado de cosas (en línea con otros estudios sobre fenómenos perifrásticos definidos como epistémicos; *vid.* Cornillie, 2004, 2005).

Es importante señalar que las lecturas negativas asociadas al sentido doxático, si bien parecen solaparse con las lecturas de tipo ‘grupo preposicional con *por*’ –las denominadas lecturas de carencia; ejs. (18) a (19)–, no están gramaticalizadas (22a), sino que son cancelables (22b):

- (22) a. Yo estoy por privatizar la sanidad
(+> *no lo está*)
b. Yo estoy por privatizar la sanidad, así que, por mí, que siga siendo privada
(+> *lo está*)

Al igual que son cancelables las lecturas más o menos convencionalizadas (GDLE: § 51.3.2.7) que en un contexto dado pueden generarse del sentido de inminencia (23b):

- (23) a. El vídeo que estás por ver
(+> *pero no verás*)
b. El vídeo que estás por ver
(+> *y verás*)

A diferencia de los usos de *estar por* + INF como grupo preposicional (*tus hijos están por civilizar*), cuya lectura negativa es indisociable de su estructura (cf. 10–11), ni los usos perifrásticos con sentido inminente, ni los usos con sentido doxático, implican necesariamente negación. Son, eso sí, usos de los que pueden inferirse lecturas negativas –cf. (22a) y (23a)–, pero estas lecturas responden a una implicatura, al menos en la medida en que (22b) y (23b) se prestan a una cancelación. Finalmente, la cancelabilidad de dichas lecturas permite también distinguir los

14. La única interpretación posible de esta oración es la de *Si yo legalizaría las drogas o no* como cláusula completiva.

sentidos inminente y doxático de otro uso analizable desde una perspectiva discursiva: *estar por* + INF en su funcionamiento aproximativo.

4. SOBRE EL USO APROXIMATIVO DE LA CONSTRUCCIÓN «ESTAR POR + INFINITIVO»

Puede sumarse, a los sentidos inminente y doxático de *estar por* + INF, un sentido aproximativo. Volviendo sobre el ejemplo (9), aquí reenumerado:

- (24) [un amigo a otro en un bar, esperando a un camarero que tarda en atenderlos]
 A: Madre mía, ¡cuánto tarda el camarero!
 B: (Oye/mira) (yo) *estoy por* irme...

Nuestro planteamiento considera que este uso puede describirse en términos de proximidad y polaridad:

- (24') (yo) *estoy por* irme (*estoy-por*—P)
 a. COMPONENTE PRÓXIMO: proximidad a P
 b. COMPONENTE POLAR: (pero) (aún) ¬P

Caben dos pruebas formales que perfilan este funcionamiento aproximativo de *estar por* + INF, frente a los usos con sentido inminente o doxático. La primera tiene que ver con el componente próximo: el funcionamiento aproximativo de *estar por* + INF, al denotar una cercanía tal a la realización de P, admite otros elementos escalares, como el adverbio de foco *hasta*. Por ejemplo:

- (25) Yo *hasta* *estoy por* irme

Esta combinación con *hasta* (25) se adecua también al sentido doxático (26), pero no parece gramatical con el sentido inminente (27):

- (26) Luis *hasta* está por privatizar la sanidad
 (27) ??El vídeo que *hasta* estás por ver

Mientras que en el sentido inminente resulta extraño (27)¹⁵, los sentidos aproximativo (25) y doxático (26) admiten una focalización escalar con *hasta*;

15. Otra cuestión sería la de una lectura con los condicionantes enunciativos del uso aproximativo (*Yo hasta estoy por ver el vídeo*, en un contexto que presuponga no poder o no deber verlo), en cuyo caso no cabría plantear la lectura de inminencia.

ello haría equivalentes ambos usos, de no ser por una segunda prueba formal, relativa al componente polar: las lecturas negativas que puedan desprenderse de la construcción *estar por* + INF responden a implicatura en el sentido doxático, mientras que la formulación de un ejemplo como (24) presupone necesariamente dicha lectura negativa:

- (24") Yo estoy por irme
 (→ pero aún no me voy)
 (→ #pero me voy)

Esto no ocurre con el sentido doxático, que no presupone una negación, de ahí que las lecturas negativas de este puedan, en principio (ejs. 22a y 22b), cancelarse:

- (22') Luis está por privatizar la sanidad
 (→ la sanidad aún es privada)
 (→ la sanidad ya es privada)

Al binomio próximo-polar hay que añadirle una expectativa temporal (ya asociada a otras construcciones aproximativas; Amaral y Del Prete, 2010; Pons Bordería y Schwenter, 2011) que proyecta hacia el futuro la realización del 'irse':

- (28) (yo) estoy por irme (*estoy-por*-P)
 a. COMPONENTE PRÓXIMO: 'proximidad a P'
 b. COMPONENTE POLAR: '(pero) aún ¬P'
 c. TEMPORALIDAD: ¬P en tiempo₀ → P en tiempo₊₁

Consiste en un uso especializado discursivamente en la proyección de eventos en el tiempo: se niega la realización de P en el momento de la enunciación (*tiempo*₀) y remite a su posible realización en el futuro (*tiempo*₊₁). Así, en esta estructura, el componente próximo, orientado argumentativamente hacia P, sirve al hablante para expresar un punto de vista favorable a 'marcharse' (18a) pero, al enunciarse, implica automáticamente una lectura negativa (28b), por la que se interpreta una concesión a cambiar de intención. De esta manera, el sentido aproximativo de *estar por* + INF se revela como un uso a caballo entre el valor modal-intencional (lo doxático, proyectado en un *tiempo*₊₁) y una negación de lo dicho (lo inminente, en un *tiempo*₀).

Cabe preguntarse, entonces, por las condiciones que propician este uso aproximativo (§ 4.1), así como por los contextos en que este funcionamiento de *estar por* + INF emerge (§ 4.2).

4.1. Significado aproximativo y enunciación

En este apartado indagamos algunas ocurrencias de la construcción *estar por* + INF, relevantes para un primer acercamiento cualitativo (si bien futuros trabajos requerirán de un análisis de corpus más amplio). Diversos ejemplos¹⁶ apuntan a que el funcionamiento aproximativo depende de unas condiciones determinadas en la enunciación:

- (29) Tutú.- *Estoy por asegurar* que se trata de una de esas casas con bicho que Él y su socio compran por dos duros y, una vez eliminado el gusano, venden a precio de oro. (Roel, Lola: «Las ratas van a la oficina». Aullidos y humareda. Las ratas van a la oficina. Madrid: Fundamentos, 2010)
- (30) PACÍFICO.- ¿Se puede ya? (Silencio. Perpetua sigue cantando, como una letanía) Pues será que sí que se puede. (Entra y también se sienta, pegado a la pared) *Estoy por irme a hacer un cafelito...* (silencio) Con bollo maimón para mojar, que nos lo ha traído una que le encontraste novio. (Silencio) ¿Nadie gusta? (Silencio) Que el café está muy rico, que es del de el otro lado de la raya, del que trae el camellito pintado
(Ripoll, Laila: Santa Perpetua. Madrid: Huerga y Fierro Editores, 2011)

En primer lugar, es común encontrar la formulación aproximativa de *estar por* + INF en primera persona del singular del presente de indicativo; así, tanto (29) como (30) se pueden parafrasear en términos próximo-polares:

- (31) a. Estoy cerca de asegurarlo (*P* en *tiempo*₊₁), aunque no lo aseguro ($\neg P$ en *tiempo*₀)
b. Estoy cerca de tomar un café (*Q* en *tiempo*₊₁), aunque no lo tomo ($\neg Q$ en *tiempo*₀)

En segundo lugar, esta formulación en primera persona del singular tiene lugar en contextos dialogales, en los que la interacción permite al resto de interlocutores relacionar la enunciación de *estoy por* con una concesión a lo dicho mediante esta construcción ('estoy por tomarme un café, aunque no lo he tomado aún').

16. Todos los ejemplos provienen de la base de datos *CORPES XXI* (<<http://www.rae.es/recursos/banco-de-datos/corpes-xxi>>; consultado el 11/01/2021) a partir de una búsqueda de la forma *estoy por*, filtrada de los casos componenciales identificados en § 3 y en la variedad peninsular. El resultado es nuestro en todos los ejemplos de la sección.

El terreno de la oralidad es propicio para la aparición de *estoy por* y es frecuente encontrar contextos que reproducen marcas propiamente coloquiales, adyacentes a esta construcción aproximativa en la escritura:

- (32) -Deja, deja, *Ruth*... ¿Qué me vas a agradecer, imbécil? Al fin y al cabo, yo he estado viviendo en tu casa, y aunque no hubiera sido así... Para mí es un honor que duermas aquí. Es más, *estoy por pedirte* que te quedes a pasar unos días aquí. Así tú te sentirías mejor y podríamos trabajar juntos en el guión, que bastante retrasado lo llevamos, y mataríamos dos pájaros de un tiro.

(Etxebarria, Lucía: *De Todo lo Visible y lo Invisible. Una novela sobre el amor y otras mentiras*. Madrid: Espasa Calpe, 2001)

- (33) [...]ni que lo regalaran, el pescado, aguarda un fisco que cierro la ventana porque con el rebumbio que arma la camioneta no te oigo... ¿*Rosi*? Sí, ya cerré... No, no te apures, que ella salió de compras y no creo que llegue tan luego... *Oye*... *Que* desde el otro día *estoy por llamarte* pa contarte, m'hija, que me pasó una cosa que cuando me acuerdo, entodavía me daño toda, y de'que llegué hoy lunes ca's de Sita Celia no he hecho sino acechar a ver si se iba la señora pa llamarte...

(Criado, Ana: «Hay que ve'lo pa cre'lo». *El ruido de las miradas*. Madrid: Lengua de trapo, 2001)

Ambos fragmentos, (32) y (33), reproducen un fluir conversacional que presupone un interlocutor; como se colige¹⁷ de las pausas (representadas mediante puntos suspensivos) para la cesión o mantenimiento del turno, las instrucciones formulativas (*que* no subordinante) y los elementos fático-apelativos (tales como vocativos, *Ruth*/¿*Rosi*?, o los marcadores del discurso interpersonales, *Oye*). Asimismo, en línea con otras construcciones aproximativas canónicas (Pardo Llibrer, 2022: §3.4.2), *estoy por* puede conllevar la reformulación de un segmento discursivo previo:

- (34) No es fácil entender siempre lo que, en otro tiempo, preferí llamar amor homogéneo. Pero creo que este amor entre iguales, plural al tiempo como debe ser, alcanza su mayor veracidad, *estoy por decir su cumplimiento mejor*, cuando esos iguales se aceptan como iguales y en cierto modo, y a la par, como iguales contrarios.

(De Villena, Luis Antonio: *La nave de los muchachos griegos*. Madrid: Alfaguara, 2003)

17. Sobre la mimesis de la oralidad en lo escrito, véase López-Serena (2007 –esp. § 4–) y López-Serena y Sáez Rivera (2018).

En definitiva, las restricciones enunciativas para el funcionamiento aproximativo de *estar por* + INF apuntan hacia una formulación en primera persona del singular (*estoy por*) mediante la cual el hablante expresa un punto de vista proclive al contenido de lo dicho (Ducrot, 2000). Al tiempo que *estoy por* sirve de estructura marco para encauzar un posicionamiento del hablante (componente próximo), se niega su realización inmediata en el tiempo_o de la enunciación (componente polar) y, como correlato comunicativo, interpela el posicionamiento del otro hablante.

4.2. Enunciación aproximativa: relación estructural con el contexto discursivo

Se pueden resumir como siguen las cuatro restricciones enunciativas básicas para el funcionamiento aproximativo de *estar por* + INF:

- a. Expresa un punto de vista en primera persona del singular favorable a la consecución de lo dicho en un *tiempo*₊₁ (esto es, la expectativa de futuro).
- b. Se da en marcos discursivos que contraponen este punto de vista y la situación de la enunciación.
- c. Persigue una reacción del otro hablante (ya que el componente polar tiene como correlato una concesión al punto de vista del interlocutor).
- d. Predomina en la oralidad (aunque es igualmente reproducible en la escritura).

Estos cuatro factores para la interpretación aproximativa de *estar por* + INF invitan a vincular la enunciación con el contexto discursivo. Para explicar esta vinculación, es posible relacionar estas condiciones pragmático-discursivas con ciertas restricciones estructurales. Tanto *-a-* el punto de vista del hablante como *-b-* su contraposición con la situación enunciativa, fundamentales para la interpretación próximo-polar, guardan relación con la aparición de *estar por* + INF en las posiciones iniciales del enunciado. Así, mientras que un ejemplo como (35) tiene una lectura aproximativa, esta no es posible en (36):

- (35) *Estoy por irme* mañana a Santander a verla otra vez (mientras escribo esto todavía quedan entradas). Y que le vayan dando a Barbra Streisand.

(Belmonte, Rosa: «Liza, la estrella no fugaz de la noche». ABC.es. Madrid: abc.es, 2007-08-14)

- (36) Padre, me deja usted muy sorprendida. *Estoy por preguntarle* ahora a usted lo mismo que le he preguntado a él. Padre, ¿es usted creyente?


(Lindo, Elvira: Una palabra tuya. Madrid: Seix Barral, 2005)

El primer ejemplo (35) da pie a una interpretación aproximativa ('estoy por ir a Santander, aunque no vaya de momento') difícilmente replicable en el segundo ejemplo (36): por una parte, no se da una lectura negativa, como constata el hecho de que finalmente se realice la pregunta (*Padre, ¿es usted creyente?*); por otra parte, *estoy por* no ocupa aquí una posición inicial desde el punto de vista discursivo. Quiere esto decir que la construcción aparece gráficamente después de un punto, pero, en el proceso de mimesis oral, (36) reproduce una intervención en la que *estoy por* no encabeza el enunciado (en todo caso, la posición inicial la ocupa el vocativo *Padre* seguido de una coma). Por añadidura, con (36) es plausible una interpretación en los términos del sentido de inminencia, lo que no resulta tan viable con (35):

- (35) ??{Voy a ir / iré} mañana a Santander a verla otra vez (mientras escribo esto todavía quedan entradas). Y que le vayan dando a Barbra Streisand.
 (36) Padre, me deja usted muy sorprendida. {Voy a preguntarle / le preguntaré} ahora a usted lo mismo que le he preguntado a él. Padre, ¿es usted creyente?

En cuanto a las restricciones concernientes a *-c-* la reacción del interlocutor y *-d-* la oralidad como lugar por defecto de aparición de esta construcción, cabe remarcar cómo *estar por* + INF, en su funcionamiento aproximativo, aparece sistemáticamente en encadenamientos discursivos. En las pocas ocurrencias estrictamente orales –y no miméticas de la oralidad– que nuestra muestra arroja, la enunciación de *estoy por* se sigue de un enunciado anterior, pero también determina la continuidad del discurso. Este tipo de intervenciones, en los estudios sobre análisis conversacional, se ha etiquetado como intervención reactivo-iniciativa (abreviado *Iri*; vid. Briz y Grupo ValEsCo, 2003; Grupo ValEsCo, 2014; Espinosa Guerri, 2016, 2019). Se trata de intervenciones que, en el desarrollo de un diálogo, constituyen una reacción que responde necesariamente a algo dicho anteriormente y que, al mismo tiempo, motivan la continuación de la conversación, es decir, son la reacción a una intervención iniciativa *-Ii-* previa y enlazan con la intervención reactiva *-Ir-* subsiguiente). Póngase por caso:


- (37) [concluyendo una entrevista radiofónica]
 A: sí sí sí
 B: lo que lo que nos comentaba Javier Urra
 A: es genial

- 
 [Ii] —B: al que le invitamos para dentro de muy poquito
 [Iri] —A: estoy por quedarme
 [Ir] —B: pues yo encantada / yo encantada
 A: oye muchísimas gracias Mariángeles
 B: gracias Javier / hasta muy prontito


(rtve.es: Últimas preguntas. Javier Urrea)

En calidad de intervención reactivo-iniciativa, la enunciación de *estoy por quedarme* en (37) no puede iniciar por sí misma un diálogo, sino que depende de un marco discursivo previamente articulado, encadenado con otro segmento requerido argumentativamente ('invitación a volver pronto al programa' → 'estoy por quedarme, aunque no pueda'). La aparición de *estar por* + INF en contextos conversacionales del tipo reactivo-iniciativo parece ser una restricción estructural para la enunciación aproximativa de esta construcción, como prueba lo incongruente de conmutar (37) por un sentido ya inminente (37'), ya doxático (37'')

(37')

- 
 [Ii] — B: al que le invitamos para dentro de muy poquito
 [Iri] — A: ?{voy a quedarme / me quedaré}
 [Ir] — B: pues yo encantada / yo encantada

(37'')

- 
 [Ii] — B: al que le invitamos para dentro de muy poquito
 [Iri] — A: ?{estoy a favor de / tengo la intención de quedarme}
 [Ir] — B: pues yo encantada / yo encantada

En el caso de emplear una equivalencia con el sentido inminente (37'), la intervención de A no es congruente con la reacción de B, dado que formula un punto de vista del interlocutor ('yo encantada') no aludido por las formas tempoaspectuales (*ir a* + INF o futuro simple); del mismo modo, una equivalencia con el sentido doxático (37''), si bien puede generar la reacción de B ('tengo intención de quedarme' → 'yo encantada'), no es coherente con la intervención iniciativa del discurso previo.

5. CONSIDERACIONES FINALES

En síntesis, la construcción *estar por* + INF presenta una enunciación (*estoy por*) que involucra un punto de vista y un contexto vinculados con ciertas restricciones

estructurales (posición inicial y consecución de intervenciones iniciativo–reactivas en la oralidad), lo que contribuye a distinguir el funcionamiento aproximativo de los otros sentidos. Estos tres sentidos se relacionan entre sí en torno a un mismo esquema:

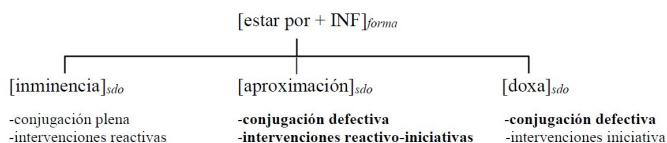


Figura 2. Esquema: forma y significado en *estar por* + INF

El uso inminente puede, en principio, formularse con el verbo *estar* en todas las personas de la conjugación, al igual que sus equivalentes perifrásticos (*ir a* + INF) o sintéticos (futuro simple), y puede constituir una reacción conversacional –cf. *supra* ej. (37')– sin continuar necesariamente el discurso. El uso doxático presenta una conjugación defectiva (tiende a estar formulado en primera persona) y –cf. *supra* ej. (37'')– puede dar lugar a otra intervención: enlaza con el componente próximo ascendente del uso aproximativo, pero sus lecturas negativas, cuando se dan, son cancelables. Por último, el uso aproximativo es claramente defectivo (*estoy por*)¹⁸: especializado discursivamente en posiciones iniciales, encadena –en tanto que intervención reactivo–iniciativa– dos puntos de vista, contraponiendo el del hablante al asumido en el discurso previo, esto es, el componente polar.

Tal y como han señalado algunos autores con respecto a otras formas con sentido de inminencia (Carrasco Gutiérrez, 2006; Giammatteo, Marcovecchio y Albano, 2011), y en línea con los valores modales (Giammatteo, 2014; *GDLE*: § 44.3.1.2) que, como el sentido doxático, desarrollan ciertos usos perifrásticos, se vislumbran algunas conexiones entre las diferentes realizaciones de *estar por* + INF. Luego, encontramos un sentido inminente, emparentado con otras perífrasis o formas verbales tempoaspectuales, de donde parten nuevos sentidos más especializados discursivamente. Dichos sentidos serían el aproximativo y el doxático: el primero presenta un componente próximo semejante al rasgo semántico de ‘inminencia’ y el segundo, ciertas lecturas negativas cancelables que se revelan, a modo de componente polar, automatizadas cuando *estoy por* se ajusta a las restricciones enunciativas pertinentes.

18. A este respecto, se nos han señalado en la revisión dos posibles casos en que *estoy por* puede funcionar en plural: el primer caso, un plural colaborativo (*Estamos por irnos si no nos atienden*) o un uso –en términos del revisor– no performativo (*Luis nos dijo que estaba por abandonar la carrera*); estos casos son, sin lugar a duda, plausibles y quedan pendientes para futuros análisis.

- Cornillie, B. (2004). The shift from lexical to subjective readings in Spanish *prometer* 'promise' and *amenazar* 'threaten'. A corpus-based account. *Pragmatics: A Quarterly Journal of the International Pragmatic Association* 141, 1-30.
- Cornillie, B. (2005). A paradigmatic view of Spanish *amenazar* 'to threaten' and *prometer* 'to promise'. *Folia Linguistica: Acta Societatis Linguisticae Europaeae* 39, 385-415.
- Ducrot, O. (2000). Quelques raisons de distinguer 'locuteurs' et 'énonciateurs'. *Les polyphonistes scandinaves/ De skandinaviske polyfonister* 3, 19-41.
- Eklund, M. (2001). Supervaluationism, vagueifiers, and semantic overdetermination. *Dialectica* 55, 363-387.
- Eklund, M. (2005). What vagueness consists in. *Philosophical Studies: A International Journal for Philosophy in the Analytic Tradition* 1, 27-60.
- Espinosa Guerri, G. (2016). Dientes de sierra: una herramienta para el estudio de la estructura interactiva del discurso dialógico. *Normas* 6, 13-27.
- Espinosa Guerri, G. (2019). Dibujar la conversación en ELE: la enseñanza de la toma de turnos del español, *Foro de Profesores de E/LE* 15, 105-125.
- Fernández Marín, P. (2014). Hacia un prototipo cuantitativo de perífrasis verbal en español. *Lenguas Modernas* 44, 31-54.
- Garachana Camarero, M. (2017). Los límites de una categoría verbal híbrida. Las perífrasis verbales. En M. Garachana Camarero (Eds.), *La gramática en la diacronía: la evolución de las perífrasis verbales modales*, pp. 35-80. Madrid-Fráncfort del Meno: Iberoamericana-Verwuert.
- García-Medall, J. (1993). Sobre *casi* y otros aproximativos. *DICENDA. Cuadernos de Filología Hispánica* 11, 153-170.
- García Negroni, M.^a M. y M. Libenson (2020). La evidencialidad desde el enfoque dialógico de la argumentación y de la polifonía. Un estudio contrastivo de los empleos inferencial y citativo del marcador evidencial *así que*. En Ó. Loureda Lamas, M. Rudka y G. Parodi (Eds.), *Marcadores del discurso y lingüística contrastiva en las lenguas románicas*, pp. 41-62. Madrid-Fráncfort del Meno: Iberoamericana-Verwuert.
- García Negroni, M.^a M. (2009). Negación y descalificación: a propósito de la negación metalingüística. *Ciências & Letras* 45, 61-82.
- García Negroni, M.^a M. (2016). Argumentación y puntos de vista evidenciales citativos: acerca de la negación metadiscursiva en el discurso político, *Oralia* 21, 223-242.
- García Negroni, M.^a M. (2017). El enfoque dialógico de la argumentación y la polifonía, puntos de vista evidenciales y puntos de vista alusivos. *RILCE* 35, 521-549.
- García-Page, M. (1991). Breves apuntes sobre el adverbio en *-mente*. *Thesaurus* 46, 183-224.
- García-Page, M. (1995). El adverbio de exactitud y aproximación. *Anuario de Lingüística Hispánica* 11, 169-182.
- García-Page, M. (2007). *Introducción a la fraseología española: estudio de las locuciones*. Barcelona: Anthropos.
- Giammatteo, M. y A. M.^a Marcovecchio. (2010). Perífrasis verbales: una mirada desde los universales lingüísticos. *Sintagma* 21, 21-38.

- Giammatteo, M., A. Marcovecchio y H. Albano. (2011). Dos dominios en intersección: habitualidad y posibilidad. Su manifestación en las perífrasis verbales. En J. Cuartero, L. García Fernández y C. Sinner (Eds.) *Estudios sobre perífrasis y aspecto*, pp. 139-157. Múnich: Peniopo.
- Giammatteo, M. (2014). La modalidad volitiva y los verbos de actitud intencional. *Traslaciones. Revista Latinoamericana de Lectura y Escritura* 1, 116-141.
- Gómez Torrego, L. (1988). *Perífrasis verbales. Sintaxis, semántica y estilística*. Madrid: Arco-Libros.
- González Rodríguez, R. (2008). Sobre los modificadores de aproximación y precisión. *ELUA. Estudios de Lingüística Universidad de Alicante* 28, 111-128.
- Grupo Val.Es.Co (2014). Las unidades del discurso oral. La propuesta Val.Es.Co. de segmentación de la conversación (coloquial). *Estudios de Lingüística del Español* 35, 13-73.
- Greenberg, Y. y M. Ronen (2013). Three approximators which are almost/more or less/be-gadol the same, ponencia presentada en el 28.º IATL.
- Horn, L. (2002). Assertoric inertia and NPI licensing. En M. Andronis, E. Debenport, A. Pycha y K. Yoshimura (Eds.), *Proceedings from the panels of the thirty-eighth meeting of the Chicago Linguistic Society* 38, pp. 55-82. Chicago: Chicago Linguistic Society.
- Horn, L. (2009). *Almost et al.: Scalar adverbs revisited*. En C. Lee, F. Kiefer y M. Krifka (Eds.), *Contrastiveness in information structure, alternatives and scalar implicatures*, pp. 283-304. Ámsterdam: Springer.
- Horn, L. (2011). *Almost forever*. En Y. Etsuyo, T. Bagchi y K. Beals (Eds.), *Pragmatics and autolexical grammar, in honor of Jerry Sadock*, pp. 1-21. Ámsterdam: John Benjamins.
- Jayez, J. y L. Tovená (2007). Discourse inference and the meaning of *presque*. En R. Nouwen y J. Dorlacil (Eds.), *Proceedings of the 2007 ESSLLI Workshop on Quantifier Modification (Dublin)*, pp. 36-41. Edición en línea: <http://perso.ens-lyon.fr/jacques.jayez/doc/presque.pdf>
- Jayez, J. y L. Tovená (2008). *Presque and almost: how argumentation derives from comparative meaning*. En O. Bonani y P. Cabredo Hofherr (Eds.), *Empirical issues in syntax and semantics*, pp. 217-239. Edición en línea: http://www.cssp.cnrs.fr/eiss10/index_en.html
- Jayez, J. (1987). Sémantique et approximation: le cas de *presque* et à *peine*. *Linguisticae Investigationes* 11, 157-196.
- Kovacci, O. (1999). El adverbio. En I. Bosque y V. Demonte (Eds.), *Gramática descriptiva de la lengua española*, pp. 705-786). Madrid: Espasa-Calpe.
- Kuteva, T. (1998). On identifying an evasive gram: Action narrowly averted. *Studies in Language* 22, 113-160.
- Li, C. (1976). A functional explanation for an unexpected case of ambiguity (S or ~S). En A. Juilland (Eds.), *Linguistics studies offered to Joseph Greenberg on occasion of his sixtieth birthday (Studia Linguistica et Philologica)*, pp. 527-535. Saratoga (California): Anma Libri.
- López-Serena, A. y D. M. Sáez Rivera. (2018). Procedimientos de mimesis de la oralidad en el teatro español del siglo XVIII. *Estudios Humanísticos: Filología* 40, 235-273.

- López-Serena, A. (2007). *Oralidad y escrituralidad en la recreación literaria del español coloquial*. Madrid: Gredos.
- Moreno Cabrera, Juan Carlos. (1984). Observaciones sobre la sintaxis de *casi*, *DICENDA. Cuadernos de Filología Hispánica* 3, 239-245.
- Olbertz, H. (2007). ¿Perífrasis verbales con función de marcador de discurso? A propósito del Diccionario de perífrasis verbales. *Verba* 34, 381-390.
- Pardo Llibrer, A. (2018). Los usos atenuantes del aproximativo *casi*: relación entre contenido proposicional y unidades del discurso. En A. García y M.^a A. Soler (Eds.), *ELUA: Estudios de atenuación en el discurso* (Anexo IV), pp. 33- 56.
- Pardo Llibrer, A. (2022). *Casi* y el significado aproximativo en español. Propuesta teórica y estudios de caso. Berlín: Peter Lang.
- Pavon Lucero, M.^a V. (1999). Sintaxis básica de las clases de palabras. En I. Bosque y V. Demonte (Eds.), *Gramática descriptiva de la lengua española*, pp. 565- 656. Madrid: Espasa-Calpe.
- Pons Bordería, S. y S. Schwenter (2005a). *Por poco no*: explicación sincrónica y diacrónica de sus componentes de significado, *Linguística Española* 27, 131-156.
- Pons Bordería, S. y S. Schwenter (2005b). Polar meaning and *expletive* negation in approximative adverbs: Spanish *por poco (no)*. *Journal of Pragmatics* 6, 268-282.
- Pons Bordería, S. y S. Schwenter (2011). Los significados próximo y polar de *casi*. En R. González y C. Llamas (Eds.), *Gramática y discurso*, pp. 159-188). Pamplona: EUNSA.
- Real Academia Española (2009). *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe. [NGLE]
- Ruiz Gurillo, L. (1997). Aspectos de fraseología española. *Cuadernos de Filología* 24 (anejos), Valencia: Universitat de València (Departamento de Filología Española).
- Sadock, J. (1981). *Almost*. En P. Cole (Eds.), *Radical pragmatics*, pp. 257-271. Nueva York: New York Academic Press.
- Sánchez, L. (2011). La interfaz léxico-morfosintáctica: una perífrasis de inminencia en el español andino del Perú. *Cuadernos de la ALFAL* 3, 216-228.
- Sauerland, U. y P. Stateva (2007). Scalar vs. epsitemic vagueness: Evidence from approximators. En M. Gibson y T. Friedman (Eds.), *Proceeding of SALT* 17, pp. 228-245. Ithaca: CLC Publications.
- Sauerland, U. (2011). Vagueness in language: The case against fuzzy logic revisited. En P. Cintula, C. G. Fermüller, L. Godo y P. Hájek (Eds.), *Understanding vagueness: Logical, philosophical and linguistic perspectives*. *Studies in logic*, pp. 185-198. Londres: College Publications.
- Sevi, A. (1998). *A semantics for 'almost' and 'barely'*, Tesis de máster, Universidad de Tel Aviv.
- Vera Luján, A. (1979). La problemática gramatical/funcional en una tipología categorial: el adverbio en español. *Lexis* 111, 171-194.
- Wierzbicka, A. (1986). Precision in vagueness: The semantics of English *approximatives*. *Journal of Pragmatics* 10, 597-614.
- Ziegeler, D. (2000). What can *almost* reveal about counterfactual inferences. *Journal of Pragmatics* 32, 1743-1776.

- Ziegeler, D. (2010). Running the gauntlet on the approximatives debate: A response to recent challenges. *Journal of Pragmatics* 42, 681-704.
- Ziegeler, D. (2016). Intersubjectivity and the diachronic development of counterfactual *almost*. *Journal of Historical Pragmatics* 17, 1-25.

Adrià Pardo Llibrer
Departamento de Lengua Española
Facultad de Filología
Universidad de Salamanca
Plaza de Anaya s/n
37008 Salamanca

SECCIÓN MONOGRÁFICA

FENÓMENOS DE LA DIGLOSIA ÁRABE:
DE LA EDAD MEDIA A LA EDAD CONTEMPORÁNEA


FENÓMENOS DE LA DIGLOSIA ÁRABE: DE LA EDAD MEDIA A LA EDAD CONTEMPORÁNEA¹

Montserrat BENÍTEZ FERNÁNDEZ²
Escuela de Estudios Árabes CSIC

María Ángeles GALLEGO³
Instituto de Lenguas y Culturas del Mediterráneo y el Oriente Próximo, CSIC

El origen de este monográfico se remonta al debate científico desarrollado en el seno de la quincuagésima edición del *Simposio de la Sociedad Española de Lingüística* –celebrado en el Centro de Ciencias Humanas y Sociales del CSIC, entre los días 5 y 8 de abril de 2022, en Madrid– y, más concretamente, a la sesión monográfica dedicada a *Contacto lingüístico y diglosia*, coordinada por Montserrat Benítez Fernández, María Ángeles Gallego y Jesús de la Villa Polo. En dicha sesión se plantearon diversas cuestiones relacionadas con el contacto lingüístico y la diglosia, con especial protagonismo de las lenguas semíticas (árabe y hebreo) y las lenguas clásicas (griego y latín), que nos permitieron contrastar no sólo los fenómenos lingüísticos planteados, sino también las metodologías y tradiciones de estudio en torno a este tema en los campos de la Filología Clásica y la Filología Semítica. Aunque este monográfico se centra en la lengua árabe, las discusiones que surgieron en dicha sesión sobre las otras lenguas y ámbitos de estudio respectivos han enriquecido las contribuciones y están en gran medida incorporadas a los artículos, todos ellos presentados originalmente como comunicaciones en dicho *Simposio*.

1. Parte de los resultados presentados en este artículo están incluidos en el trabajo desarrollado por M. A. Gallego dentro del proyecto de I+D+i *Ciencia y religión en el judaísmo medieval* (PID2020-118688GB-I00), financiado por MCIN/AEI/10.13039/501100011033/ y del realizado por M. Benítez Fernández en el marco del proyecto *Variación diastrática en las variedades habladas del árabe vernáculo de Marruecos*, financiado por MICIN/AEI/FEDER (FFI2017-87533-P).

2. montsebenitez@eea.csic.es.  <https://orcid.org/0000-0002-8103-1428>

3. mariangeles.gallego@cchs.csic.es.  <https://orcid.org/0000-0001-9987-5680>

El concepto central de aquel panel y de este monográfico es, por lo tanto, el de «diglosia», que los investigadores llevan más de un siglo analizando desde que Psichari lo aplicase en 1885 para referirse a la situación de la lengua griega. Si nos centramos en el uso del término aplicado a la lengua árabe, fue William Marçais (1930) quien lo aplicó por primera vez a esta lengua. Sin embargo, el concepto lleva estando en el centro de los debates lingüísticos desde que Ferguson lo popularizara en su polémico artículo publicado en *Word* en 1959. Para teorizar sobre la cuestión, Ferguson aplica el concepto a cuatro comunidades de lengua diferentes entre las que se encuentran, como no podía ser de otra manera, el griego y el árabe, pero también otras situaciones diglósicas como las que se producen en Suiza y en Haití. En aquel trabajo el fenómeno se define de la siguiente manera:

(...) una situación relativamente estable en la que, además de los dialectos primarios (que pueden incluir un estándar o estándares regionales), se encuentra una variedad superpuesta muy diferente y altamente codificada (a menudo gramaticalmente más compleja), vehículo de un corpus amplio y respetado de literatura escrita —bien de un período anterior o de otra comunidad de lengua—, ampliamente aprendida a través de la educación formal y usada principalmente por escrito o en situaciones formales orales y no en conversaciones ordinarias por ningún sector de la comunidad. (Ferguson, 1959, p. 336)⁴

Ferguson llega a esta definición tras describir 9 rasgos que considera característicos de la diglosia y que se producen en los cuatro estudios de caso de su trabajo. Estos rasgos son los siguientes: 1) la función, ya que tanto la variedad *H*[*igh*] como la variedad *L*[*ow*] ejercen funciones que les son propias; 2) el prestigio, ya que tradicionalmente la variedad *H* es más respetada que la variedad *L*, llegando incluso a negarse la existencia de esta última; 3) el patrimonio literario que se expresa normalmente en la variedad *H*; 4) la adquisición, que se realiza mediante la socialización primaria para la variedad *L* y a través de la enseñanza para la variedad *H*; 5) la estandarización, que se traduce en la amplia tradición de estudios gramaticales relativos a la variedad *H*; 6) la estabilidad, ya que es una situación que perdura durante varios siglos; 7) la gramática, que diferencia a las variedades *H* y *L* porque la primera suele disponer de categorías gramaticales que no existen en la segunda; 8) el léxico, ya que la variedad *H* debe incluir expresiones técnicas que no siempre tienen su equivalente en la variedad *L*; 9) la fonología, porque el inventario fonológico de ambas variedades forman parte de un único sistema. En

4. La traducción es nuestra.

ese sistema, los sonidos de la variedad *H* que no se encuentran en la variedad *L* a menudo son remplazados por los de esta última variedad.

Desde aquel trabajo seminal, numerosos son los estudios que se han centrado en esta cuestión desde diferentes perspectivas. Hay investigadores que han tratado de completar las lagunas del trabajo de Ferguson (1959) añadiendo la noción de bilingüismo al concepto de diglosia (Fishman, 1967) o que la han concebido como producto del contacto de lenguas (Sayahi, 2014). Centrándonos exclusivamente en el caso de la lengua árabe –ya que tratar sobre otras lenguas excedería el marco temático y espacial de este artículo–, se ha mostrado que, raramente, la diglosia concierne exclusivamente a dos niveles (Badaoui, 1973; Youssi, 1983; Moscoso García, 2010) y la investigación ha tratado de explicar en qué consiste cada uno de los estratos que se encuentran en el concepto de la diglosia, especialmente las variedades mixtas que se encuentran entre el árabe clásico y las variedades vernáculas (Mitchel, 1986; Youssi, 1992; Mejdell, 2011 y 2012). Los estudios han determinado, incluso, que las diferentes variedades de la lengua árabe no son las distintas caras de una moneda, sino que forman parte de un único sistema, lo que se ha dado en llamar continuo lingüístico (Hary, 1996). Por otro lado, se ha puesto de manifiesto la existencia y convivencia de varias normas prestigiosas (Abdel Jawad, 1986); se han analizado las funciones que las variedades alta (*H*) o baja (*L*)⁵ ejercen en la sociedad (Haeri, 2003; Bassiouney, 2010) y cómo funciones propias de la variedad *H* no son ya exclusivas del árabe moderno estándar (Miller, 2013; Caubet, 2017; Achour, 2022; Ferrando, 2012). De la misma manera, los investigadores han reflexionado sobre la posibilidad de que este concepto –la diglosia– sea fruto, en realidad, de la ideología lingüística dominante en la que la lengua clásica es un ideal inalcanzable (Ferguson, 1996; Brustad, 2017; Hachimi, 2013).

Es importante señalar, en cualquier caso, que mucho antes de que surgiera el concepto de diglosia como tal y la discusión que hemos resumido más arriba, existió en el mundo árabe conciencia sobre la existencia de dos grandes variedades de la lengua árabe. Desde la época medieval se diferenció entre ambas con las denominaciones genéricas de *al-luġa al-fuṣḥà* (= lengua correcta/elocuente, a la que solemos referirnos como «árabe clásico») y *al-luġa al-‘āmmiyya* (= lengua del vulgo/de la gente común, a la que solemos referirnos como «árabe vulgar»). El origen de

5. En el caso del árabe, la variedad alta o *H*, es la llamada «árabe clásico», «literario», «moderno», «estándar», entre otros términos, mientras que la variedad baja, o *L*, es denominada «árabe vernáculo», pero también «dialecto», «árabe dialectal», «árabe vulgar», «árabe oral», etc. En esta publicación trataremos todos estos conceptos, tanto los relativos a la variedad *H* como a la variedad *L*, como sinónimos.

esta división se atribuyó tradicionalmente a la corrupción o deterioro de la única variedad que se reconoce como auténtica lengua árabe, es decir, *al-luġa al-fuṣṣḥà*, a consecuencia de su contacto con las lenguas de los pueblos no-árabes⁶. Dentro de esta narrativa, el contacto que supuestamente desencadenó el proceso de corrupción lingüística tuvo lugar en la etapa de la gran expansión del islam entre los siglos VI y VIII e.c., en la que grandes comunidades no-arabófonas y practicantes de religiones distintas del islam, fueron incorporadas al nuevo imperio islámico⁷. Así lo expresa en diversos pasajes de su *Introducción a la historia universal (Muqaddima)* el famoso intelectual musulmán Ibn Jaldún (1332-1406):

La lengua de la población de oriente se diferencia en alguna medida de la lengua de la gente de al-Magrib, y lo mismo ocurre con la de la gente de al-Andalus con respecto a ambas. Todos ellos consiguen expresar, cada uno en su lengua, lo que desean comunicar (...) La carencia de vocales casuales no les supone impedimento ninguno, como dijimos a propósito de los árabes beduinos de esta época. Y en cuanto a que está más alejada de la lengua árabe primitiva que la de los árabes beduinos actuales, ello se debe a que el alejamiento de la lengua procede siempre del contacto con poblaciones extranjeras, de manera que quien más se mezcla con ellas hace que su lengua se aleje más de la original (...) Esto es algo que se puede observar en las ciudades de Ifriqiya, de al-Magrib, de al-Andalus y de oriente. En Ifriqiya y en al-Magrib los árabes se mezclaron con los beréberes que allí había, de habla no árabe (...) Por eso las formas extranjeras se impusieron en la lengua árabe que ellos emplean y se formó una lengua mezclada en la que lo no árabe predomina, como ya dijimos, por lo que se trata de una lengua alejada de la lengua primitiva. Y lo mismo ocurre en oriente, donde los árabes dominaron a persas y turcos y se mezclaron con ellos. (...) La lengua se fue corrompiendo al alterarse su uso y acabó convirtiéndose en otra lengua. Y lo mismo ocurrió en al-Andalus por el contacto con extranjeros gallegos y francos [= cristianos]. (Ruiz Girela, 2008, pp. 687-688)

6. La realidad es que muy posiblemente existieran variedades neo-árabes ya en época preislámica (véase Corriente, 1971). Aunque no existe una única teoría consensuada, en la actualidad se entiende que el origen del neo-árabe o «árabe vulgar» en sus múltiples variedades responde a una serie de factores que tienen que ver con la propia evolución interna de la lengua, junto con fenómenos de adstrato, substrato y superestrato que sí estarían relacionados con el contacto con lenguas no-árabes pero también con el contacto entre los distintos dialectos árabes. Véase un resumen sobre las diversas teorías en Versteegh (2001, pp. 93-113).

7. Así, por ejemplo, lo refiere Ibn Jaldún (Ruiz Girela, 2008, p. 687) al hablar de las características lingüísticas de la gente de las ciudades. Según el pensador musulmán, estos grupos de población que tras la conquista islámica adoptaron la lengua árabe fueron un factor determinante en su pérdida de desinencias y corrupción general.

Según la tradición islámica de época medieval, la necesidad de corregir los errores y evitar la contaminación lingüística que había producido el contacto con los no-árabes estuvo en el origen de la composición de la primera gramática árabe. Este propósito estaba, además, estrechamente relacionado con el ámbito religioso puesto que la preservación y conocimiento del árabe clásico resultaba esencial para una correcta comprensión del Corán. Estos aspectos quedan claramente reflejados en el relato legendario transmitido por Abū Bakr Muḥammad ibn al-Anbārī (m. 1181)⁸ en torno al nacimiento de la ciencia gramatical. Según al-Anbārī, en esa primera etapa de expansión del islam, el gobernador de Basora, preocupado por el efecto pernicioso que tenía en la lengua árabe el elevado número de extranjeros en su territorio y que había llevado a su deterioro y corrupción, solicitó a Abū al-Aswad al-Duʿalī (603-689) que compilara y describiera los principios de la lengua sagrada. Inicialmente, el sabio musulmán rechazó la propuesta al considerar que no estaba a la altura de poder formular las normas de la lengua de la Revelación. Sin embargo, posteriormente, al escuchar por la calle los errores que se cometían al recitar el texto árabe del Corán, decidió finalmente aceptar el encargo.

Esta visión del neo-árabe como una versión contaminada o vulgar del árabe clásico impidió que existiera una literatura escrita en las diversas variedades habladas, con contadas excepciones como los refraneros o algún otro tipo de literatura popular, tal y como sucede en las situaciones diglósicas. Podemos reconstruir, sin embargo, este estadio de desarrollo del neo-árabe en la época medieval a través de diversos mecanismos. Uno de ellos es utilizar los testimonios recogidos en un género conocido como *Laḥn al-ʿamma* o *Tratado de errores lingüísticos cometidos por la gente común*. El primer tratado que nos ha llegado de este género fue obra del andalusí Abū Bakr al-Zubaydī (928-989) quien justifica su obra con los consabidos argumentos que describen la lengua árabe vulgar o *ʿammiyya*: «Se trata de alteraciones debidas al vulgo, que ha modificado la pronunciación o adaptado significados, siendo seguidos en esta práctica por mucha gente, hasta el punto de que estos usos incorrectos se han infiltrado en la obra de poetas y hasta los más eminentes escritores y funcionarios los incluyen en su correspondencia y utilizan expresiones depravadas en sus conversaciones»⁹.

Otro de los procedimientos habituales para obtener información sobre el neo-árabe es el estudio de los diversos niveles de lengua que reflejan obras pretendidamente escritas en árabe clásico pero que, de hecho, contienen abundantes rasgos

8. Traducción adaptada del texto inglés de Versteegh (1997, p. 3).

9. Cita recogida en Pellat (2023).

que no se corresponden con la variedad elevada. Como suele ocurrir en las situaciones diglósicas, la estricta división entre «variedad alta» y «variedad baja» que existe en el imaginario de los hablantes no se corresponde a la realidad lingüística. En la mayor parte de los escritos medievales que habitualmente adscribimos lingüísticamente al árabe clásico se pueden hallar numerosas interferencias del árabe vulgar, así como otra serie de rasgos que surgen de un conocimiento limitado de la variedad elevada como, por ejemplo, las pseudocorrecciones¹⁰. En relación con esta línea de trabajo contamos en este monográfico con el artículo de Estefanía Valenzuela Monchón, «La variación lingüística en fuentes árabes de al-Andalus: el Manual de ‘Abd al-Ra’ūf (Córdoba, s. X)», en el que la autora analiza los rasgos no-estándar que aparecen en dos manuscritos de un manual de *hisba* andalusí del siglo X en los niveles ortográfico, fonético, léxico, morfológico y sintáctico. Identifica, además, una serie de factores que intervienen en la expresión de la variación lingüística que indicarían la existencia de patrones regulares en el empleo de esas formas no-estándar en una aproximación novedosa en este campo. Otro aspecto abordado en este monográfico en relación con el uso de la lengua árabe en al-Andalus y, más concretamente, con los préstamos léxicos del árabe andalusí al romance castellano, es el estudio realizado por María Ángeles Gallego y Patricia Giménez Eguibar en torno a «A/Al- inicial como identificador de arabismos en la *Recopilación* de Diego de Guadix» En esta contribución las autoras analizan este conocido diccionario del siglo XVI y describen específicamente los mecanismos utilizados para el establecimiento de etimologías a través de las entradas que comienzan por *a/al-*, así como la percepción general de Diego de Guadix y de algunos contemporáneos suyos sobre la lengua árabe y la existencia de sus dos grandes variantes (la «mala» y la «buena» algarabía).

Como vemos, en el mundo arabófono ya se producían debates en torno a la diglosia incluso antes de que se acuñara el término, pero, en la actualidad, la coexistencia de lenguas y culturas ha aumentado exponencialmente con respecto a otros periodos del pasado en un mundo hiperconectado como el actual. De hecho, varios factores extralingüísticos inciden en la hiperconectividad y el contacto entre lenguas, variedades, etc. Entre esos factores vamos a mencionar cuatro que nos

10. Los rasgos de la lengua coloquial o que no se corresponden al árabe clásico suelen ser más habituales en escritos de autores judíos o cristianos ya que estas comunidades no comparten la percepción islámica del árabe como lengua sagrada. Por otra parte, el desconocimiento no es en su caso necesariamente la causa de no seguir fielmente los paradigmas de la variedad elevada por lo que el propio concepto de pseudocorrección no es aplicable en estos casos. Véase Gallego (2010).

parecen especialmente paradigmáticos e influyentes en el contacto de variedades. En primer lugar, en las últimas décadas se ha producido un especial desarrollo de los medios de comunicación en el sector audiovisual del mundo arabófono. Esto se debe a la liberación del panorama audiovisual en numerosos países árabes, a la proliferación de canales de TV por satélite y de cadenas de radio de proximidad, a la aparición de plataformas de visionado en *streaming*, etc. En segundo lugar, ha aumentado la movilidad, tanto geográfica como social de los individuos debido a movimientos migratorios tanto dentro como fuera del mundo arabófono. En tercer lugar, se ha desarrollado, paulatinamente y dependiendo de los países, el nivel de escolarización y, con ello, se ha producido un aumento de la tasa de alfabetización. Por último, se han multiplicado los cauces a través de los que se establecen las relaciones sociales hoy en día (en vivo, a través de RRSS) lo que ha favorecido una mayor familiarización con procesos de escritura. Estos factores han beneficiado el contacto de variedades lingüísticas a diferentes niveles: el contacto de variedades altas (H-H) –árabe clásico y otras lenguas europeas o árabe clásico y estándares nacionales/regionales todas ellas consideradas lenguas/variedades prestigiosas–, el contacto árabe clásico-árabe dialectal (H-L) –ya que el mayor grado de alfabetización de los hablantes favorece que un número más importante de ellos tenga acceso al árabe clásico o un contacto más cotidiano con esta variedad– o el contacto entre variedades habladas (L-L) –tradicionalmente consideradas variedades no prestigiosas–. Todo ello está reconfigurando las relaciones entre variedades lingüísticas y las jerarquías que se establecen entre ellas.

En esta sección monográfica nos hemos centrado en tres debates que, como decimos, muestran esa «reconfiguración» de las jerarquías lingüísticas entre las que tradicionalmente se han considerado «variedad alta» (árabe clásico) y «variedad baja» (árabes vernáculos). De esta manera, observamos la emergencia de un nuevo término, el «árabe blanco» (o *White Arabic*), cuya particularidad principal reside en haber sido acuñado por los propios hablantes. El «árabe blanco», según explican Iriarte Díez, Laaber, van Kampen y Benítez Fernández en este volumen, es un término complejo, que puede hacer referencia a diferentes realidades. Por un lado, puede tratarse de una variedad no marcada que utilizan los hablantes de un país para evitar variantes dialectales estigmatizadas, una suerte de estándar nacional y, por tanto, una variedad prestigiosa que puede utilizarse en situaciones semiformales como las intervenciones televisivas. Por otro lado, puede ser concebido como una variedad mixta en la que se mezclan elementos del árabe clásico con formas de distintos dialectos que se utiliza en contextos transnacionales como estrategia de

acomodación y en situaciones igualmente semiformales, como los debates entre universitarios, en concurrencia directa con lo que los lingüistas llaman *Educated Spoken Arabic* –término acuñado por Mitchel (1986)–. Otro de los aspectos en los que hemos puesto el foco, mediante el artículo elaborado por Pennisi, es el análisis del registro formal escrito. En este trabajo, la autora examina diferentes artículos del periódico marroquí en línea *Goud*. Al tratarse de un texto escrito y de una situación formal, se espera que la variedad elegida sea el árabe clásico, sin embargo, Pennisi demuestra que se trata de una variedad mixta, cuyos rasgos son heterogéneos. Por último, Benítez Fernández elabora un amplio estado de la cuestión sobre la noción de diglosia y aplica los últimos avances en torno a esta cuestión a la situación lingüística marroquí. En su trabajo, Benítez Fernández analiza la emergencia de estándares regionales, el desarrollo de funciones de las variedades vernáculas y, debido a ese desarrollo de funciones, el peligro de atrición o pérdida de rasgos en variedades marcadas o estigmatizadas por entrar en competencia con los estándares regionales.

Lo que perseguimos en esta sección monográfica es reflexionar de nuevo sobre el concepto de diglosia desde perspectivas tan diferentes como la diacrónica y la sincrónica, así como desde distintas cotas, el nivel macro –que llevamos a cabo en el análisis de registros, el estudio de nuevos conceptos o la observación de la ideología lingüística que se desprende de la noción de diglosia– y el nivel micro –que realizamos mediante el examen de rasgos no estándar–, el análisis del contacto de lenguas o de situaciones de habla. Igualmente, hemos tratado de hacer especial hincapié en los nuevos contextos en los que se encuentran las variedades H y L. Nos referimos tanto a una relación más frecuente con las primeras, gracias a una mayor alfabetización de la población, como a un desarrollo de funciones de las segundas, por su aparición en medios de comunicación audiovisuales y escritos que conlleva, a su vez, una cierta desacralización –en sentido literal y figurado– de la variedad H.

REFERENCIAS

- Abdel Jawad, H. R. (1986). The emergence of an urban dialect in the Jordanian urban centres, *International Journal of the Sociology of Language* 61, 53-63.
DOI: <https://doi.org/10.1515/ijsl.1986.61.53>
- Achour, M. (2022). La diglossie traversée: La littérature en tunisien et le tunisien dans la littérature, *International Journal of the Sociology of Language* 278, 23-50.
DOI: <https://doi.org/10.1515/ijsl-2022-0018>

- Badawi, E. A. (1973). *Mustawayat al-lughā al-‘arabiyya al-mu‘asira fi miṣr* [Levels of Contemporary Arabic in Egypt]. Cairo: Dar al-Ma‘arif.
- Bassiouney, R. (2010). Identity and code-choice in the speech of educated women and men in Egypt: Evidence from talk shows. En R. Bassiouney (Ed.), *Arabic and the Media*, pp. 97-121. Londres: Brill. DOI: <https://doi.org/10.1163/ej.9789004182585.i-303.50>
- Brustad, K. (2017). Diglossia as Ideology. En J. Høigilt y G. Mejdell (Eds.), *The Politics of Written Language in the Arab World. Writing Change*, pp. 41-67. Leiden: Brill. DOI: https://doi.org/10.1163/9789004346178_004
- Caubet, D. (2017). Morocco: An Informal Passage to Literacy in dārija (Moroccan Arabic). En J. Høigilt y G. Mejdell (Eds.), *The Politics of Written Language in the Arab World. Writing Change*, pp.116-141. Leiden: Brill. DOI: https://doi.org/10.1163/9789004346178_007
- Corriente, F. (1971). On the functional yield of some synthetic devices in Arabic and Semitic morphology, *Jewish Quarterly Review* 62, 20-50. DOI: <https://doi.org/10.2307/1453861>
- Ferguson, Ch. A. (1959). Diglossia, *Word* 15: 2, 325-340. DOI: <https://doi.org/10.1080/00437956.1959.11659702>
- Ferguson, Ch. A. (1996). Epilogue: Diglossia revisited. En A. Elgibali (Dir.) *Understanding Arabic: Essays in Contemporary Arabic Linguistics in Honor of El-Said Badawi*, pp. 49-67. Cairo: American University of Cairo Press.
- Ferrando I. (2012). El árabe marroquí en la novela del siglo XXI. El caso de Salem Himmich y su obra *فتنة الرووس والنسوة*, en A. Barontini, Ch. Pereira, Á. Vicente y K. Ziamari (Eds.), *Dynamiques langagières en Arabophonies: variations, contacts, migrations et créations artistiques. Hommage offert à Dominique Caubet par ses élèves et collègues*, pp. 415-426. Zaragoza: Área de Estudios Árabes, Universidad de Zaragoza.
- Fishman, J. A. (1967). Bilingualism With and Without Diglossia; Diglossia With and Without Bilingualism, *Journal of Social Issues* 23: 2, 29-38. <https://doi.org/10.1111/j.1540-4560.1967.tb00573.x>
- Gallego, M. Á. (2010). The Impact of Arabic Diglossia among the Muslims, Jews and Christians of al-Andalus. En F. Cabo Aseguinolaza, A. Abuín González y C. Domínguez (Eds.), *A Comparative History of Literatures in the Iberian Peninsula* 1, pp. 351-365. Ámsterdam: John Benjamins. <https://doi.org/10.1075/chle.xxiv.17ang>
- Hachimi, A. (2013). The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world, *Journal of Sociolinguistics* 17: 3, 269-296. DOI: <https://doi.org/10.1111/josl.12037>
- Haeri, N. (2003). *Sacred language, ordinary people: Dilemmas of culture and politics in Egypt*. Nueva York: Palgrave Macmillan.
- Hary, B. (1996). The importance of the language continuum in Arabic multiglossia, en A. Elgibali (Ed.), *Understanding Arabic: Essays in contemporary Arabic linguistics in honor of El-Said Badawi*, pp. 69-90. Cairo: American University of Cairo Press.
- Marçais, W. (1930). La diglossie arabe, *L'Enseignement Public – Revue pédagogique* 104: 12, 401-409.

- Mejdell, G. (2011-2012). Diglossia, Code Switching, Style Variation, and Congruence: Notions for Analyzing Mixed Arabic, *Al-'Arabiyya* 44-45, 29-39.
DOI: <https://www.jstor.org/stable/43208722>
- Miller C. (2013). Évolution des usages linguistiques dans les nouvelles radios marocaines. En M. Benítez Fernández, C. Miller, J. J. de Ruiter y Y. Tamer (Eds.), *Évolution des pratiques et des représentations linguistiques dans le Maroc du XXIème siècle*, pp.14-59. París: L'Harmattan.
- Mitchell, T. F. (1986). What Is Educated Spoken Arabic?, *International Journal of the Sociology of Language* 61: 1, 7-32. DOI: <https://doi.org/10.1515/ijsl.1986.61.7>
- Moscoso García, F. (2010). La pentaglosia en Marruecos. Propuesta para la estandarización del árabe marroquí, *Miscelánea de estudios árabes y hebraicos. Sección árabe e islam* 59, 45-61.
- Pellat, Ch. (2023). Laḥn al-ʿĀmma. En P. Bearman, Th. Bianquis, C.E. Bosworth y E. van Donzel (Eds.), *Encyclopaedia of Islam*, Second Edition. Consultado online on 10 August 2023. DOI: http://dx.doi.org/10.1163/1573-3912_islam_SIM_4613
- Psichari, J. (1885). Essais de grammaire historique néo-grecque, *Annuaire de l'Association pour l'Encouragement des Études Grecques en France* 19, 1-228.
- Ruiz Girela, F. (Ed. y trad.) (2008). Ibn Jaldun, A. *Introducción a la historia universal (al-Muqaddima)*. Córdoba: Al-Muzara.
- Sayahi, L. (2014). *Diglossia and Language Contact: Language Variation and Change in North Africa*. Cambridge: Cambridge University Press.
- Versteegh, K. (1997). *The Arabic linguistic tradition. Landmarks in linguistic thought III*. Londres: Routledge.
- Versteegh, K. (2001). *The Arabic language*. Edimburgo: Edinburgh University Press.
- Youssi, A. (1983). La triglossie dans la typologie linguistique, *La Linguistique* 19: 2, 71-83.
- Youssi, A. (1992). *Grammaire et lexique de l'arabe marocain modern*. Casablanca: Wallada.

A/AL- INICIAL COMO IDENTIFICADOR DE ARABISMOS EN LA *RECOPIACIÓN* DE DIEGO DE GUADIX¹

María ÁNGELES GALLEGO²

Consejo Superior de Investigaciones Científicas

Instituto de Lenguas y Culturas del Mediterráneo y Oriente Próximo CCHS - CSIC

Patricia GIMÉNEZ EGUÍBAR³

Western Oregon University - IEMYRhd (USAL Spain)

Resumen

La tendencia a identificar como arabismos términos del castellano que empiezan por *a/al-* se puede encontrar en toda la literatura científica y de ficción de los Siglos de Oro. En este artículo examinamos el origen de este fenómeno y sus características a través del análisis de la *Recopilación de algunos nombres arábigos* de Diego de Guadix centrándonos en los falsos arabismos que genera la utilización de este principio clasificatorio. Dicho análisis está acompañado por un estudio más amplio sobre la visión de Guadix de la lengua árabe y sus principios de análisis etimológico.

Palabras clave: arabismos; contacto lingüístico; Guadix; lengua hebrea; lexicografía


INITIAL A/AL- AS AN IDENTIFIER OF ARABISMS IN THE *RECOPIACIÓN* BY DIEGO DE GUADIX

Abstract

The tendency to identify as Arabisms those Castilian terms starting with *a/al-* can be traced in the Spanish scientific and fictional literature of the Golden Centuries. In this article we examine the origin of this phenomenon through the analysis of the *la Recopilación de algunos nombres arábigos* by Diego de Guadix with a focus on the false Arabisms that

1. Los resultados presentados en este artículo forman parte del trabajo desarrollado por M. A. Gallego dentro del proyecto de I+D+i *Ciencia y religión en el judaísmo medieval* (PID2020-118688GB-I00), financiado por MCIN/AEI/10.13039/501100011033/.

2. mariangeles.gallego@cchs.csic.es.  <https://orcid.org/0000-0001-9987-5680>

3. gimenezp@wou.edu.  <https://orcid.org/0000-0002-1871-2239>

are included in his work due to the employment of this classification principle. This analysis is accompanied by a wider study on Guadix's perception of the Arabic language and his own principles of etymological analysis.

Keywords: Arabisms; Language Contact; Guadix; Hebrew language; Lexicography

RECIBIDO: 03/07/2023

APROBADO: 02/08/2023

1. INTRODUCCIÓN

La tendencia generalizada a identificar como arabismos términos del castellano que comienzan por *a/al-* se puede rastrear en toda la literatura de los siglos XVI y XVII, tanto en la de carácter lingüístico como en la de ficción. En este artículo examinamos el origen y características de este fenómeno a través del estudio de la primera obra de carácter gramatical en la que no sólo se refleja este fenómeno a través de una desmesurada inclusión de supuestos arabismos, sino que se ofrecen también explicaciones de naturaleza lingüística. Como veremos, la entrada de arabismos en romance castellano con el artículo árabe aglutinado es un factor crucial en esta percepción generalizada de la época y que el propio Guadix tendrá en cuenta en sus análisis sobre el origen de los lemas incluidos en su obra. Más allá del análisis estrictamente lingüístico, se aprecian ya en esta obra tendencias normativas con trasfondo sociolingüístico que quedarán asentadas en la literatura posterior como es el destierro de léxico con *a/al-* inicial que queda irremediamente asociado al mundo arabo-islámico.

2. GUADIX Y SU *RECOPILACIÓN* DE ARABISMOS

El franciscano fray Diego de Guadix escribe *La Recopilación de algunos nombres arábigos* en Roma a finales del siglo XVI. El único testimonio de la obra, que se encuentra en la Biblioteca Colombina de Sevilla, ha permanecido en estado manuscrito hasta su primera edición en 2005 que corrió a cargo de Bajo y Maíllo junto con una segunda edición independiente que ve la luz dos años después, de Moreno Moreno (2007)⁴.

Su obra cuenta con el mérito de ser uno de los primeros diccionarios monolingües: en concreto, una recopilación de arabismos con una explicación etimológica de

4. Todas las citas de la *Recopilación* de Guadix utilizadas en este trabajo proceden de la edición de Bajo y Maíllo (2005). Las citas del trabajo de Moreno Moreno (2007) corresponden exclusivamente a su estudio introductorio.

nombres comunes, propios y topónimos de presunto origen árabe en el romance castellano. Se trata de una obra señera y original en su planteamiento tanto en la micro como la macroestructura, ambiciosa en cuanto al léxico presentado, ideada para definir y explicar la etimología de palabras, expresiones, antropónimos, topónimos, orónimos e hidrónimos de origen árabe en varias lenguas, aunque con especial atención a aquellas usadas en España. Una obra fascinante bajo cualquier punto de vista, prácticamente desconocida, que supone además una fuente metalingüística para los regionalismos, especialmente para aquellos del antiguo Reino de Granada, aunque no exclusivamente. Las entradas de su diccionario proveen un sin fin de hipótesis interpretativas que dejan sorprendido al lector actual, por la riqueza de datos, algunos de ellos inverosímiles, que nos permiten entrever la intuición lingüística de Guadix (Corriente, 2005, pp. 94-95, 112). En su obra queda plasmada, además, la imagen social de las diversas lenguas a finales del siglo XVI y principios del XVII.

En cuanto a la vida de Diego de Guadix, se conocen únicamente pinceladas: autor de linaje desconocido que desarrolla su actividad en el último cuarto del siglo XVI y primeros años del siguiente. Hace carrera eclesiástica al servicio de los marqueses de Jódar. En 1587 se le nombra intérprete de lengua arábica en el Tribunal de la Inquisición de la ciudad de Granada y su Reino, en atención a su saber en dicha lengua, que le garantiza su llamamiento y estancia en Roma como experto en lengua árabe en 1590, etapa en la que compondrá *La Recopilación de algunos nombres arábigos*. Vuelve a España como adocrinador de moriscos en Granada y guardián de convento en San Francisco en Córdoba (Bajo y Maíllo, 2005, pp. 15-18; Moreno Moreno, 2007, pp. XIII-XXXII). A pesar de que en varias entradas de su *Recopilación* hace alusión a su condición de cristiano viejo («porque todos los *christianos viejos que, en el reyno de Granada, vivíamos entre moriscos, quasi todos hablábamos la lengua arábica*» s.v. *moçarabe*), Corriente (2005, pp. 101-102) señala, a partir de datos de su *Recopilación* que denotan amplios conocimientos del hebreo y la cultura judía, que probablemente se trate de un converso del judaísmo con educación rabínica pero cuyo contacto con los moriscos le proporciona conocimientos del árabe que sabrá utilizar para la «promoción de su carrera».

La apreciación de la figura y la obra de Guadix ha variado según la óptica adoptada: los estudiosos de la lexicografía hispánica consideran la *Recopilación* el segundo repertorio lexicográfico monolingüe más destacado del Siglo de Oro español (cf. Bajo y Maíllo, 2005; Moreno Moreno, 2007). De hecho, es de consulta obligada para los estudios de dialectología por su heterogéneo inventario no

solo de voces comunes, sino también de otras poco conocidas o con restricción diatópica y diastrática. En cada una de las entradas el franciscano aplica y explica lo que entiende ser su etimología, así como su significado y su historia, más algún comentario particular, propio de su cultura, pero que, en todo caso, proporcionan detalles dialectales nada desdeñables. La información provista refleja el sistema de creencias de la época, ajustada a los conocimientos e intereses del autor como lexicógrafo, religioso e intérprete de la lengua árabe en el Santo Tribunal de la Inquisición de Granada (Moreno Moreno, 2011a, p. 139).

Los que se han acercado a la *Recopilación* como fuente documental del árabe andalusí, tal y como explica Corriente (2005, pp. 110-111), se encuentran ante un volumen que ofrece fiabilidad en aproximadamente la mitad de las entradas: un porcentaje nada desdeñable dada la naturaleza de obra, el tipo de léxico contenido, la amplitud de sus objetivos y del innegable carácter pionero de la misma. La inexactitud de la mitad de las etimologías no debería, en nuestra opinión, conducir a desdeñar su labor como lexicógrafo (cf. García Arenal, 2018, p. 368) máxime cuando se trata de una obra sin precedente alguno en la época, que abre camino a la investigación etimológica y que, a pesar de sus errores, no deja de ser un punto de inflexión y un avance significativo en la materia. Aunque numerosas entradas carecen de sistematicidad y precisión, el diccionario supone una riquísima fuente de información sobre las diversas percepciones de la lengua árabe en el contexto de la España de la época (Gallego y Giménez-Eguíbar, 2020, pp. 121-143). Una de las mayores sorpresas del diccionario es la actitud vanguardista de Guadix quien, en una época reacia a valorar el componente árabe de la lengua española, entiende su obra como una aportación máxima al estudio del léxico español, tal y como afirma en el proemio, a la par que desafía los ánimos doctrinales propios de la orden franciscana.

En cuanto a las bases ideológicas de su diccionario, Moreno Moreno (2011a, p. 387) ha caracterizado la obra como el producto intelectual de un franciscano, «defensor a ultranza de la cristiandad y del catolicismo, monárquico y antisemita, que nos explica el mundo y la lengua filtrada a través de la sombra de un varón blanco, culto y cristiano». Si bien esta afirmación puede ser matizada, es evidente que Guadix comparte en lo esencial y defiende activamente la mentalidad que predomina en su época con respecto al islam y a lo árabe en general. Aun así, como se verá en las próximas secciones, manifiesta una gran sensibilidad hacia lo diferente de la cultura propia y, lejos de simplificar o aceptar explicaciones populares, se inclina a la matización y al entendimiento de la complejidad de los fenómenos lingüísticos.

3. GUADIX COMO ARABISTA E HISPANISTA

En lo que respecta al conocimiento del árabe por parte de Guadix, se podría caracterizar como de un bilingüe imperfecto «con conocimiento activo limitado», tal y como concluye Corriente (2005, p. 94), cuyas nociones del árabe clásico parecen ser más escasas y asistemáticas que las del árabe dialectal granadino. De nuestro propio estudio se desprende que, a pesar de esa asistematicidad en el conocimiento del árabe clásico, tiene una visión global de la lengua y está bien informado sobre algunos de sus principios fundamentales. Su posible origen judío y claros conocimientos de la lengua hebrea probablemente fueron de gran utilidad en su dilucidación de la base gramatical del árabe clásico. De hecho, Guadix considera al árabe como una versión «corrupta» de la lengua hebrea,⁵ lo que le lleva ocasionalmente a establecer etimologías que parten directamente del hebreo, como es el caso de *morriar* ('esquilar') que él deriva de la voz hebrea *mora* (= navaja) o de la palabra *jardín* que deriva del nombre del río que, según afirma, tanto en hebreo como árabe se denomina *charden* (= Jordán) aunque lo cierto es que esa denominación sólo existe en hebreo.

Llama la atención la fina percepción sociolingüística que Guadix muestra en su caracterización de la diglosia árabe distinguiendo entre el árabe clásico, «antigua y fina algarabía» en sus palabras (s.v. *alere* o s.v. *alcabala* o *alcauala*) y el árabe vulgar o «mala o corrupta algarabía» según quedan reflejadas en las entradas *alacrán* o *cáliz*:

1. Dizen en Latin para significar lo que, en castellano, sustentar. Fue deriuado y tomado d'este verbo, *ªala*, que, en fina y antigua algarabía significa lo mesmo, i., sustentar (s.v. *alere*).
2. Y aun a esta pobre algarabía la rinden y lleuan al aula de gramática y le hazen pasar de la carrera de los nombres latinos de la primera declinación, pues que la declinan *gabela gabele*. Aduiértase qu'este nombre *cabala* es deduzido d'este verbo *quebel*, que, en fina y antigua algarabía, significa rescebir (s.v. *alcabala* o *alcauala*).
3. [...] esto digo porque no se engañe nadie, por el nombre vulgar *ªacra*, de que usaban los árabes en la mala o corrupta algarabía que en España, Siçilia, partes de Italia y França hablaron (s.v. *alacrán*)
4. [...] porque en la mala algarabía que los álabes [sic] hablaron en España es *quiz* (s.v. *cáliz*).

5. «[...]la lengua arábiga gana en antigüedad a las demás lenguas del mundo, porque es la lengua hebrea, aunque corrupta» (s.v. *Primera aduertencia*).

Este dominio sociolingüístico se muestra también en la distinción, a diferencia de gran parte de sus coetáneos, entre lengua árabe e islam, tema al que dedica detalladas explicaciones y cálculos para demostrar la anterioridad de la lengua árabe frente a la «doctrina mahometana»:

Aduerta el docto y discreto lector que no andan a una el ser árabes o arábigos y el ser moros, como queda dicho en la primera advertencia del principio de esta parte, porque difiere mucho la lengua de la religión. [...] Esto es dicho para que nadie sea tan ignorante que le parezca ser todo una pieza o andar todo a una: el hablar, en alguna tierra, lengua árabe y el ser aquella gente descendiente de mahometanos (s.v. *aguadinace*).

El lector curioso que se pusiere bien en estas computaciones de tiempos y quisiere restar y computar unos tiempos de otros, tomará entera noticia y sabrá cuántos siglos es más antigua la lengua hebrea que la árabe, y la árabe que la griega y la griega que la latina (s.v. *algarabía*).

Pues, si Mahoma y su seta fueron después de la encarnación de Xhristo 592 años y la lengua latina comenzó 600 o 700 años antes de la encarnación, luego la lengua latina comenzó 1192 ó 1292 años antes de Mahoma y de su maldita seta, y hallamos la lengua latina aver tomado muchos verbos y vocablos de la lengua árabe (como se verá en sus lugares), luego 1192 o 1292 años antes que viese en el mundo Mahoma ni seta de moros, ya se usaban en el mundo y estaua el mundo lleno de verbos y vocablos arábigos (*Primera advertencia*).

Con este tipo de aclaraciones, logra independizar la lengua del pueblo que la trae a la Península, es decir, «la maldita seta de Mahoma». Además, hace alusión en diversos lugares al uso de la lengua árabe por parte de otras comunidades religiosas como son los cristianos:

[...] da mucha luz y favorece mucho a esta verdad ver que los naturales de la isla de Malta, y otros millones de ch[r]istianos que habitan en Assia y Turquía son árabes de nación y hablan la lengua árabe, y no por eso son descendientes de moros ni tienen que ver con moros, porque comenzaron [sic] a ser christianos desde la predicación de los apóstoles, que es como seiscientos años antes que el maldito Mahoma naciese en el mundo ni su maldita seta se publicase. Esto es dicho para que nadie sea tan ignorante que le parezca ser todo una pieza o andar todo a una: el hablar, en alguna tierra, lengua árabe y el ser aquella gente descendiente de mahometanos (s.v. *aguadinace*).

A los cristianos de al-Andalus (mozárabes) no duda en atribuirles una profunda arabización, fruto de su integración en la sociedad andalusí:⁶

Este nombre *moçárabe* [...] significa este adgetiuo árábigo, .i., hombre que sabe o habla la lengua árábiga, [...] Por ventura los llamaron los moros por este nombre, porque los tales christianos sabían o hablaban la lengua árábiga, del gran trato o por el grande, continuo y familiar trato que tenían con los moros [...] En tan larga comunicación y trato con los moros, ya la lengua árábiga les devía de ser lengua materna, pues los llamaron *arábigos* o *moçárabes* (s.v. *moçárabe*).

También en el caso de los judíos hace referencia a la inmersión lingüística en su entorno más allá de la identificación generalizada de una lengua con una religión específica. Señala, por ejemplo, la distinción que se establece en Marruecos entre los judíos autóctonos de lengua árabe y los judíos llegados tras la Expulsión de España en 1492 que eran ya de lengua romance (aunque originalmente hubieran procedido de territorios andalusíes araboparlantes):

[...] los judíos antiguos [...] todos hablaun y tenían por lengua materna, la lengua árábiga y, quando los reyes católicos, de felice, recordación, echaron de España a los judíos, muchos de los dichos judíos se pasaron a las dichas ciudades y reynos de Fez y Marruecos, los quales, como recién ydos d'España, no sabían la lengua árábiga y así, para diferenciarse los judíos de los otros, llamaron y llaman oy en día a aquellos judíos que antiguamente estauan en aquellas ciudades y reynos, *moriscos*, .i., *judíos moriscos* y a los judíos que nuevame[n]te avían ydo d'España, llamaron y llaman oy en día, *españoles*, .i., *judíos españoles* (s.v. *moçárabe*).

Ofrece información sobre la evolución de los arabismos en las distintas variedades romances y la mayor o menor fidelidad al término original árabe, como comprobamos en la entrada *albalá* (del árabe andalusí *al-bará*)⁷

Es la mesma algarabía, .i., *albara*, y significa lo mesmo, .i., la cédula; y, corrompido, dizen *albalá*; y, en la parte d'España a que llaman reyno de Aragón usan d'este mesmo nombre árábigo, aunque en menor corrupción, porque dizen *albarán* (s.v. *albalá*).

6. Fueron precisamente estos cristianos arabizados que emigran a territorios cristianos los principales agentes de la introducción de arabismos en el romance castellano (Lapesa, 1990, pp. 129-130; Maíllo Salgado, 1991, pp. 495 y sig.; Corriente, 1996, p. 5).

7. Corriente, Pereira y Vicente (2019, p. 59).

Del mismo modo, Guadix recoge y define por primera vez en la lexicografía hispánica un nutrido número de préstamos de las lenguas amerindias o indígenas, que el lexicógrafo considera arabismos probablemente en tanto que términos alejados del mundo cristiano conocido y sus lenguas de tradición. De este modo da entrada a lemas como *arcabuco*, *canti*, *guaca*, *hamaca*, *nigua* así como los topónimos *cequinala*, *Cuba*, *La Havana*, *Lima*, *Maru*, *Mexico*, *Paraginahay*, *Parramos*, *Parrihalchay*, *Tayrona* o *La Sierra de Tayrona* (cf. Moreno Moreno 2011, pp. 140-145).

En lo que respecta a la lingüística aplicada, Guadix es también sensible a los retos a los que se enfrentan los aprendientes del árabe como segunda lengua, de ahí las minuciosas descripciones de los sonidos árabes con mayor dificultad para los «cristianos españoles» como es el caso del *ʿayn* /ʕ/.

Y los christianos españoles, *no sabiendo ni pudiendo pronunçiar esta gutural arábiga*, hazen esta corrupçión: açofeyfo (s.v. *açofeyfo*).

Asimismo, es consciente de que la evolución lingüística sigue unos patrones generales, y lejos de llegar a auténticos modelos, intuitivamente reproduce la reconstrucción del étimo (Moreno Moreno, 2011, p. 398). Para establecer sus etimologías, Guadix parte del nombre actual para intuir el árabe que pudo darle origen. No concibe los cambios como evoluciones a las leyes fonéticas, sino en términos de corrupción, en consonancia con la teoría de las primeras obras castellanas. Esto le lleva a aceptar cambios radicales, infrecuentes y asistemáticos que él achaca al «capricho del vulgo». Del total de 4336 entradas que comprende la macroestructura, 996 (el 23%) corresponden a la letra A. A continuación, analizamos el artículo aglutinado como marcador sociolingüístico para el reconocimiento de voces árabes. El hecho de que Guadix incluya en su obra supuestos arabismos, responde, entre otras razones, a la creencia generalizada en la época de considerar que las palabras que empiezan con *a-* o *al-* provienen de la lengua árabe. A este aspecto, entre otros, dedicaremos las páginas que siguen.

5. EL ARTÍCULO DETERMINADO EN LOS PRÉSTAMOS LÉXICOS DE LA LENGUA ÁRABE AL CASTELLANO

Como es bien sabido, el prefijo árabe *a(l)-* es el mecanismo habitual para marcar la determinación de los sustantivos en la lengua árabe, así por ejemplo *kitāb* equivale al español ‘un libro’ mientras que *al-kitāb* equivale a ‘el libro’. Un hecho destacable del artículo determinado árabe es que, si bien es invariable en su grafía e

independiente del género y número de la palabra determinada, sufre una transformación fonética delante de determinadas consonantes. Se trata de 14 consonantes, denominadas «solares» en árabe, y con las que se asimila, a saber: interdental, dento-alveolares, alveolares y prepalatales. De esta manera, por ejemplo, la expresión ‘el sol’ en árabe escrito quedaría reflejada como *al-šams* mientras que se pronunciaría *aššams*. La asimilación fonética del artículo a las consonantes solares es precisamente una característica que permite diferenciar los préstamos de la lengua árabe que han sido introducidos a través de la lengua escrita, por traductores con unos conocimientos limitados de esta lengua, frente a los préstamos que han entrado a través de la lengua oral y que reflejan la pronunciación real con asimilación. Por ello, cuando nos encontramos en romance un arabismo en el que se ha preservado el fonema /l/ en situaciones de asimilación fonética, podemos asumir que se trata de un término importado desde la lengua escrita. En gran parte de los casos estos arabismos han entrado en el romance a través de las traducciones de textos de tipo científico. Así, por ejemplo, sucede con *aldemanel* (‘forúnculo’) procedente del árabe *addamāmīl*, o el nombre de las estrellas *Aldebarán* y *Altair*, cuyo origen son las voces árabes *addabarān* y *attā’ir* respectivamente.⁸

La mayor parte de los arabismos en castellano procede, como cabría esperar, de la lengua hablada, es decir, de la variedad árabe andalusí. Por ello, en el establecimiento de las etimologías es fundamental tomar como punto de partida este dialecto árabe y no el árabe clásico, un procedimiento y metodología que ya aplicó Joan Coromines en su *Diccionario crítico etimológico de la lengua castellana* (1954) y que Federico Corriente ha desarrollado y completado a través de sus múltiples y valiosísimas contribuciones a este tema.⁹ Este principio se aplica también, lógicamente, al artículo determinado cuya pronunciación delante de consonante solar en árabe andalusí no siempre coincidió con la del árabe clásico y, por ejemplo, las interdental pasaron a ser pronunciadas como dentales. En cualquier caso, en relación con el empleo del artículo determinado, una de las cuestiones que más ha llamado la atención es el hecho de que haya entrado aglutinado al lexema en los préstamos que se han introducido en el romance castellano y, en menor medida, el portugués. Si bien el número de arabismos con el artículo aglutinado no es tan elevado como se ha creído durante mucho tiempo, lo cierto es que supera el 50% de

8. Véase Corriente, Pereira y Vicente (2019, pp. 95, 179 y 94).

9. Véase, por ejemplo, Corriente (1999, pp. 17-65), en donde el autor describe minuciosamente los procedimientos de entrada de arabismos en las lenguas romances y los diversos mecanismos de adaptación y asimilación.

los arabismos totales en el castellano (855 con artículo frente a los 840 sin artículo), según las estadísticas aportadas por Winet (2006).

Esa elevada proporción de arabismos con artículo frente a arabismos sin artículo en castellano es llamativa en sí misma y en comparación con otras lenguas romances como el catalán, el francés o el italiano, por razones que no terminan de estar claras. Ejemplos de préstamos con artículo aglutinado en castellano y portugués pero que no lo tienen en otras lenguas romances son el caso del término árabe *al-sukar* (pronunciado *assukar*), que da origen a ‘azúcar’ en castellano y ‘açúcar’ en portugués pero ‘sucre’ en francés y catalán y ‘zuccherò’ en italiano o el término árabe *al-quṭn* que da origen al castellano ‘algodón’ y el portugués ‘algodão’ frente al francés ‘coton’, al catalán ‘cotó’ y al italiano ‘cotone’. El propio Guadix señala ocasionalmente estas diferencias y, por ejemplo, en el caso de «alcanfor» (del árabe *al-kāfir*) comenta: «En Italia usan d’esta mesma algarabía aunque sin el artículo al, porque llaman a esta droga *canphora*» (s.v. *alcanfor*).

Desde que a comienzos del siglo xx diversos estudiosos llamaran la atención sobre el número tan elevado de arabismos que han entrado en castellano y en portugués con el artículo árabe aglutinado, son numerosas las teorías que se han propuesto para explicar este fenómeno, todas ellas resumidas y comentadas por Winet (2006) en un amplio análisis y estado de la cuestión. Una de las hipótesis que parece contar con mayor apoyo es la avanzada por Elcock (1960) y desarrollada por F. Corriente (1999, pp. 57-62), conocida como «tesis bereber». De acuerdo con dicha teoría el uso de la lengua árabe por parte de los bereberes, grupo mayoritario entre los musulmanes de al-Andalus en un estadio inicial, estuvo influido por su lengua de origen que cuenta con dos características diferenciadoras frente al árabe, a saber, la no existencia de un artículo determinado y la marcación de sustantivos con un prefijo específico. Estas dos características tendrían un impacto significativo y se reflejarían en el árabe empleado por la población bereber dando lugar a un uso abusivo y distorsionado del artículo árabe como marca nominal que quedaría reflejado posteriormente en los préstamos de esta lengua a las variedades romances con las que se produjo un mayor nivel de contacto en la Península Ibérica. Partiendo en lo esencial de esta teoría, Noll (2019) ha añadido nuevos elementos de análisis como son el bilingüismo y las situaciones de contacto lingüístico para ofrecer un marco más amplio que nos permita entender los mecanismos que influyen en que en unos casos los arabismos entren con el artículo árabe aglutinado y en otros no ocurra lo mismo.

Independientemente de las causas de este fenómeno, lo cierto es que la aglutinación del artículo al lexema estuvo en el origen de una percepción generalizada tanto a nivel popular como entre los gramáticos de la temprana Edad Moderna de que aquellas palabras que en castellano comenzaban por *al-* o *a-* eran de origen árabe, tal y como ilustraremos en la siguiente sección.

6. EL ARTÍCULO AGLUTINADO *AL-* COMO IDENTIFICADOR DE ARABISMOS EN EL ROMANCE CASTELLANO EN LOS SIGLOS XVI Y XVII

La metodología utilizada por Guadix, así como por la mayoría de los lexicógrafos de los siglos XVI y XVII, para identificar un término como arabismo se basó fundamentalmente en el parecido fonético entre un término romance y un término árabe. Más allá de este principio básico, el criterio generalizado y rápido para la adscripción de una etimología árabe a un término castellano estuvo ligado, como ya se ha mencionado, al hecho de que una palabra comenzara por *al-* o *a-*. De esta identificación se derivó toda una serie de fenómenos de tipo sociolingüístico tales como añadir artificialmente el artículo árabe a nombres latinos o romances para arabizarlos cuando la lengua árabe era considerada una lengua de prestigio como sucede con el término *almenara* (⟨al+minae) o el topónimo *Alpuente* (⟨Al+puente) (Müller, 2004). De la misma manera, pero en mucha mayor medida, cuando la lengua árabe quedó desprestigiada en tanto que lengua del enemigo musulmán, observamos el rechazo sistemático a los términos que empezaban por *a/al-* por su (supuesta) procedencia árabe (Giménez Eguíbar, 2016, p. 367).

Esta asociación de *a(l)-* al comienzo de una palabra con la lengua árabe y, por extensión, con el islam, se convirtió en una percepción generalizada, tal y como se refleja en la literatura de la época (Giménez Eguíbar, 2016). Para ejemplificar esta asociación, de entre los muy numerosos testimonios que hallamos en la literatura de la Edad Moderna, destacan las afirmaciones de un autor anónimo de mediados del siglo XV en sus *Anotaciones al Inferno* de Dante, en donde aborda una serie de temas lingüísticos como la variación léxica de las distintas zonas de la Península Ibérica con especial atención a Andalucía y Toledo, en cuyas variedades del castellano abundan los arabismos. Es preciso reparar en que dentro de los cinco ejemplos que ofrece, cuatro de ellos llevan el artículo aglutinado y los identifica como moriscos,

[...] porque yn Ytalia, commo quier que todos se entiendan, pero ay infinitos vocablos que unos non dizen commo otros, nin los usan, commo en Castilla los asturianos y gallegos, y de cada parte ay sus diferencias, commo del Andalucía a

Castilla Vieia o de Toledo a Çamora donde non saben qué quiere dezir *alamud nin azendoque nin çati nin abóndiga nin alçuscucu nin otros infinitos, porque estos son no[n]bres moriscos* que nos quedaron de los moros, quando por nuestros pecados poseyeron nuestras tierras. (Webber, 1962, p. 34)

En las páginas siguientes de este breve tratado lingüístico se advierte sobre la generalización de la regla de la que venimos tratando,

[...] Quantos *vocablos que tenemos que escomiençan en al son moriscos y non latinos, commo alhonbra, algodón, alcalde, aldayra, almud, alcaýde, alcaçar, almiherez, alcántara, almadraque, almohada, alquilé, almizcle, alcatara, alfiler, alhaja, alcusa, alhoça, alholí, alcubilete, alcuscuçu, almojáuana, albóndiga, almorí, almodrote* y otros infinitos; y otros que comiençan en gua-, commo nonbres de ríos: Guadalquiuir, Guadaxox y [commo] Guadalajara, Gu[a]dameçil, Guadalupe y (de) otros *infinitos vocablos moriscos que tenemos* (Webber, 1962, p. 35).

La lexicografía de la época presenta igualmente numerosos ejemplos de empleo de este principio básico, como los que se encuentran en el *Vocabulario del romance en latín* (c. 1495) de Nebrija:

Alfaqui en *aravigo*. casi sacerdote.
Alaton *morisco*. orichalcum .i.
Añafil de *moros*. tuba punica. (Nebrija c. 1495)

Se desprende también la aplicación de este principio en el sinónimo que Rodrigo Fernández de Santaella ofrece de la voz *cantinum* en su obra *Vocabularium seu Lexicom Ecclesiasticum*, 1499, y que además adscribe a Andalucía:

[...] o lo que los andaluzes por vocablo *aravigo* llaman *almofia* (Santaella 1499, s.v. *cantinum*).

En el *Diálogo de la Lengua* (1535) de Valdés, género radicalmente distinto al de los diccionarios, su autor se permite una mayor flexibilidad para la explicación y reflexión sobre los rasgos léxicos del castellano,

Casi siempre son arávigos los vocablos que empiezan en al, como almohada, alhombra, almohaça, alhareme y los que comienzan en az, como azaguán, azar, azagaya (Lope Blanch, 1984, p. 19).

Juan Francisco Ayala Manrique en una muy plástica reflexión sobre el arabismo *alfaqín* nos regala otro ejemplo del principio básico en su *Tesoro de la lengua Castellana*,

Por lo mismo que gobernador, Zurita, Lib. 3 de los Anales, Cap. 30. *Ya se ve que es voz arábica* (1693: s.v. *alfaquin*).

En las comedias de Lope de Vega particularmente en las que se caricaturiza el dialecto de los moriscos, el madrileño omite sistemáticamente el artículo árabe *al* o *a*: *ceite, calde, caide, guacil* (Case, 1982, p. 597). En esta misma línea, también en la literatura, Cervantes hace referencia a este principio básico de que todos los nombres que comienzan por *al-* proceden de la lengua árabe:

Y este nombre *albogues es morisco, como lo son todos aquellos que en nuestra lengua castellana comienzan en al*, conviene a saber: almohaza, almorzar, alhombra, alguacil, alhucema, almacén, alcancía, y otros semejantes, que deben ser pocos más (Cervantes, 1615, p. 67).

7. A/AL- COMO IDENTIFICADOR DE ARABISMOS EN LA *RECOPILACIÓN* DE GUADIX

Dentro de esta misma tendencia ya comentada de asociación de *a/al-* inicial en castellano con una etimología árabe, Diego de Guadix identificó un número importante de arabismos a través de dicho principio que, en muchas ocasiones, resultó ser erróneo como se verá más adelante. Pero más allá de ofrecer etimologías, encontramos en la *Recopilación* interesantes explicaciones de tipo gramatical y fonético con las que Guadix describe el origen, uso o funcionamiento de esta *a/al-* inicial.

Así, por ejemplo, en la «octava advertencia» de la introducción describe el comportamiento fonético y la asimilación regresiva del artículo ante «letras solares»:

Lo octauo: presupongo que es regla de gramática arábica y muy practicada y usada entre árabes que una letra a que en su alfabeto llaman *lam* como nuestra .l., latina, aunque escriua en la dición, *no se a de leer, ni a de sonar la pronun-ciación, quando se le siguiese alguna letra de las que los árabes llaman gemcias, .i., solares*, que son las siguientes ط ظ ن ذ ز د ث ث ر ض ص ث ث el sonido de las quales corresponde a el sonido de nuestra .t., .r., .z., .d., .c., .x., .n., y aun los arabes d'España, digo, los antiguos naturales de los Reynos de Granada y Valençia, en sus pronunçaciones, / an mostrado ser también letra xemcia otra letra de su alfabeto qu'es la sig[u]jiente ح cuyo sonido, quatro más a menos, assimila a esta sílaba nuestra .ch., y, porque [a] algún curioso lector que fuere arábigo, le parecerá cosa dura asignar o dezir que ay otra letra xemçia o solar más de aquellas arriba dichas, que son las que se hallan asignadas por xemçias en todas las gramáticas xarquias o levantiscas, dígame o pronúncieme estos

nombres: *alchar*, *alchoray*, *alchuza*, y verá que, en buena pronunciación árábica, a de pronunciar *axar*, *axoray*, *acheuza*.

Es muy de importancia la inteligencia d'esta reglilla de gramática árábica para saber pronunciar un nombre árábigo en verdadera y legítima pronunciación árábica.

Nuestro autor hace referencia también a las fórmulas de determinación e indeterminación en árabe y advierte sobre el error de añadir el sufijo de nominativo indeterminado *-un* a un nombre que está determinado por el artículo *-al* tal y como aprecia en el caso de *almarbón* ('escoplo'):

[...] así que, todo junto, *almurab[a]a* significa el cuadrado, y el *on* es un *un*, que es terminación de un caso a que, en la gramática árábica, llaman *mubteddé* [sic], que corresponde al nominativo de nuestra gramática latina. [...] Y bien parece que el vulgo hizo o compuso estas gramaticaciones, porque está aquí cometido vn solecismo contra vna regla de gramática árábica, que enseña que en el nombre o dición a quien precediere el artículo *al*, no puede aver tanuin, i., estar terminaciones *un*, *in*, *an* (s.v. *almarbón*).

En ocasiones, Guadix identifica erróneamente la *a-* inicial de palabras árabes con el artículo determinado, lo que le lleva a optar por su omisión si acompaña a un nombre propio como es el caso de 'Ayyūb (Job). En su reconstrucción del topónimo Calatayud, Guadix lo incluye con la forma pseudocorrecta Yub:

Consta de *calat*, que, en árábigo significa castillo, i., que haze frontera, y de *yub*, que significa Job [...] (s.v. *Calatayud*).

Las explicaciones de Guadix en relación con esa *a-* inicial característica de los arabismos no se limitan, sin embargo, al artículo determinado. Elabora también una teoría sobre el origen de la *a-* inicial en las formas verbales. Según nuestro autor, no sólo aparece en los arabismos como tales, sino que su influencia ha llegado a las formas verbales del castellano. Aunque sin hacer una referencia explícita, Guadix alude a lo que conocemos como Forma IV de los verbos en árabe, cuyo sentido es habitualmente causativo y que se caracteriza morfológicamente por añadir el prefijo *'a-* a la raíz verbal. El autor muestra aquí sus nociones del árabe clásico, es decir, la existencia de la susodicha Forma IV y sus características morfológicas y semánticas. No es consciente, sin embargo, de que en dialecto andalusí, como en la mayor parte del neoárabe, prácticamente desaparece esta forma verbal (Corriente, 2012, p. 84), lo que invalida su teoría sobre la influencia del árabe en el romance

castellano en este aspecto concreto. Para Guadix, esa *-a* inicial de carácter causativo daría lugar a formas híbridas árabo-castellanas que él relaciona con usos del «vulgo»:

Sea la sexta y última consideración saber que en la lengua árábica ay muchas voces que comiençan en esta letra *.a.*: la tienen antepuesta a sus letras substanciales, y aun hasta en esto quiso la comunicación con los árabes ponernos en su imitación, y el bulgo lo rescibió y el uso lo confirmó; y así, a muchos de nuestros verbos castellanos o españoles, les anteponevamos vna *.a.*, y dezimos: *asserrar*, *amolar*, *asentar*, etc., y yo e oýdo a hombres curiosos començar a dar de mano a esta *.a.*, cognoçiéndola o tiniéndola por forastera y aduenediça en verbos castellanos o españoles, y assí dizen y an de decir, *serrar*, *molar*, *sentar*, etc. (s.v. *a*).

Como se puede apreciar en esta cita, Guadix constata también el fenómeno de rechazo y estigmatización de este prefijo en tanto que elemento «forastero» y «advenedizo». En realidad, esta tendencia normativa de destierro de formas verbales con prefijo *a-* en los registros más formales estaba vigente ya desde finales del siglo xv, sustentada en gran medida por el filtro estandarizador que ejercieron los editores de la época a través de la imprenta (Tejedo Herrero, 2009, pp. 321-325). Dicha tendencia se aplicó también a otras formas de afijación distintas a la *a-* inicial, sin relación con la lengua árabe, y que han quedado asociadas hasta hoy en día con variedades de registro bajo (Del Barrio de la Rosa, 2021, pp. 95-125). En el caso específico de *a-* no podemos descartar, sin embargo, que dentro de esa tendencia normativa haya existido un condicionante sociolingüístico de identificación con lo arabo-islámico, tal y como demostraría la cita de Guadix.

Independientemente del origen atribuido o real de *a/al-* inicial de palabras en castellano, su utilización como principio de identificación de arabismos llevó a Guadix a incluir un elevado número de términos que no necesariamente procedían de la lengua árabe. De los 742 arabismos incluidos en la sección dedicada a la letra «a» hallamos que 131, es decir, un 17%, son en realidad falsos arabismos.¹⁰

El grueso de los casos de falsos arabismos son verbos (de la primera conjugación casi exclusivamente) y suman un total de 63 entradas:

abahar, *abaxar*, *abarcas*, *abarraganar*, *abarrancar*, *abarrar*, *abasallar*, *abatir*, *abezar*, *abocar*, *abotonar*, *abrigar*, *acabar*, *acaudillar*, *achicar*, *acodar*, *acorar*, *acortar*, *acosar*, *acudir*, *adilgar*, *adobar*, *afear*, *afeytar*, *aflechar*, *aforrar*, *agasajar*, *agramar*, *aguardar*, *aguijar*, *aguzar*, *ahajar*,

10. Solo se han considerado para este recuento los sustantivos y formas verbales.

ahincar, ahitar, ahogar, ahorcar, alabar, albergar, alcançar, alçar, alexar, alindar, allanar, allegar, almiar, amaynar, amancebar, amblar, andar, anegar, aporrear, arrojar, atafagar, atapar, atar, ataviar, atorar, atrever, aventajar.

El otro gran grupo de falsos arabismos son sustantivos y, en menor medida, adjetivos y pronombres, sumando un total de 52 entradas:

abad, abadessa, abexaruco, abelú, abieto [abeto], abuna, academia, adelantado, afrecho, agalla, agengibre, agua sentada, aguadinace, agüelo [abuelo], alabarda, álamo, alarido, alatón, alcabuco o arcabuco, alcabuz, alere, alesna, alguien, alguno, aliança, almario, almeja, almendra, almofrodita, almuerzo, alnado, alquitrabe, ama, amo, anascote, anca, anciano, añino, arenga, arnés, arras, arbol, arroyo, ascua, aspa, atajador, atino, aula, azada, acero.

Desde un punto de vista etimológico, entre las voces con atribuida etimología árabe en la *Recopilación* se encuentran setenta y tres voces patrimoniales:

abad, abadessa, abaxar, abarcar, abarraganar, abarrar, abatir, abexaruco, abezar, abieto, abrigar, acabar, academia, acaudillar, achicar, acodar, acorar, acortar, acosar, acudir, adelantado, afear, afeitar, afrecho, agalla, agengibre, agua sentada, aguadinace, agüelo, aguzar, ahajar, ahincar, ahitar, ahogar, ahorcar, aina, al margen, alabar, alatón, alcanzar, alçar, alejar, alguien, alguno, alianza, alindar, allanar, allegar, almafrodita, almario, almendra, almiar, almuerzo, alnado, amancebar, amblar, amen, andar, anegar, aporrear, arras, arbol, arrojar, arroyo, atajador, atar, aula, aunque, aviesso, ayer, azada, aziago.

Les siguen doce galicismos: *abismo, abotonar, adorbar [adobar], aflechar, aforrar, alemán, anascote, anciano, arcabuz, arnés, atafagar, aventajar*; nueve voces pertenecen al gótico: *agasajar, álamo, albergar, amainar, aspa, atapar, ataviar, atorar, atrever*; seis germanismos: *aguardar, alabarda, alesna, ama, amo, añino*; dos occitanismos: *anca, arenga*, un italianismo: *alquitrabe*; un lusismo: *almeja*; una voz prerromana: *abarrancar*; un catalanismo: *adrede*; un celtismo: *avasallar*; y un indigenismo: *alcabuco* o *arcabuco*. Otras palabras con la atribuida etimología árabe tienen un origen incierto en su *Recopilación*: *ababar [avabar], abocar, adilgar [endilgar], agramar, aguijar, alarido, ascua, asido, atino.*

Se confirma, pues, la ya mencionada tendencia que prevalecerá en toda la época moderna de identificación generalizada de términos que comiencen por *a-*/*al-* con la lengua árabe llegando, como en el caso de Guadix, a forjar imaginativas etimologías

que justifiquen esta adscripción. Llama la atención, en este sentido, la clasificación como arabismo del indigenismo ‘*alcabuco*’, que refleja la asimilación con lo árabe de lo desconocido o de aquello que sale de la esfera del mundo latino/romance:¹¹

Llaman en Las Indias Occidentales a el bosque de árboles o breña de matorrales. Consta de *al*, que, en arábigo, significa el, y de *caa* que significa asiento o sitio, y de *abuq*, que significa tu padre; así que, todo junto, *alca^aabuq* significa el asiento de tu padre o el sitio de tu padre; y, corrompido, dizen *alcabuco*. y otros que lo corrompen más *arcabuco* (s.v. *alcabuco*).

Otro factor que interviene en la asignación de una etimología árabe es la onomasiología. Es bien sabido que en la historia del español han predominado los arabismos que se refieren a actividades y a la designación de seres concretos, en campos como la arquitectura y el urbanismo, la agricultura, el vestido y ornamentos, alimentos y coquinaria, zoónimos, comercio y propiedades, entre otros. En menor medida se han transmitido, no obstante, términos que designan categorías abstractas propias del discurso técnico y científico, reflejo del papel relevante que desempeñó el islam en estas áreas tanto por su labor de transmisión como por su desarrollo propio y original. Así, junto a términos agrícolas, nombres de las profesiones y mercaderías, se unen al catálogo de arabismos voces de la ciencia, la filosofía, las matemáticas, la geografía o la medicina.

En la *Recopilación* de Guadix observamos que a la tendencia a clasificar como arabismos términos que comienzan por *a/al-*, se une la de clasificar como tales muchos que provienen de esos campos específicos. Lematiza, por ejemplo, préstamos del gótico, en un campo onomasiológico como el de la agricultura, caracterizado por la abundancia de arabismos (cf. Giménez-Eguívar, 2020, pp. 33-64). Este es el caso de *álamo*:

Llaman en España a un árbol infructuoso que, de ordinario, naze y se cría cerca de los ríos o fuentes. Consta de *al*, que, en arábigo, significa el, y de *olmo*, que, en castellano, significa el dicho árbol (s.v. *álamo*).

Incluye asimismo voces dentro de este campo de la agricultura que, según el autor, son híbridos entre el árabe y el latín, utilizando una vez más el principio de

11. La aparición de indigenismos en la literatura gramatical de la época ha sido objeto de diversos estudios, como el famosísimo caso de *canoá* del *Tesoro* de Covarrubias. Menos atención han recibido los que aparecen en la *Recopilación* de Guadix (cf. Moreno Moreno, 2011, p. 140) y que no hemos incluido en este trabajo al limitarnos a los términos que comienzan por *a*.

identificación de términos que comienzan por *al-* con la lengua árabe como es el caso de *almendra*:

Llaman en España a la fruta del árbol almendro, que es de todos bien conocida. Consta de *al*, que, en arábigo, significa la, y de *migdala*, que, en latín, significa este dicho fruto de almendro [...] (s.v. *almendra*).

Con la etimología árabe aparecen también voces patrimoniales como *abeto*, *afrecho*, *jengibre*, *alindar* y *almiar*.

Dentro de los objetos concretos relativos a ciertas profesiones –campo onomasiológico caracterizado también por la presencia de arabismos,– se halla el germanismo *alesna*, un instrumento de los zapateros:

Llaman en España a una herramientilla con que cosen los çapateros. Consta de al que –en arábigo– significa ‘la’ y de exna que significa ‘la sangradera de surco’. De suerte que todo junto alexna significa la sangradera de surco’. Y corrompido dizen alesna (s.v. *alesna*).

Otro instrumento profesional mencionado en la *Recopilación* es *alegra* (*DLE* 2014: s.v. *legra*), instrumento de cirujanos:

Llaman en España a un instrumentillo o herramientilla con que suelen hazer cucharas, y aun curar, los cirujanos, algunas heridas de cabeça. Consta de *al*, que, en arábigo, significa el, y de ‘*aacra*, que significa escorpión, .i., alacrán; de suerte que, todo junto, al‘aacra significa el escorpión, .i., el animalejo ponçoñoso a que, en España, llaman *alacrán*; y, corrompido, dizen *alegra* (s.v. *alegra*).

Con la etimología árabe aparecen también dentro de este campo onomasiológico *azada* y *alatón*.

Aunque en menor medida que los latinismos, Guadix lematiza muchos galicismos, algunos de ellos pertenecientes al campo onomasiológico de las mercaderías y telas como *anascote*, adjetivos como *alemán* o el topónimo *Alemania* cuyo parecido formal al patrón del que venimos hablando es indudable. Lo considera proveniente de la (inexistente) expresión árabe *al-īmān-nā* ‘la fe nuestra’ alterando el sufijo *-nā* y ofreciendo una forma agramatical con doble determinación a través de artículo y pronombre posesivo (Corriente 2005, p. 95):

Es el nombre de una región austral de todos muchas vezes oýda. Consta de *al*,

que, en arábigo, significa la, y de *yman* que significa fidelidad o creencia, y el nia es afixo de primera persona del plural y significa nuestra (s.v. *Alemania*).

Teniendo en cuenta la masiva entrada de arabismos en el campo onomasiológico militar y dada la aparente presencia del artículo árabe, se encuentra el galicismo *arcabuz*,

Llaman en Italia y en España al arma e instrumento bélico de fuego que todos sabemos.

Consta de *al*, que, en arábigo, significa el y de *caa'*, que significa asiento, y de *buz*, que significa, este imperativo, besad [...] (s.v. *alcabuz*).

En esta misma situación se encuentra el galicismo *arnés* y *alabarda*, este último del francés o tal vez del italiano.

Finalmente, es necesario señalar que, aunque Guadix se guíe en buena medida por el inicio *a/al-* para su clasificación de términos romances como arabismos, es consciente y refleja la existencia de alternancia entre la presencia o ausencia de afijación en la base. Muchas de sus entradas se presentan con el doblete como *axarave* o *xarave*, («[...] corrompido, dicen *axarabe* y otros lo pronuncian sin el artículo al, diciendo *xarabe* y estos hablan más propriamente») *aceña* o *seña* («[...] digo que en el reyno de Valencia usan d'este mesmo nombre, aunque en mayor corrupçión, porque dizen *seña*»), o en una entrada lexicográfica se remite a la forma sin prefijación, como en el caso de *abentajar* «Dizen en España para significar el acto de particularizarse o de pasar adelante vna cosa de otra. Véase el nombre *ventaja*».

8. CONCLUSIONES

Un fenómeno cuyo origen no parece completamente aclarado como es el de la introducción de gran número de préstamos de la lengua árabe con el artículo determinado aglutinado en romance castellano estuvo en el origen de una identificación generalizada de palabras que comenzaban por *a/al-* con la lengua árabe. Esta percepción que hemos documentado a través de testimonios de la literatura de los siglos XVI y XVII sirvió como herramienta de lingüistas para la clasificación y análisis de arabismos desde comienzos del siglo XVI; más aún: sería de utilidad posterior en el establecimiento de tendencias normativas encaminadas a su desuso y destierro.

La *Recopilación* de arabismos de Guadix es una obra fundamental para comprender el proceso de identificación como arabismos de palabras romances con *a/al-* inicial. Guadix analiza este fenómeno en múltiples entradas de su

Recopilación no sólo a través del uso del artículo determinado, sino que aventura incluso una teoría sobre la influencia de la Forma IV de los verbos árabes en la lengua castellana. El protagonismo del artículo determinado en todas sus etimologías será, en cualquier caso, constante en su obra y conducirá incluso a la distorsión puntual de sus análisis como en la aféresis de *a-* en un nombre propio frente a la posible *inconsistencia* lingüística.

Pese a todas las deficiencias propias de una obra pionera en su género, la *Recopilación* revela una visión global de la lengua árabe con toda su complejidad. Llama la atención, por ejemplo, que el autor no sólo reconoce y formula en sus propias palabras el fenómeno de la diglosia, sino que desvincula formalmente religión musulmana y lengua árabe, aspectos ambos que aparecen tratados por primera vez en la literatura gramatical castellana. Fruto de su probable origen judío más que de su formación bíblica parece ser su percepción del árabe como versión corrupta de la lengua hebrea que le lleva incluso a establecer etimologías directamente desde esta lengua.

Finalmente, y como reflejo de la incertidumbre inicial que ocasiona la naturaleza de los indigenismos que empiezan a asentarse en castellano, Guadix los incluirá y analizará en su *Recopilación* como arabismos. La conexión con la lengua árabe no parece proceder más que de una percepción de dichos términos como parte de la otredad y de la lejanía que representan frente al mundo cristiano conocido y sus lenguas de tradición.

REFERENCIAS

- Ayala Manrique, J. F. (1693-1729). *Tesoro de la Lengua Castellana. En que se añaden muchos vocablos, etimologías y advertencias sobre el que escribió el doctísimo Sebastián de Cobarruvias*. Empeçóse esta obra a 8 de mayo, día de la aparición de S. Miguel, del año de 1693.
- Bajo Pérez, E. y F. Maíllo Salgado (Eds.) (2005). Ver Guadix (1593)
- Case, Th. E. (1982). The significance of Morisco Speech in Lope's plays. *Hispania* 65, 594-600.
- Corriente, F. (1992). *Árabe andalusí y lenguas romances*. Madrid: Editorial Mapfre.
- Corriente, F. (1996). Novedades en el estudio de los arabismos en iberorromance. *Revista Española de Lingüística* 26 (1), 1-13. DOI: <https://doi.org/10.31810/RSEL.26.1>
- Corriente, F. (2005). Notas lingüísticas acerca de la «Recopilación de algunos nombres arábigos» de Diego de Guadix. *Estudios de Dialectología Norteafricana y Andalusí* 9, 93-114.
- [Corriente, F.] /Institute of Islamic Studies of the University of Zaragoza (2012). *A descriptive and comparative grammar of Andalusí Arabic*. Brill: Leiden.

- Corriente, F., Pereira, Ch. y Vicente, Á. (2019). *Dictionnaire des emprunts ibéro-romans. Emprunts à l'arabe et aux langues du Monde Islamique*. Berlin-Boston: De Gruyter.
- Del Barrio de la Rosa, F. (2021). La prefijación «inexpresiva» en el español rural: verbos prefijados denominales con a-. *Revista de Filología Española*, CI (1), 95-125. DOI: <https://doi.org/10.3989/rfe.2021.004>
- Real Academia Española y Asociación de Academias de la Lengua Española (2014). *Diccionario de la lengua española* [en línea], 23a ed., Madrid, Espasa. <http://dle.rae.es/>
- Fernández de Santaella, R. (1499). *Vocabularium ecclesiasticum*. En línea: http://alfama.sim.ucm.es/dioscorides/consulta_libro.asp?ref=B18136436&idioma=0
- Gallego, M. Á. y Giménez-Eguíbar, P. (2020). He who loses his language loses his law: The politics of language in Medieval and Early Modern Iberia. *Revista Ideação* 22 (2), 121-143.
- García Arenal, M. (2018). Castilian and Arabic: The debates about the natural languages of Spain. En S. Schmidtke (Ed.), *Studying the Near and Middle East at the Institute for Advanced Study, Princeton, 1935-2018*, pp. 363-370. Piscataway, NJ: Gorgias Press.
- Giménez-Eguíbar, P. (2016). Attitudes toward lexical Arabisms in 16th Century Spanish texts. En S. Sessarego y F. Tejado Herrero (Eds.), *Spanish language and sociolinguistic analysis*, pp. 363-380. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/ihll.8.16gim>
- Giménez Eguíbar, P. (2020). Hacia una elaboración del tecnolecto agrícola: los arabismos léxicos en el Libro de Agricultura (1513-1539) de Gabriel Alonso de Herrera. En S. Kania y C. Kauffeld (Eds.), *Hispano-Romance Historical Linguistics and Lexicography: A tribute to John J. Nitti*, pp. 33-64. Nueva York: Hispanic Seminary of Medieval Studies.
- Guadix, D. de (1593). *Recopilación de algunos nombres arábigos que los árabes pusieron a algunas ciudades y a otras muchas cosas*, edición, introducción, notas e índices de E. Bajo Pérez y F. Maíllo Salgado. Gijón: Trea, Bibliotheca araboromanica et islamica, 2005.
- Lapesa, R. (1991). *Historia de la lengua española*. Madrid: Gredos.
- Lope Blanch, J. M. (Ed.). (1984). Juan de Valdés, *Dialogo de la lengua*. Madrid: Castalia.
- Llorca Ibi, F. X. (1992). L'article al- en els arabismes valencians. *Sharq al-Andalus* 9, 183-186. DOI: <https://doi.org/10.14198/ShAnd.1992.9.13>
- Maíllo Salgado, F. (1983). *Los arabismos del castellano en la Baja Edad Media. Consideraciones históricas y filológicas*. Salamanca: Universidad de Salamanca.
- Moreno Moreno, M. Á. (Ed.). (2007). *Diccionario de arabismos*. Jaén: Servicio de Publicaciones de la Universidad.
- Moreno Moreno, M. Á. (2011a). Lexicografía histórica y subjetividad. La obra lexicográfica (1593) de Diego de Guadix. En F. San Vicente, C. Garriga y H. Lombardini (Coords.), *Ideolex. Estudios de lexicografía e ideología*, pp. 385-402. Monza: Polimetrica International Scientific Publisher.

- Moreno Moreno, M. Á. (2011b). Las voces americanas de los diccionarios generales del español (siglos XV-XVII). *Revista Argentina de Historiografía Lingüística* III (2), 133-151.
- Müller, B. (2004). Die Arabisierung romanischer Wörter im Spanischen des Mittelalters. En A. Gil, D. Osthus, y C. Polzin-Haumann (Eds.), *Romanische Sprachwissenschaft. Zeugnisse für Vielfalt und Profil eines Faches. Festschrift für Christian Schmitt zum 60. Geburtstag*. I, pp. 203-211. Fráncfort del Meno: Lang.
- Noll, V. (2019). The agglutinated Arabic article in Ibero-Romance. *Iberorromania* 90, 185-196. doi: <https://doi.org/10.1515/iber-2019-0017>
- Real Academia Española (2001). *Nuevo Tesoro Lexicográfico de la Lengua Española*, Madrid, Espasa. En línea: <http://ntlle.rae.es/ntlle/SrvltGUILoginNtllle>
- Sayahi, L. (2003). La conservación del artículo definido 'al' en las palabras españolas de origen árabe. *Languages, Literatures and Cultures Faculty Scholarship* 19, 105-112.
- Solá-Solé, J. M. (1968). El artículo al- en los arabismos del iberorrománico. *Romance Philology* 21 (3), 275-285.
- Tejedo Herrero, F. (2009). Condicionamientos sociolingüísticos en la productividad afi-jal de finales del siglo xv. *Bulletin of Hispanic Studies* LXXXVI (3), 309-341.
- Vidal Castro, F. (2008). Los diccionarios español-árabe: más de cinco siglos de lexicografía bilingüe. *Philologia Hispalensis* 22, 319-345.
doi: <https://doi.org/10.12795/PH.2008.v22.i01.11>
- Webber, Edwin J. (1962). A Spanish linguistic treatise of the fifteenth century. *Romance Philology* 16 (1), 32-40.
- Winet, M. (2006). *El artículo árabe en las lenguas iberorrománicas (Aspectos fonéticos, morfológicos, y semánticos de la transferencia léxica)*. Córdoba: Servicio de publicaciones de la Universidad de Córdoba.

María Ángeles Gallego
CCHS - CSIC
Calle Albasanz 26-28
28037 Madrid

Patricia Giménez
Western Oregon University 345 Monmouth Ave N.
Monmouth, OR 97361 USA

VARIACIÓN LINGÜÍSTICA EN FUENTES ÁRABES DE AL-ÁNDALUS: EL TRATADO DE ‘ABD AL-RA’ŪF (CÓRDOBA, S. X)

Estefanía VALENZUELA MOCHÓN¹
Escuela de Estudios Árabes, CSIC

Resumen

Este trabajo explora la noción de variación lingüística en fuentes escritas árabes de la época premoderna. Más específicamente, examina los dos manuscritos (ms. M-O-9 y ms. d3643) que nos han llegado de la obra *Risāla fī Ādāb al-Ḥisba wa-l-Muḥtasib* de ‘Abd al-Ra’ūf, un tratado de *ḥisba* andalusí del siglo x, con el objetivo de describir el uso de rasgos no estándares e identificar los factores que motivan este tipo de alternancias. La descripción lingüística de los textos de *ḥisba* muestra que el uso de formas no estándares ocurre en todos los niveles: ortográfico, fonético, léxico, morfológico y sintáctico. Además, indica que hay, al menos, cuatro factores internos y externos que parecen activar la aparición de la variación lingüística: convenciones ortográficas de fuentes escritas en variedades no clásicas o no estándares, el entorno dialectal en el que se compone el documento, el contexto sociolingüístico y, finalmente, aspectos semánticos y pragmáticos. El análisis revela que el uso de formas no estándares en estos documentos desempeña funciones similares a las identificadas en otras fuentes orales y escritas del árabe. Estos resultados confirman que la variación lingüística en los mss. M-O-9 y d3643 no es arbitraria, sino que sigue patrones regulares del uso de la lengua, lo que representa un camino prometedor para futuras investigaciones sobre variedades no estándares del árabe escrito en fuentes del Occidente islámico.

Palabras clave: manuscritos árabes; sociolingüística histórica; variación lingüística; variedades no estándares del árabe escrito y dialectología

1. estefaniavalenzuela@eea.csic.es.  <https://orcid.org/0000-0003-3553-8938>

LINGUISTIC VARIATION IN ARABIC SOURCES FROM AL-ANDALUS:
THE MANUAL OF ‘ABD AL-RA’ŪF (CORDOVA, 10TH C.)

Abstract

This paper explores the notion of linguistic variation in Arabic written sources from the premodern period. More specifically, it examines the two surviving manuscripts (ms. M-O-9 y ms. d3643) of *Risāla fī Ādāb al-Ḥisba wa-l-Muḥtasib* by ‘Abd al-Ra’ūf, an Andalusī *ḥisba* manual from the 10th c. CE, with the goal of describing the use of non-standard features and identifying the factors that drive this type of alternation. The linguistic description of the *ḥisba* texts shows that the use of non-standard forms occurs at all levels: orthographic, phonetic, lexical, morphological, and syntactical. In addition, it suggests that there are, at least, four internal and external factors that seem to motivate the appearance of linguistic variation: orthographic conventions of non-classical or non-standard written sources, the dialectal environment surrounding the document’s composition, sociolinguistic context, and, finally, semantic and pragmatic aspects. The analysis reveals that non-standard forms in these documents perform similar functions to those attributed in other oral and written varieties of Arabic. These findings confirm that the linguistic variation shown in ms. M-O-9 and in ms. d3643 is not random; that is, it follows regular patterns of language use, which represents a promising path for future research on non-standard written varieties of Arabic from the Islamic West.

Keywords: Arabic manuscripts; historical sociolinguistics; linguistic variation; non-standard written (NSW) Arabic, and dialectology

RECIBIDO: 06/03/2023

APROBADO: 07/08/2023

INTRODUCCIÓN

La variación lingüística es un fenómeno que ocurre de forma natural en las lenguas: se trata de una «característica intrínseca» de las mismas (Hernández *et al.*, 2015, p. 405). Sucede cuando «se encuentra una «diferencia específica y analizable» entre dos estructuras que, *grosso modo*, tienen el mismo significado y se usan en los mismos contextos» (Hernández *et al.*, 2015, p. 404). Estas diferencias, denominadas alternancias o variantes, afectan a todos los niveles de la lengua: el fonético, fonológico, el léxico, el morfológico y el sintáctico. El uso de una alternancia o variante determinada puede estar motivado tanto por factores internos o lingüísticos como por factores

externos, es decir, aquellos relacionados con el contexto histórico, social, cultural o individual del hablante (Hernández *et al.*, 2015). La variación lingüística se produce como resultado del uso que los hablantes hacen de las lenguas y, por consiguiente, se manifiesta tanto en sus variantes orales como escritas.

Este artículo plantea el estudio de la variación lingüística en fuentes árabes de al-Ándalus a fin de ampliar nuestro conocimiento de la historia del árabe escrito en el Occidente islámico. En árabe, son muchos los documentos escritos de la época temprana, medieval y premoderna, aproximadamente hasta finales del siglo XIX y principios del siglo XX, que muestran la alternancia de rasgos lingüísticos (clásicos, dialectales y no estándares) de forma relativamente sistemática. Sin embargo, el análisis de estos rasgos no siempre ha ocupado un lugar destacado en los estudios de lingüística histórica, especialmente, debido a razones de índole ideológica, como se explicará más adelante.

El presente trabajo propone analizar la variación lingüística de una importante obra andalusí del siglo X, compuesta supuestamente en Córdoba: el tratado de *hisba*² de ‘Abd al-Ra’ūf. Las dos ediciones que existen hasta la fecha de este breve tratado sobre el almotacenazgo, o supervisión del mercado, (Lévi-Provençal, 1955; Chalmeta, 2019) hacen referencia a sus particularidades lingüísticas; dicho de otro modo, a la elección de rasgos no pertenecientes a la variedad clásica, o estándar, en alternancia con formas del árabe clásico. Partiendo de estos estudios, pretendemos examinar las dos copias que nos han llegado de esta obra, el ms. M-O-9 y el ms. d3643, con los siguientes objetivos: primero, ofrecer una descripción inicial del tipo de variación lingüística que muestran los textos; segundo, contextualizar los rasgos identificados en relación con su uso en otras variedades escritas y orales del árabe, y tercero, explorar qué factores, lingüísticos y extralingüísticos, parecen motivar la alternancia de formas clásicas, dialectales y no estándares en estos textos.

El artículo está organizado según se indica a continuación: en la sección 1 ofrecemos una breve descripción de las variedades del árabe escrito, explicando el lugar dominante que la variedad clásica ha ocupado tradicionalmente en este ámbito y prestando especial atención al estudio de las variedades no estándares. La sección 2 presenta la obra objeto de nuestro estudio, el tratado de ‘Abd al-Ra’ūf, y hace un breve recorrido por las investigaciones pioneras que sacaron a la luz los cuatro tratados de *hisba* occidentales. La sección 3 se dedica en exclusiva al análisis y explicación de la variación lingüística en el ms. M-O-9 y el ms. d3643 desde un punto de vista ortográfico, fonético, léxico,

2. El término *hisba* hace referencia a la obligación de todo musulmán de «ordenar el bien y prohibir el mal» *al-‘amr bi-l-ma’rūf wa-l-nahy ‘an al-munkar* y, también, al *muhtasib* como cargo público responsable de garantizar la aplicación de esta norma en el mercado (Chalmeta, 1973; Cahen *et al.*, 2012).

morfológico y sintáctico. Por último, la sección 4 señala las conclusiones más destacables de nuestro estudio y expone su relevancia con vistas a futuras investigaciones sobre las variedades no clásicas del árabe escrito.

1. VARIEDADES DEL ÁRABE ESCRITO

Al abordar el estudio de la variación lingüística en fuentes árabes escritas es imprescindible preguntarse qué es el árabe escrito y qué realidades lingüísticas representa. Pues bien, el árabe escrito se suele identificar generalmente con una variedad en concreto de esta lengua: el árabe clásico, o *fushā*, en árabe, cuyo significado literal quiere decir ‘árabe más elocuente’. Esta variedad, que representa la norma estándar codificada durante los siglos VII-X en la tradición gramática árabe, posee un gran valor cultural en las sociedades islámicas y arabófonas (Ferrando, 2007; Versteegh, 2014). El árabe clásico se percibe, o se ha percibido hasta un periodo relativamente reciente, como una única variedad normalizada, estática y uniforme en los territorios arabófonos a través de los siglos.

En el marco de la diglosia, según la define Ferguson en su influyente estudio (Ferguson, 1959), el árabe clásico ocuparía el lugar de la variedad «alta», en contraste con la variedad «baja» que estaría representada por los dialectos orales³. De este modo, el árabe clásico estaría asociado con las nociones de prestigio, patrimonio literario, homogeneidad, estandarización y, por supuesto, corrección lingüística. Sin embargo, numerosas investigaciones han demostrado que el concepto de diglosia en árabe está más relacionado con la creación de un ideal lingüístico, cuyo máximo exponente sería el árabe clásico, que con la existencia de una relación jerárquica entre las diferentes variedades de esta lengua (véanse Brustad, 2017; Benítez Fernández, en este volumen).

En su estudio sobre la noción de diglosia como ideología, Brustad (2017, p. 43) señala que la idea de que la gramática y los estándares del árabe clásico apenas han cambiado en los últimos 1400 años es prácticamente un axioma en la disciplina del arabismo. Como consecuencia, el estudio de la variación lingüística en la lengua árabe se ha centrado durante mucho tiempo en las variedades orales, es decir, los dialectos nativos de los hablantes. No fue hasta el siglo XX cuando los especialistas, principalmente europeos, empezaron a mostrar interés por el análisis de fuentes escritas de las épocas temprana y medieval que no reflejaban la norma prescriptiva

3. Ferguson emplea los términos «high variety» o «H» y «low variety» o «L» para definir las variedades lingüísticas que caracterizan a las comunidades de habla diglósicas (1959, p. 327).

del árabe clásico. Fueron precisamente estas investigaciones pioneras sobre documentos que no se ajustaban al modelo estándar las que pusieron de manifiesto que las comunidades arabófonas de este periodo histórico no empleaban exclusivamente la variedad estándar en sus escritos, sino que se servían de otras variedades de la lengua, las cuales coexistían paralelamente con el árabe clásico. Durante esta fase inicial en los estudios del árabe escrito no estándar, hay dos variedades que recibieron una atención especial por parte de los arabistas; nos referimos más concretamente al judeo árabe y al árabe cristiano.

El primero designa un tipo de árabe usado por las comunidades judías de diferentes épocas y territorios que se diferencia del resto de variedades del árabe, entre otras cosas, por estar escrito en alfabeto hebreo (Khan, 2007); mientras que el segundo hace referencia a un corpus de textos cristianos tempranos⁴, copiados en los monasterios de Palestina del sur⁵, así como a otro grupo de textos más tardíos compuestos por las comunidades cristianas después del año mil (Grand' Henry, 2007). Joshua Blau marca el punto de partida en los estudios del judeo árabe y el árabe cristiano con la publicación de dos obras que ofrecen por primera vez una descripción sistemática y detallada de las características lingüísticas de estas dos variedades: *The Emergence and Linguistic Background of Judaeo-Arabic* (Blau, 1981) y *A Grammar of Christian Arabic Based Mainly on South-Palestinian Texts from the First Millenium* (v.1)⁶ (Blau, 1966).

No obstante, un par de décadas después, ya encontramos importantes contribuciones al conocimiento de las variedades no estándares del árabe escritas por otras comunidades. Un buen ejemplo de ello es el estudio seminal de Hopkins en el que analiza la lengua de un corpus de papiros datados antes del año 300 de la hégira y 912 de la era común (Hopkins, 1984). Asimismo, empiezan a publicarse, aunque en menor número y por motivos que se explicarán más adelante, algunos estudios lingüísticos de obras de autores musulmanes como el poemario *Kitāb al-'Aḡānī* de al-'Iṣfahānī (Corriente, 1975) o la conocidísima edición de *Las Mil y Una Noches* (Mahdi, 1984).

Esta breve, pero ilustrativa, sinopsis de las diferentes variedades del árabe escrito pone en evidencia que, lejos de representar una entidad homogénea y estática, la lengua escrita se caracteriza por incorporar múltiples registros y variantes que

4. La fecha de composición de algunos de estos documentos se remonta al s. IX y al s. X.

5. Entre ellos, cabe destacar el renombrado Monasterio de Santa Catalina del Monte Sinaí, situado en la península del Sinaí, en Egipto.

6. Los volúmenes 2 y 3 de esta gramática se publicaron poco tiempo después en 1967.

reflejan una realidad lingüística diversa y llena de matices. En cuanto al aspecto terminológico, si como hemos señalado anteriormente el árabe clásico se considera la variedad estándar de esta lengua, la expresión comúnmente aceptada para referirse a esta diversidad tipológica de textos compuestos en registros que no reflejan las normas gramaticales del árabe clásico es, de manera más concreta, árabe medio.

1.1. *¿Árabe medio o variedades no estándares del árabe escrito?*

Lentin (2008a) define el árabe medio como una variedad lingüística y estilísticamente diversa, empleada en un gran número de textos escritos y cuya característica principal consiste en la alternancia de «rasgos estándares y coloquiales con otros de un tercer tipo»⁷ (p. 216) que no pertenecen, estos últimos, ni al árabe clásico ni a las variedades dialectales. Se trata, según su explicación, de una variedad intermedia y multiforme. A pesar del importante grado de variación que existe en cuanto a la alternancia y combinación de rasgos lingüísticos de un tipo u otro en los textos, Lentin asegura que los documentos escritos en árabe medio exhiben una gran cantidad de características comunes, especialmente con respecto a los rasgos «del tercer tipo», que parecen ser específicos de estas variedades del árabe escrito. Por ello, nos advierte que no sería adecuado describir el árabe medio como un uso «errático»⁸ de la lengua, ya que uno encuentra usos bien establecidos y jerarquías estilísticas entre variantes.

Pese a que el término «árabe medio» se ha venido usando desde hace ya algunas décadas para designar las variedades arriba mencionadas, esta denominación no está, ni mucho menos, exenta de problemas y prejuicios ideológicos. La expresión «árabe medio» se ha utilizado a lo largo de su historia para identificar múltiples realidades lingüísticas e históricas que no siempre ofrecían una descripción rigurosa de este tipo de documentos. Por ejemplo, Lentin (2008a) señala que el adjetivo «medio» ha recibido interpretaciones muy diversas, entre ellas «históricamente medio, sociolingüísticamente intermedio, lingüísticamente mixto e incluso hay quien lo ha usado para referirse a la calidad mediocre de los textos»⁹ (p. 215). Por ello, y a fin de evitar ambigüedades, aquí se usarán las expresiones «árabe escrito no

7. Traducción mía del texto original en inglés: «It combines standard and colloquial features with others of a third type, neither standard nor colloquial».

8. Lentin utiliza el término «wild».

9. Traducción mía del texto original en inglés: «Historically middle, sociolinguistically intermediate, linguistically mixed – not to mention the middling quality of texts, in the opinion of some people».

estándar» y «variedades no estándares del árabe escrito», siguiendo la terminología adoptada por Brustad (en prensa)¹⁰.

Una de las nociones más extendidas en los estudios de textos escritos en árabe medio, especialmente en sus comienzos, tiene que ver con la idea de que estos textos recogen un conjunto de errores cometidos por autores y copistas que no tenían la suficiente competencia lingüística en árabe clásico y, como resultado, eran incapaces de adecuarse a las normas establecidas en esta variedad. Lentin (2008a, pp. 216–217) cuestiona esta hipótesis argumentando que son muchos los autores celebrados por su dominio de la variedad estándar que nos han dejado, asimismo, textos en los que emplean intencionadamente formas lingüísticas que no pertenecen al árabe clásico. Por lo tanto, concluye, la elección de estos autores de emplear formas del árabe medio no puede justificarse como reflejo de su desconocimiento del árabe clásico o como un mero uso accidental de la lengua, sino que debía responder a otros factores como, por ejemplo, los destinatarios de dichos textos. Partiendo de esta afirmación, Mejdell (2017, p. 73) menciona la obra *Kitāb al-ʿIṭbār* de ʿUsāma bin Munqid (1095–1188) como un «caso temprano» de uso deliberado del árabe medio por parte de un autor versado en la variedad clásica. Por otra parte, Brustad (2017, p. 44) examina una carta escrita y firmada por Buṭrus al-Bustānī (m. 1883) en la que el afamado intelectual de la *nahḍa* y entusiasta defensor de la lengua árabe, particularmente del árabe clásico, no parece tener ningún reparo a la hora de incluir expresiones que no reflejan un uso clásico o estándar de la lengua. Estos dos ejemplos demuestran que las prácticas de escritura de las sociedades arabófonas premodernas, tal y como concibe Lentin, revelan una situación sociolingüística compleja, más allá de la adquisición de la variedad clásica, que necesita ser estudiada con detenimiento en base a los datos que nos proporcionan las fuentes originales árabes.

En un principio, el estudio de las variedades de árabe escrito no estándar se conceptualizó como una característica propia de las comunicadas cristianas y judías, bajo la premisa de que estos autores estaban menos apegados ideológicamente a las normas estilísticas del árabe clásico y, por lo tanto, más dispuestos a utilizar formas no prescriptivas en sus textos (Blau, 1966–1967, 2002; Wagner, 2010; Khan, 2007). Sin embargo, esto se debe en parte al gran interés que los textos escritos por estas comunidades despertaron en los especialistas durante una fase inicial en los estudios de árabe medio y, no menos importante, «al hecho de que de muchas obras

10. «Non-Standard Written» (NSW) en inglés.

de autores musulmanes se estudian en ediciones normalizadas artificialmente»¹¹ (Lentin, 2008a, p. 218); dicho de otro modo, se estudian mediante ediciones «estandarizadas», cuya lengua se modificaba en conformidad con el canon clásico.

Pese al gran número de documentos que exhiben rasgos atribuibles al árabe medio, Lentin (2008a) señala que «la función e importancia de estas variedades intermedias han sido frecuentemente rechazadas por árabes y arabistas, especialmente debido al estatus excepcional del árabe clásico y por razones de purismo y elitismo»¹² (p. 219): en otras palabras, por razones ideológicas. Sin embargo, Lentin señala que el árabe medio representa una parte importante de la producción escrita en árabe, ha sido instrumento de comunicación en la lengua escrita durante siglos y, muy posiblemente, haya conservado rasgos lingüísticos muy antiguos que desaparecieron de la variedad clásica. Por todo ello, defiende que «el árabe medio es de una importancia fundamental para entender correctamente la naturaleza y la historia de la lengua árabe»¹³ (p. 219). Una vez contextualizado el estudio del árabe escrito y aclarados los aspectos relacionados con la terminología, nos disponemos a presentar los textos objeto de nuestro análisis lingüístico.

2. EL TRATADO DE *HISBA* DE ‘ABD AL-RA’ŪF

A partir del siglo X, un nuevo género de literatura especializada, la de los tratados de *hisba*, empieza a florecer a través de los territorios del imperio islámico. Estas obras se conciben a modo de referencia o guía profesional con el objetivo de orientar al almotacén o inspector del mercado, en árabe, *muhtasib* o *ṣāhib al-sūq*, en sus quehaceres diarios y facilitarle la labor (Chalmeta, 1973; Constable, 1996). Los tratados de *hisba* proporcionan información muy práctica y específica sobre la vida diaria en los mercados de importantes núcleos urbanos de la época; en términos generales, y salvando las diferencias de contenido y enfoque que puedan existir entre ellos: regulan conductas comerciales y sociales en el mercado y otros lugares públicos de la ciudad, describen los fraudes más comunes que se llevaban

11. Traducción mía del texto original en inglés: «It might be also due to the fact that many texts by Muslim authors are studied in artificially normalized editions».

12. Traducción mía del texto original en inglés: «The function and importance of these intermediate varieties has often been neglected, both by Arabs and Arabists, mainly because of the outstanding status of Classical Arabic, and for reasons of purism and elitism».

13. Traducción mía del texto original en inglés: «Middle Arabic is of fundamental importance for a correct understanding of the nature and history of the Arabic language».

a cabo en los diferentes gremios, explican los métodos de detección más efectivos y establecen sanciones específicas según la falta o el delito cometidos.

El tratado de ‘Abd al-Ra’ūf pertenece a un grupo de cuatro obras de *ḥisba* creadas en el Occidente islámico entre los siglos X y XIV¹⁴. Se trata del primero de su categoría conocido en al-Ándalus¹⁵, solamente precedido, en el caso de los tratados orientales, por el *Kitāb al-‘Ḥtisāb* del imam zaydí al-Nāṣir al-Ḥasan b. ‘Alī al-‘Uṭrūš (m. 917). Siguiendo las crónicas de Ibn ‘Idārī y de Ibn Ḥayyān, Chalmeta (1973, 2019) establece que nuestro autor, cuyo nombre completo parece haber sido ‘Aḥmad b. ‘Abd Allāh ‘Abd al-Ra’ūf, podría haber ocupado el puesto de almotacén en la capital del Califato Omeya entre los años 926–929 o 314–317 según el calendario islámico. Además de la evidencia textual que aportan las fuentes históricas, Chalmeta se sirve del análisis crítico de los contenidos para determinar el lugar y la fecha de composición de la obra (véase Chalmeta, 1973, pp. 382–383 y Chalmeta, 2019, pp. 15–25).

Desde el punto de vista temático, el tratado de ‘Abd al-Ra’ūf puede dividirse en dos partes. La primera se dedica a garantizar el correcto cumplimiento de ciertas obligaciones y prácticas religiosas: la oración, los funerales, el ayuno, el azaque, el matrimonio y los habices. Según Chalmeta (2019), el tratamiento de estos temas resulta llamativo por no estar presente en ningún otro tratado de *ḥisba*. Además, el carácter normativo de este tratado lo diferencia de obras de *ḥisba* posteriores, exclusivamente dedicadas a «la detección de fraudes y malas prácticas» (pp. 12). Chalmeta atribuye este hecho al momento histórico y social en el que se gesta la obra de ‘Abd al-Ra’ūf, esto es, una sociedad, la cordobesa de la época califal, que todavía no estaba completamente islamizada. En la segunda parte del tratado, ‘Abd al-Ra’ūf se encarga de regular y describir asuntos propiamente relacionados con el funcionamiento del mercado y los gremios de artesanos que allí ejercían sus respectivas ocupaciones.

En cuanto a las fuentes que nos han llegado de este tratado de *ḥisba*, en la actualidad, solamente se conocen dos copias: el ms. M-O-9 y el ms. d3643. El primero,

14. Presentados en orden cronológico, los cuatro tratados de *ḥisba* occidentales son: *Risāla fī ‘Ādāb al-Ḥisba wa-l-Muḥtasib* de ‘Abd al-Ra’ūf (Córdoba, X), *Risāla fī l-Qaḍā’ wa-l-Ḥisba* de Ibn ‘Abdūn (Sevilla, XI-XII), *Kitāb fī ‘Ādāb al-Ḥisba* de al-Saqāṭī (Málaga, XIII) y *Risāla fī l-Ḥisba* de al-Jarfīsī (Guercif?, XIII-XIV).

15. En este sentido, y aunque no se trate de un tratado de *ḥisba* propiamente, no se puede dejar de mencionar la obra precursora más directa de los tratados de *ḥisba* en el Occidente islámico. Nos referimos al *Kitāb ‘Aḥkām al-Sūq* (Ordenanzas del zoco) de Yahyā b. ‘Umar (828–901). Se trata de un compendio del siglo IX escrito en forma de «responsae» que recopila las sentencias del autor, natural de Jaén y afincado en Qayrawān, acerca de numerosas cuestiones relacionadas con el mercado (García Gómez, 1957).

ms. M-O-9, perteneció a Lévi-Provençal y, tras su muerte en 1956, pasó a formar parte de la colección de manuscritos árabes del Instituto Egipcio de Estudios Islámicos de Madrid. El segundo, ms. d3643, se conserva en la Biblioteca Nacional del Reino de Marruecos (BNRM). No obstante, en su pionera edición del tratado de ‘Abd al-Ra’ūf (1955), Lévi-Provençal utilizó un tercer manuscrito, identificado en las notas como copia «A», que le sirvió como texto base de su edición. Desafortunadamente, esta copia se halla, desde el fallecimiento del arabista, perdida o en paradero desconocido (Chalmeta, 2019, p. 6). Tampoco existe, o al menos no se ha descubierto a día de hoy, una copia autógrafa del tratado, lo que, desde el punto de vista lingüístico, representa una ausencia muy significativa, ya que nos impide atestiguar qué rasgos de la lengua reflejan usos coetáneos al autor y cuáles fueron introducidos posteriormente. De hecho, los textos que han sobrevivido revelan que la obra, en su forma presente, no fue escrita por ‘Abd al-Ra’ūf. El uso repetido de la fórmula قال عبد الرؤف *qāl(a)* ‘*Abd al-Ra’ūf*’ dijo ‘Abd al-Ra’ūf’ para introducir las opiniones y juicios del autor es una clara muestra de ello, tal y como señala Chalmeta (1973, p. 383).

A pesar de la escasez de datos que tenemos sobre las fuentes originales, el estudio de los rasgos lingüísticos que exhiben estos textos ofrece información valiosa acerca de la situación del árabe escrito en el Occidente islámico durante la época premoderna, así como de su conexión con otras variedades orales y escritas de las regiones orientales. Las páginas que siguen están dedicadas a contextualizar los antecedentes históricos en los que se cimienta nuestra investigación de las copias existentes del tratado de ‘Abd al-Ra’ūf. Partiendo de esta base, la sección 2.1 presenta una visión general de los estudios que dieron a conocer los tratados de *ḥisba* occidentales a la comunidad científica durante el siglo XX, prestando especial atención a la importancia que dichas investigaciones otorgaron a la presencia de elementos dialectales en los textos.

2.1. Estudios previos de los tratados de *ḥisba* occidentales

La diversidad lingüística que presentan los tratados de *ḥisba* occidentales, es decir, el uso de formas no pertenecientes a la variedad clásica de la lengua árabe, no pasó desapercibida a los primeros especialistas que descubrieron estos documentos medievales y se interesaron por su estudio. De hecho, los trabajos de investigación que dieron a conocer las obras de *ḥisba* del Occidente islámico, publicados entre los años 1931 y 1955, tienen un claro enfoque lingüístico, además de histórico.

En este aspecto, es imprescindible destacar la gran labor realizada por dos figuras clave del arabismo de esta época, Georges Séraphin Colin y Évariste Lévi-Provençal. En 1931, estos dos célebres arabistas franceses inauguran los estudios de los tratados de *ḥisba* occidentales con la publicación de una edición inédita del tratado de al-Saqatī (Málaga, s. XIII) (Colin y Lévi-Provençal, 1931). En la introducción, los autores señalan que la importancia de esta obra no solo radica en el valor de los datos históricos que aporta, sino en el componente dialectal que muestra la lengua de los textos. Por lo tanto, concluyen que el estudio de esta obra puede ser de interés tanto para historiadores de la civilización musulmana occidental como para especialistas en historia de la lengua árabe¹⁶. El empeño de estos dos autores por documentar la presencia de rasgos dialectales en la obra del autor malagueño queda evidenciada en la frecuente mención de usos y voces del «árabe-hispánico», más tarde conocido como «árabe andalusí»¹⁷, en las notas lingüísticas del texto y en el glosario que acompañan al texto en árabe.

A esta obra pionera le sucedió una edición del hasta entonces desconocido tratado de Ibn ‘Abdūn (Sevilla, s. XII), (Lévi-Provençal, 1934). Siguiendo la estructura de la obra anterior, el texto árabe aparece acompañado de una introducción y un glosario. Finalmente, Lévi-Provençal culmina la etapa inicial de presentación de los tratados de *ḥisba* occidentales con la publicación de *Documents arabes inédits: sur la vie sociale et économique en Occident musulman au Moyen Âge. Trois traités hispaniques de Hisba (texte arabe)* (Lévi-Provençal, 1955). Este trabajo incluye una nueva reproducción del texto árabe del tratado de Ibn ‘Abdūn y la edición de dos textos de *ḥisba* inéditos hasta entonces: el tratado de ‘Abd al-Ra’ūf (Córdoba, s. XIV) y el tratado de al-Jarfīsī (Guersif?, s. XIV).

En la introducción, Lévi-Provençal vuelve a incidir en la importancia filológica de estos textos como una rica fuente de voces y expresiones procedentes de los dialectos árabes occidentales y, también, algunas de origen romance. Asimismo, revela su intención de crear un glosario extensivo de los tratados y la futura publicación de nuevos textos árabes relacionados con la vida social y económica

16. El texto original en francés aparece en la introducción del trabajo tal y como sigue a continuación: «Il nous a paru susceptible d’intéresser à la fois les historiens de la civilisation musulmane occidentale et les historiens de la langue arabe».

17. La denominación «árabe andalusí» fue acuñada en los años 80 por el mayor especialista de este dialecto, Federico Corriente. El término «hispano-árabe», usado hasta entonces para designar el dialecto árabe de al-Ándalus, resultaba problemático por atribuir erróneamente una importancia excesiva al componente hispano, cuya influencia es, como se ha demostrado en estudios posteriores, limitada (Ferrando, 2012, p. 222).

de al-Andalus y el Magreb en la Edad Media. Desgraciadamente, el inesperado fallecimiento del arabista tan solo un año después interrumpe bruscamente su deseo de expandir esta línea de investigación. En las décadas que siguen a los trabajos precursores de Colin y Lévi-Provençal, el creciente interés por los tratados de *ḥisba* occidentales se ve reflejado en la publicación de nuevas ediciones, traducciones y revisiones de los textos. Sin embargo, el análisis lingüístico de los mismos pasa a un segundo plano en estas nuevas investigaciones que adoptan un enfoque más centrado en el análisis histórico, social, económico y crítico de las obras.

El presente trabajo representa un modesto intento de retomar el camino iniciado por Colin y Lévi-Provençal hace ya casi un siglo. En particular, aquí se pretende abordar el estudio de la variación lingüística que se observa en las dos copias que nos han llegado del primer tratado de *ḥisba* andalusí y explorar los posibles factores que motivan esta variación; todo ello con el objetivo fundamental de contribuir al conocimiento de las variedades no estándares del árabe escrito, especialmente aquellas que se utilizaban en las regiones más occidentales del mundo arabófono con anterioridad al siglo xx.

3. VARIACIÓN LINGÜÍSTICA EN EL TRATADO DE ‘ABD AL-RA’ŪF

Como ya hemos anticipado, el tratado de ‘Abd Ra’ūf pertenece a una tradición manuscrita que en lo que se refiere al uso de la lengua no se ciñe exclusivamente a las normas prescriptivas de la variedad clásica, o *fushā* en árabe, sino que incorpora otro tipo de rasgos lingüísticos que pertenecen a lo que aquí denominamos variedades no estándares¹⁸. Dicho de otro modo, la lengua de los manuscritos que nos han llegado de esta obra muestra una interesante combinación o alternancia de elementos clásicos, dialectales y otros específicos de las variedades no estándares de la lengua árabe, todo lo cual aporta una gran riqueza lingüística al texto y plantea, por otra parte, muchas preguntas. La variación lingüística, entendida en este sentido, es una característica que comparten, en mayor o menor medida, los cuatro tratados de *ḥisba* creados en el Occidente islámico entre los siglos x-xiv y que los diferencia, entre otras cosas, de sus homólogos orientales.

La descripción lingüística de este tipo de textos, así como el análisis contextualizado de sus características, constituye una tarea esencial para poder entender las prácticas lingüísticas de las sociedades arabófonas de la época premoderna,

18. Las variedades no estándares del árabe escrito reciben más comúnmente la denominación de «árabe medio», véase sección 2.

especialmente, teniendo en cuenta que nos permite documentar usos y variantes de la lengua que fueron desapareciendo en los procesos de edición y corrección de textos que proliferaron a partir del siglo XIX, eliminando, de esta forma, una parte importante de la historia del árabe escrito (Lentin, 2008a; Brustad, 2017). Sin embargo, el trabajo con manuscritos árabes no está exento de obstáculos que limitan nuestra capacidad de interpretar y evaluar las características lingüísticas de estos documentos.

Una de las cuestiones más complejas a las que se enfrentan los lingüistas a la hora de analizar estas fuentes deriva de la falta de información que existe sobre el contexto de producción y transmisión del manuscrito. En el caso del tratado de 'Abd Ra'ūf, de las dos únicas copias conocidas que se conservan de esta obra, solamente el ms. d3643 menciona datos que proporcionan la posible identidad del copista responsable del trabajo y la fecha de composición, más allá de las características caligráficas de los textos, de estilo magrebí, y la propia materialidad de los documentos. Los detalles de la posible relación que une a ambos manuscritos es otro aspecto que sigue sin conocerse y necesita ser investigado. En definitiva, el profundo desconocimiento de los factores extralingüísticos que motivaron y condicionaron algunas de las elecciones lingüísticas de los textos tiene un considerable impacto en el análisis de los mismos, tal y como veremos en la siguiente sección. En este complejo y nada revelador contexto, el estudio lingüístico de las fuentes primarias cobra, aún más si cabe, una importancia crucial para extraer cuanta información sea posible y dotar de sentido a los escasos datos que nos han llegado.

A continuación, se procede a la descripción detallada de los rasgos lingüísticos más relevantes de los dos manuscritos objeto de estudio desde una perspectiva comparativa, entre los propios documentos y, también, en relación con la existencia de usos similares en otras variedades escritas y orales del árabe, tanto de la época premoderna como contemporánea.

3.1. Descripción lingüística: ms. M-O-9 y ms. d3643

Los cuatro tratados de *hisba* occidentales representan a grandes rasgos, y dejando a un lado momentáneamente las diferencias lingüísticas y de contenido que existen entre ellos, un corpus de textos escritos en árabe no estándar relativamente homogéneo. Además de la temática, comparten unas coordenadas geográficas y una extensión cronológica delimitadas que los convierten en un objeto de estudio ideal en el ámbito de la lingüística histórica. Por ello, la descripción de las características lingüísticas del tratado de 'Abd al-Ra'ūf, que como ya se ha indicado

fuere el primero de su género en al-Ándalus, cronológicamente hablando, ofrece un prometedor punto de partida para la investigación lingüística de estas fuentes y el conocimiento de las variedades no estándares del árabe escrito en Occidente islámico.

3.1.1. *Ortografía y fonética*

Desde el punto de vista ortográfico, uno de los rasgos más extendidos y fácilmente reconocibles en las variedades no estándares del árabe escrito corresponde a la frecuente omisión de la consonante oclusiva glotal sorda /ʔ/, *hamza* en árabe, representada con el grafema ء. En los dos manuscritos del tratado de 'Abd al-Ra'ūf, este fenómeno se puede observar tanto en posición inicial > احد *aḥad* 'uno' como final > سوداء سودا *sawdā* 'negra' > أخطاء اخطا *aḥṭā* 'errores'.

La escritura de la *hamza* sobre el soporte de la vocal larga /ī/ o consonante palatal fricativa sonora *yā* /y/ ي es otra peculiaridad ortográfica de nuestros textos. Según la norma clásica, la representación de este grafema está condicionado por el entorno vocálico que lo rodea: cuando va seguida de /a/, se escribe sobre el soporte de *alif* أ /'a/, mientras que cuando va seguida o precedida de /i/ o /y/, se escribe sobre el soporte de *yā* ي /'i/'. Sin embargo, nuestros textos muestran un interesante caso en el que no se mantiene esta convención ortográfica. El verbo *preguntar* يسأل *yas'alu* '[él] pregunta', derivado de la raíz consonántica س *sīn* - ء *hamza* - ل *lām*, presenta la escritura invariable de *hamza* sobre el soporte de *yā* cuando le sucede la vocal corta /a/ en ambos manuscritos: > يسألون > يسألون *yas'alūn* '[los hombres] mendigan'¹⁹, > يسأل *tas'al* '[la mujer joven] mendiga'²⁰, > يسألون *yus'alūn* 'se les preguntará [a los hombres]', > يسأل *yus'al* 'se le preguntará [a esas gentes]', > فيسألهم *fa-yas'alu-hum* 'y preguntará a los [primeros]'²¹.

En otras ocasiones, la *hamza* tampoco se representa sobre el soporte de la *alif* أ /'a/, como dicta la norma clásica, sino que aparece escrita de forma independiente > آذان *adān* 'llamada a la oración' y > أكل *akl* 'comida'; aunque en su gran mayoría, esta grafía refleja de forma específica la secuencia /'ā/, expresada normalmente por el grafema آ, conocido en la gramática árabe con el nombre de *'alif mamdūda* 'prolongada': > رأى *ra'ā* 'vio', > آخر *'āḥir* 'último', > آمنوا *'āmanū* '[ellos] creyeron', > أتوا *'ātū* 'pagad [vosotros]', > آيات

19. Siguiendo la traducción de Chalmeta (2019, p. 48).

20. Siguiendo la traducción de Chalmeta (2019, p. 48).

21. Siguiendo la traducción de Chalmeta (2019, p. 77).

'āyāt 'versículos o aleyas del Corán' y آلات > آيات 'ālāt 'utensilios' y مرآة > مراة
mir'āt 'espejo(s)'.
A pesar de las muchas similitudes que muestran los dos manuscritos, las preferencias ortográficas de ambos no siempre coinciden. Una diferencia significativa entre ellos atañe la elisión de la vocal larga ¹ /ā/ en la grafía de algunos términos, principalmente de carácter religioso, en el ms. del Instituto Egipcio; la copia de Rabat, en cambio, muestra las variantes con /ā/, *scriptio plena*, de forma sistemática: تعالى vs تعلى ta 'ālā 'el Altísimo', ملك vs. مالك mālīk 'Malik', اسبغ vs. اسباغ isbāg 'ablución' شيطان vs. شيطان sayṭān 'demonio' y ذلك vs. ذلك dālīka 'aquel'.

Según Wright (1896–1898, vol.1, pp. 2–3), este rasgo, es decir, la ausencia de la vocal larga ¹ /ā/, refleja un estilo de escritura temprana en fuentes árabes que aparece con más frecuencia en mss. de origen magrebí que en documentos de otras regiones arabófonas. En consecuencia, podría resultar tentador asumir que el ms. M-O-9 fue escrito en una fecha anterior al ms. d3643. No obstante, y teniendo en cuenta que los dos documentos comparten características de lo que se conoce como caligrafía magrebí (Van den Boogert, 1989; Bongianino, 2017), la única hipótesis que podemos aventurar por el momento es que los textos fueron posiblemente copiados por manos diferentes, cuya formación e individualidad queda plasmada en este tipo de prácticas ortográficas.

Asimismo, la escritura de los sustantivos de la forma nominal فَعْلَةٌ fa 'ala(t), con و wāw como tercera consonante radical, delata la adhesión a convenciones ortográficas distintas en cada manuscrito. En este tipo de sustantivos, la consonante و /w/ se suele representar gráficamente con una 'alif ¹ /ā/, aunque en algunas ocasiones se mantiene la ortografía etimológica con /w/ (Wright, 1896–1898, vol.1, p. 121). Nuestros manuscritos ilustran perfectamente el uso de ambas variantes en términos que indican dos de los cinco pilares de la doctrina islámica: الصلاة al-ṣalā(t) 'la oración' y الزكاة al-zakā(t) 'el azaque'. La copia de Rabat muestra sistemáticamente las variantes con 'alif, الصلاة و الزكاة; mientras que la copia del Instituto Egipcio alterna entre las variantes con 'alif y con wāw الصلاة – الصلوة و الزكاة – الزكوة.

El registro consonántico también presenta una serie de alternancias que revelan cierto nivel de divergencia con el árabe clásico. Este fenómeno se encuentra de forma más concreta en la representación de algunas consonantes (inter)dentales y sibilantes. Aunque este tipo de cambios son bastante frecuentes en árabe escrito no estándar (Lentin, 2008a), el análisis comparativo de la distribución de las variantes consonánticas en el ms. M-O-9 y el ms. d3643 arroja algo de luz acerca del entorno

dialectal en el que se crearon los textos y sugiere que las alternancias consonánticas responden a patrones específicos:

$\underline{t} > t$: El conocido cambio de la fricativa interdental \underline{t} por la oclusiva dental t se observa en ambos manuscritos, v. gr., ثروب > ثروب *turūb* ‘redaños’ و يستائر > يستائر *yasta ‘tir* ‘reservarse algo’; aunque esta tendencia parece estar algo más marcada en el manuscrito de Rabat: ورتة > ورتة *warāta* «herederos». La sustitución de las interdentes fricativas por las dentales oclusivas correspondientes ($\underline{t} > t - \underline{d} > d - \underline{z} > d$) es una característica distintiva de una gran parte de los dialectos urbanos (Palva, 2006, pp. 605–606) que también está documentada en algunas fuentes andalúsies (Corriente, 2013, p. 21) y en variedades no estándares del árabe escrito (Hopkins, 1984, p. 33; Lentin 2008a, p. 220). Algo más inusual es, sin embargo, el cambio inverso, es decir, la sustitución de t por \underline{t} que se produce con el término تمر *tamr* ‘dátiles’ en la copia de Rabat. Curiosamente, éste aparece representado gráficamente por la variante interdental \underline{t} , ثمر *tamr*, en las tres ocasiones en las que se menciona en el tratado.

$\underline{d} > d$: La consonante oclusiva dental sonora d también reemplaza ocasionalmente a su variante fricativa interdental \underline{d} en nuestros manuscritos. Sin embargo, en este caso, el uso de la variante dental está restringido a términos muy concretos, específicamente aquellos derivados de la raíz ق *qaf* - ذ *dāl* - ر *rā*’ y de la raíz ذ *dāl* - ب *bā*’ - ب *bā*’ : فقرة > فقرة *qadira* ‘sucia’, متقدرين > متقدرين *mutaqaddirin* ‘sucisimos’, اقدار > اقدار *aqdār* ‘suciedad’ و ذباب > ذباب *dubb* ‘moscas’ و يذاب > يذاب *yudāb* ‘se disuelve’.

$\underline{d} > \underline{d}$: Una particularidad notable del manuscrito del Instituto Egipcio consiste en la representación de la oclusiva dental d como fricativa interdental \underline{d} . Como ya hemos mencionado, la sustitución de las consonantes interdentes fricativas por las dentales oclusivas es un rasgo bastante más común en muchas variedades del árabe que el fenómeno contrario (dental > interdental). La diferencia que existe en este aspecto entre ambos manuscritos puede apreciarse en ejemplos como تمادی *tamādā* vs. تماندی *tamādā* ‘persistiendo’ o مناديل *manādīl* vs. مناديل *manādīl* ‘paños’. Curiosamente, la variante interdental afecta de forma especial y recurrente a las derivaciones léxicas de la raíz triconsonántica ء *hamza* - د *dāl* - ب *bā*’ : يودب > يودب *yū‘addab* ‘[él] será castigado’, ادبوا > ادبوا *uddibū* ‘[ellos] fueron castigados’, ادباً > ادباً *adaban* ‘castigo’, etc.

En árabe andalusí, la realización interdental del fonema *d* se explica como una consecuencia del sustrato romance en este dialecto (Corriente, 2013, p. 15). Por supuesto, en nuestro caso, no es posible determinar si la alternancia *d* > *d̪* que presenta el ms. M-O-9 refleja necesariamente una procedencia andalusí sin más evidencias que apunten hacia esa dirección, pero representa, sin duda alguna, un dato adicional que pone de relieve la sistematicidad que existe en la elección de ciertas variantes lingüísticas en las fuentes escritas de la época premoderna.

ʃ > s - s > ʃ: La velarización y desvelarización de las sibilantes, según explica Corriente (2013, p. 25) es un rasgo bastante extendido en numerosas variedades del árabe, incluido el dialecto andalusí, que ya aparece documentado en las primeras gramáticas de esta lengua. Los dialectos orientales modernos tampoco son ajenos a este fenómeno y revelan casos de dicha alternancia (véase, a modo de ejemplo, Berlinches (2016, pp. 38–39) para el dialecto de Damasco).

En las dos copias del tratado de ‘Abd al-Ra’ūf, la silbante alveolo-predorsal velarizada *ʃ* aparece gráficamente representada por la variante sorda *s* en dos términos derivados de la raíz ص *sād* - ب *bā’* - غ *ḡayn*: تصبيغ > تصبيغ *tasbīg* ‘teñir’ y صبغ > صبغ *sibg* ‘tinte’. Por otra parte, la velarización de la silbante alveolo-predorsal sorda *s* > *ʃ* ocurre cuando esta va seguida por la vocal larga *ā* y la oclusiva dental sorda velarizada *ṭ*: مصاطب > مصاطب *maṣāṭib* ‘bancos’ y صاطور > صاطور *ṣāṭur* ‘trinchante’²². En este contexto, la velarización de la sibilante *s* se entiende fácilmente desde el punto de vista fonológico por la presencia adyacente de la oclusiva dental sorda velarizada *ṭ*.

En definitiva, y a modo de resumen, la mayoría de los rasgos ortográficos y fonéticos identificados en las dos copias existentes del tratado de ‘Abd al-Ra’ūf se encuentran con bastante frecuencia en fuentes árabes de la época premoderna escritas en variedades no estándares: omisión de *hamza* (Blau, 1966–1967, pp. 83–104; Schen, 1972, pp. 235–236; Hopkins, 1984, pp. 19–33; Lentin, 2008a, p. 220), el cambio fonético interdental > dental (Blau, 1966–1967, pp. 107–108; Hopkins, 1984, p. 36; Lentin, 2008a, p. 220) y la (des)velarización de las consonantes sibilantes (Blau 1966–1967, pp. 109–113; Hopkins, 1984, pp. 37–38; Lentin, 2008a, p. 220).

22. Siguiendo la traducción de Chalmeta (2019, p. 86).

Además de estas particularidades lingüísticas que ambos documentos comparten con otras muchas variedades del árabe, escritas y orales, nuestros dos manuscritos presentan algunas divergencias que aportan datos relevantes acerca del proceso de composición de las copias; más concretamente, sugieren que los documentos fueron copiados por manos diferentes. De particular importancia son, en este aspecto, las prácticas ortográficas arcaizantes o más conservadoras, omisión de ^l /ā/ y escritura de la variante etimológica /w/ en ciertos términos, así como la tendencia al cambio *d* > *ḍ* que exhibe el manuscrito del Instituto Egipcio en contraste con la ausencia de estos rasgos en el ms. de Rabat.

Por otra parte, es necesario advertir que la presencia de variantes dialectales (*ṭ* > *t* – *ḍ* > *d* – *ṣ* > *s* – *s* > *ṣ*) en los dos manuscritos del tratado de ‘Abd al-Ra’ūf no implica que éstas sean un reflejo inequívoco del dialecto nativo del escriba; hecho prácticamente imposible de determinar sin tener información sobre el mismo ni conocer el método o los métodos de transmisión de estos textos. Más bien, podría interpretarse de forma tentativa como la aparición de características propias de la variedad o las variedades dialectales presentes en el entorno de creación de cada documento, así como de las prácticas de escritura adoptadas por diferentes tradiciones manuscritas. En cualquier caso, lo que sí parece quedar claro es el papel esencial de los copistas en la inserción y transmisión de variantes lingüísticas a nivel ortográfico y fonético en estos textos.

3.1.2. *Léxico*

Como se ha señalado en 2.1, la riqueza léxica de los tratados de *ḥisba* occidentales atrajo desde un primer momento la atención de los especialistas, más concretamente la de los reconocidos arabistas franceses Colin y Lévi-Provençal, quienes consideraban el uso de términos romances en estos documentos de una gran importancia histórica y, por supuesto, filológica. En su edición del tratado de ‘Abd al-Ra’ūf, Chalmeta (2019, p. 29) también menciona la presencia de algunas palabras romances en el breve comentario que ofrece sobre las características lingüísticas de este tratado de *ḥisba* andalusí.

En nuestros textos, el ms. M-O-9 y el ms. d3463, los términos romances se utilizan para designar carnes y pescados de consumo, como es el caso de قنينيات *qunnīniyāt* ‘conejos’, del Latin *cuniculus*, y سردين *sardīn* ‘sardinias’, para referirse a prácticas o técnicas culinarias, شرمولة *šarmūla*, ‘salmuera’, o, incluso, para indicar el nombre de plantas الاستب *al-istibb* ‘estepa’. Aunque son escasos en número, estos

ejemplos nos ofrecen información valiosa en cuanto nos permiten documentar el uso de voces de origen romance en las obras de *ḥisba* del Occidente islámico y, por extensión, en sus mercados medievales.

En otro de los tratados de *ḥisba* occidentales, el de al-Saqaṭī (Málaga, s. XIII), Colin y Lévi-Provençal (1931, p. 9) identifican un total de nueve préstamos de origen romance. Cinco de ellos hacen referencia a utensilios de trabajo: اقليال *iqliyāl* ‘clavo para fijar el hierro de una bestia’, كوب *kūb* ‘cubo’, قبطال *qubṭāl* ‘reglón o regla de albañilería’, طلي *ṭalyo* ‘cuchillo con el que el carnicero corta la carne’ y شيرة *šayra* ‘cestas para transportar almendras’. Otros tres están directamente con relacionados con el ámbito de la alimentación: بلاجة *balāġa* ‘un tipo de paté preparado con los órganos del carnero²³’, قراشيل *qarāšīl* ‘tercer producto del cernido’ y شوابي *šawābī* ‘sepia.PL’. Finalmente, encontramos la palabra قليبيرة *qal(a)bīra* que se utiliza en este tratado para indicar una parte del cuerpo humano, el ‘cráneo’ más concretamente, manteniendo su significado actual.

Por lo tanto, se puede concluir que uno de los factores que parece motivar la variación lingüística en nuestros manuscritos y en los manuscritos del tratado de al-Saqaṭī, al menos a nivel léxico, está íntimamente relacionado con el contexto sociolingüístico de la obra, entendido aquí no como la aportación lingüística de un grupo reducido de hablantes, sino en un sentido más amplio. Es decir, la diversidad social y lingüística que parece haber existido en los centros urbanos más importantes de al-Ándalus aparece aquí reflejada en las voces de origen romance que exhiben los documentos originales.

3.1.3. Morfología y sintaxis

La morfología y la sintaxis de los dos manuscritos del tratado de ‘Abd al-Ra’ūf también incluyen, como era de esperar, algunos de los rasgos lingüísticos que han sido identificados en textos árabes escritos en variedades no estándares. Estos usos de la lengua, que no se ajustan a la norma prescriptiva del árabe clásico, ocurren, en mayor o menor medida, en alternancia con las formas clásicas y no siempre aparecen conjuntamente en los dos documentos. Entre ellos, cabe mencionar los siguientes:

En algunos casos, se emplea la terminación *-īn* بين para marcar un sustantivo masculino plural en función nominativa en lugar de la flexión clásica *-ūn* -ون *ūn* *ūmar al-qallāyīn* ‘se ordenará a los freidores...’.

23. (Chalmeta, 2019, p. 120, nota 410).

Ocasionalmente, también se mantiene la terminación *-n* del plural masculino en estado constructo *بياعين الجبن bayyā 'in al-ġubn* ‘vendedores de queso’ (ms. M-O-9). El pronombre relativo masculino singular *الذي alladī* ‘(el) que’ puede modificar sustantivos femeninos plurales en ambas copias del tratado como se observa en *تنظيف ساحاتهم الذي يكونون فيها tanzīf sāḥāt(a).F.PL-hum alladī.M.SG yakūnūn(a) fī-hā* ‘limpiar los puestos donde (en que) ejercen su oficio’²⁴.

En la variedad clásica, los verbos concuerdan en género pero no en número cuando preceden a un sujeto plural; sin embargo, la concordancia natural en estos casos, es decir, de género y número, es común en muchas otras variedades de esta lengua como el árabe escrito no estándar y los dialectos modernos. En las copias del tratado de ‘Abd al-Ra’ūf, encontramos algún que otro ejemplo de concordancia natural²⁵: *يطالبون العاملون yuṭālabūn(a).3M.PL al-‘āmilūn* ‘se requerirá a los fabricantes...’²⁶ و *فيسئلوا اهل الرفقة fa-yus‘alū.3M.PL ahl al-rifqa* ‘se preguntará a los viajeros’²⁷ (ms. M-O-9). Las relaciones sintácticas entre sustantivo y adjetivo también muestran usos similares que reflejan la multifuncionalidad y complejidad de la concordancia en árabe.

Según la norma estándar, el adjetivo toma la forma femenina singular cuando modifica a un sustantivo plural no humano; sin embargo, el uso del adjetivo plural no es inusual en muchas otras variedades del árabe. La concordancia plural con adjetivos de la forma morfológica *فعال fī ‘āl*, que suelen denotar atributos permanentes del sustantivo al que modifican, es muy productiva en este aspecto. De hecho, representa un rasgo muy antiguo de la lengua árabe que ya se documenta en fuentes preislámicas (Belnap, 1991; Belnap y Shabaneh, 1992). También se encuentran casos de este tipo de concordancia en árabe andalusí con el adjetivo *كبار kibār* ‘grande-PL’ (Corriente, 2013, p. 170) y en manuscritos levantinos de la época premoderna (1600-1890): *مراكب زغار marākib ziġār* ‘barcos pequeños’ y *حيل ملاح hayl milāḥ* ‘caballos bonitos’ (Lentin, 1997, p. 805). Un claro ejemplo de este tipo de uso semántico no normativo de la concordancia puede apreciarse en el siguiente ejemplo extraído de nuestros textos: *الاعدال الثقال al-a ‘dāl al-tiqāl* ‘las cargas pesadas’.

24. Siguiendo la traducción de Chalmeta (2019, p. 81).

25. En la gramática árabe, este tipo de concordancia se conoce con la expresión *luġatu akalūnī al-barāġīl* o «dialecto de aquellos que dicen ‘me comieron las pulgas’». Aquí la flexión masculina plural del verbo, *akalūnī*, contrasta con la forma femenina singular *akalatnī*, cuyo uso establece la norma del árabe clásico (Corriente, 2008, nota 18).

26. Siguiendo la traducción de Chalmeta (2019, p. 78).

27. Siguiendo la traducción de Chalmeta (2019, p. 65).

En términos de variación lingüística, los patrones de concordancias y el uso de los modos verbales en los dos manuscritos objeto de nuestro estudio son particularmente interesantes, ya que parecen estar motivados por factores semánticos y pragmáticos. A diferencia de las peculiaridades ortográficas y fonéticas descritas en el apartado anterior, la presencia de estas variantes lingüísticas en nuestros textos no parece estar relacionada con prácticas de escritura determinadas ni derivar del contexto dialectal en el que se compuso el texto. Por el contrario, la elección de ciertas variantes morfológicas y sintácticas, como veremos a continuación, introduce matices que expresan significados concretos. Dicho de otro modo, se trata de un tipo de variación que cumple funciones comunicativas específicas. Por ejemplo, la concordancia femenina de adjetivos con la terminación $\text{ة} -a(t)$ se utiliza para indicar la forma singulativa del sustantivo ذهب *dahab* ‘oro’, es decir, para diferenciar su significado incontable del contable. Véase el contraste entre la concordancia adjetival masculina en الذهب الطيب *al-dahab al-ṭayyib*.ADJ.M.SG ‘el oro bueno’ y la femenina en الذهب الناقصة الوزن *al-dahab al-nā-qīṣat*.ADJ.F.SG *al-wazn* ‘las monedas faltas de peso’²⁸.

Otro caso notable se da con el morfema femenino plural $\text{ات} -āt$ que emerge en nuestros textos como una marca de concordancia adjetival con referentes individualizados. En árabe, las nociones de colectividad e individuación han sido identificadas como un factor decisivo en la concordancia con sustantivos plurales, humanos y no humanos, en numerosas variedades orales y escritas²⁹. Los sustantivos cuyo referente es percibido por los hablantes como parte de un todo, esto es, como un colectivo, concuerdan con la forma femenina singular; mientras que los sustantivos plurales más individualizados, aquellos que denotan un referente más específico en función de sus propiedades sintácticas, semánticas y/o discursivas, asignan la concordancia plural.

Por su parte, el árabe estándar moderno establece que los adjetivos que modifican sustantivos plurales no humanos deben tomar la forma femenina singular. En nuestros dos manuscritos, los patrones de concordancia parecen regirse generalmente por esta norma, pero en otros casos siguen los patrones de colectividad e individuación

28. Siguiendo la traducción de Chalmeta (2019, p. 71).

29. Véase, entre otros: Wright (1896–1898, vol. 2, pp. 273–274, 291 y 296–297) para árabe clásico, Cowell (1964, pp. 423–426) para el dialecto sirio de Damasco; Belnap (1991, pp. 61–85) para el dialecto egipcio del Cairo, Belnap (1991, pp. 116–142) y Belnap y Shabaneh (1992) para árabe preislámico, coránico, clásico y estándar moderno; Brustad (2000, pp. 52–62) para los dialectos sirio, egipcio y marroquí; Bettega 2017 para el dialecto omaní, D’Anna (2017) para los dialectos libios de la región de Fezzan, Procházka y Gabsi (2017) para el dialecto urbano de Túnez, Ritt-Benmimoun (2017) para el dialecto tunecino de la región de Nifzāwa y, finalmente, Bettega y D’Anna (2022) para una visión histórica y diacrónica de la concordancia en árabe.

mencionados anteriormente. Tomemos el ejemplo del sustantivo incontable *šuhūm* ‘grasas’ para ilustrar los diferentes tipos de concordancia que puede recibir un mismo sustantivo en árabe. En contextos donde *šuhūm* denota un referente genérico toma la forma femenina singular *الشحوم المحملة الخالصة: -a(t)* *al-šuhūm al-muḥammala(t)*. *ADJ.F.SG al-ḥāliṣa(t)*. *ADJ.F.SG* lit. ‘las grasas repletas puras’ o más idiomáticamente ‘las grasas blancas’³⁰.

Sin embargo, cuando el referente en cuestión representa una entidad específica, es decir, recibe un valor más individualizado y contextualmente relevante, entonces, el adjetivo toma el morfema femenino plural *-āt*. Nótese la discordancia entre el demostrativo femenino singular y el adjetivo femenino plural en el siguiente ejemplo: *هذه الشحوم المحرّمات* *hāḍiḥi*. *DEM.F.SG al-šuhūm al-muḥarramāt*. *ADJ.F.PL* ‘estas grasas prohibidas’³¹. Sintácticamente, el sustantivo *شحوم šuhūm* ‘grasas’ aparece doblemente modificado por el artículo definido invariable *ال* *al* ‘el’ y por el demostrativo *هذه* *hāḍiḥi* ‘esta’, lo cual ya sugiere un importante grado de individuación del mismo. Por otro lado, el adjetivo *محرّم muḥarram* ‘prohibido’ parece ser un factor clave en la activación de la concordancia marcada al tratarse de un concepto religioso bien definido e identificable por los hablantes y cuya carga semántica, como expresión lingüística de la autoridad de *Allāh*, le confiere una relevancia discursiva indudable (Valenzuela Mochón, 2020).

Los modos verbales también parecen adquirir valores marcados en base a una serie de consideraciones de índole semántica y pragmática en los dos manuscritos de ‘Abd al-Ra’ūf. Este tipo de usos responden a lo que Lentin (2008a) considera rasgos de un tercer tipo, propios de las variedades no estándares del árabe escrito. En árabe clásico los modos verbales se expresan morfológicamente a través de una serie de vocales y morfemas que se unen a la forma imperfectiva del verbo. Aquí nos interesan más concretamente dos usos no clásicos de los morfemas de segunda y tercera persona plural: *-ūn(a)* (indicativo) y *-ū* (yusivo).

La terminación verbal *-ūn(a)* ha sido identificada por Lentin (2008a, 2008b) como una variante estilística y/o sintáctica marcada en textos árabes premodernos tanto de las regiones arabófonas orientales como occidentales. Los mss. M-O-9 y d3643 también recogen una construcción en la que el morfema plural del modo indicativo en árabe clásico adquiere un valor marcado relacionado con la expresión de intencionalidad (Valenzuela Mochón, 2020):

30. Siguiendo la traducción de Chalmeta (2019, p. 90).

31. Siguiendo la traducción de Chalmeta (2019, p. 90).

يمنعون ان لا ينتفعوا بثياب الخام حتى تبلى ثم يقصرونها بعد ذلك ويبيعونها فيظن الجاهل انها جديدة

<i>yumna</i> -' <i>ūn(a)</i>	<i>an</i>	<i>lā</i>	<i>yantafi</i> -' <i>ū</i>	<i>bi-tiyāb</i>	<i>al-ḥām</i>	<i>ḥattā</i>
prohibir-IND.3PL	que	NEG	utilizar-SBJV.3PL	con-ropas	DEF-algodón.crudo	hasta
<i>tablā</i>	<i>tumma</i>		<i>yuqaṣṣir-ūn(a)-hā</i>	<i>ba'da</i>	<i>dālika</i>	
desgastar.IND.3SG	luego		limpiar-IND.3PL-la	después	eso	
<i>wa-yabī</i> -' <i>ūn(a)-hā</i>	<i>fa-yazunn(-u/-a)</i>		<i>al-ḡāhil</i>	<i>anna-hā</i>	<i>ḡadida</i>	
y-vender-IND.3PL-la	CONJ-piensa(-IND/-SBJV.3SG)		DEF-ignorante	que-la	nueva	

'Se les prohibirá utilizar ropas de algodón crudo hasta que estén raídas y luego limpiarlas y venderlas, creyéndose el ignorante que son nuevas.'³²

En este ejemplo, la conjunción coordinativa *tumma* '(y) luego/después' introduce el verbo indicativo *yuqaṣṣirūn(a)* 'limpian [ellos]' que sintácticamente, según dicta la variedad estándar, debería tomar la flexión del modo subjuntivo, ya que está en régimen de coordinación con el verbo subjuntivo *yantafi* 'ū' 'utilicen [ellos]'. Por lo tanto, se esperaría que los verbos que aparecen tras la conjunción, *yuqaṣṣirūn(a)* y *yabī* 'ūn(a), tomaran asimismo el modo subjuntivo. Aquí, los verbos en modo indicativo transmiten la intención de un grupo de artesanos de realizar un acto engañoso para cometer un fraude. A modo de precaución, 'Abd al-Ra'ūf dispone que se debe prohibir a los vendedores de telas reutilizar las ropas de algodón para engañar a la gente, haciéndoles creer que son nuevas.

Otra función semántico-pragmática del modo verbal que encontramos en nuestras fuentes consiste en el empleo del yusivo con verbos semánticamente vinculados a éste en contextos que gramaticalmente requieren el indicativo. En árabe clásico y en árabe estándar moderno, el modo yusivo va precedido de una serie de partículas negativas, obligativas y condicionales que rigen su uso. Es decir, está regulado principalmente por factores sintácticos. Sin embargo, en nuestros manuscritos, el morfema plural de tercera persona que indica el modo yusivo -*ū* aparece en al menos tres ocasiones, en lugar del modo indicativo, con dos verbos que expresan obligación, يأمر *ya'mur* 'ordenar' y ينهاي/ينهو *yanhuw/yanhā* 'prohibir', sin que esté introducido por ninguna de las partículas de uso preceptivo en la variedad estándar (Valenzuela Mochón, 2020): ذلك عندها ينهاي *yunhaw* 'an *dālika* 'se les prohibirá eso[estas prácticas]³³ بان يومروا بان يتخذوا *yūmarū bi-an yattahidū a 'wādan* 'se les ordenará que tengan un tajo'³⁴ و بان يتولى الذبح *yūmarū an*

32. Siguiendo la traducción de Chalmeta (2019, p. 71).

33. Siguiendo la traducción de Chalmeta (2019, p. 51).

34. Siguiendo la traducción de Chalmeta (2019, p. 86).

yatawallā al-dabḥ ‘se les ordenará encargar el degüello a’³⁵. En estos casos, el yusivo no estaría justificado de ninguna manera según la norma estándar del árabe escrito ni tampoco se consideraría gramaticalmente correcto. Por el contrario, el empleo del modo yusivo en los dos textos de ‘Abd al-Ra’ūf sugiere que el valor semántico de ciertos verbos también puede activar la elección de este modo en las variedades no estándares del árabe escrito.

4. CONCLUSIÓN

El análisis lingüístico de las dos copias existentes del tratado de ‘Abd al-Ra’ūf, el ms. M-O-9 y el ms. d3463 confirma, que estos textos exhiben un cierto grado de variación lingüística en los planos ortográfico, fonético, léxico, morfológico y sintáctico. Lejos de reflejar usos ocasionales o aislados de la lengua, los ejemplos de variación identificados en nuestro estudio sugieren que la alternancia de formas responde a patrones determinados. En algunos casos, se corresponden con rasgos documentados en las variedades dialectales y no estándares del árabe escrito, p. ej., la omisión de *hamza* en todas las posiciones, el mantenimiento de la terminación *n* en estado constructo, el cambio *t > t* o la influencia del entorno consonántico en la desvelarización de la sibilante *ṣ > s* con lexemas de la raíz *ṣād - bā’ - ḡayn*: تصيبغ *taṣbīḡ* > تسبيغ *tasbīḡ* ‘teñir’. En otras ocasiones, estos usos no normativos reflejan rasgos característicos de las variedades no estándares, como, por ejemplo, el valor marcado de los modos verbales.

Los manuscritos también muestran divergencias en el uso de ciertas formas que sugieren que estos documentos fueron copiados por manos diferentes. Especialmente relevante es la tendencia, reflejada solamente en el manuscrito del Instituto Egipcio, de sustituir la oclusiva dental *d* por la fricativa interdental *ḍ* con la raíz *hamza - dāl - bā’*: *yūaddab* > *yūaḍḍab* [‘él] será castigado’ y de omitir la vocal larga *ā* /ā/ en términos específicos تعالى vs. شيطان *ta ālā* ‘el Altísimo’ o شيطان vs. شيطان *šayṭān* ‘demonio’. Se trata de un indicio prometedor que apunta a que el estudio lingüístico de estas fuentes puede contribuir a elucidar algunos detalles sobre el contexto de creación del manuscrito.

En cuanto a los factores que motivan la variación en nuestros textos todo parece señalar a que existe una relación estrecha entre ellos y el nivel de la lengua en el que se encuentra la variante lingüística. Por ejemplo, a nivel ortográfico y fonético, la aparición

35. Siguiendo la traducción de Chalmeta (2019, p. 87).

de alternancias lingüísticas está determinado por las convenciones ortográficas adoptadas en las variedades no estándares del árabe escrito, así como por el entorno dialectal en el que se transmite o se copia el texto. En cuanto al nivel léxico, la elección de voces de origen romance en los dos manuscritos puede explicarse como consecuencia del contexto histórico y sociolingüístico en el que surge la obra. Finalmente, la variación de algunos aspectos morfológicos y sintácticos de los textos está relacionada con rasgos lingüísticos propios de las variedades no estándares y dialectales, por una parte, y con aspectos de índole semántica y pragmática, por otra. Este último es el caso de las concordancias marcadas con el morfema plural de género femenino y los usos no clásicos del indicativo y el yusivo.

En definitiva, nuestro análisis de las características lingüísticas del tratado de ‘Abd al-Ra’ūf arroja un poco de luz sobre el uso de algunas variantes empleadas en estos documentos y de los factores que intervienen en la expresión de la variación lingüística en los diferentes niveles de la lengua. Pero sin duda alguna, este acercamiento al estudio lingüístico de los documentos árabes de la época premoderna no representa más que un punto de partida en la descripción detallada de estas variedades funcionales del árabe escrito en el Occidente islámico. En futuras investigaciones, se pretende llevar a cabo un análisis exhaustivo de los manuscritos que se han conservado de los otros tres tratados de *hisba* occidentales con el objetivo de ofrecer una visión más esclarecedora y detallada de la variación lingüística en estas fuentes históricas y complementar, de este modo, los estudios ya existentes sobre variedades no estándares del árabe escrito.

AGRADECIMIENTOS

Este trabajo ha sido posible gracias al apoyo recibido por la organización *Social Science Research Council* mediante la beca *International Dissertation Research Fellowship (IDRF)*, financiada por la fundación Andrew W. Mellon Foundation. Asimismo, me gustaría expresar mi agradecimiento a los dos revisores anónimos, cuyas acertadas sugerencias y correcciones han contribuido a mejorar el manuscrito significativamente. Como no podría ser de otro modo, asumo la responsabilidad de cualquier error que pueda aparecer en estas páginas.

REFERENCIAS

- Belnap, K. (1991). *Grammatical agreement variation in Cairene Arabic*. Tesis doctoral. University of Pennsylvania.
- Belnap, K. y O. Shabaneh. (1992). Variable agreement and nonhuman plurals in Classical and Modern Standard Arabic. En E. Broselow, M. Eid y J. McCarthy (Eds.). *Perspectives*

- on Arabic linguistics*, vol. 4, pp. 245–262. Ámsterdam: John Benjamins Publishing Company.
- Berlinches, C. (2016). *El dialecto árabe de Damasco (Siria)*. Zaragoza: Prensas de la Universidad de Zaragoza.
- Bettega, S. (2017). Agreement with plural controllers in Omani Arabic: Preliminary remarks. En S. Bettega y F. Gasparini (Eds.). *Linguistic studies in the Arabian Gulf*, 153–174. Turín: Università di Torino.
- Bettega, S. y L. D'Anna. (2022). *Gender and number agreement in Arabic*. Boston: Brill.
- Blau, J. (1966–1967). *A grammar of Christian Arabic: Based mainly on south Palestinian texts from the first millennium*, vols. 1–3. Louvain: Universitatis Catholicae Americae Universitatis Catholicae Lovaniensis.
- Blau, J. (1981). *The emergence and linguistic background of Judaeo-Arabic: A study of the origins of Middle Arabic*, 2ª ed. Jerusalén: Ben-Zvi Institute for the Study of Jewish Communities in the East.
- Blau, J. (2002). *A handbook of early Middle Arabic*. Jerusalén: The Max Schloessinger Memorial Foundation, The Hebrew University of Jerusalem.
- Bongianino, U. (2017). *The origin and development of Maghribī round scripts. Arabic palaeography in the Islamic West (4th/10th-6th/12th centuries)*. Tesis doctoral. University of Oxford.
- Brustad, K. (2000). *The syntax of spoken Arabic: A comparative study of Moroccan, Egyptian, Syrian, and Kuwaiti dialects*. Washington, D.C.: Georgetown University Press.
- Brustad, K. (2017). Diglossia as ideology. En J. Høigilt y G. Mejdell (Eds.). *The politics of written language in the Arab world*, pp. 41–67. Leiden, Boston: Brill.
- Brustad, K., ed. (en prensa). *A grammar of Arabic*. Londres: Routledge.
- Cahen, Cl., M. Talbi, R. Mantran, A.K.S. y A. S. Bazmee Ansari. (2012). Ḥisba. En P. Bearman, Th. Bianquis, C.E. Bosworth, E. van Donzel, W.P. Heinrichs, *Encyclopaedia of Islam*, 2ª ed.
- Chalmeta, P. (1967). El «Kitāb fī ādāb al-ḥisba (Libro del buen gobierno del zoco) de al-Saqāṭī. *Al-Andalus* 32(1), 359–397.
- Chalmeta, P. (1973). *El «señor del zoco» en España: Edades media y moderna: Contribución al estudio de la historia del mercado*. Madrid: Instituto Hispano-Árabe de Cultura.
- Chalmeta, P. (2010). *El zoco medieval. Contribución al estudio de la historia del mercado*. Almería: Fundación Ibn Tufayl de Estudios Árabes.
- Chalmeta, P. y F. Corriente. (2014). *Al-Saqāṭī al-Mālaqī. El buen gobierno del zoco*. Almería: Fundación Ibn Tufayl de Estudios Árabes.
- Chalmeta, P. (2019). *Ibn 'Abd al-Ra'ūf. Córdoba a mediados del siglo X*. Almería: Fundación Ibn Tufayl de Estudios Árabes.
- Colin, G. S. y É. Lévi-Provençal. (1931). *Un manuel hispanique de ḥisba: Traité d'Abū 'Abd Allāh Muḥammad B. Abū Muḥammad As-Saqāṭī de Malaga*. Paris: Librairie Ernest Leroux.
- Constable, O. R. (1996). *Trade and traders in muslim Spain: The commercial realignment of the Iberian Peninsula (900-1500)*. Cambridge: Cambridge University Press.

- Corriente, F. (1975). Marginalia on Arabic diglossia and evidence thereof in the *Kitab al-Agani*. *Journal of Semitic Studies* 20(1), 38–61.
- Corriente, F. (2013). *A descriptive and comparative grammar of Andalusí Arabic*. Boston: Brill.
- Cowel, M. 1964. *A reference grammar of Syrian Arabic*. Washington, D.C.: Georgetown University Press.
- D'Anna, L. (2017). Agreement with plural controllers in *Fezzānī* Arabic. *Folia Orientalia* 54, 101–122.
- Doss, M. y H. Davies. (2013). *Al-‘āmmiyya al-Misriyya al-Maktūba*. Cairo: GEBO.
- Ferguson, C. A. (1959). Diglossia. *Word* 15(2), 325–340.
doi: <https://doi.org/10.1080/00437956.1959.11659702>
- Ferrando, I. (2007). History of Arabic. En K. Versteegh (Ed.). *Encyclopedia of Arabic Language and Linguistics*, vol. 2, pp. 261–268. Leiden, Boston: Brill.
- Ferrando, I. (2012). Los Estudios sobre el árabe andalusí: Avances y estado de la cuestión. En M. Fierro, J. Martos, J. P. Monferrer y M. J. Viguera (Eds.). *711-1616, de árabes a moriscos: una parte de la historia de España*. Córdoba: Al-Babtain Foundation.
- García Gómez, E. (1957). Unas «Ordenanzas del Zoco» del siglo IX. *Al-Andalus* 22(2), 253–316.
- GrandHenry, J. (2007). Christian Middle Arabic. En K. Versteegh (Ed.). *Encyclopedia of Arabic language and linguistics*, vol. 2, pp. 383–387. Leiden, Boston: Brill.
- Hernández, E., P. Martín Butragueño, R. Barriga, V. Demonte, P. García Mouton, R. Gutiérrez Bravo, E. Herrera Zendejas, I. Pérez-Jiménez y J. Pozas Loyo. (2015). ¿Qué es la variación lingüística? En E. Hernández y P. M. Butragueño (Eds.) *Variación y diversidad lingüística*. México D.F.: El Colegio de México.
- Hopkins, S. (1984). *Studies in the grammar of Early Arabic: Based upon papyri datable to before 300 AH/912 AD*. Oxford: Oxford University Press.
- Khan, G. (2007). Judaeo-Arabic. En K. Versteegh (Ed.). *Encyclopedia of Arabic language and linguistics*, vol. 2, pp. 526–536. Leiden, Boston: Brill.
- Lentin, J. (1997). Recherches sur l'histoire de la langue arabe au Proche-Orient à l'époque moderne. Thèse de doctorat d'état. University of Paris III.
- Lentin, J. (2008b). Unité et diversité du Moyen Arabe au Machreq et au Maghreb. En J. Lentin y J. GrandHenry (Eds.). *Moyen arabe et variétés mixtes de l'arabe à travers l'histoire: actes du premier colloque international (Louvain-la-Neuve, 10–14 mai 2004)*, pp. 305–319. Lovaina: Université Catholique de Louvain.
- Lentin, J. (2008a). Middle Arabic. En K. Versteegh (Ed.). *Encyclopedia of Arabic Language and Linguistics*, vol. 3, pp. 215–224. Leiden, Boston: Brill.
- Lévi-Provençal, É. (1934). Un document sur la vie urbaine et les corps de métiers à Séville au début du XII^e siècle: Le traité d'Ibn ‘Abdūn. *Journal Asiatique* 224, 177–299.
- Lévi-Provençal, É. y E. García Gómez. (1948). *Sevilla a comienzos del siglo XII. El tratado de Ibn ‘Abdūn*. Madrid: Moneda y Crédito.
- Lévi-Provençal, É. (1955). *Trois traités hispaniques de ḥisba*. Cairo: Imprimerie de L'Institut Français D'Archéologie Orientale.
- Mahdi, M. (1984). *The thousand and one nights = Alf Layla wa-Layla: From the earliest known sources*, (vol. 1–2). Leiden: Brill.

- Mejdell, G. (2017). Changing norms, concepts and Practices of written Arabic A 'long Distance' Perspective. En J. Høigilt y G. Mejdell (Eds.). *The politics of written language in the Arab world*, pp. 68–89. Leiden, Boston: Brill.
- Palva, H. (2006). Dialects: classification. En K. Versteegh (Ed.). *Encyclopedia of Arabic language and linguistics*, vol. 1, pp. 604–613. Leiden, Boston: Brill.
- Procházka, S. y I. Gabsi. (2017). Agreement with plural heads in Tunisian Arabic: The Urban North. En V. Ritt-Benmimoun (Ed.). *Tunisian and Lybian Arabic dialects, common trends, recent developments, diachronic aspects*, pp. 239–260. Zaragoza: Prensas de la Universidad de Zaragoza.
- Ritt-Benmimoun, V. (2017). Agreement with plural heads in Tunisian Arabic: The Bedouin south. En V. Ritt-Benmimoun (Ed.). *Tunisian and Lybian Arabic Dialects, Common Trends, Recent Developments, Diachronic Aspects*, pp. 261–287. Zaragoza: Prensas de la Universidad de Zaragoza.
- Schen, I. (1972) Usāma Ibn Munqidh's memoirs: Some further light on muslim Middle Arabic (Part I). *Journal of Semitic Studies* 17(2), 218–236.
- Schen, I. (1973) Usāma Ibn Munqidh's memoirs: some further light on muslim Middle Arabic (Part II). *Journal of Semitic Studies* 18(1), 64–97.
- Valenzuela Mochón, E. (2020). Revisiting the western *hisba* manuscripts: Contribution to a history of written Arabic in al-Andalus and the Maghreb. Tesis doctoral. University of Texas at Austin.
- Valenzuela Mochón, E. (2022). Modality in Arabic: The multiple functions of the (non)-indicative markers *-ūn* and *b-*. *Languages* 7(272).
DOI: <https://www.mdpi.com/2226-471X/7/4/272>
- Van den Boogert, N. (1989). Some notes on Maghribi script. *Manuscripts of the Middle East* 4, pp. 30–43.
- Versteegh, K. (2014). *The Arabic Language*, 2^a ed. Edinburgo: Edinburgh University Press.
- Wagner, E. M. (2010). *Linguistic variety of Judaeo-Arabic in letters from the Cairo Genizah*. Leiden: Brill.
- Wright, W. (1896–1898). *A grammar of the Arabic language, translated from the German of Caspari, and edited with numerous additions and corrections*, 3rd ed. Cambridge: Cambridge University Press.

Estefanía Valenzuela Mochón
Escuela de Estudios Árabes, CSIC
Cuesta del Chapiz, 22
Granada 18010

WHAT IS WHITE ARABIC? NEW LABELS IN A CHANGING ARAB WORLD¹

Ana IRIARTE DÍEZ²
University of Vienna

Claudia LAABER³
University of Vienna

Nina VAN KAMPEN⁴
INALCO, CERMON associated with IFPO

Montserrat BENÍTEZ FERNÁNDEZ⁵
Escuela de Estudios Árabes, CSIC

Abstract

Arabic has been traditionally described as a canonically diglossic language (Ferguson, 1959), with Standard Arabic (SA) as the high variety and spoken vernaculars as low varieties. Further research has proven that the actual linguistic landscape in Arabic speaking countries does not reflect this dichotomy, but rather a layered continuum where different varieties (and sometimes languages) interact, fulfilling different communicative functions and carrying multiple symbolic values.

In this sea of varieties, the metalinguistic label «White Arabic» has gained prominence in the last decade, coinciding with the emergence of an increasingly interconnected Arab world. Although the notion of WA was treated peripherally in previous scientific studies (Al-Rojaie, 2020; Dufour, 2008; Germanos, 2009; O'Neill, 2017), none of these deal with the term as a main object. There seems to be no clear consensus regarding the definition of White Arabic – data hints at different understandings of the concept in Lebanon, Jordan,

1. We owe a huge debt of gratitude to Stephan Procházka, Mahmoud Al-Batal and Kristen Brustad for their valuable comments on earlier versions of this manuscript. Also, we sincerely thank Julie Haslé for participating in the discussions that shaped this article and Montserrat Benítez Fernández for making this collaboration possible in the frame of her research project *Arabic Sociolinguistics and Linguistic Anthropology* (LINKB20056), funded by CSIC. All the remaining errors are solely the authors' responsibility.

2. ana.iriarte.diez@univie.ac.at.  <https://orcid.org/0000-0002-7014-6713>

3. claudia.laaber@univie.ac.at.  <https://orcid.org/0000-0003-1259-8292>

4. ninavankampen@inalco.fr.  <https://orcid.org/0000-0003-4205-2282>

5. montsebenitez@eea.csic.es.  <https://orcid.org/0000-0002-8103-1428>

UAE, Saudi Arabia, Yemen, Egypt, Tunisia, Algeria and Morocco. Thus, this study aims at exploring the speakers' understandings and perceptions of this notion by analyzing metalinguistic comments made by speakers in qualitative interviews, as well as in media (podcasts, newspapers, blogs, etc.) and social media platforms (facebook, youtube, twitter, etc.). The data is complemented by findings from qualitative questionnaires conducted among speakers from five of the abovementioned Arabic speaking countries.

Keywords: White Arabic; standardization; koineization; accommodation; diglossia; Arabic high varieties; Arabic non standard varieties

¿QUÉ ES EL ÁRABE BLANCO?

ETIQUETAS NUEVAS EN UN MUNDO ÁRABE CAMBIANTE

Resumen

El árabe ha sido descrito tradicionalmente como uno de los ejemplos canónicos de lenguas afectadas por el fenómeno de la diglosia (Ferguson, 1959), con el árabe estándar actuando como variedad alta y las variedades vernáculos habladas como bajas. Sin embargo, investigaciones más recientes han demostrado que la situación lingüística actual de los países arabófonos no refleja esta dicotomía, sino, más bien, un continuo estratificado en el que diferentes variedades –y a veces lenguas– interactúan cumpliendo diferentes funciones comunicativas y portando múltiples valores simbólicos.

En este mar de variedades, la etiqueta metalingüística «White Arabic» ('árabe blanco', a partir de ahora WA) ha ganado importancia en la última década, coincidiendo con el aumento de interconexión en el mundo árabe. Aunque la noción de WA ha sido tratada de forma tangencial en investigaciones previas (Al-Rojaie, 2020; Dufour, 2008; Germanos, 2009; O'Neill, 2017), ninguna de ellas trata la cuestión como objeto principal de estudio y no parece haber un consenso claro en la definición del término. De hecho, los datos apuntan hacia diferentes formas de entender el concepto en Líbano, Jordania, Emiratos Árabes Unidos, Arabia Saudí, Yemen, Egipto, Túnez, Argelia y Marruecos.

Por tanto, el objetivo de este estudio consiste en explorar cómo entienden y perciben los hablantes esta noción. Para ello, se ha llevado a cabo un análisis metalingüístico de entrevistas, intervenciones y comentarios realizados por hablantes nativos en medios de comunicación tradicional (periódicos y revistas) y en línea (podcasts, blogs, vídeos, etc.) y en plataformas de redes sociales (facebook, youtube, twitter, etc.). Estos datos se han complementado con los resultados obtenidos mediante el análisis de cuestionarios cualitativos distribuidos en línea entre hablantes de cinco de los países árabes mencionados anteriormente.

Palabras clave: Árabe blanco; estandarización; koinización; acomodación; diglosia; variedades altas del árabe; variedades no estándares del árabe

RECIBIDO: 31/10/2022

APROBADO: 02/06/2023

1. INTRODUCTION

Arabic has been traditionally described as a canonically diglossic language (Ferguson, 1959) with Standard Arabic (SA) as the high variety and spoken vernaculars as low varieties⁶. Further research has proven that the actual linguistic landscape in Arabic speaking countries does not reflect this dichotomy, but rather a layered continuum where different varieties (and sometimes languages) interact, fulfilling different communicative functions and carrying multiple symbolic values (e.g., Badawi, 1973; Brustad, 2017; Mejdell, 2017, etc.).

In this sea of varieties, the metalinguistic label «White Arabic» (*al-luġa al-bayḍāʾ*? [‘The White Language’] or *al-lahġa al-bayḍāʾ*? [‘The White Dialect’])⁷ has gained prominence in the last decade, coinciding with the emergence of an increasingly interconnected Arab world that bears witness to the appearance of superdiverse communicative settings. Although White Arabic (WA) is generally identified to be used among speakers of different Arabic dialects, there seems to be no clear consensus regarding its linguistic characterization – data hints at different understandings of the concept in Lebanon, Jordan, UAE, Saudi Arabia, Yemen, Tunisia and Algeria.

Mentions of the notion of White Arabic appear scarcely and unevenly distributed along the literature across different disciplines dealing directly or indirectly with the Arabic language. However, although WA has been treated peripherally in a number of previous scientific studies (e.g., Al-Rojaie, 2020; Dufour, 2008; Germanos, 2009; O’Neill, 2017, etc.), none of these works discusses the term as a main object of study. Besides relying on a scarce body of literature, the current state-of-the-art on White Arabic can be defined as rather disorienting. This is, partly, because the label of WA does not seem to correlate with a specific fixed set of linguistic

6. We use «Standard Arabic» (SA) as a wide label which includes «Modern Standard Arabic» (MSA), «Classical Arabic» (CA), and the «Arabic» term *fuṣḥā*. Similarly, we use *ʿāmmiyya* (general term for vernaculars), spoken varieties and dialects interchangeably to refer to Arabic vernaculars.

7. In our data, speakers use both *al-luġa* (lit. language) and *al-lahġa* (lit. dialect, variety) qualified by the adjective «white». Although the use of these two labels in our data does not consistently correlate to any specific characterization of WA, further study would be necessary to establish a potential difference in the use of these two metalinguistic labels. For the purpose of this article, we use the term «White Arabic», encompassing both terms.

1.1. *White Arabic: a new label or a new variety?*

Mentions to WA in the academic literature suggest that WA is considered to be relatively new. In Yemen, Dufour (2008, p.141) estimates that it must have emerged around the beginning of the 21st century, a fact that seems parallel to the situation in Saudi Arabia, where young respondents are reported to be more familiar with the concept (Al-Rojaie 2020, p.42). Wikipedia also affirms that youth expressions (*at-ta'ābīr al-šabābiyya*) are part of WA, and that the appearance of this concept is directly related to the emergence of social media and globalization (Wikipedia: لهجة بيضاء [*lahǧa bayḍā?*]). However, the literature also hints at that the concept is not exclusively known among young generations. Germanos' work (2009) registered mentions to WA within the qualitative interviews she conducted in Beirut in 2005, some of them made by a 72-year-old informant (Germanos, 2009, p. 104-5), which could suggest that the use of this term would not be so recent, at least not in the Lebanese context.

In fact, the idea of one variety that can be understood by any speaker of Arabic has long been an ideological motif among Arabic speakers. SA was often portrayed to fulfill the function of interdialectal communication among speakers from different Arab countries, and while that could partially be a reality among educated speakers in the written sphere, it surely did not match the reality of interdialectal oral interactions (e.g., Abu-Melhim, 1992; S'hiri, 2002; Chakrani, 2015; Bassiouney, 2015; Attwa, 2019; Soliman, 2015; Schulties, 2015).

The early decades of the 20th century witnessed the emergence of similar meta-linguistic labels such as «the third language» (*al-luǧa al-tālīṭa*) and «the middle language» (*al-luǧa al-wuṣṭā*), mainly by the hand of writers and intellectuals such as Tawfiq al Hakim, Taha Hussein and Naguib Mahfouz among others. «The intermediate» Arabic they would use was generally understood as «a type of style in written prose which, though adhering to the basic norms of classical Arabic, is easily understood by any speaker of Arabic, and is not far removed from the vocabulary, structure, and rhythm of spoken dialects» (Somekh, 1981, p. 74). In fact, during the second half of the 20th century many labels to refer to «intermediate forms of the language» (Ferguson, 1959, p. 332) appeared. To name a few: *luǧa ʕarabiyya mutawassiṭa* or «Intermediary Arabic» (ʕUbayd 1964), *ʕarabiyyat al-muṭaqqafīn* or «Educated Arabic» (Badawi 1973) *fuṣṣḥāmmiyya* or «standardialect» (Rosenbaum 2000), *al-ʕarabiyya al-muyassara* or «simplified Arabic» (Frayha 1953), *al-fuṣṣḥā al-muxaffafa* or «lightened *fuṣṣḥā*» (Farah Antun in Badawi, 1973, p. 69-70), *Arabe Marocain Médian* (Youssi, 1986), *Arabe*

Marocain Moderne (Youssi, 1992), and «Educated Spoken Arabic» (Mitchell, 1986) among others.

Despite not being identical, these labels generally designate, functionally speaking, «a pragmatic, communicative response of educated speakers to a situation where the basic functional distribution of High (H) and Low (L) variety, *fushā* and *ʿāmmiyya*, Standard Arabic and the vernacular, is challenged, and therefore code choice is diffuse» (Mejdell, 2011). We may then, for the purpose of this article, conclude that a myriad of metalinguistic labels were created to refer to an intermediate point (or range) within the spectrum between the two poles of *fushā* and *ʿāmmiyya* used by speakers to adapt their registers when facing semi-formal or formal interactions.

«White Arabic» seems to be different from the aforementioned labels in this aspect. Although the label WA may also cover the aforementioned notion of «intermediary stage» between *fushā* and *ʿāmmiyya* (see section 3.1.), our data indicates that its use is by no means restricted to it. As we will see, WA is a versatile label that is also often used to refer to other codes, practices or varieties resulting from different processes of accommodation taking place within different communicative settings (see sections 3.2. and 3.3.). We find the «new»-and-versatile nature of the WA label a matter not only worthy of research but also indicative of a change of landscape that the Arabic language is currently witnessing.

The goal of this article is therefore twofold. First, we attempt to describe the versatility of WA's label through three working definitions that group the perceptions of WA reported in our data according to well-established and relevant sociolinguistic theories. Secondly, we attempt to understand why, despite the myriad of metalinguistic labels available in the literature, there is still a need among some Arabic speakers for a new label that designates a neutral, simple form of Arabic that any Arab can understand. For this purpose, in the discussion section (see section 4 in this article) we engage in a debate of the plausible explanations and potential implications of the emergence and seemingly fast spread of the metalinguistic label of White Arabic.

2. THEORETICAL FRAMEWORK AND METHODOLOGY

The present work is situated on the interface of different sociolinguistic and sociological theories and concepts. As we mentioned in the introduction, WA is a complex linguistic concept, which emerges mainly from Arabic native speakers'

perceptions to describe current developments in their linguistic realities. This study therefore generally draws on perceptual dialectology, also known as folk linguistics – a discipline that investigates speakers' beliefs, feelings and perceptions of a certain language or variety (Albury, 2014, p. 86-87). Analyzing non-linguists' beliefs and views on language is crucial for the understanding of language change, as these beliefs and views highly influence actual linguistic behavior (Preston, 2017). This work also benefits from already well-established definitions of sociolinguistic concepts and processes such as those of koineization (e.g., Miller, 2011), leveling (e.g., Trudgill, 1986), accommodation (e.g., Giles & O'Garra, 2007), diglossia (e.g., Ferguson, 1959), etc., which are treated in more detail in the following sections of this article.

Notwithstanding the usefulness of the aforementioned concepts and theories, given the intrinsic complexity and dynamicity of the notion of WA, we found it necessary to complement these with alternative theoretical notions, mainly anchored in sociology and communication theory. An underlying theme and notion that helps capture the complexity of the communicative settings in which WA is reported to be used is that of «superdiversity».⁹ The concept of «superdiversity» recognizes the importance of considering multiple variables when addressing the complexity of our modern society, which is heavily influenced by globalization and the resulting migration movements. The notion of superdiversity, along with that of linguistic repertoires, were mainly developed within studies dealing with mobility, multiethnic and multilingual urban districts (Jørgensen *et al.*, 2011; Blommaert & Backus, 2011; Blommaert & Rampton, 2011, 2016) and have been increasingly used in sociolinguistic research investigating the emergence of complex communicative settings resulting from globalization (e.g., Creese & Blackledge, 2018).

As for the data, our study considers several kinds of sources dealing with the notion of WA. On the one hand, we have collected and analyzed characterizations of WA as they appear in academic works in the fields of Arabic (socio)linguistics, media and communication studies. On the other hand, we have analyzed non-academic sources (i.e., articles, essays or posts published on websites, online newspapers, social media and podcasts, personal communications) where the term «White Arabic» is used mostly by non-linguist native speakers, either explaining what it is, and/or how, when, and why it is employed.

9. This term was coined by Vertovec (2007) in a sociological study on immigrant communities in London. In this study the author pled for the need to avoid focusing exclusively on ethnicity as a variable, given the heterogeneity characterizing the communities under study.

Our data analysis intends to systematically categorize the strongly varying perceptions on the nature of WA. To do so, we grouped these perceptions into three different working definitions of WA, whose purpose is to serve readers as an analytical framework and help them grasp the manifold idiosyncrasies of WA according to the speakers' perceptions.

3. ANALYSIS AND RESULTS

The current section presents three working definitions that try to synthesize, in an organized manner, speakers' various perceptions on WA: (1) Educated Spoken Arabic – an intermediate form between *fushhā* and *ʿāmmiyya*; (2) Dialectal urban koine – a developing national spoken standard; and (3) a set of dialectal accommodation practices used in pan-Arab communication.

Some remarks about the data are due before we delve into the individual definitions. First, we must keep in mind that speakers' perceptions on WA, when reported, show a high variability and may overlap considerably, even within one single speaker. Identifying and describing WA is neither something all speakers can do – many speakers do not identify WA as a label and seem not to have reflected on its status, nature or properties – nor something those speakers who do may find an easy task. The proof is that, as Germanos (2009, p. 105) noticed, and as we have also observed in our data, speakers tend to define WA by what it is not rather than by what it is.¹⁰ For these reasons, the three working definitions we explain in the following sections should not be treated as rigid categories, but rather as flexible descriptive tendencies of what WA represents in the mind of Arabic speakers.

Secondly, we have noticed that, regardless of the definition they align with, speakers' perceptions on the status of WA are by no means consensual or rigid, but rather, it ranges on a spectrum between «fluid sets of practices» and «a stable variety». That WA is often perceived as a nuanced, dynamic, and continuous process is evident in the description of some speakers of their use of WA through the verb تَبَيُّض [tabayyaḍ] 'to grow or become white' – e.g., in affirmations such as: «My language gets whiter the closer I get to Damascus».¹¹ This indicates that the perceptual status of WA is, thus, similar to that of other perceived varieties,

10. «It seems to be easier for speakers to notice the absence of specific salient features than to provide a list of features that is actually used while speaking WA» (Germanos 200, p. 105).

11. Arabic original: «luḡti btatbayyaḍ kəll-ma bʿarrəb ʿa-š-šām». Personal communication from a 26-year-old speaker from Swēda (Syria) who studied and resided in Damascus and later in Beirut.

such as *luġa wuṣṭā* ‘the middle language’, which is also treated by some as a separate (though variable) variety, and by others as a dynamic product of constant mixing of elements from different varieties (Mejdell, 2011).

In this light, we provide the reader with the three working definitions of WA that resulted from our analysis. By shedding light on their distinct yet overlapping nature, we hope to contribute, however slightly, to the already vivid debate on these (and other) sociolinguistic concepts.

3.1. White Arabic as Educated Spoken Arabic

In the first working definition of WA emerging from our data, WA is seen as a hybrid form of Arabic arising from a diglossic situation. In fact, as we have mentioned above, a number of terms (e.g., *luġa wuṣṭā*, *al-luġa al-tālīya*, *ʿāmmiyyat al-muṭaqqafīn*) have previously been used by linguists, academics and writers to describe intermediate mixed varieties between SA and Arabic spoken vernaculars. The term «Educated Spoken Arabic» (ESA) (Mitchell, 1986) refers to a higher-register, «mixed» spoken form of speech used by educated Arabic speakers. This section will show how these terms, and more specifically the label of «Educated Spoken Arabic», are homologous to this working definition of WA.

WA is frequently defined in our sources as a mix between *fuṣṣḥā* and local vernaculars (Wikipedia لهجة بيضاء [*lahġa bayḍāʾ*?], Abdel Nasser, 2017; Abdel Hamid, 2015). In an online article from the UAE newspaper *Al-Bayan* dealing with journalists’, writers’ and scholars’ attitudes towards WA, it is also referred to as «the daughter of *fuṣṣḥā*» (Abdel Hamid, 2015).¹² Similarly, in an online article of the magazine *Al-Majalla*, WA is defined as *fuṣṣḥā muxaffafa*, ‘lightened’ or ‘diluted’ *fuṣṣḥā*, fused with colloquial terms known in most colloquial dialects» (Al-Felou, 2020).¹³

A particularly interesting source for the exploration of speakers’ perceptions on WA is a video entitled *The Unity of the Arabic Language* by Dr. Alexander Argüelles, professor at the American University in Dubai at the time (henceforth Argüelles 2018). In this video, the professor discusses the topic of Arabic diglossia with two female students from Algeria and Syria, who, despite their different origins, also describe WA

12. Arabic original: «تعتبر البيضاء ابنتها» [*tuṣṭabar al-bayḍāʾ ibnata-hā*].

13. Arabic original: «المختفة المدموجة بمصطلحات عامية معروفة عند أغلب اللهجات العامية الفصحى» [*al-muxaffafatu l-madmūġa bi-muṣṭalaḥātīn ʿāmmiyyatin maʿrūfatin ʿinda ʾaġlabi l-lahġāti l-ʿāmmiyyati al-fuṣṣḥā*].

consistently as a «mix of *fushā* and vernacular Arabic». ¹⁴ This understanding of WA has been described in studies dealing with the UAE (Hopkyns *et al.*, 2021, p. 165) and Jordan (Alfaisal & Aljanada, 2019, p. 110).

While previously mentioned metalinguistic labels (i.e., *luġa wustā*, *al-luġa al-tālita*, *ġāmmiyyat al-muṭaqqafīn*) referred, at least partly, to written language, Educated Spoken Arabic, as its name indicates, refers almost exclusively to the spoken realm. According to our analysis, our sources on WA seem to echo the idea that WA is, also, mainly a spoken form. It is not a surprise, then, that all the above-mentioned articles (Wikipedia لهجة بيضاء [*lahġa bayḍā*?]; Abdel Hamid, 2015; Abdel Nasser, 2017; Al-Felou, 2020) associate WA to speech-related terms such as «كلام» [*kalām*] ('speech'), «يتحدث» [*yataḥaddat*] ('to speak'), «ينطق / يلفظ» [*yalfiḥ / yantiq*] ('to pronounce'), among others. Hopkyns *et al.* (2021, p. 165) explicitly categorize WA as spoken and informal, as does Argüelles in the abovementioned video when he says: «Are there books written in WA? No, all books are written in [standard] Arabic» ¹⁵. As we will see below, the sources that associate WA to the domain of the media also seem to restrict their analysis to spoken media forms. ¹⁶

In fact, the data associating WA with written language is very scarce. Abdel Nasser (2017) briefly mentions the use of WA on online forums and in reactions of media platforms to readers. In Abdel Hamid (2015), an Emirati poet affirms that «despite his love for *fushā*, in which he wrote 70% of his poems, he devoted the remaining 30% to *Nabaṭi* poetry, which he wrote in WA» ¹⁷. It is thus worth noting WA is only associated with written language either in informal written social media, which is

14. Arabic original:

Sana: «اللغة البيضاء هي عبارة عن القليل من اللغة الفصحى مع القليل من اللغة العامية المستخدمة يوميًا». [*l-luġa l-bayḍā? hiyya ʿibāra ʿan l-qalīl min al-luġa l-fuṣṣḥā maʿa l-qalīl min al-luġa l-ʿāmmiyya l-mustaxdama yawmiyyan*].

and Fatīḥa: «هذه هي اللغة البيضاء، يعني هي مزيج لهجة مع العربية الفصحى». [*hādīhi l-luġa l-bayḍā?, yaʿnī hiyya maziġ lahġa maʿa l-ʿarabiyya l-fuṣṣḥā*].

15. Arabic original: «هل موجود كتب باللغة البيضاء؟ لا، كل كتاب في اللغة العربية. لا يوجد كتب». [*hal mauġūd kutub bi-l-luġa l-bayḍā? lā. kull kitāb fī l-luġa l-ʿarabiyya. lā yūġad kutub*].

16. The use of WA in written newspapers is, for example, never mentioned in our data.

17. Arabic original:

الشاعر عبد الله الهدية رغم عشقه للفصحى التي كتب بها 70% من قصائده، إلا أنه خصص الـ30% الباقية للشعر النبطي الذي كتبه باللهجة البيضاء، وأشار إلى أن البيضاء هي لغة السهل الممتنع التي يستعين بها الإعلام المكتوب إذ إنها تتبعد عن المفردات الفصحية التي قد يعتبرها العامة معقدة وغير مفهومة.

[*l-šāʿiru ʿabd allah l-hadiyya raġma ʿaṣqi-hi li-l-fuṣṣḥā allatī kataba bi-hā 70% min qaṣāʾi-di-hi, ʿilla ʿanna-hu xaṣṣa al-30% l-bāqiyya li-l-šīʿri l-nabaṭī alladī kataba-hu bi-l-lahġati l-bayḍā?, wa-ʿaṣāra ʿilā ʿanna al-bayḍā? hiyya luġatu l-sahl l-mumtaniʿi allatī yastaʿīn bi-hā l-ʿilāmu l-maktūbu ʿid ʿinna-hā tabtaʿidu ʿan l-mufradāti l-faṣṭiḥati allatī qad yuṣṭabaru-hā l-ʿāmmatu muṣaqqadatan wa-ġayr mafhūma*].

often perceived as a prolongation of spoken communication (online forums), or in poetry, a written genre also intimately related to spoken styles, whose aesthetics relies mainly on auditive features like meter, rhythm and rhyme.

As for the contexts in which WA as a hybrid standard/vernacular form is perceived to be used, Arabic spoken media – radio and television – (Argüelles, 2018¹⁸; Abdel Hamid, 2015¹⁹; Abdel Nasser, 2017²⁰) and advertising (Al-Felou, 2020²¹) resonate strongly as the main contexts for the use of WA. Here again, this understanding of WA matches closely with the notion of ESA, which «in recent usage [...] refers more and more frequently to unscripted spoken Arabic used in the Arabic broadcast media [...] in interviews and in spontaneous commentary situations» (Ryding, 2011). In Germanos (2009), two Lebanese informants also link the use of WA specifically to the Arab broadcast media in Lebanon – in advertisements or used by presenters.²² This claim is supported by a third informant, a director at a Lebanese television channel who reported to have trained presenters

18. Sana: «Today, there is the «white language», which is used a lot in the media».

Arabic original: «يوجد اللغة البيضاء اليوم، نستخدمها كثيرا في الإعلام». [yūğadu l-luğa l-baydā? l-yawm, na-staxdimu-hā kaṭīran fī l-?iṣlām].

19. «I use WA in my daily radio program «live Broadcast», which is broadcasted by Noor Dubai».

Arabic original: «أستخدم البيضاء في برنامجي الإذاعي اليومي (البث المباشر) الذي يذاع من قناة نور دبي». [?astaxdim al-baydā? fī barnāmiğ-i l-?idāṣī l-yawmī (al-baṭṭ al-mubāšir) alladī yuḍā? min qanā' nūr dubay] and «I often use WA to ensure that a wider segment of the audience can follow».

Arabic original: «في كثير من الأحيان أستخدم اللهجة البيضاء لضمان متابعة شريحة أعرض من الجمهور المتابعين». [fī kaṭīr mina l-?aḥyān ?astaxdimu l-lahğa l-baydā? li-ḍamān mutāba'at' šarīḥa ?aṣrağ mina l-ğumhūrī l-mutābišīn].

20. «[WA] can be used in many audio(visual) media, such as in television interviews, especially ones with artists, online forums, songs, reactions to readers, press interviews, etcetera».

Arabic original:

وقد تستعمل في العديد من وسائل الإعلام والاستماع والاطلاع مثل المقابلات التلفزيونية وخاصة الفنية، المنتديات من خلال «الانترنت، الأغاني، الرودود على القراء والمقابلات الصحفية وغيرها»

[wa-qad tusta?mal fī l-?adīdī min wasā?li l-?iṣlām wa-l-istimā? wa-l-iṭṭilā? miṭla l-muqābalāt l-tilīfīzyūniyya wa-xāṣṣatan al-fanniyya, l-muntadayāt min xilāl al-intirnit, al-?ağānī, l-rudūd ?alā l-qirā? wa-l-muqābalāti l-šaḥāfiyya wa-ğayri-hā].

21. «When a company speaks to thousands of people from different regions, which vernacular would they choose? And what would that [choice] make users of other vernaculars feel? The sensible solution that companies resort to is the use of the 'white vernacular'».

Arabic original:

وتطرح (...) هنا إشكالية بصطدم بها المحتوى الإعلاني العلمي، بالقول: «عندما تتحدث شركة لآلاف الأشخاص من مناطق مختلفة، فأى لهجة عامية ستختار؟ وماذا سيُشعر ذلك مستخدمى اللهجات الأخرى؟»، أما الحل المعقول الذي تلجأ إليه الشركات، فهو: استخدام «اللهجة البيضاء».

[Wa-tuṭraḥu (...) hunā ?iškāliyyatu yaṣṭadimu bi-hā l-muḥtawā l-?iṣlāmī l-?āmmī, bi-l-qaw l-?indamā tataḥaddatu šarikatun li-?alāfi l-?ašxāš min manāṭiqin muxtalifa, fa-?ayya lahğatan ?āmmiyyatan sa-taxtār? wa-mādā li-?uṣṣā? l-?iṣlāmī l-lahğati l-?uxrā? ?amma l-ḥallu l-maṣ?ūlu alladī tağā? Pilay-hi l-šarikāt fa-huwwa stixḍāmu l-lahğati l-baydā?].

22. French original: «Chez deux des informateurs qui parlent de *lāhze bayda* (F30M et F39S), cette expression est utilisée pour décrire non pas un vernaculaire, mais plutôt une variété dont l'usage est lié

of the channel in using WA, which he defined «as a mixed variety of Arabic (standard/vernacular), devoid of any local dialectal features that creates in the audience a feeling of «distance» between them and the presenter»²³ (Germanos, 2009, p. 103-4, our translation).

An especially relevant development that seems to be related to the ‘emergence’ of WA is the increasing supra-national orientation of the Arab media industry and the emergence of pan-Arab satellite channels. Kraidy (2006) notes that national and local programs have been challenged by Arab satellite channels tending to produce programs that appeal to viewers from all over the Arab world. It is in this context that the author describes the emergence of «what is now known as «white Arabic», a media compatible, simplified version of Modern Standard Arabic that is becoming a *lingua franca* for regional public discourse» (Kraidy, 2006, p. 11). Schulthies’ study (2015), which explores linguistic practices in pan-Arab talent shows, echoes this definition of WA, broadening the scope of its use also to the advertising industry (Schulthies, 2015, p. 61).²⁴

As reported by speakers, the emergent use of WA as ESA can be explained on the grounds that (1) WA is perceived to be easily understandable by all speakers of Arabic, (2) WA is perceived to be less «heavy» and therefore more attractive than *fushā*, whose reportedly «complicated» nature mainly stems from the marking of mood and case endings (Argüelles, 2018, min. 14:18)²⁵, and that (3) WA, as an intermediary/mixed variety, integrates «the best of both worlds» when it comes

à un contexte bien spécifique (télévisuel ou radiophonique: utilisation dans la publicité chez F30M, et par les présentateurs, chez F29S)» (Germanos, 2009, pp. 103-4).

23. French original: «une variété d’arabe ‘mixte’ (arabe standard/arabe dialectal) dénuée de tout trait dialectal local qui engendrait chez l’auditeur ou le téléspectateur un sentiment de ‘distance’ ou plutôt ‘d’étrangeté’ vis-à-vis du présentateur» (Germanos, 2009, pp. 103-4).

24. It should be noted that, although in Schulthies (2015) speakers’ perceptions may portray MSA or WA as a sort of ‘light MSA’ to be the norm for interaction within these talent shows, a linguistic analysis of these interactions actually show that accommodation often does not necessarily take place in the direction of MSA, but rather through a variety of strategies such as the subtitling of lesser understood vernacular varieties, or accommodation via Mashreqi media forms. In this point, as well, we observe considerable overlap with our third working definition of WA as a set of accommodative strategies for pan-Arab communication.

25. «When we’re sitting with friends, for example, we don’t speak *fushā* because it’s heavy. Unfortunately, nowadays a lot of people can’t speak *fushā* because of the mood and case endings, and the grammatical rules, and a lot of grammatical rules are linked to the mood and case endings. That’s why it’s hard to speak *fushā*, and that’s why WA was created». Arabic original:

عندما نجلس مع الأصدقاء مثلاً، لا نتكلم الفصحى لأنها ثقيلة، للأسف في يومنا هذا، نسبة كبيرة من الناس لا يستطيعون التكلم بالفصحى بسبب التشكيل والقواعد، هناك العديد من القواعد المرتبطة بذلك. لذلك يصعب التحدث بها. لذلك قمنا بإيجاد اللغة البيضاء.

[ʕīndamā naḡlisu maʕa l-ʔaʕdiqā? maʕalan, lā natakalamu l-fuṣḥā liʔanna-hā ʔaqīla, li-l-ʔasaf fi yawmi-nā hādā, nisba kabīra min al-nās lā yastaʕīnūna l-takallum bi-l-fuṣḥā bi-sababi l-taškīl

to the tone and symbolic value of code choice, being perceived as a variety that is adequate in register, yet suitable «to approach the audience in a simple language that is closer to theirs»²⁶ (Al-Felou, 2020). These motivations also explain WA's growing popularity among media and advertisement producers, who, via WA, do not only aim to avoid the reported «slight barrier in the psyche of the viewer» generated by the use of *fushā* in these contexts,²⁷ but also score higher ratings and obtain more profit as they cater to a broader audience throughout the Arab world.

3.2. *White Arabic as an urban dialectal koine (and/or a developing national standard)*

A (dialectal) koine is generally defined as the stabilized mixed variety resulting from koineization, i.e., a process of interdialectal contact leading to an amount of linguistic restructuring (Miller, 2011). The process of koineization consists of the mixing and subsequent leveling of features of varieties which are mutually intelligible, such as regional or social dialects (Siegel, 2001, in Kerswill, 2003). One of the main characteristics of leveling is the removal or reduction of marked forms from dialects over a period of time (Trudgill, 1986, p. 98). By «marked» here we, following Trudgill, refer to features that are used by a lesser number of speakers or whose use is restricted regionally.²⁸

In our data, WA is often described as a koine (or as a variety under the process of koineization) for it is perceived as a neutral variety which does not show regional or marked features. Sources from Saudi Arabia, for instance, describe WA as being «less affected by tribal or rural vocabulary» (Alfaisal & Aljanada, 2019, pp. 109-110). In the same line, a Saudi female schoolteacher exemplifies this point by

wa-l-qawāf'id, hunāka l-ṣadīd min al-qawāf'idi l-murtabiṭa bi-dalik. Li-dalik qumnā bi-ṭiḡād al-luḡa al-bayḏā?

26. Arabic original: أعتقد أنه من اللطيف أحياناً الدمج بين العامية والفصحى. فللعامية وقع حميم ومؤثر، لكن ليس بإمكانها أن تحل محل الفصحى بشكل كامل (...). والهدف من وراء اللهجة البيضاء، كما أعتقد، هو التقرب من الجمهور بلغة بسّية أقرب للغة.

[ʔaʕtaqidu ʔanna-hu min al-laṭīfi ʔaḥyānan al-damḡu bayna l-ʕāmmiyyati wa-l-fuṣṣā. fa-li-l-ʕāmmiyyati waqīʕun ḥamīmun wa-muʔattirun, lākin laysa bi-ʔimkāni-hā ʔan taḥilla mahall al-fuṣṣā bi-ṣaklin kāmilin wa-l-hadafu min warāʔa l-lahḡati l-bayḏā?, ka-mā ʔaʕtaqid, huwwa l-taqarrub min al-ḡumhūri bi-luḡatin baṣṭatin ʔaqrab li-luḡati-hi].

27. Arabic original: «هذا يخلق حاجزاً بسيطاً في نفسية المشاهد، قد لا تحقق من ورائه الوصول لهدف رسالتك» [hādā yaxlaq ḥāḡizan baṣṭan fī nafsīyyati l-muṣāhidi, qad lā tuḥaqqiq min warāʔi-hi l-wuṣūlu li-hadafin risālati-k] see Bassiouney 2010 on the use of MSA and vernacular in advertisements.

28. Here we echo Trudgill's definition of markedness. See Haspelmath (2006) for an overview of the many possible uses of «markedness» as reported in the literature.

saying: «When I use the white dialect, I don't use affricated sounds like *ts* or *dz*²⁹» (Al-Rojaie, 2020, p. 45). It is seemingly due to this «canceling» or leveling of local features that WA receives the qualifier of «white». El-Hage explains it as follows in the Lebanese context: «By 'white' we mean 'neutral', this is, the opposite of regional [...] it is, therefore, a variety understood by all Lebanese, which does not have a regional color» (El-Hage, 2017, p. 30; our translation).³⁰

After mixing and leveling, koineization finishes with the process of «simplification» (Trudgill, 1986, p. 127), which associates to «either an increase in regularity or a decrease in markedness» (Siegel, 1985, p. 358, quoting Mühlhäusler, 1980, p. 21). The loss of marked features, and therefore the «simplification» of the variety, results in a new variety that, in the view of some speakers, is «neutral» or lacks a specific character. One of Germanos' (2009) Lebanese informants uses these precise words when identifying the lack of regional character of a linguistic variety that she labels as «white»: «It is possible, of course, there must be a dialect, let's say, white, that does not have a precise character» (Germanos, 2009, p. 103).³¹

As we will see below, it is clear that for some, the use of a white variety seems to bear clear communicative benefits. For others, however, this «simplicity» or convergence carries with it the lack of a defined identity, a fact that sometimes triggers negative attitudes in some speakers who perceive WA as the main cause behind the erasure of local diversity. In an episode of the Lebanese podcast *Sarde* (henceforth Azouri & Jaber, 2021), the Tripolitan rapper El Rass illustrated the process of formation of Lebanese WA with a rather critical tone, affirming that WA was created on the grounds of the unification of sects and regions within Lebanon³² (Azouri & Jaber, 2021, min. 92:30), and explains the deliberate creation of WA as the result of a totalitarian approach, that «attempts to harmonize different elements by erasing its idiosyncrasies and melting them all in a single pot»³³ (El Rass, in Azouri & Jaber, 2021, min. 92.35 translation ours).

29. Typically marked as Bedouin in Saudi Arabia.

30. French original: «par 'blanc' nous voulons dire 'neutre' donc le contraire de 'régional' [...] Il s'agit donc du parler compris par tous les Libanais, du parler qui n'a pas de couleur régionale» (El-Hage, 2017, p. 30).

31. French original: «F45GO: c'est possible, c'est possible, bien sûr, il doit y avoir un dialecte... disons, blanc, si tu veux, qui n'a pas un caractère précis».

32. Arabic original (oral statement): *hayy xturišət ?āl, domən mantə? tawḥīd ət-ṭawāyif w-əl manāṭə?* (El Rass, min. 92:30).

33. Arabic original (oral statement): *šuf ha-l-ša?liyye l-tōtālītariyye... kīf bətxalli ?əšya tətnēgam bən baš'da bi-?ənnak təlge xašā?iša w-tdawwəba kəlla bə-xāle waḥde.*

The concept of koineization has also been used to describe the changes that are occurring in many contemporary dialects following movements of population and urbanization. In fact, the most prominent koines we know of in the Arab world are those that emerged during the last century within urban centers, as a result of different urbanization processes (Miller, 2004, 2007). Hence, most contemporary Arabic urban vernaculars can be considered to have emerged from dialect contact, convergence, variation and change (Miller & Falchetta, 2021, p. 724).³⁴ In these cases, urbanization generally implied a rural/urban migration towards the main urban centers of each Arab country, often capital cities, but also to other urban centers more recently developed. For this reason, it is not surprising that the linguists, bloggers, journalists, etc., who mentioned WA in their works describe it as being either the variety of a main urban center within their countries (often the capital), or as being heavily based on urban vernaculars – i.e. in the latter, WA would be identified as a koine emerging from the main vernacular elements of the cities, while still avoiding the most localized features, therefore not in all cases completely overlapping with the main city vernacular.

In fact, the concept of «centralization» (*markaziyya*) represents a common theme in our data, and often appears connected to the notions of «education» and «open-mindedness» In an article published in the blog *Raseef22* entitled «Beirut, the White Dialect» (Qarout, 2018), the author refers to WA as the new dialect of the Lebanese capital, Beirut. In her opinion, this variety, which emerged from speakers' agreement to use a common variety to ease communication, reflects Beirut's diverse and multicultural character: «Beirut is an open and multicultural city. It is the capital, and the center of the state, and therefore its residents had to agree among themselves on a language that everyone understands, which we call «the white dialect» (Qarout, 2018; our translation).³⁵ Worthy of mention is the explicit differentiation the author makes between the old variety of Beirut (also

34. Despite their shared processes, it is important to point out, however, that degrees of koineization and leveling certainly depend on each city's history and on the rate of rural/urban migration. Therefore, there is neither a single model nor a common linear development (Miller & Falchetta, 2021).

35. Arabic original:
بيروت مدينة منفتحة ومتعددة الثقافات، هي العاصمة ومركز الدولة وبالتالي على القاطنين فيها الاتفاق فيما بينهم على لغة مفهومة من الكل، هي ما نسميه "اللهجة البيضاء" (Qarout, 2018)

[*Bayrūt madīnatun munfatihātun wa mutaʿaddidatu l-ṭaqāfāt, hiyya l-ʿāšimatu wa-markazu l-dawlati wa bi-l-tālī ʿalā l-qāṭinīn fī-hā l-ʿittifāqu fī-mā bayna-hum ʿalā luġatin mafhūmatin mina l-kull, hiyyā mā nusammī-hi "l-lahġa l-bayḍāʿ"*] (Qarout, 2018).

known as *bäyrüte*³⁶) and the new white variety. In her opinion, the original inhabitants of Beirut became educated and then realized that their «heavy» dialect was not appropriate anymore for their social status, which also kept them from teaching this variety to their kids, hence leading to the disappearance of the old vernacular dialect of Beirut.³⁷ In a blunt display of linguistic attitudes, the author qualifies the suspected disappearance of old *bäyrüte* as «a good thing», and as a justified and natural development given that «language has a communicative purpose» (Qarout, 2018).

The stigmatization of regional dialects (such as old *bäyrüte*) in the growing urban context of the capital was, in fact, one of the reported reasons that led young speakers living in Beirut to use WA: «when he would speak his dialect (a kid from the mountain who moved to Beirut), we would make fun of him (...) then the next generation started speaking a white dialect, let's call it like this, which is neither popular beirutite nor his village's dialect» (Germanos, 2009, p. 105).³⁸ Interestingly, the aforementioned two sources hint at the fact that WA as an urban dialectal koine is perceived to have become a nativized variety for some of the speakers from recent generations living in Beirut.³⁹ The native character of WA in this case seems to be unique to the definition of WA as an urban dialectal koine, given that references to WA aligning with definitions one and three (WA as an Educated Spoken Arabic and WA as a set of accommodative strategies for pan-Arab communication) always

36. Beirut's population rose from fewer than about ten thousand at the beginning of the nineteenth century to 426,861 by the end of the twentieth century [...] One of the consequences of this relatively recent expansion is the most frequently cited distinction in Beirut social representations between native residents and those considered non-native, even though they may have been born in Beirut and their families may have been living there for several generations (cf. for example Tarazi-Fawaz, 1983, p. 1). This distinction is reflected in the linguistic psyche: the term *bäyrüte* (Beirutite) refers specifically to the dialect of the native residents of the city (Germanos, 2011, p. 45).

37. Arabic original:

أيضاً برأيي أن أهل بيروت تعلموا، لأن الجامعات تأسست فيها قبل المناطق الأخرى، بسبب المركزية، ولم يعد ملائماً لطبيقتهم الاجتماعية استعمال اللهجة القديمة، الثقيلة بكل الأحوال. هم لم يستعملوها خلال الحديث مع أولادهم، فلم يورثوهم بها ومن الطبيعي أن تزول. (Qarout, 2018).

[*ʔayḏan bi-raʔy-ī ʔanna ʔahla bayrūt taʔallamū, liʔanna l-ḡāmiʕāti taʔassasat fī-hā qabla l-manāʕiqi l-ʔuxrā, bi-sababi l-markaziyyati, wa-lam yaʔud mulāʔiman li-tabaqati-him al-ʔiḡtimāʕiyyati istiʕmālū l-laḡḡati l-qadīmati, l-ḡaʕilati bi-kulli l-ʔaḥwāl. hum lam yastaʕmilū-hā xilāl l-ḡadīti maʕa ʔawlādi-him, fa-lam yūrithū-hum yā-hā wa-mīna l-ḡabʕī ʔan tazūl*] (Qarout, 2018).

Contrary to the perception of the author, the original dialect of Beirut has not completely disappeared but rather is still spoken by some groups, although its use has been increasingly restricted to familiar and close-circle interactions (Germanos, 2009).

38. French original: «lorsqu'il parlait avec son dialecte, [l'enfant originaire de la montagne qui s'installait à Beyrouth], on se moquait de lui (...) alors la génération suivante a commencé à parler un dialecte blanc, appelons-le comme ça, qui n'est ni le beyrouthin populaire, ni le dialecte de son village.» (Germanos, 2009, p. 105).

39. This has been also confirmed by ten years of participant observation in Beirut and several qualitative interviews for a variety of sociolinguistic studies (Iriarte Díez, 2021)

refer to WA as an acquired, even «invented» variety that is decidedly nobody's mother tongue.

The aforementioned data, therefore, indicates that WA is not perceived to be *any* variety of old traditional cities – given that some old traditional varieties (e.g. old Beirut or Tripoli in Lebanon; old Tetouan or old Fez in new urban centers of Morocco⁴⁰) are, in fact, as stigmatized as other rural or regional dialects, and are therefore regarded as «too marked» to be labeled as or included in WA. Instead, WA seems to be associated and/or highly built upon features from new/contemporary urban varieties resulting from «recent» urbanization processes.

A relatively contemporary notion that played an essential role in the development and the perception of Arabic dialects is the concept of nation-state. As feelings of nationalism grew across the Arab world the need to agree (even if purposelessly) on a «national dialect» became a sociolinguistic reality in many places of the Arabic speaking world. In fact, various were the sociolinguistic studies that noticed, back in the late 1980s, that many of the urban vernaculars of the capital cities in the Arab world were «de facto functioning as prestigious non-official national standards» (Miller & Falchetta, 2021, p. 726). Features from these varieties were considered «standard urban features», which granted them a high degree of overt prestige. These are precisely the kind of prestigious and dialectal standard features that our data on this definition of WA strongly correlates with.

Along these lines, WA often appears characterized in our sources as a national koine/unofficial national spoken standard which does not reflect regional particularities and therefore serves as a leveled variety between urban-rural speakers or rural-rural speakers from different areas. One of the simplest yet rather telling indicators of the WA-national spoken standard association are the numerous references to the diverse «nationalities» of WA. We find specific mentions to Lebanese WA «*hayy l-lahža l-bayḍa l-labnāniyye*» (El Rass, in Azouri & Jaber, 2021, min. 92:26), to Saudi WA «*al-lahḡa al-saʿūdīyya al-bayḍāʾ*?» (Al-Barrak, 2018), even to «White Algerian terms», described as «common Algerian dialect words that are used by all Algerians» (Bougrine *et al.*, 2017, p. 143). Along the same lines, in Jordan, Alfaisal and Aljanada consider WA to be «the fifth dialect of the country» (Alfaisal & Aljanada, 2019, pp. 109-110).

The idea of a national variety is intimately connected to the characteristics of a contemporary dialectal koine, namely to its aforementioned «urban» and

40. See Germanos (2009) on Lebanon. See Hachimi (2007) and Vicente (2021) on Morocco.

«neutral» nature, given that using such a variety would allow any citizen in a specific country to communicate without revealing his/her geographical origins or social and religious background. We cannot but assume that this was, in a historical period marked by massive processes of migration and urbanization, a rather useful tool for some native speakers of less prestigious varieties or from stigmatized communities who moved their lives to the city. Al-Rojaie (2020, pp. 40-42) points out this idea in his study about the emergence of a national koine in Saudi Arabia, which, according to his data, is associated with the Riyadh dialect on the basis of its clarity, simplicity and lack of marked features. So does El-Hage (2017, p. 30) in Lebanon, who defines WA as a neutral non-regional variety used in Great Beirut by newly arrived non-Beirutis.⁴¹

Deep feelings of pride for the nation are often also transferred to the variety that represents that nation in the speakers' perception. In our data, this correlates with mentions of WA in examples such as the following, illustrated in the words of Al-Rojaie (2020, p. 45): «By speaking the white dialect you get a sense of national feeling, whether inside or outside Saudi Arabia. It truly represents our nation in general». Contrarily, situations of revolt where citizens consider their nation a failed state, naturally render not such positive attitudes towards a national WA. This was the case in Lebanon, where, according to Iriarte Díez (2021), during the October Revolution in 2019, speakers vindicated the role of the revolution in celebrating the cultural production of local varieties and against the dialect leveling behind the formation of Lebanese WA: «Our revolution is [also] for songs in the regional dialects. Down with the insipid white variety!» (Iriarte Díez, 2021, p. 25).⁴²

In summary, this section illustrated how some of the mentions of WA in our data fit the main aspects that characterize Arabic contemporary dialectal koines as described in the literature, i.e., being the result of a leveling process; having an urban character; and, in some cases, being used as national spoken standards. Nevertheless, and before moving on, it is important to clarify that the perception

41. French original : « Le parler arabe libanais blanc est en fait le parler utilisé dans le Grand Beyrouth par les non-Beyrouthins, c'est-à-dire par les gens venus des différentes régions du Liban pour s'installer à Beyrouth et qui ont essayé d'adapter leur parler à celui des Beyrouthins pour qu'ils puissent être compris par ceux-ci » (El-Hage, 2017, p. 30).

42. In this article, the author already pointed at the complex and overlapping perception of WA among Lebanese speakers: «Although 'white variety' or 'al-lahze l-bayda' commonly refers to the koineized urban variety that I call 'Beirut koine' when I inquired the author of this post about his understanding of the meaning of 'white variety', he provided me with two different definitions that would be equivalent to 'Modern Standard Arabic' and the 'Beirut koine' respectively» (Iriarte Díez, 2021, p. 24).

of WA as an urban dialectal koine in our data remains flexible and therefore sometimes overlaps and coexists with our other two working definitions.

3.3. *White Arabic as a set of dialectal accommodation practices used in pan-Arab communication*

In our data, White Arabic also refers to a set of dialectal accommodation strategies that aims at facilitating mutual understanding among speakers of different Arabic varieties across the Arab world.

Communication accommodation, the research object of Communication Accommodation Theory (CAT), is defined as the «adjustments individuals make to create, maintain, or decrease social distance in interaction [...] It explores the different ways in which we accommodate our communication, our motivations for doing so, and the consequences» (Giles & Ogay, 2007, p. 16). Communication in globalized superdiverse settings resulting in complex and fluid communicative networks raised the need for strategies that enable mutual understanding between speakers coming from different parts of the Arab world. Research conducted in the framework of CAT distinguishes between different so-called accommodation strategies, one of these being «convergence» – i.e., the effort to adjust to your interlocutor (Giles & Ogay, 2007, p. 294-5). As the data shows, the main accommodation strategy behind the use of WA in this third definition is convergence between speakers from Arabic varieties that are perceived by them as unintelligible.

In the following extract taken from an online article, the author theorizes about the creation of WA and the communicative motivations behind it, highlighting its shared character (Abdel Nasser, 2017; our translation): «It [WA] became necessary due to the space opened by new communication and social media, for a lot of people found it difficult to understand some of the local dialects of the Arab world, which made communication hard. This is why White Arabic was invented, and it was named like this because it does not carry any specific identity, for it is a dialect that belongs to all»⁴³.

43. فأصبحت الحاجة إليها بسبب الافتتاح الكبير توفره وسائل الاتصال الجديدة، فالكثير من الناس وجدوا بعض الصعوبة في اللهجات المحلية لبعض الدول العربية فأصبح التواصل صعب، فلذلك اخترعوا "اللهجة البيضاء" وقد سميت بهذا الاسم لأنها لا تحمل أي هوية معينة فهي لهجة ملك الجميع.

fa-ʔaʃbaḥat al-ḥāga ʔilayhā bi-sababi l-infītāḥi l-kabīri tuwaffiruhu wasāʔilu l-ittiṣāli l-ḡa-dīda , fa-l-kaṭīr mina l-nās waḡadū baʃḍa l-ʃuʃūbati fī l-laḡāti l-maḥalliyyati li-baʃḍi l-duwali l-ʃarabiyyati fa-ʔaʃbaḥa l-tawāʃul ʃaʃb, fa-li-dālika xtarʔū „al-laḡa l-bayḍāʔ“ wa qad sammiyat bi-ḥāda l-ism li-ʔannahā lā taḥmilu ʔayy hawiyya muʔayyana fa-ḥiyya laḡa mulk al-ḡamīʔ.

The author specifically mentions the «creation» of WA in the context of superdiverse communication through social media, pointing at the latter as the main trigger for the emergence of new supra-national, pan-Arab communication networks. Communication within these new broader networks prompted new situations of dialect contact that were previously unusual. The emergence of WA in the context of media is also explicit in an article published in *Al-Bayan* newspaper⁴⁴, where the author specifically describes WA as a solution to presenters not being widely understood by the audience in their local dialects (Abdel Hamid, 2015).

Many are the mentions of WA in our data that highlight the accommodative function of WA within supra-national settings. For instance, the Wikipedia article on WA (Wikipedia لهجة بيضاء [lahġa bayḍāʔ]) states that it is a mixture between the local dialects of the involved speakers, youth expressions and English words, that, therefore cannot be ascribed to a specific country.⁴⁵ The article also provides an example of a pan-Arab communicative setting in which WA would be used: «If you invite three people from Egypt, Tunisia and Saudi Arabia, they will have to use common words and terms in order to understand each other».⁴⁶

44. Arabic original: .

ويواجه الطلبة اشكالية الفهم من المعلمين الذين يتحدثون العامية، وكذلك المشاهدون لا يفهمون كلمات عامية ينطقها مذيعون من أقطار عربية مختلفة.

[*wa-yuwāġihu l-ṭalabatu iškāliyyata l-fahmi mina l-muʿallimīn alladīna yataḥaddatūna l-ʿāmmiyya, wa-kaḍālika l-muṣāhidūna lā yafhamūna kalimātin ʿāmmiyyatin yanṭiqu-hā muḍīfūna min ʿaqtārin ʿarabiyyatin muxtalifa*].

45. Arabic original:

[*al-lahġatu al-bayḍāʔu ʔaw l-luġatu l-bayḍāʔu hiyya ʔarīqatun fī l-kalāmi takūn fī-hā l-luġatu* اللهجة البيضاء أو اللغة البيضاء هي طريقة في الكلام تكون فيها اللغة الأم هي الأساس، كأن تكون العربية مزوجة بطريقة محلية في الكلام، وغالبا ما تكون عبارة عن خليط من لهجات مختلفة مع وجود كلمات أجنبية و عبارات يستحدثها الشباب؛ لأن الحاجة إليها أصبحت ماسة عند الناطقين بها، بسبب عوامل التفاعل والانفتاح والتواصل العريض؛ فهي مزيج بين اللغة العربية الفصحى وبين اللهجة المحلية، ولا يمكن أن تصنفها تحت بلد معين وإنما هي خليط وهي واضحة إلى حد ما بحيث يسهل على مختلف الجنسيات العربية فهمه والتحدث بها.

[*l-ʔummi hiyya l-ʔasās, ka-ʔanna takūn l-ʿarabiyyatu mamzūġatan bi-ʔarīqatin maḥaliyyatin fī l-kalāmi, wa-ġāliban mā takūn ʔibāratan ʔan xalīṭin min lahġatin muxtalifatin maʿa wuġūdin kalimātin ʔaġnabiyyatin wa-ʔibāratin yastahḍiṭu-hā l-ṣabāb; liʔanna l-ḥāġata ʔilay-hā ʔaṣbaḥat māssatan ʔinda l-nāṭiqīn bi-hā, bi-sababi ʔawāmili l-tafāʔuli wa-l-infītāhi wa-l-tawāṣuli l-ʔarīḍ: fa-hiyya mazīġun bayna l-luġati l-ʿarabiyyati l-fuṣḥā wa-bayna l-lahġati l-maḥaliyyati, wa-lā yumkinu ʔan tuṣannifa-hā taḥta baladin muʿayyanin wa-ʔinna-mā hiyya xalīṭun wa-hiyya wāḍiḥa ʔilā ḥaddin mā bi-ḥayṭu yashulu ʔalā muxtalifi l-ġinsiyyāti l-ʿarabiyyati fahmu-hā wa-t-taḥadduṭ bi-hā*].

46. Arabic original:

وإذا قدمت الدعوة لثلاثة أشخاص , من جمهورية مصر العربية ومن المملكة العربية السعودية ومن تونس , سيكون عليهم التحدث بكلمات ومصطلحات مشتركة يفهم بعضهم البعض.

[*wa-ʔidā qaddamta l-daʿwata li-ṭalāṭati ʔaṣxās, min ġumhūriyyati miṣra l-ʿarabiyyati wa-mīna l-mamlakati l-ʿarabiyyati l-saʿūdīyyati wa-min tūnis, sa-yakūn ʔalay-him at-taḥadduṭ bi-kalimātin wa-muṣṭalahātīn muṣṭarakatin bi-fahmi baʔaḍi-him l-baʔḍ*].

In this context, worthy of mention is the situation of WA in the UAE (O'Neill, 2017; Hopkyns *et al.*, 2021; Argüelles, 2018; Abdel Hamid, 2015). As a growing economic hub, the UAE has experienced enormous migration influxes in the recent decades from different Arab countries as well as from multiple non-Arabic speaking countries. The coexistence of immigrants from all over the Arab world has made dialect contact in the UAE a daily reality. Within this socially and linguistically superdiverse context, WA is often described as a form of Arabic resulting from the mix of different dialects, SA and English (Hopkyns *et al.*, 2021, pp. 178-179). This also applies to educational contexts there, often international, where WA is reported to be used both (1) in lectures – in order to overcome the students' difficulties to understand the professors' native dialects (Abdel Hamid, 2015) – ⁴⁷ and (2) in everyday communication among students of different Arabic nationalities (Argüelles, 2018).⁴⁸

In his article on multilingual diversity in Dubai, O'Neill (2017) sheds light on the influence of linguistic superdiverse settings on individual speakers through the experience of Shaikha, an individual born and raised in the UAE with Emirati-Moroccan origins. Shaikha identifies WA as one of the varieties of Arabic she speaks, and describes it as a form of Arabic that includes «all accents and dialects» (O'Neill, 2017, pp. 225-226). Interestingly, although Shaikha claims to have acquired WA mainly in her working environments – where new dialect contact situations often took place – her use of WA is not restricted to professional settings only, but rather spreads also onto more intimate communication settings involving family and friends (O'Neill, 2017, p. 231).

Although as we previously mentioned, this definition of WA generally relates to supra-national contexts, our data also points at two specific national contexts where the term WA also aligns with the present definition: Jordan and Yemen. WA is listed by Alfaisal & Aljanada (2019) as one of the dialects of Jordan, which is used when speakers of «considerably different» varieties spoken in Jordan communicate with each other (Alfaisal & Aljanada, 2019, p. 110). The same is claimed about Yemen, where Dufour (2008) observed that WA is used when mutual understanding is endangered by the perceived linguistic distance between local Yemeni dialects (Dufour, 2008, p. 141). He also explicitly states that WA in Yemen is neither a

47. Arabic original: «ويواجه الطلبة اشكالية الفهم من المعلمين الذين يتحدثون العامية»، [wa-yuwāḡih al-ṭalaba iškāliyyata l-fahmī mina l-muʿallimīna alladīna yataḥaddaṭūna l-ʿāmiyya].

48. Here, Prof. Argüelles specifically refers to the students of his university, the American University of Dubai.

fixed variety, nor the variety of the city, and that it is not the mother tongue to any speaker (Dufour, 2008, p. 141).

The fact that WA in Yemen and Jordan is reported to be used to enable communication among speakers from within the national borders of the respective countries may, at first sight, contradict the previous supra-national pan-Arabic understanding of WA described in this section. However, if we consider that Yemen's and Jordan's linguistic landscapes encompass a myriad of distinct varieties that are often perceived by their speakers as mutually unintelligible, then this data would confirm that, regardless of national borders, WA is understood to be used more generally in a variety of contexts – both national and supra-national – that present increasing communication among speakers who deem their varieties as unintelligible to their interlocutors.

The long prevailing assumption that speakers of different Arabic varieties use only SA to enable communication has already been refuted by several studies focusing on accommodation (e.g., S'hiri, 2002; Miller, 2005; Chakrani, 2015, etc.). Due to globalization, the resulting increased mobility and the spread of social media, speakers are currently exposed to a wide range of Arabic varieties. In this context, as our data confirms, Arabic speakers tend to use and mix different varieties that are familiar to them and that they consider widely intelligible in superdiverse communication settings.

It is worth noticing that speakers' linguistic choices in these superdiverse settings highly depend on prevailing language ideologies. Previous research has shown (Hachimi, 2013, p. 278; S'hiri, 2002; Schulthies, 2015, among others) that the accommodative burden in communicative settings between Maghrebi and *Mashreqi*⁴⁹ speakers is usually carried by Maghrebi speakers, for the vernaculars of the latter are often regarded as «not pure Arabic» and therefore as unintelligible (Hachimi, 2013, p. 290). In this context, Chakrani (2015, p. 10) observes that the «social capital» that is assigned to certain varieties may influence the speakers' accommodative behavior. In his study on the influence of attitudes on interdialectal communication in a diaspora setting in the US, he shows that the use of features of Egyptian and Levantine dialects is favored in accommodative processes because of the higher social capital of these two varieties – which, in comparison to the

49. *Mashreqi* means «from the Middle East area» and it is the counterpart of Maghrebi for «North African».

recently arrived Sudanese and Maghrebi communities, are better established in the US (Chakrani, 2015, p. 7).

The influence of language ideologies and attitudes on speaking behaviors that favor the use of Levantine and Egyptian varieties for accommodative purposes is also confirmed in our data on WA. A participant in O'Neill's study (2017, p. 226), for example, mentions that she uses reflexes of the verb *yrūh* 'to go' when she speaks with Lebanese speakers – which have been attested to be widely used in Levantine dialects (Behnstedt & Woidich, 2014, map 312a, p.14 – *Bewegungsverben*) – rather than those of *ysīr* – which she perceives as the usual form in UAE Arabic (O'Neill, 2017, p. 226). When inquired about her motivations, the participant appealed to the broadly shared nature of *yrūh* throughout the Arabic-speaking world.

In another relevant study, that investigates accommodation between speakers from Baghdad and Tunis (Laaber, 2021), a 26-year-old Tunisian speaker stated he used WA mainly to facilitate communication with speakers from Baghdad. He defined WA as a variety that was «invented» by Maghrebi speakers for communication with speakers of the Mashreq region and reported that the main strategies involved in the use of WA are (1) the avoidance of typically Tunisian features and (2) the mixing between Egyptian Arabic and Syrian Arabic – as these varieties are widely used in movies, series and social media – and some lexical elements of SA.

In sum, this section showed that WA is also described as a set of accommodation practices often used among Arabic speakers at the supranational level. This phenomenon follows the emergence of increasing superdiverse contexts in which speakers, above all, pursue intelligibility with a wide variety of Arabic speakers. The speakers' choices on the linguistic resources used in this accommodative process rely heavily on prevailing language attitudes and ideologies. Our data shows that the main accommodative practice involved in WA consists of replacing features that are regarded as unintelligible by a mix of features from different dialectal varieties (e.g., Levantine, Egyptian)⁵⁰ that are perceived to be widely understood due to prevailing ideologies, along with the occasional adoption of SA and English lexical elements.

50. These varieties do not necessarily need to be the native varieties of either of the interlocutors involved in the communicative act.

4. SUMMARY

The analysis of the data provided in this study suggests that White Arabic is a new metalinguistic label that is increasingly used across different countries of the Arab world. Notwithstanding the fact that speakers' perceptions on WA may be oftentimes overlapping and even linguistically contradictory, our data indicates that they may be summarized along three notions: (1) WA as Educated Spoken Arabic, (2) WA as an urban dialectal koine and (3) WA as a set of dialectal accommodation practices used in pan-Arab communication. The diversity found in our results suggests that there is not a fixed consensual definition of WA that all speakers agree with, but rather, that the speakers' perceptual definition of WA is dynamic and context-dependent.

The results of this study also suggest that the relatively recent and increasing spread of the label WA across the Arab world may be linked to its conceptual novelty. We can conclude that, when compared to previous or contemporary labels for analogous concepts, WA stands out for three important characteristics:

(1) WA seems to have emerged from the speakers themselves. Labels such as 'the third language' (*al-luġa al-ṭālīta*), 'the middle language' (*al-luġa al-wuṣṭā*) or Middle Arabic, were created by linguists, writers and experts in Arabic language to designate what they considered interesting linguistic phenomena, while WA, where present, has emerged from speakers themselves. To this point, we only found two written academic sources aiming to describe WA,⁵¹ while the majority of references to WA are rather attestations of the speakers' attempts to elucidate the meaning that the label already carries. This suggests that speakers find WA a useful label to reflect their perceived linguistic reality and their current communicative and ideological needs.

(2) WA seems to have a predominantly (almost exclusively) spoken character. It refers predominantly to spoken varieties and/or communication strategies, contrary to aforementioned previous labels, that emerged mainly through the

51. To the best of our knowledge there is only a textbook entitled *اللغة البيضاء طريقك إلى التحدث بلغة عربية واضحة* *al-luġa al-bayḍā? ṭarīqu-ka/ki ʔilā l-taḥadduṭ bi-luġatin ʕarabiyyatin wāḍiḥatin* 'The White dialect. Your way towards speaking in a clear Arabic language' (Al-Ajami, 2019), whose main aim is to teach Spoken Arabic without any «social» color, and a linguistic study describing the phonetic and phonological features of Lebanese WA, entitled *Etude phonétique-phonologique de parler arabe libanais blanc* 'A phonetic and phonological study of Lebanese White Arabic' (Naboulsi, 2013). The recent publication of these works, together with our data on speakers' perceptions confirm our hypothesis that this label emerged from speakers.

creation and/or the analysis of written texts. In this sense, we believe the spread of WA may have been fueled in the last decade by the emergence of technological advancements that have enabled and facilitated transnational oral communication (e.g., online conference platforms, video calls, voice messages, etc.)

(3) WA is not necessarily «a mix of *fushā* and *ʿāmmiyya*» and therefore it is not restricted to semi-formal or formal interaction. Although, as the first definition indicates, WA may echo the notion of Educated Spoken Arabic, the other two definitions of WA found in our data suggest that WA, unlike the aforementioned previous labels, is not exclusively an intermediary stage between *fushā* and *ʿāmmiyya*. Instead, this label also points at different codes, practices or varieties that may be practically «devoid» of MSA elements, since they do not necessarily result from the mix of a regional dialect and MSA, but rather from the mix of different regional dialects. These results suggest that the creation of new metalinguistic labels for Arabic varieties echoes the claim that variation and changes in spoken Arabic involve the interplay between the local dialects and the emerging regional standards independently of Classical Arabic or Modern Standard Arabic (Miller & Falchetta, 2021, p. 724).

5. DISCUSSION: ARE NEW LABELS INDICATIVE OF A NEED FOR CHANGE IN PARADIGM?

Our study confirmed that the existence of White Arabic remains a self-evident, solid reality for some Arabic speakers. The emergent use of the label WA is indicative of a series of current linguistic behaviors, processes, and phenomena that result from the complexity of today's Arab world. However, the relatively «new» metalinguistic label of WA presents a multifaceted, dynamic, and versatile nature that simply cannot be explained within diglossic frameworks, at least not in all its complexity.

Although this traditional diglossic framework has, as we have mentioned, been challenged by several studies changing the binary, dichotomous model to the understanding of diglossia as a continuum – and while Arabic linguistics has witnessed the emergence of many sociolinguistic works analyzing Arabic varieties in other ways, from variationist, interactional and linguistic anthropologist perspectives –, the ideology of a diglossic framework stays pervasive both in non-specialists' views and in the field of Arabic studies. This is evident, for instance, from teaching practices both inside and outside Arabic-speaking countries, where

fushā and the spoken vernaculars are, very often, still taught as perfectly distinct varieties, each having their own limited set of grammar rules and vocabulary.

In our view, a framework that does not fit the speakers' reality is simply bound to be updated and, ultimately, changed. Considering that many of the notions structuring theoretical frameworks of Arabic linguistics find their origins in the middle of the 20th century, we believe a joint attempt to continue a critical reflection upon some of the already established sociolinguistic notions and/or variables is necessary. In the following paragraphs, we discuss what could be possible leads on our path to a change of paradigm.

EDUCATION

In the (traditional) diglossic framework, the notion of education was associated with the vast or partial knowledge of speakers of SA. This was evident from the fact that the abovementioned labels that situate themselves somewhere on the SA-vernacular continuum and refer to educated speakers, such as «*ʿġāmmiyyat al-mutaqqafīn*» or «Educated Spoken Arabic» always implied a partial use and knowledge of SA. The fact that this variety (SA) is acquired through education and is never a mother tongue reinforces the symbolic linkage of SA with education. Previous research suggests that the educational instruction of *fushā* to Arabs is actually not increasing the speaker's competency as much as it is reinforcing its ideological status (Brustad, 2017). The «clash of overt and covert norms and expectations» that students face when learning SA (Parkinson, 1993, p. 72), along with the limited communicative contexts covered by SA are rendering the use of this variety less popular among the young generations.

Both SA's loss in popularity among the younger generations and the growing importance and systematic instruction of foreign languages have resulted in knowledge of SA not being anymore a *sine qua non* requirement to consider a speaker as «educated». In this way, SA gradually stopped being the main criterion upon which a speaker's level of education can be measured. Instead, many young Arabs seem to now consider the language of the media as «educated» and tend to reproduce it in formal communicative contexts –irrespective of the myriad of linguistic varieties that may be considered to be «Media Arabic» as well as of how «linguistically close» they may actually be to SA. This fact is especially relevant

since the Arabic used in the media seems to be turning more and more into regional and spoken standards.⁵²

In the current superdiverse and hyperconnected Arab world, where institutions are not the exclusive holders of information, the level of education stopped being strictly measurable by number (or level) of degrees. As has been shown in Al-Wer (2002), the notion of education as a variable for analysis in sociolinguistics needs to move away from the idea that higher education equals more use of SA, since it is a *proxy* variable pointing out other significant social patterns: «it is not level of education *per se* which correlates with linguistic usage, rather that level of education is actually an indicator of the nature and extent of the speakers' social contacts» (Al-Wer, 2002, p. 42).

Currently, as Miller & Falchetta (2021) pointed out, educated speakers of Arabic tend to experience greater mobility and have larger networks, which results in these speakers' increased «exposure to different social values and the need to adopt common features shared by a wider number of people and not indexed with localness» (Miller & Falchetta, 2021, p. 726), this is, the type of features involved in the emergence of urban koines and other accommodative practices (see sections 3.2 and 3.3). For this reason, education represents a key concept for the understanding of WA, and, in general, for the study of any form of accommodation among speakers of Arabic, regardless of its label. Given that «educated speakers appear to be leading the changes, most often in the direction of urban and koineized regional standards» (Al-Wer, 1997, p. 259), now more than ever, accurately understanding, defining and applying education as a sociolinguistic variable is of paramount importance to the field.

For all these reasons, we believe the notion of education as a variable for analysis in sociolinguistics needs to be considered in other terms than number of degrees or proficiency in SA. This means that, on the one hand, we should broaden the notion of education to incorporate/integrate the notions of «exposure» and «speaker's networks» (next section), and on the other hand, complement it with additional (yet independent) variables that study relevant yet often neglected factors such as «speaker's (meta)linguistic awareness».

52. This could be a reason for our first and second definitions of WA to overlap in speaker's perceptions (see section 3.1. and 3.2. of this article) – both definitions tend to the notion of “Standard” (see subsection «Standardization» of this discussion).

SPEAKERS' NETWORKS

Speech communities have generally been described as rather fixed and stable groups of speakers, and often associated to a specific geographic location, whose ascribed unity relies on one or a few shared linguistic features. In today's Arab world, marked by speakers' increasing mobility and their resulting broadened exposure to different linguistic varieties, the notion of «speech community» bears the danger of (1) neglecting the plurality of resources that are fluidly used to enable communication in various superdiverse communicative settings and (2) being limited to communities cohabitating in a specific territory.

As our results show, WA cannot hardly be ascribed to be spoken by a specific community or group of speakers confined within a specific territory⁵³ or across an ethnic or religious community. Nevertheless, we believe that the use and emergence of WA could be explained through more flexible notions, such as that of «speakers' networks» (Milroy, 1987). Milroy's network is understood as an «aggregate of relationships contracted with others, a boundless web of ties which reaches out through social geographical space linking many individuals, sometimes remotely» (Milroy & Gordon, 2003, p. 117).

In fact, the diverse and overlapping understandings of WA could be partially analyzed through this notion. In definition 2 (WA as an urban dialectal koine) the speakers' network is geographically tied to a specific location – i.e. cities like Beirut and Riyadh – which leads to the systematic repetition of certain communicative settings that results in a fixation of speaking behaviors. In this case, the high density of the network is both relevant and explanatory of the leveling and koineization processes that take place as these varieties become more stable. Alternatively, definition 3 (WA as a set of accommodation practices for pan-Arab communication) could be the result of a prominently fluid network. Here, in a context where WA is used to describe linguistic behavior in dynamic superdiverse settings that undergo constant change, speakers' networks would be more numerous and dynamic, but probably have a lower constant density and a higher degree of instability.

Aside from WA, we believe that applying the notion of speakers' networks systematically could be beneficial to understand currently emerging shifts in

53. The only exception to this would be WA when understood as an urban dialectal koine (see section 3.2.)

ideologies and/or linguistic behavior resulting from fluid groups (both at the national and transnational levels) united by shared feelings of social and political unrest.

STANDARDIZATION

Standardization traditionally referred to the institutionalization of a written standard undertaken by language planners, normally representing institutions that are holders of economic, political, religious, intellectual or social authority (Haugen, 1966, p. 933). The result of this standardization process made SA the official written standard along the Arab world. Nevertheless, the institutions that granted the standard with its status failed to address the changing reality of the Arabic language, creating a growing gap between prescriptive norms and linguistic practices.⁵⁴ This conflict confirms the idea that standard is «an idea in the mind rather than a reality – a set of abstract norms to which actual use may conform to a greater or lesser extent» (Milroy & Milroy, 2012, p. 19) and that standardization is an ideological process that often results from political, economic and social changes.

The strongly conflicting ideologies associated respectively with SA and spoken varieties of Arabic have left many speakers feeling they lack mastery of their own language. Foreign languages appeared here as attractive alternatives, not only because they allow speakers to be connected to broader networks, but also because they cover a wider range of registers, and hence, have also earned a role in the emergence of current spoken standards.

Rather than fixed and permanent, standardization processes are complex and continuous. In fact, several standardization processes may take place simultaneously within one language. In the Arabic-speaking world, for instance, despite SA being established as the «common official written standard», Arabic language is witnessing many different processes of standardization/destandardization, especially in the spoken realm. As Miller and Falchetta rightly pointed out, the term standardization may be now applied to any process of «unofficial focusing, koineization and conventionalization resulting from ‘spontaneous’ linguistic choices that operate at the spoken (eventually also at the written) level» (Miller & Falchetta, 2021, p. 716).

In this line, Milroy and Milroy remind us that, beyond institutional prescriptive norms, the main and ultimate function of a standard is communicative efficiency, this is, «that everyone should use and understand the language in the

54. A good example is that of institutions of higher education, which generally did not succeed in maintaining SA as a language of instruction and academic production along the Arab world.

same way with a minimum of misunderstanding and the maximum of efficiency» (Milroy & Milroy, 2012, p. 19), and that the status of a standard is mainly subject on speakers' acceptance of a certain variety as such. In this light, the diversity of definitions of WA could be explained via its perception by speakers as a new emerging standard. Not only is WA unanimously perceived to fulfill the purpose of «facilitating communication», but also, as we could see throughout the data, it often awakens the type of attitudes – both positive and negative – usually held towards standard varieties (i.e. it is professional, it belongs to all speakers, it minimizes and/or erases local differences, etc.).

We hereby insist on the necessity to acknowledge that standardization is a complex process that happens simultaneously in a vertical and a horizontal manner (Miller & Falchetta, 2021), for such an understanding of standardization effectively sheds light on the emergence of new standards that, like speakers' networks, may go beyond geographical and social borders.

The question, however, remains: how could we possibly fit these new emerging standards in the traditional diglossic *fushā* and *ʿāmmiyya* dichotomy?

SHIFTING THE LINGUISTIC MODEL: FROM LANGUAGE-BASED TO SPEAKER-BASED

Before suggesting possible alternatives, it is necessary to establish that the current Arabic linguistic model – heavily influenced by a traditional understanding of diglossia – is a language-based model, meaning it establishes two theoretically perfectly distinguishable linguistic entities – i.e., H and L; SA and spoken dialects – as opposite poles of a continuum. Moreover, this language-based model is reflective of ideological concepts rather than of linguistic behaviors, since it defines «ideal imaginary» monolithic varieties that exist as theoretical constructs, but not as lived realities.



Representation of the current language-based model (SA-spoken vernaculars; H-L variety)

In such a model, the choices of the speakers are significantly limited. The inadequacy of this dichotomy became manifest in the emergence of labels such as *al-luġa al-tālīta* and *al-luġa al-wuṣṭā*, which were used to cover various practices of Arabic between the poles of *fushā* and the spoken vernaculars (Mejdell, 2011). In the Fergusonian extension of the term, they represented primarily a compromise strategy for speech in semi-formal interactions. However, even when we acknowledge the existence

of a continuum with intermediary stages between these two poles, a variety-based continuum typology remains insufficient to account for speakers' complex linguistic choices as well as for the social, interactional and identitarian motivations behind them.

Alternatively, for more than 20 years, sociolinguists had already developed speaker-based models (e.g., Agha, 2006; Coupland, 2007; Garrett, 2010; Busch, 2015, 2021) which have already been used as frameworks for studies on accommodation and convergence in Arabic, e.g. Shi'iri's (2002). Regrettably, these models have not, in our opinion, received the attention they deserve within the field of Arabic sociolinguistics, probably due to the fact that the internal Arabic diversity is not perceived or treated as multilingual by most experts on the field.

In such a communicative speaker-based model, varieties are not regarded as monolithic rigid entities the speaker may reach or approach to – for they are neither references for normativity nor objectives –, but rather available sets of linguistic features and strategies encompassing sets of strategies that can be used partially, according to the speakers' wishes upon evaluation of the different factors relevant to a specific communicative situation (i.e. communicative priorities, communicative actors, communicative settings, etc.). In a communicative speaker-based model, speakers are conceived to be complex actors situated at the receiving ends of communication. Every speaker may belong to a variety of networks, and these different networks are brought together by the speaker themselves in their own set of repertoires, forming their social and linguistic identity. Acknowledging the superdiversity of networks and repertoires a speaker counts on, and how their use may adapt to specific communicative situations, provides us with a more flexible framework for the study of the emergent, context-dependent linguistic realities that speakers witness in today's Arab world.

The metalinguistic label of White Arabic cannot find its place in a rigid language-based model because its dynamic nature springs from speakers and their superdiversity. However, the seemingly contradictory and overlapping definitions on WA would not pose a problem when envisioned in a speaker-oriented framework. For instance, at the receiving ends of WA as Educated Spoken Arabic (definition 1), we could envision presenters, and a pan-Arab audience; at the receiving ends of WA as an urban dialectal koine (definition 2), we could for example envision speakers from different villages/towns within the same country meeting in an urban context (e.g. a speaker from Tripoli and a speaker from Saida meeting in Beirut); and at the receiving end of WA as a set of accommodative practices in pan-Arab communication (definition 3) we could envision speakers of different regional dialects of the Arab world (e.g. Iraqi and Moroccan) who perceive their varieties as mutually unintelligible.

- D8%A7-%D8%A7%D9%84%D9%84%D8%BA%D8%A9-%D8%A7%D9%84%D8%A8%D9%8A%D8%B6%D8%A7%D8%A1
- Albury, N. (2014). Introducing the folk linguistics of language policy. *International Journal of Language Studies* 8, 85–106.
- Alfaisal, A. and Aljanada, R. (2019). Diglossia in Arabic: Views and opinions. *The International Journal of Humanities & Social Studies* 7(5), 108–111.
- Al-Felou, M. (2020). لماذا يتأرجح المحتوى الإعلاني السعودي بين الفصحى... و«البيضاء»... والعامية؟ [limādā yataʿarġaḥ l-muḥtawā al-ʿiḥḥānī sl-saʿūdī bayna l-fuṣṣḥā ā wa-“l-bayḍā?” wa-l-ḥammīyya?]. *al-Majalla*, 31 August 2020. <https://arb.majalla.com/node/101306/%D9%84%D9%85%D8%A7%D8%B0%D8%A7-%D9%8A%D8%AA%D8%A3%D8%B1%D8%AC%D8%AD-%D8%A7%D9%84%D9%85%D8%AD%D8%AA%D9%88%D9%89-%D8%A7%D9%84%D8%A5%D8-B9%D9%84%D8%A7%D9%86%D9%8A-%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A-%D8%A8%D9%8A%D9%86-%D8%A7%D9%84%D9%81%D8%B5%D8%AD%D9%89-%D9%88%D8%AB%D8%A7%D9%84%D8%A8%D9%8A%D8%B6%D8%A7%D8%A1%D8%BB-%D9%88%D8%A7%D9%84%D8%B9%D8%A7%D9%85%D9%8A%D8%A9%D8%9F>
- Al-Rojaie, Y. (2020). The emergence of a national koiné in Saudi Arabia. In R. Bassiouney and K. Walters (Eds.), *The Routledge handbook of Arabic and identity*, pp. 26–50, London-New York: Routledge.
- Al-Wer, E. (1997). Arabic between reality and ideology». *International Journal of Applied Linguistics* 7(2) 251–265.
- Al-Wer, E. (2002). Education as a speaker variable. In A. Rouchdy (Ed.), *Language contact and language conflict in Arabic. Variations on a sociolinguistic theme*, pp. 41–53. London: Routledge Curzon.
- Argüelles, A. (Dir.) (2018). *The unity of the Arabic language – وحدة اللغة العربية* [waḥda' l-luġa l-ʿarabiyya]. American University of Dubai: Youtube. <https://www.youtube.com/watch?v=Pcc-eE7OrXU>
- Attwa, M. F. (2019). On the way to understand the pan-Arab voice. In C. Miller et al. (Eds.), *Studies on Arabic dialectology and sociolinguistics, Proceedings of the 12th International Conference of AIDA held in Marseille from May 30th to June 2nd, 2017*. Aix en Provence: IREMAM. <https://books.openedition.org/iremam/4449>
- Azouri, M. and J. Mouin (2021). EL RASS: Rap, identity, and the Arabic language. *Sarde After Dinner Podcast* (podcast), 10 October 2021.
- Badawi, E. S. M. (1973). [Mustawayāt al-ʿarabiyyah al-muʿāṣira fī miṣr. baḥṯ fī ʿalāqāt al-luġa bi-l-ḥaḍāra]. Cairo: Dār al-maʿārif.
- Bassiouney, R. (2010). *Arabic and the media: Linguistic analyses and applications*. Leiden: Brill.
- Bassiouney, R. (2015). Dialect and stance-taking by non-Egyptian celebrities in Egypt. *Open Linguistics* 1, 614–633.
- Behnstedt, P. and Woidich, M. (2014). *Wortatlas der Arabischen Dialekte: Band III: Verben, Adjektive, Zeit und Zahlen. Handbook of Oriental studies. Section 1, The Near and Middle East 100*. Leiden-Boston: Brill.

- Blommaert, J. and Backus, A. (2011). Repertoires revisited: 'Knowing language' in superdiversity. *Working Papers in Urban Language and Literacies* 67. sc: wpull. https://wpull.org/wp-content/uploads/2022/04/WP67_Blommaert_and_Backus_2011_Repertoire.pdf
- Blommaert, J. and Rampton, B. (2011). Language and superdiversity. *Diversities* 13(2), 1-21.
- Blommaert, J. and Rampton, B. (2016). Language and superdiversity. In K. Arnaut, J. Blommaert, B. Rampton and M. Spotti (Eds.), *Language and Superdiversity*, pp. 21-48. New York: Routledge.
- Bougrine, S., Chorana, A., Lakhdari, A. and Cherroun, H. (2017). Toward a web-based speech corpus for Algerian dialectal Arabic varieties. In N. Habash, M. Diab, K. Darwish, W. El-Hajj, H. Al-Khalifa, H. Bouamor, N. Tomeh, M. El-Haj, W. Zaghouni (Eds.), *Proceedings of the Third Arabic Natural Language Processing Workshop*, pp. 138-146. Stroudsburg: ACLAntology.
- Brustad, K. (2017). Diglossia as Ideology. In J. Hoigilt and G. Mejdell (Eds.), *The politics of written language in the Arab world: Writing change*, pp. 41-67. Leiden: Brill.
- Busch, B. (2015). Expanding the notion of the linguistic repertoire: On the concept of Spracherleben – The lived experience of language. *Applied Linguistics*, July, 340-358.
- Busch, B. (2021). *Mehrsprachigkeit. 3., Vollständig aktualisierte und erweiterte Auflage*. UTB Sprachwissenschaft 3774. Vienna: Facultas.
- Chakrani, B. (2015). Arabic interdialectal encounters: Investigating the influence of attitudes on language accommodation. *Language & Communication* 41, 17-27.
- Coupland, N. (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.
- Creese, A. and Blackledge, A. (Eds.) (2018). *The Routledge handbook of language and superdiversity: An interdisciplinary perspective*. London-New York: Routledge
- Dufour, J. (2008). Attitudes linguistiques et identités locales à Şan'â. *Revue des Mondes Musulmans et de La Méditerranée* 121-122, 133-144.
- El Hage, A. (2017). *L'informatique au service des sciences du langage: La conception d'un programme étudiant le parler arabe libanais blanc*. Paris: Sorbonne Paris Cité.
- Ferguson, Ch. A. (1959). Diglossia. *Word* 15(2), 325-40.
- Frayha, A. (1953). *Essentials of Arabic: A manual for teaching Classical and Colloquial Arabic*. Beirut: Rue Bliss.
- Garrett, P. (2010). *Attitudes to language*. Cambridge: Cambridge University Press.
- Germanos, M. A. (2009). *Identification et emploi de quelques stéréotypes, traits saillants et autres variables sociolinguistiques à Beyrouth (Liban)*. Paris: Université de Paris III – Sorbonne Nouvelle.
- Germanos, M.-A. (2011). Représentations linguistiques et contact dialectal : remarques sur l'évolution de cinq variantes régionales à Beyrouth. *Langage et société* n. 138 (4): 43-58.
- Giles, H. and Ogay, T. (2007). Communication Accommodation Theory. In W. Samter and B.B. Whaley (Eds.), *Explaining communication: Contemporary theories and exemplars*, pp. 293-310. Mahwah-New Jersey: Lawrence Erlbaum Associates.

- Hachimi, A. (2007). Becoming Casablančan: Fessis in Casablanca as a case study. In C. Miller, E. al-Wer, D. Caubet, and J. C.E. Watson (Eds.), *Arabic in the city: Issues in dialect contact and language variation*, pp. 97–122. London: Routledge.
- Hachimi, A. (2013). The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world. *Journal of Sociolinguistics* 3(17), 269–296.
- Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics* 42(1), pp. 25–70.
- Haugen, E. (1966). Dialect, language, nation. *American Anthropologist* 68(4), 922–935.
- Hopkyns, S., Zoghbor, W. and Hassall, P. J. (2021). The use of English and linguistic hybridity among Emirati millennials. *World Englishes* 40(2), 176–190.
- Iriarte Díez, A. (2021). Language and revolution: Arabic in Lebanon after the October Revolution as a case study. *Journal of Arabic and Islamic Studies* 21, 5–37.
- Jørgensen, J.N., Karrebæk, M. S., Madsen, L. M. and Mølle, J.S. (2011). Polylinguaging in superdiversity. *Diversities* 13, 2.
- Kerswill, P. (2003). Dialect levelling and geographical diffusion in British English. In D. Britain and J. Cheshire (Eds.), *Social dialectology. In honour of Peter Trudgill*, pp. 223–243. Amsterdam-Philadelphia: John Benjamins.
- Kraidy, M. (2006). *Hybridity, or the cultural logic of globalization*. Philadelphia: Temple University Press.
- Laaber, C. (2021). *Communication accommodation between speakers of Baghdad and Tunis*. Unpublished MA thesis, Vienna: Universität Wien.
- Mejdell, G. (2011). Luġa Wusṭā. In L. Edzard and R. de Jong (Eds.), *Encyclopedia of Arabic language and linguistics*. Leiden: Brill. Online Edition. Consulted online on 24 October 2022.
http://dx-doi-org.uaccess.univie.ac.at/10.1163/1570-6699_eall_EALL_SIM_000004
 First published online: 2011. First print edition: ISBN: 9789004177024, 20090831.
- Mejdell, G. (2017). Changing norms, concepts and practices of written Arabic: A «long distance» perspective. In J. Hoigilt and G. Mejdell (Eds.), *The politics of written language in the Arab world: Writing change*, pp. 68–89. Leiden: Brill.
- Miller, C. (2004). Variation and change in Arabic urban vernaculars. In M. Haak, R. de Jong, and K. Versteegh (Eds.), *Approaches to Arabic dialects: A collection of articles presented to Manfred Woidich on the occasion of his sixtieth birthday*, pp. 177–206. Leiden: Brill.
- Miller, C. (2005). Les Saʿīdis au Caire. Accommodation dialectale et construction identitaire. In J.L. Arnaud (Ed.), *L'urbain dans le monde musulman de Méditerranée*, pp. 175–194. Paris: Maisonneuve & Larose.
- Miller, C. (2007). Arabic urban vernaculars: Development and change. In C. Miller, E. al-Wer, D. Caubet, and J. C.E. Watson (Eds.), *Arabic in the city: Issues in dialect contact and language variation*, pp. 1–31. London: Routledge.
- Miller, C. (2011). Arabic urban vernaculars. In S. Weninger (Ed.), *The Semitic languages: An international handbook* (pp. 982–990). Berlin-Boston: De Gruyter Mouton.
- Miller, C., and Falchetta, J. (2021). Standardization and new urban vernaculars. In W. Ayres-Bennett and J. Bellamy (Eds.), *Language standardization*, pp. 713–740. Cambridge: Cambridge University Press.

- Milroy, L. (1987). *Language and social networks* (2nd ed.) *Language in Society* 2. Oxford-New York: Blackwell.
- Milroy, J., and Milroy, L. (2012). *Authority in language: Investigating standard English*. Abingdon-Oxon-New York: Routledge.
- Milroy, L., and Gordon, M. J. (2003). *Sociolinguistics: Method and interpretation. Language in Society* 34. Malden, MA: Blackwell Pub.
- Mitchell, T. F. (1986). What is educated spoken Arabic? *International Journal of the Sociology of Language* 61(1), 7–32.
- Mühlhäusler, P. (1980). Structural expansion and the process of creolization. In A. Valdman and A. Highfield (Eds.), *Theoretical orientations in Creole studies*, pp. 19–55. New York: Academic.
- Naboulsi, R. (2013). *Étude phonétique/phonologique du parler arabe libanais blanc*. Beirut: Editions Aleph Ya.
- O'Neill, G. T. (2017). «It's not comfortable being who I am»—Multilingual Identity in Superdiverse Dubai. *Multilingua* 36(3), 215–245.
- Parkinson, D. B. (1993). Knowing standard Arabic: Testing Egyptians' MSA abilities. In C. Holes and M. Eid (Eds.), *Perspectives on Arabic linguistics V: Papers from the Fifth Annual Symposium on Arabic Linguistics*, pp. 47–73. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Preston, D. R. (2017). Perceptual dialectology. In C. Boberg, J. Nerbonne, and D. Watt (Eds.), *Handbook of dialectology*, pp. 177–203. Hoboken, NJ: Wiley-Blackwell.
- Qarout, R. (2018). «بيروت، اللهجة البيضاء» [“Bayrūt, l-lahǧa l-bayḏāʿ?]. *Raseef* 22, 20 June 2018. <https://raseef22.net/article/153181-%D8%A8%D9%8A%D8%B1%D9%88%D8%AA%D8%8C-%D8%A7%D9%84%D9%84%D9%87%D8%AC%D8%A9-%D8%A7%D9%84%D8%A8%D9%8A%D8%B6%D8%A7%D8%A1>.
- Rosenbaum, G. M. (2000). «Fuṣḥāmmiyya»: Alternating style in Egyptian prose. *Zeitschrift Für Arabische Linguistik* 38, 68–87.
- Ryding, K. C. (2011). Educated Arabic. In L. Edzard and R. de Jong (Eds.), *Encyclopedia of Arabic language and linguistics*. Leiden: Brill. Online Edition. Consulted online on 01 December 2023. http://dx-doi-org.uaccess.univie.ac.at/10.1163/1570-6699_eall_EALL_COM_0095. First published online: 2011. First print edition: ISBN: 9789004177024, 20090831.
- Schulthies, B. (2015). Do you speak Arabic? Managing axes of adequation and difference in Pan-Arab talent programs. *Language & Communication* 44, 59–71.
- Shiri, S. (2002). Speak Arabic please!: Tunisian Arabic speakers' linguistic accommodation to Middle Easterners. In A. Rouchdy (Ed.), *Language contact and language conflict in Arabic. Variations on sociolinguistic themes*, pp. 149–174. London-New York: Routledge.
- Siegel, J. (2001). Koine formation and Creole genesis. In N. Smith and T. Veenstra (Eds.), *Creolization and contact*, pp. 175–197. Amsterdam: John Benjamins.
- Siegel, J. (1985). Koines and koineization. *Language in Society* 14(3), 357–378.
- Soliman, R. (2015). *Arabic cross-dialectal conversations with implications for the teaching of Arabic as a second language*. Leeds: University of Leeds.
- Somekh, S. (1981). The concept of «third language» and its impact on modern Arabic poetry. *Journal of Arabic Literature* 12, 74–86.

- Tarazi-Fawaz, L. (1983). *Merchants and migrants in nineteenth-century Beirut*. Cambridge, Mass: Harvard University Press.
- Trudgill, P. (1986). *Dialects in contact*. (in *Society* 10) Oxford- New York: Blackwell.
- Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies* 30(6), 1024–1054.
- Wikipedia. لهجة بيضاء [lahǧa bayḏāʔ]. https://ar.wikipedia.org/wiki/%D9%84%D9%87%D8%AC%D8%A9_%D8%A8%D9%8A%D8%B6%D8%A7%D8%A1
- Youssi, A. (1986). *L'arabe marocain médian, Analyse fonctionnaliste des rapports syntaxiques*. Thèse d'état, Paris: Paris III.
- Youssi, A. (1992). *Grammaire et lexique de l'arabe marocain moderne. Préface de D. Cohen*. Casablanca: Wallada.
- ʔUbayd, ʔĪ. (1964). مجموعة قصص مصرية معاصرة: إحسان حنيم [Maǧmūʔat qiṣaṣ maṣriyya muʔāṣira: Iḥsān Ḥanīm]. Cairo: Ad-dār al-qawmiyya li-ṭ-ṭibāʔa w-an-naṣr.
- Vicente, Á. (2021). *l-hədra b-əl-qāla* (ou la réalisation [ʔ] de /q/) dans les parlers arabes du Nord-ouest du Maroc. In N. Comolli, J. Dufour and M.-A. Germanos (Eds.). *Libellules arabes, sémitiques, italiennes, berbères - études linguistiques et littéraires offertes à Jérôme Lentin par ses collègues, élèves et amis*, pp. 415-430. Paris: Geuthner.

Ana Iriarte Díez
 Institut für Orientalistik
 University of Vienna
 Spitalgasse 2
 1090 Viena

Claudia Laaber
 Institut für Orientalistik
 University of Vienna
 Spitalgasse 2
 1090 Viena

Nina van Kampen
 INALCO, CERMON associated with IFPO
 Inalco
 65 rue des Grands Moulins
 75214 Paris Cedex 13

Montserrat Benítez Fernández
 Escuela de Estudios Árabes, CSIC
 Cuesta del Chapiz, 22
 Granada 18010

VARIACIÓN LINGÜÍSTICA O LA EMERGENCIA DE NUEVAS VARIEDADES ALTAS DE LA LENGUA ÁRABE: MARRUECOS COMO ESTUDIO DE CASO¹

Montserrat BENÍTEZ FERNÁNDEZ²
Escuela de Estudios Árabes, CSIC

Resumen

El concepto de diglosia ha sido aplicado, desde sus orígenes, a la lengua árabe entre otras, siendo los trabajos fundacionales los realizados por Marçais (1930) y Ferguson (1959). Tradicionalmente, a nivel ideológico, las variedades vernáculas del árabe se han mostrado en una posición de subordinación frente al árabe clásico. Es decir, se ha propuesto el uso de un marco dicotómico para clasificar las variedades del árabe a partir de sus funciones y del prestigio de dos polos opuestos, que son el árabe clásico y los dialectos. En los últimos 50 años, diferentes estudios han mostrado lo inapropiado del concepto y han propuesto una ruptura de ese paradigma planteando distintas soluciones más ajustadas a la realidad arabófona.

Partiendo de una revisión bibliográfica en torno al concepto de diglosia, este artículo propone el análisis de otras ideologías lingüísticas que pueden aplicarse a la lengua árabe y muestra, a partir de la situación lingüística de Marruecos, cómo la competencia entre variedades prestigiosas, diferentes del árabe clásico, o estigmatizadas puede favorecer la variación y el cambio lingüístico en las diferentes variedades vernáculas de Marruecos.


Palabras clave: diglosia; continuum lingüístico; variación lingüística; Marruecos

LINGUISTIC VARIATION OR THE EMERGENCE OF NEW HIGH VARIETIES OF THE ARABIC LANGUAGE: MOROCCO AS A CASE STUDY

Abstract

Since its origins, the notion of Diglossia has been applied, among others, to the Arabic language. Marçais (1930) and Ferguson (1959) are the foundational works on this topic. In order to

1. Parte de los resultados presentados en este artículo se enmarcan en la investigación desarrollada por M. Benítez Fernández en el seno del proyecto *Variación diastrática en las variedades habladas del árabe vernáculo de Marruecos* (FFI2017-87533-P), financiado por AEI/FEDER.

2. montsebenitez@eea.csic.es.  <https://orcid.org/0000-0002-8103-1428>

classify Arabic varieties, they proposed a dichotomous framework based on functions and prestige in which Classical Arabic and the vernacular varieties represent two opposite poles. From an ideological perspective, Arabic vernaculars held a subordinated position to Classical Arabic, considered the ideal language. However, during the last 50 years, research has shown the inappropriateness of the concept, claiming a rupture of that paradigm and proposing different solutions more adjusted to the Arab-speaking reality.

This paper has a twofold purpose. On the one hand, I review the literature on the concept of diglossia and on other linguistic ideologies that can be applied to the Arabic language. On the other hand, and based on the linguistic situation in Morocco, I show how the competition between prestigious – different from Classical Arabic – and stigmatized varieties favor variation and linguistic change.

Keywords: Diglossia; Linguistic continuum; Linguistic variation; Morocco

RECIBIDO: 18/11/2022

APROBADO: 03/08/2023

1. INTRODUCCIÓN: DIGLOSIA SEGÚN FERGUSON

El primero en aplicar el término diglosia a la lengua árabe fue William Marçais (1930, p. 401), que proponía dos niveles de lengua: «el árabe escrito», que es una lengua escrita en todas partes y en todas las épocas y que, aún en la época de Marçais, se consideraba la lengua exclusiva para redactar todo tipo de textos; y «los idiomas hablados», que no han sido nunca escritos. Esta primera teorización sobre la noción de diglosia pasa bastante desapercibida en los ambientes científicos dedicados a las lenguas semíticas, siendo el trabajo que ha suscitado, y suscita aún, más reacciones el que publicó Ferguson en 1959 en la revista *Word*. En aquel trabajo, la lengua árabe se concebía como un ejemplo evidente de lo que se entiende como diglosia, es decir,

... is a relatively stable language situation in which, in addition to the primary dialects of the language (which may include a standard or regional standards), there is a very divergent, highly codified (often grammatically more complex) superposed variety, the vehicle of a large and respected body of written literature, either of an earlier period or in another speech community, which is learned largely by formal education and is used for most written and formal spoken purposes but is not used by any sector of the community for ordinary conversation.³ (Ferguson, 1959, p. 336)

3. «...es una situación relativamente estable en la que, a los dialectos (que pueden incluir un estándar nacional o varios estándares regionales)– se superpone una variedad, muy divergente y altamente

De acuerdo con esta definición, lo que se espera encontrar en el mundo arabófono es un conjunto de diferentes variedades vernáculas de árabe –incluyendo, a veces, un estándar nacional y, otras veces, varios estándares regionales–, a las que se superpone una variedad culta, escrita, prestigiosa, que cumple con los requisitos que Ferguson (1959) propone en la definición precedente. Sin embargo, como otros autores se han encargado de demostrar en numerosas investigaciones, la realidad es bastante más compleja.

En este sentido, lo que se presenta en este artículo es una síntesis de la crítica que la noción de diglosia aplicada a la lengua árabe ha recibido en los últimos años. A la luz de estas críticas, mostraré otras ideologías lingüísticas que se producen en el mundo árabe y tomaré, como caso de estudio, la situación lingüística de Marruecos para ejemplificar lo que ocurre en las diferentes variedades del árabe cuando alejamos el foco de la problemática planteada por Ferguson («variedad alta» frente a «variedad baja»). He decidido centrar el artículo en ese país porque concentra varios fenómenos que, de manera parcial, ejemplifican las situaciones lingüísticas que se producen en otros países de la región arabófona (competencia con lenguas coloniales, varias lenguas vernáculas, aparición de varios estándares regionales, etc.). Así pues, la cuestión principal a la que pretendo dar respuesta en este artículo es si el marco dicotómico que presenta la diglosia sigue siendo válido o, por el contrario, deberíamos analizar esta cuestión desde otros puntos de vista.

2. CRÍTICA DE LA DIGLOSIA Y APARICIÓN DE OTROS CONCEPTOS

A pesar de que su artículo más citado sobre este asunto es el de 1959, fue el propio Ferguson, en «Diglossia revisited» (Ferguson 1996), el que se encargó de matizar su posición inicial sobre este fenómeno. En ese texto, Ferguson pone de manifiesto cuáles son las debilidades de su trabajo inicial y propone nuevos marcos de estudio que podrían haber completado el suyo. El autor se corrige en la forma en la que aplicó el concepto a la lengua árabe, proponiendo este término no tanto como una característica del árabe, sino como una particularidad de la comunidad de habla/lingüística⁴. Sea como fuere,

codificada (a menudo gramaticalmente más compleja), que vehicula un corpus amplio y muy respetado de literatura escrita, –procedente de un periodo anterior, o de otra comunidad lingüística–, que es aprendida mediante la educación formal y usada, principalmente, por escrito o en situaciones orales de tipo formal, pero no en conversaciones cotidianas.» (Traducción de la autora)

4. Ferguson (1996, p. 55) propone utilizar los términos de «comunidad de habla» y «comunidad lingüística» de manera intercambiables con el fin de tener en cuenta las nociones de estructura y uso, relacionadas con «comunidad de habla», y las de actitudes y creencias, ligadas a «comunidad lingüística».

considerar a todo el mundo arabófono como perteneciente a una misma comunidad de habla/lingüística parece ser el error inicial, ya que las situaciones lingüísticas son muy diferentes de un país al otro: en algunos puede encontrarse una variedad vernácula que cumple la función de estándar nacional, más o menos bien aceptada; otros disponen de un nivel de alfabetización muy alto y, por tanto, de un dominio de la lengua clásica muy avanzado; en otros hay varias lenguas vernáculas (arameo, persa, turco, *amazigh*⁵...) conviviendo con el árabe –en cualquiera de sus variedades–, etc. De la misma manera que Ferguson (1959) no tiene en cuenta el conjunto del mundo germanófono para aplicar el concepto de diglosia y se concentra en la situación del alemán de Suiza, debería haber limitado su estudio a una comunidad de habla/lingüística que cumpliera con su propia definición de comunidad: «grupo social que comparte características en términos de estructura, de usos y de actitudes lingüísticas y que funciona como una unidad sociolingüística en la variación y o en el cambio lingüístico» (Ferguson, 1996, p. 55). Otra de las debilidades de Ferguson (1959) es no tener en cuenta a comunidades bilingües en las que las funciones de la variedad alta pueden ser ejercidas por una lengua, mientras que las de la variedad baja por otra (Fishman, 1967).

Las debilidades –a las que el mismo Ferguson hace referencia en su artículo de 1996– generaron también bastantes desacuerdos entre la comunidad investigadora, que no percibía la lengua árabe como diglósica, o más bien bipolar, es decir, no como dos caras de la misma moneda, cuyas variedades tienen unas funciones muy determinadas y entre las cuales no hay nada que ocupe un lugar intermedio. Por el contrario, la concebían bien como un conjunto de variedades que cumplen diferentes funciones, lo que favoreció la aparición de nuevos términos como «triglosia», «tetraglosia», «pentaglosia», «pluriglosia», «poliglosia», etc., bien como protagonista de una situación de un continuum lingüístico (Kaye, 1994; Youssi, 1995; Moustauoui,

5. El *amazigh* es la lengua de la población autóctona del Norte de África antes de la expansión del islam. Ya dialectalizada en fechas anteriores a dicha expansión, durante siglos ha sido una lengua oral minorizada, ya que la lengua árabe en su variedad escrita era la lengua para comunicarse con la administración y la lengua de instrucción en la escuela, mientras que las variedades orales del árabe servían como lingua franca entre hablantes de distintas variedades de *amazigh* ininteligibles entre sí. Desde la segunda mitad del s. XX se han sucedido los movimientos reivindicativos sobre la posición del *amazigh* en el Norte de África lo que ha favorecido una importante toma de conciencia tanto de la población como de las autoridades dando pasos paulatinos hacia su reconocimiento –su introducción en la escuela (1999), la creación del *Instituto Real de la Cultura Amazigh* (2003), etc.–. Finalmente, la constitución de Marruecos de 2011 reconoce el *amazigh* como lengua oficial, incluyéndola en la señalética pública y favoreciendo su aprendizaje. Lamentablemente, desarrollar esta cuestión excede los límites de este artículo que se va a centrar en las variedades árabes. Para profundizar más en esta cuestión, consúltese Benítez Fernández (2010) y Moustauoui (2016), entre otros.

2005), donde los niveles pasan a ser, más bien, estratos de ese todo. En cualquiera de los dos casos, esta nueva interpretación de la lengua árabe potenció la aparición de estudios que trataban de describir las variedades intermedias de ese continuo (Badaoui, 1973). Esto genera la aparición de conceptos como *Educated Spoken Arabic* que Mitchel (1986) describe como una variedad utilizada por la población altamente alfabetizada, de manera oral en situaciones formales, o una década más tarde, el *Arabe Marocain Moderne* (AMM) que Youssi (1992) describe para la realidad marroquí de forma similar: variedad formal oral diferente del árabe clásico, utilizada en contextos universitarios (debates, reuniones entre especialistas, etc.) o en los medios de comunicación (intervenciones semi-espontáneas en televisión o radio), entre otros posibles contextos formales o semiformales.

Más recientemente, tras la emergencia de todos estos conceptos –que tratan de describir la lengua, sus variedades o sus estratos– algunas investigadoras (Haeri, 2003; Hachimi, 2013; Brustad, 2017) mueven el foco de interés de los niveles lingüísticos a los hablantes como actores del discurso y como protagonistas de la elección de una variedad u otra. Se trata de una concepción del campo más propia de la antropología lingüística, que utiliza terminología propia del ámbito ideológico y que deja de entender el objeto de estudio como un elemento a describir, para concebirlo como variedades o estratos que forman parte del registro de los hablantes.

3. IDEOLOGÍA DE LA LENGUA ESTÁNDAR

A pesar de todos esos términos que han ido surgiendo en la segunda mitad del s. xx, durante décadas, el concepto de diglosia ha dominado –y sigue estando muy presente aún hoy en día– la producción del conocimiento en torno a la lengua árabe en las comunidades arabófonas (Haeri, 2003). La idea de variedad alta frente a variedad baja ha sido entendida, tanto por los hablantes como por los investigadores dentro y fuera del mundo arabófono, como una dicotomía en la que la lengua *fushhà* (clásica) sería el ideal de perfección inalcanzable y el estándar culto, mientras que los dialectos, o variedades vernáculas, son considerados como el paradigma de la imperfección, de la corrupción o de la incultura. De hecho, por esta razón, estudios recientes la han conceptualizado como una ideología lingüística. La idea de que la diglosia sea una característica de la comunidad de hablantes/lingüística (Ferguson, 1996) apoya el hecho de que este fenómeno pueda ser considerado una ideología lingüística, ya que se trata de una creencia generalizada, probablemente favorecida también por el hecho de estar ligada a la religión musulmana, más concretamente, por ser considerada la lengua de la revelación coránica.

En este sentido, Hachimi (2013) y Brustad (2017) han postulado que la diglosia aplicada a la comunidad de hablantes arabófonos se ha entendido, tanto en el mundo árabe como fuera de él, como una ideología de la lengua estándar (*standard language ideology*), es decir, una construcción social basada en la creencia de que existe una variedad consistente que es o debe ser compartida por todos: una lengua abstracta, idealizada y homogénea que se impone y se mantiene por las instituciones dominantes. Los hablantes, influidos por esta ideología, no solo piensan que es deseable, sino que es una posibilidad real (Lippi-Green, 1997) tener una lengua estandarizada y única.

El hecho de que el árabe *fushḥa* (clásico/literario/árabe estándar) sea la única lengua que se pone por escrito –o lo haya sido hasta fechas muy recientes–, que no se utilice en conversaciones cotidianas –dos de los requisitos propuestos por Ferguson para que una comunidad lingüística «sufra» o «disfrute» de la diglosia– que esté asociada al libro sagrado del islam, lo cual la dota de distancia y autoridad, así como su estatus de lengua oficial de todos los países árabes y el aparecer como símbolo del nacionalismo pan-árabe son ejemplos claros de una ideología de la lengua estándar, según explica Hachimi (2013, p. 4).

Según Brustad (2017, p. 48), un reflejo de la voluntad de protección del árabe clásico es la aparición de las distintas academias de la lengua árabe, cuya mera existencia muestra la intención de mantener la ideología de lengua estándar. Esta autora defiende que la función principal⁶ de las academias es la de salvaguardar la lengua árabe de la corrupción y de la decadencia, entendiendo la corrupción y la decadencia que representa el desarrollo de funciones de los dialectos en distintos momentos del s. XX. Por ejemplo, cuando durante la *Nahḍa* –o resurgir cultural árabe que se produce entre finales del s. XIX y principios del s. XX– hubo un cierto beneplácito de la escritura en dialecto o cuando desde la última década del siglo pasado hasta ahora, las variedades vernáculas del Norte de África ocupan lugares –normalmente restringidos al árabe clásico– en los medios de comunicación, por escrito en las redes sociales, blogs, literatura, etc. La reacción a este desarrollo de funciones en las últimas décadas puede ser la aparición de cadenas de televisión internacionales, como *al-Jazeera*, donde la única lengua de uso es el árabe clásico,

6. Según reza en la página web de la Academia de la lengua árabe de El Cairo (fundada en 1934) <https://web.archive.org/web/20070506194650/http://www.arabicacademy.org.eg/> y en el documento fundacional del *Instituto de Estudios e Investigación para la Arabización*, fundado en Rabat en 1960, sus funciones son varias: preservar la integridad de la lengua árabe; hacerla adecuado a los requisitos de la ciencia, la literatura, las artes y a la vida moderna y desarrollada; modernizar los métodos de enseñanza; estudiar los medios de adaptación de préstamos y la creación de neologismos; investigar las distintas facetas de la lengua árabe y trabajar en su difusión.

aunque esta vía necesita más investigación. Sobre la cuestión del desarrollo de funciones de las lenguas vernáculas volveremos más adelante.

Otro ejemplo del mantenimiento de la ideología de la lengua estándar y de la voluntad de proteger el árabe clásico del peligro de corrupción, puede ser la política de protección, promoción y desarrollo de la lengua árabe clásica o *fushḥà* –también conocida como arabización– que los países árabes han llevado a cabo a lo largo de todo el s. XX. Estas políticas lingüísticas pueden entenderse tanto como reacción al temor del «avance» de las variedades vernáculas –que se había producido bien a principios del s. XX con el resurgir cultural árabe o bien durante el periodo colonial–, como un refuerzo de la ideología lingüística de la lengua estándar, mediante la cual las instituciones dominantes y las élites culturales perpetúan la reproducción social (Grandguillaume, 1983) y la creencia de la supremacía del árabe clásico.

Más concretamente en los casos de Marruecos y Argelia, la política lingüística de arabización se pone en marcha tras la independencia⁷. Los momentos previos a la consecución de las diferentes independencias de las potencias coloniales o una vez instalada la autoridad local son periodos especialmente relevantes en lo concerniente a las reivindicaciones identitarias. Frente a la autoridad extranjera, las élites locales se posicionan en la búsqueda de una lengua prestigiosa que supla la impuesta por la autoridad colonial y esa lengua no puede ser otra que la *fushḥà*. De esta manera, la situación lingüística de época colonial –bilingüe o multilingüe⁸– dio paso a una administración que primaba el uso del árabe clásico y a la enseñanza en árabe clásico a lo largo de toda la segunda mitad del s. XX. Las lenguas extranjeras siguieron enseñándose solo como lenguas extranjeras desde etapas muy precoces de la enseñanza primaria.

4. OTRAS JERARQUÍAS LINGÜÍSTICAS

Además del fenómeno de la diglosia, entendido como una ideología de la lengua estándar que, como se decía, ha dominado la producción del conocimiento durante prácticamente todo el siglo XX, numerosos estudios nos demuestran que

7. Sobre ello, consúltese Benítez Fernández (2010), Moustauoui (2007) y Benrabah (2013).

8. La administración colonial se expresaba de manera bilingüe. Tanto la documentación generada por la administración como la comunicación entre administración y administrados se hacía en francés y árabe en las zonas bajo dominio francés de Marruecos y Argelia, y español y árabe en el Protectorado español en Marruecos –zona norte y la región del Sahara–. En el sistema educativo, se hacía uso de las lenguas coloniales como herramientas de instrucción y se puso en marcha la enseñanza de lenguas vernáculas –amazigh y árabe marroquí– (vid. Benítez Fernández, 2010).

otras jerarquías lingüísticas son posibles. En esta sección se muestra cómo dos de los elementos que Ferguson (1959) utiliza para caracterizar el fenómeno de la diglosia, a saber, ser la única lengua que se pone por escrito, o ser la única lengua estándar utilizada en conversaciones formales, ya no son exclusivos del árabe clásico/estándar.

4.1. *¿Árabe clásico como única lengua escrita?*

Cuando anteriormente mencionaba el árabe clásico como única lengua escrita, según proponían Marçais (1930) y Hachimi (2013), deberíamos entender, más bien, que esta variedad es la única cuya escritura no suscita críticas o preguntas o que es la única lengua escrita visibilizada. Varios autores –podríamos citar a Blau (1981), Lentin (2008) o Zack y Schipper (2012) entre los muchos que se dedican al estudio del árabe medio– han demostrado que las lenguas vernáculas y otros rasgos no estándares se han puesto por escrito en todos los periodos. De hecho, el estudio llevado a cabo por Doss y Davis (Doss y Davis 2013) demostró que, al menos en Egipto, hay una producción literaria bastante amplia que utiliza la variedad no-estándar. Estos autores han registrado estas prácticas desde, al menos, el s. xv, por lo que defienden que la producción escrita no-estándar debería empezar a considerarse como parte del registro escrito y dejar de tratarse como un catálogo de errores.

Además de este (re-)«descubrimiento» de la escritura de variedades no estándares, es obligatorio señalar el desarrollo del paso a la escritura que está viviendo el Norte de África desde, al menos, las últimas tres décadas⁹. Tanto la liberalización del panorama audiovisual marroquí –que ha promovido la aparición de canales de radio y televisión alternativos también en el uso de la lengua–, como la llegada y posterior desarrollo de las redes sociales –mensajes de texto, foros, blogs, WhatsApp, Facebook, Twitter, etc.– o la digitalización de prensa escrita –que ha permitido la inserción de reacciones de los lectores en forma de comentarios– han favorecido una relación fluida del hablante con la escritura sin que ésta tenga, necesariamente, que hacerse en árabe clásico, redefiniendo así los valores de las variedades alta y baja. En este mismo sentido, Achour (2022) cuestiona la dicotomía de variedad alta y variedad baja de Ferguson mediante el análisis de la literatura en diferentes variedades de árabe vernáculo de Túnez. Su estudio concluye que este tipo de literatura en lengua vernácula reinterpreta las categorías tradicionalmente atribuidas

9. Sobre esta cuestión, consúltese Aguadé (2013); Benítez Fernández (2009 y 2012); Caubet (2017b entre otros), Miller (2017), Pennisi (2020), etc. Lamentablemente, debido a limitaciones de espacio este listado no puede ser exhaustivo.

a las distintas variedades. Sin embargo, constata que los estudios de literatura invisibilizan la que se escribe en árabe tunecino, ya que las antologías literarias solo proponen dos categorías: literatura en árabe clásico y literatura en francés.

4.2. *¿Árabe clásico como única lengua estándar? Otras variedades/lenguas prestigiosas*

Por otro lado, también hay autores que no solo ponen en duda los postulados de la diglosia de una variedad alta frente a una o varias variedades bajas, sino que han demostrado que el árabe clásico no es la única variedad prestigiosa –otro de los principios que proponía Ferguson (1959) para definir la diglosia. Ya en los años 80 del siglo pasado, Abdel Jawad (1986) confirmó la convivencia de varias normas prestigiosas: por un lado, el árabe clásico y por otro, la aparición de una variedad urbana propia de las grandes ciudades como Amman. A partir de su estudio los especialistas empiezan a tener en cuenta la noción de prestigio y emprenden la investigación de otras variedades prestigiosas (al-Wer, 2007; Bassiouney, 2010; Walters, 2003; Sánchez y Vicente, 2012). Se trata de estándares nacionales o regionales que emergen de los diferentes procesos de urbanización, del contacto de variedades vernáculos dentro de un mismo país provocado, principalmente, por el éxodo rural, así como de procesos de nivelación o acomodación que conllevan la emergencia de estas variedades. Es lo que Miller y Falchetta (2021) llaman vernáculos neo-urbanos (*new-urban vernaculars*) y lo que Ech-Charfi (2016) ha estudiado como «proceso de desdialectalización».

Además de la eclosión de estas variedades vernáculos neo-urbanas, también se ha observado la aparición de una jerarquía entre las variedades magrebíes, o norteafricanas, y las «mashrequíes», originarias de Oriente Medio y la Península Arábiga. Es lo que Hachimi (2013, 2022) ha llamado la «ideología Magreb-Mashreq», es decir la ideología del occidente y el oriente árabófonos. Sus trabajos se basan en el análisis de valoraciones metalingüísticas y de prácticas lingüísticas en emisiones televisivas retransmitidas a todo el mundo árabe y con participantes de distintas regiones del mismo –como por ejemplo *Arab Idol*, equivalente del programa «La Voz»–. Según sus conclusiones, en el mundo árabe existen una serie de creencias que conciernen a las distintas variedades habladas. Por ejemplo: la variedad levantina es más bella, la variedad egipcia más internacionalmente comprensible, las variedades magrebíes están más alejadas de la lengua clásica por estar fuertemente influidas por el francés o el amazigh y, por tanto, son incomprensibles o, incluso, no son árabe, entre otros muchos juicios. De esta manera, los participantes de origen

magrebí en esas emisiones suelen acomodar su discurso a las variedades de oriente medio –no al árabe clásico–, con el fin de hacerse entender mejor en la emisión, llegar a un mayor número de espectadores, etc. Hachimi demuestra una jerarquía entre las variedades vernáculas con cierta supremacía de las variedades orientales sobre las occidentales. Como puede verse, los medios de comunicación favorecen el «contacto» de estándares nacionales y están beneficiando la perpetuación de esa ideología lingüística de las variedades orientales sobre las occidentales.

5. MARRUECOS COMO ESTUDIO DE CASO

En esta sección se propone el análisis de la situación lingüística de Marruecos como modelo de estudio porque reúne varias situaciones que en otros países solo se producen parcialmente. Me refiero al surgimiento, en las últimas décadas, de varios estándares regionales, a la distinta tipología de variedades prestigiosas y estigmatizadas que conviven en la región o al hecho de contar con más de una lengua vernácula (árabe vernáculo y *amazigh*), etc.

En las páginas que siguen se expone la situación lingüística de Marruecos, señalando las variedades prestigiosas y estigmatizadas y una sucinta evaluación de cómo han llegado a esa situación (véase sección 5.1); se analiza la presión que las variedades prestigiosas pueden estar ejerciendo sobre las variedades estigmatizadas (sección 5.2) y se plantea la cuestión de si la variación de género o de edad pueden ser consideradas objeto de ciertas ideologías lingüísticas de valorización o minoración similares a la tan debatida diglosia.

Para ello se han utilizado tanto datos obtenidos por medio del trabajo de campo, llevado a cabo en la ciudad de Ouezzane y en las zonas montañosas circundantes (especialmente en la tribu de Ghzaoua) durante los años 2014 al 2021, como datos aportados por estudios de otros investigadores sobre las variedades de Marruecos y su jerarquización.

5.1. *Situación lingüística de Marruecos: prestigio y estigmatización*

En Marruecos, como en cualquier otro país arabófono, encontramos junto al árabe clásico diversas variedades de árabe vernáculo. Las diferencias entre estas variedades pueden encontrarse en todos los niveles lingüísticos y tienen su origen bien en el momento de arabización de la zona, que se produce, al menos, en dos

oleadas¹⁰, por dos comunidades de habla diferentes (sedentarios-beduinos); bien en el tipo de hábitat, que puede distinguirse entre antiguas medinas y las zonas rurales de montaña –que se arabizan en la primera oleada–, y las zonas rurales del centro y sur –que se arabizan en la segunda fase–. Entre las primeras podemos citar las variedades tradicionales de las medinas de Tetuán, Fez, Meknés o Rabat y las variedades de montaña de la región de Jbala (noroeste de Marruecos), y entre las segundas se incluyen las variedades habladas en la costa atlántica y las hablas urbanas y rurales del centro y sur de Marruecos. La tercera gran diferencia radica en el ya mencionado proceso de urbanización por el cual pequeños núcleos de población se han convertido en grandes urbes a lo largo de la segunda mitad del s. xx. Es el caso de Tánger, Casablanca, Agadir o las antiguas medinas, que han visto cómo el éxodo rural se ha multiplicado, mermando la población autóctona y, por tanto, su variedad vernácula. El aumento de población ha favorecido el contacto entre variedades y con ello, la emergencia de nuevas variedades urbanas menos marcadas, es decir variedades menos «anticuadas» –como las de las antiguas medinas– o menos rurales –como las habladas por los emigrantes del campo a la ciudad–.

En este complejo mosaico de variedades, encontramos variedades prestigiosas –las variedades neo-urbanas (Miller y Falchetta, 2021) y los dialectos de las antiguas medinas– y hablas estigmatizadas –principalmente las de las zonas rurales–. Entre las variedades neo-urbanas podríamos citar el dialecto de Casablanca, considerado por algunos autores (Miller, 2006, pp. 593-597) como caso paradigmático de estándar nacional, y otros estándares regionales como la variedad norteña o la variedad de Marrakech que se resisten al desarrollo de la variedad de Casablanca (Sánchez y Vicente, 2012). En el caso concreto de Casablanca, su importante expansión hacia otras zonas de Marruecos y la aceptación de esta variedad por otros hablantes se debe a varios factores entre los que podemos contar la presencia significativa de la misma en los medios de comunicación, ya que esta ciudad es la sede de las principales

10. La primera oleada se produce durante la expansión del islam en los ss. VII-VIII. Estos contingentes de población se asientan en los núcleos ya urbanizados en época romana localizados, principalmente, en las zonas costeras del norte de África y en algunas ciudades del interior, por ejemplo, *Volubilis* en el caso de Marruecos. Estos núcleos y las vías de comunicación entre ellos son un primer contacto con la lengua árabe y favorecerán la arabización de estas zonas. La segunda oleada se produce a partir de la expulsión de Egipto de las tribus beduinas conocidas como *Bānu Hilāl* (entre las que se incluyen los *Bānu Hilāl* propiamente dichos, los *Bānu Sulaym* y los *Bānu Maʿqil*). Su llegada al norte de África está datada en s. XII, completando la arabización iniciada en los primeros siglos de la expansión del islam, ya que se sitúan en zonas desocupadas u ocupadas por población autóctona (población *amazigh*). En Marruecos concretamente, se localizan en la costa atlántica desde el sur de Larache, en las franjas áridas del este de Marruecos y las llanuras del centro y sur del país.

cadena de radio y televisión, y el hecho de haberse convertido en un polo cultural –mediante producciones cinematográficas y audiovisuales, movimientos pictóricos y escultóricos contemporáneos– y de cultura *underground* –multiplicación de grupos de música contemporánea (rap, fusión, rock, punk, etc.), de festivales de música, etc.–. cuyas actividades se vehiculan mayoritariamente en esta variedad¹¹. Por otro lado, tanto el de Casablanca como el resto de estándares regionales mencionados representan a los hablantes de las ciudades económicamente más florecientes de Marruecos (Tánger/Tetuán, Casablanca, Marrakech), ciudades que, a su vez, cuentan con un mayor número de hablantes. Todos estos son, sin duda, factores influyentes para que estas variedades se consideren prestigiosas.

Por su parte, las variedades de las antiguas medinas han sido tradicionalmente consideradas como prestigiosas ya que sus hablantes representaban un mayor grado de refinamiento, educación y cultura. Estas variedades se corresponden con la primera oleada de arabización de Marruecos. Sus hablantes reclaman un origen andalusí, como se demuestra en Vicente y Naciri (2017), y sus variedades habladas muestran una cierta influencia del árabe andalusí, como han demostrado los estudios elaborados por Ferrando (1998) y Vicente (2010), variedad que históricamente ha gozado de prestigio entre los arabófonos.

También hay diferentes tipos de variedades estigmatizadas (en concreto, las variedades de montaña del norte de Marruecos –de tipo sedentario igual que las variedades de las antiguas medinas, pero estigmatizadas por la situación socioeconómica, el aislamiento, la pobreza y el analfabetismo de la población–) y las variedades rurales del resto del país –principalmente de tipo beduino y también estigmatizadas por, prácticamente las mismas razones que las variedades del norte–. Ejemplo de esta estigmatización son las imitaciones que estos acentos suscitan entre hablantes urbanos –sedentarios o urbanizados– de otras regiones de Marruecos e, incluso, en series de televisión, vídeos en redes sociales, etc. destinados a menospreciar o, directamente, burlarse de los hablantes que usan esas variedades, como ha quedado demostrado en Falchetta (2019a y b) y en Benítez Fernández y Guerrero (2022), entre otros estudios dedicados a profundizar en los estereotipos de las variedades habladas de Marruecos. Entre estas variedades estigmatizadas deberíamos señalar también a las *amazighes*, lenguas orales de uso cotidiano, pero hasta fechas muy recientes, sin ningún tipo de reconocimiento oficial ni nacional.

11. Sobre esta cuestión consúltese, por ejemplo, Caubet 2005, 2017a; Caubet y Miller 2013.

5.2. Jerarquización de las variedades árabes en Marruecos

En este momento, cabría preguntarse qué influencia ejercen los estándares regionales –como ejemplo de variedades prestigiosas– en el resto de variedades. No cabe duda de que la expansión de las variedades neo-urbanas o de los estándares regionales está generando una progresiva pérdida del prestigio de los dialectos de las antiguas medinas (Messaoudi, 2003). Como ha demostrado Vicente (2021), éstas últimas disfrutaban en la actualidad de una situación ambivalente, pues son consideradas variedades prestigiosas entre los miembros de la comunidad de hablantes, pero resultan anticuadas o, incluso, pedantes para otros hablantes.

Si tenemos en cuenta la variedad considerada alta (el árabe clásico), hay que señalar también una cierta competencia entre los estándares neo-urbanos y esta variedad. Esto se debe a que los estándares neo-urbanos están ocupando espacios tradicionalmente reservados al árabe clásico. Como se ha avanzado anteriormente, las variedades orales prestigiosas (tanto los estándares neo-urbanos como otras variedades mixtas –por ejemplo, el denominado *AMM*–) son las preferidas para las conversaciones orales formales, intervenciones espontáneas o debates en radio y televisión y para usos por escrito en redes sociales, en parte de la prensa escrita (Pennisi, 2020) y en parte de la literatura que se está publicando últimamente (Ferrando, 2012; Agudé, 2013, entre otros). El desarrollo de funciones que las variedades vernáculas están conociendo pone en tela de juicio la dicotomía de una única variedad alta frente a las variedades bajas y rompe con el paradigma propuesto por la diglosia.

De la misma manera, los estándares regionales están ejerciendo presión sobre las variedades no prestigiosas o estigmatizadas. Esta presión se observa, por ejemplo, en la variación, atrición y/o pérdida de diferentes rasgos que caracterizan a estas variedades menos valoradas, así como en la introducción de rasgos exógenos a esas variedades.

5.2.1. Variación

La presión de variedades prestigiosas también se observa en la aparición de rasgos propios de otras variedades que alternan con los rasgos locales. Por ejemplo, en la variedad vernácula de la región de Ghomara (oeste de Tetuán), Naciri-Azzouz (2020, pp. 149-150) señala el uso generalizado de la partícula *māši* –y sus variantes *māš* y *ma*– seguida de un verbo en imperfectivo para expresar el futuro próximo. Sin embargo, junto a esta forma, sus informantes utilizan también la partícula *ḡādi* o *ḡa*-, común en el estándar de Casablanca, con el mismo valor. Otro ejemplo de este tipo de variación podría ser la alternancia entre el enmudecimiento de la fricativa

glotal sorda /h/, que es característico en las variedades del norte. Este fenómeno ocurre, especialmente, en los pronombres sufixados de 3ª persona masculino, femenino y plural, siempre que estén anexados a una palabra que termina en consonante (Cantineau, 1960, p. 75). Sin embargo, se ha observado en el habla de Ouezzane que este enmudecimiento alterna con el mantenimiento de la fricativa glotal sorda /h/ en esos mismos contextos, que sería la norma en variedades más al sur, como puede verse en (1).

(1) *dyāl-a / dyāl-ha* ‘de ella’; *ʕāṭit-um / ʕāṭit-hum* ‘les di’

Además de la presión que las variedades estándares o prestigiosas puedan ejercer sobre los dialectos estigmatizados o no marcados, es posible que la alternancia de distintas variantes indique no tanto la influencia de una variedad sobre otra, sino, más bien la colisión de dos variedades prestigiosas: por un lado, el estándar nacional del eje Casablanca-Rabat y por otro el habla norteña¹². Esta variación puede observarse en algunos rasgos de ciudades periféricas del norte de Marruecos como Ouezzane (Benítez Fernández, 2016b; 2019) o Larache (Guerrero Parrado, 2015). En estas variedades se usa mayoritariamente la uvular sorda [q], variante característica de variedades neo-urbanas, de antiguas medinas y del árabe clásico, todas ellas prestigiosas. Sin embargo, se han introducido términos que emplean la variante velar sonora [g], propios del estándar de Casablanca. Esto ocurre, especialmente, en el verbo *gāl-ygūl* ‘él dijo-él dirá’ que en las variedades del norte debería ser *qāl yqūl*. Probablemente, aunque sería necesario desarrollar la investigación, se trate de una forma de préstamo léxico o un tipo de variación inducida por la elección del léxico, como ha ocurrido ya entre los fesíes residentes en Casablanca. A propósito de este cambio, Hachimi (2012) explica que los fesíes de Casablanca lo consideran un rasgo identificativo de la variedad casablanquesa, ligado al ingenio y la rapidez que caracteriza a estos últimos. La evaluación positiva del rasgo ha debido motivar la adopción del mismo también en los hablantes de Ouezzane y Larache.

En el caso de los determinantes, la variedad estándar norteña utiliza la forma invariable de lejanía *dīk* ‘ese/aquel-esa/aquella-esos/aquellos’, mientras que la variedad de Casablanca muestra una distinción de género y número en esta forma: *dāk* ‘ese/aquel’ *dīk* ‘esa/aquella’, *dūk* ‘esos/aquellos’. De nuevo en Ouezzane encontramos la convivencia de ambas normas en hablantes que utilizan una u otra sin

12. Sobre la llamada «habla norteña», que se considera también un estándar regional, consúltese Sánchez y Vicente (2012).

que variables como el género o la edad puedan explicar esta variación. A falta de otra explicación más convincente, debemos entender que Ouezzane, capital de provincia que se encuentra en el límite sur oeste de la región norte y, por tanto, a caballo entre dos variedades prestigiosas, está construyendo su variedad urbanizada con elementos de ambas variedades: la norteña y la de Casablanca.

5.2.2. *Atrición*

Como ejemplo de atrición, se puede citar la oclusiva glotal [ʔ] –variante de /q/–, un rasgo considerado característico de la medina en Tetuán (Vicente, 2021), de los artesanos de Fez (Caubet, 1993), de la población de Ouezzane (Khoukh, 1993; El Khomsi, 2017) y propio del repertorio *jebli* en variedades de montaña¹³. El paulatino desuso de este rasgo se ha constatado entre los hablantes de la tribu de Ghzaoua, –tribu *jebli* localizada entre el suroeste de la provincia de Chef Chauen y el noreste de la provincia de Ouezzane– y una pérdida total del mismo entre los hombres altamente alfabetizados y las mujeres de Ouezzane. Se mantiene, en variación con la uvular sorda [q], entre los artesanos de la ciudad –fundamentalmente poco o nada alfabetizados– y el grupo de mujeres especialmente ligadas a la montaña, bien por haber llegado a la ciudad en fechas recientes, bien por no mostrar –ni en su forma de vestir, ni en sus relaciones, ni en sus actividades– una completa adaptación a la vida urbana (Benítez Fernández, 2023; Benítez Fernández y Guerrero, 2022). El mismo proceso de atrición está sufriendo la velar sonora [g] en el dialecto de Oujda –noreste de Marruecos–. Esta variedad vernácula ha sido clasificada tradicionalmente como perteneciente a los dialectos de tipo beduino o *hilālī* (Colin, 1945), entre otras razones por la realización velar sonora [g] de la uvular sorda /q/ (Behnstedt y Benabbou, 2005). Sin embargo, en fechas recientes se ha observado que la velar sonora [g] se mantiene exclusivamente en los verbos *gāl-ygūl* ‘él dijo–él dirá’ y *lga yalga* ‘él encontró–él encontrará’ (Benítez Fernández, 2017). La atrición de la oclusiva glotal [ʔ] y de la velar sonora [g] benefician el empleo de la uvular sorda /q/ que es la variante generalizada en las variedades de prestigio neo-urbanas, en parte de las variedades de las antiguas medinas y en árabe clásico, todas ellas consideradas prestigiosas, como se ha dicho más arriba.

El mismo fenómeno de atrición y/o pérdida ocurre con las variantes espirantizadas ([t̪] y [d̪])¹⁴ de los fonemas dentales sordo y sonoro (/t/ y /d/). Estos alófonos

13. Consúltense los capítulos publicados en Vicente, Caubet y Naciri-Azzouz (2017).

14. Cuando se mencionan las consonantes espirantizadas [t̪] y [d̪] no se hace referencia a las consonantes interdentales tradicionales del árabe clásico, desaparecidas en la mayoría de variedades habladas, sino

interdentales son igualmente característicos de las variedades sedentarias de montaña, pero hoy en día sufren una cierta variación en beneficio de sus correspondientes fricativo y oclusivo (/t̪/ y /d/). De hecho, entre los hablantes de la tribu de Ghzaoua solo los hablantes mayores de 60 años utilizan sistemáticamente las variantes espirantizadas; el resto de informantes bien utiliza ambos alófonos, bien ha dejado de utilizar por completo la variante espirantizada.

5.2.3. *Introducción de nuevos rasgos*

Por su parte, Hachimi (2007, 2011 y 2018) ha observado la evolución de las variedades de Fez, Ben Rachid (zona rural cuya variedad es de tipo beduino que se encuentra al este de Casablanca) y Tafilalt (sur este de Marruecos), cuyos hablantes (residentes en Casablanca) han acomodado su forma de hablar a la de Casablanca –en concreto entre las generaciones de migrantes no nativas en Casablanca– o han modificado sus variedades de origen –entre los hablantes de 3ª generación y, por tanto, ya nativos en la ciudad– en beneficio de la variedad de Casablanca. Ejemplos de acomodación son, por ejemplo:

- (2) La realización uvular vibrante [ʁ] típica de Fez que ha evolucionado hacia la variante /r/ propia de Casablanca –y de otras variedades de Marruecos–, ya que los hablantes evalúan este rasgo como especialmente afeminado (Hachimi, 2011, p. 30).
- (3) La adopción de la variante oclusiva velar sonora [g], común en la variedad de Casablanca, en lugar de la oclusiva uvular sorda [q] o de la oclusiva glotal [ʔ] propias de la variedad de Fez, ya que los hablantes califican la [g] como un rasgo que aporta dureza (Hachimi, 2011, p. 35).
- (4) Las distintas formas que las tres comunidades (Fez, Ben Rachid y Tafilalt) tienen de adaptarse a la distinción de género en la 2ª persona singular de la variedad de Casablanca (Hachimi, 2018, p. 72 y ss).

5.2.4. *Valoración metalingüística*

No solo las prácticas lingüísticas muestran cierta presión de las variedades neo-urbanas sobre las variedades menos prestigiosas. Los juicios metalingüísticos de los hablantes revelan un cierto menosprecio de sus propias hablas en beneficio

a una espirantización de segundo grado a partir de sus correspondientes alófonos africados que ocurre en variedades de Marruecos y Argelia, probablemente por influencia del sustrato bereber (Vicente, 2000, p. 41). Sobre las interdentales y espirantizadas consúltese Guerrero (2021 y 2023).

de las variedades prestigiosas. Ejemplos de estos juicios metalingüísticos que favorecen a variedades neo-urbanas son:

- (5) Se explicó a un participante en la investigación el tema de estudio y el interés por las variedades vernáculas. La respuesta del informante fue «sí, soy de aquí, pero mi habla está muy influenciada por el habla de Tánger»¹⁵.
- (6) Se preguntó a una informante nativa de la ciudad de Ouezzane¹⁶ por el uso de la variante oclusiva glotal que previamente se había constatado, mediante observación participante, entre miembros varones de su familia. La informante respondió que ese rasgo no era local y que solo lo realiza la gente del campo.
- (7) Se preguntó a 3 informantes varones¹⁷ si podían narrar cuentos o refranes con el fin de documentar esa parte del patrimonio oral marroquí. Durante la grabación los informantes mencionaban los refranes, recreaban un posible contexto de enunciación (Benítez Fernández, 2016a) y hacían otro tipo de comentarios relativos al supuesto origen de los refranes (el refrán de los holgazanes, los refranes de las bodas, etc.). En más de una ocasión los informantes identificaban los refranes como originarios de la región montañosa de Jbala. En esas ocasiones, uno de los hablantes –recordemos que se trata de hablantes altamente alfabetizados y cuya variedad de habla es de tipo neo-urbano–, aprovechaba para imitar¹⁸, de forma jocosa, los rasgos más característicos de los hablantes de la montaña y mostrar lo que para ellos era el catálogo de particularidades a evitar.

Estos ejemplos demuestran, por un lado, una clara inseguridad lingüística (Labov, 1972; Bennis, 2003; Benítez Fernández y Guerrero, 2022) de los hablantes hacia su propia variedad y, por otro, la consciencia lingüística de los hablantes a propósito

15. El participante es hombre, en el momento de la entrevista tenía 39 años, es analfabeto y residente en una pequeña aldea al sur de la provincia de Chaouen. Aunque ha pasado breves temporadas de su vida en la ciudad de Tánger (algunos periodos vacacionales motivados por bodas u otros eventos familiares), ha vivido toda su vida en esa misma ubicación.

16. Informante de unos 40 años, mujer y analfabeta. Ha residido principalmente en la ciudad de Ouezzane salvo durante un periodo de 5 años en el que ella y su marido se trasladaron a la ciudad de Tánger por razones laborales.

17. Se trata de 2 jóvenes varones (alrededor de 20 años) y un hombre (aproximadamente 30 años), todos ellos altamente alfabetizados. En el momento de la grabación solo uno de los jóvenes no había salido aún de Ouezzane, los otros dos estaban estudiando en la Universidad de Kenitra (costa atlántica de Marruecos) donde pasaban parte del año.

18. Según Trudgill (1984, p. 12) la imitación se produce en diferentes eventos discursivos (contar chistes, teatralizar a diferentes personajes, etc.) durante los que afloran los rasgos que los hablantes consideran como más evidentes de la variedad imitada.

de la existencia de variedades o rasgos estigmatizados y el importante papel que ello juega en la variación y el cambio lingüístico (Trudgill, 1984, p. 11; Hachimi, 2018).

Por otro lado, frente a estos casos de expansión de variedades neo-urbanas en variedades menos prestigiosas o de colisión de estas variedades prestigiosas, algunos estudios muestran también un cierto movimiento de resistencia. Por ejemplo, Hachimi (2018, pp. 76-79), en su investigación ya citada sobre varias comunidades de migrantes (de Fez, Berrachid y Tafilalt) residentes en Casablanca, muestra que las 4 generaciones de origen fesí todavía conservan algunos rasgos propios de su dialecto de herencia, entre otros, la confusión de género en el pronombre sujeto y en la conjugación verbal de segunda persona singular¹⁹, a lo largo de 4 generaciones. Otra muestra de resistencia podría ser la reinterpretación social de la realización africada con apéndice chicheante [tʃ] de la dental oclusiva /t/, tal y como se muestra en un estudio sobre el argot juvenil en el norte de África. En esa publicación, Ziamari, Caubet, Miller y Vicente (2020) señalan que la africación (con apéndice sibilante) de dentales sordas es la realización habitual en las variedades vernáculas del norte de Marruecos, mientras que es un rasgo considerado popular o vulgar en grandes ciudades como Casablanca o Meknés. Sin embargo, otro tipo de africación, la realizada con apéndice chicheante, está presente en el habla propia de jóvenes varones participantes de la cultura *underground* lo que puede considerarse un elemento de construcción de la identidad de grupo y proporciona una especie de prestigio encubierto a este rasgo, favoreciendo su diseminación²⁰.

De la misma manera, se podría señalar la evolución de rasgos fonéticos que no siguen forzosamente las normas de las variedades prestigiosas. Me refiero al debilitamiento de la fricativa velar sonora /g/ en fricativa faringal sonora [ʕ], en fricativa glotal sorda [ħ] y finalmente su desaparición.

- (8) *ġā[di] > ʕā[d] > ʕa > a* ‘marca de futuro próximo’; *ġayr > ġay > ʕay > ʕe > he* ‘solamente’

Este rasgo se ha observado en diferentes lugares tanto pequeñas ciudades del norte como Ouezzane (Benítez Fernández, 2019) o Larache (Guerrero Parrado, 2015), como

19. Me refiero aquí a que, de manera general, las variedades árabes distinguen género en la 2ª persona tanto en los pronombres sujeto, en los pronombres complemento como en la conjugación verbal (en perfecto, imperfectivo e imperativo). Es decir que hay dos formas diferentes de decir «tú» (*nta* en masc. y *nti* en fem.), ‘para ti’ (*la-ka* en masc. y *la-ki* en fem.) o ‘escribiste’ (*katabta* en masc. y *katabti* en fem.). Sin embargo, algunas variedades han evolucionado hacia una pérdida de la distinción de género en esa persona como es el caso de Fez y las variedades sedentarias del norte de Marruecos entre otras.

20. Sobre este asunto, además del ya mencionado trabajo de Ziamari, Caubet, Miller y Vicente (2020), consúltese Benítez Fernández (2023).

en la comunidad virtual de Casablanca²¹, por lo que podría deberse a una evolución fonética y no a causas extralingüísticas.

6. CONCLUSIÓN

Este artículo ha tratado de sintetizar las aportaciones recientes que los especialistas de la lengua árabe han hecho a la noción de diglosia. La principal cuestión a la que se pretendía dar respuesta era la oportunidad de mantener o no el uso del paradigma dicotómico de la diglosia.

Es bien sabido que han corrido ríos de tinta en torno a esta cuestión y que el trabajo de Ferguson (1959) ha sido fuertemente criticado, al menos en lo que concierne a la lengua árabe. De ahí que numerosos estudios hayan proporcionado nuevas visiones de lo que ocurre a propósito de las variedades de la lengua árabe –entendida, más bien, como un continuo– o se hayan centrado en la descripción de los niveles intermedios existentes entre la variedad alta y la baja, aportando nueva terminología –como «triglosia», «pentaglosia», «pluriglosia», «*Educated Spoken Arabic*», «*Arabe Marocain Moderne*», entre otras muchas–²².

Según el estado de la cuestión elaborado en este estudio, parece evidente que el concepto diglosia debe ser entendido como una ideología lingüística de la lengua estándar (sección 3), ya que se basa en la creencia generalizada de la supremacía de una lengua frente a otra(s) u otra(s) variedad(es). Por otro lado, las prácticas lingüísticas expuestas en este estudio –analizadas tanto por la autora como por otros investigadores– han demostrado que la ideología de la lengua estándar no es la única ideología lingüística que puede aplicarse a la lengua árabe –en todo su conjunto de variedades–, ni la única a la que se está confrontando esta lengua. De hecho, estudios sobre documentos escritos en árabe medio o sobre producción literaria contemporánea (sección 4.1) han demostrado la reinterpretación de las variedades alta y baja²³. Estas investigaciones han señalado cómo otras variedades no estándares son igualmente utilizadas como posibles formas escritas, minorizando la exclusividad del árabe clásico como única lengua escrita. Las investigaciones sociolingüísticas llevadas a cabo en diferentes centros urbanos han demostrado que

21. Agradezco aquí a la Dra. D. Caubet que nos puso en la pista de este rasgo durante su comunicación en el congreso de la *Asociación Internacional de Dialectología Árabe* que tuvo lugar en Doha en 2013. Lamentablemente su conferencia no está publicada, por lo que me veo en la imposibilidad de citar el texto completo.

22. Sobre todo ello, consúltese Iriarte Díez *et al.* en este volumen.

23. A propósito de este asunto consúltese, igualmente, Valenzuela Mochón en este volumen.

el árabe clásico no es ni la única variedad estándar ni la única prestigiosa y que otras variedades vernáculas pueden ser más apropiadas para registros orales formales o semi-formales como pueden ser los medios de comunicación o los intercambios orales entre profesionales, etc. (sección 4.2). También se ha mostrado cómo del análisis del contexto pan-árabe emerge una jerarquía entre variedades vernáculas de oriente y de occidente que, de nuevo, en ningún momento plantea al árabe clásico como vehículo de comunicación (sección 4.2). Finalmente, (sección 5) se ha puesto de manifiesto por qué unas variedades vernáculas gozan de prestigio mientras otras sufren la estigmatización; cómo la urbanización ha favorecido el contacto entre variedades vernáculas beneficiando la aparición de estándares nacionales/regionales que los hablantes dotan de prestigio o cómo las variedades prestigiosas pueden ejercer cierta presión sobre el resto de variedades o, incluso, pueden poner en peligro la diversidad lingüística –en el estudio de caso analizado sobre Marruecos–. Los resultados expuestos apuntan que, por el momento, parece que tanto el contacto de variedades como la colisión de variedades prestigiosas favorecen la variación y el cambio lingüístico lo que se traduce, por ahora, en procesos de variación, acomodación o nivelación –aunque también de atrición o pérdida– pero que las variedades vernáculas (prestigiosas o no) todavía están completamente vivas y hacen frente a la supremacía única de cualquier variedad.

Tanto el estudio de las ideologías lingüísticas, como el análisis de la variación demuestran, por tanto, una clara obsolescencia del término diglosia en beneficio de otros procesos como la urbanización, la aparición de variedades vernáculas neo-estándar, que apuntan, más bien a una situación de estándar (o estándares) con dialectos.

REFERENCIAS

- Abdel Jawad, H. R. (1986). The emergence of an urban dialect in the Jordanian urban centres. *International Journal of the Sociology of Language* 61, 53-63.
DOI: <https://doi.org/10.1515/ijsl.1986.61.53>
- Achour, M. (2022). La diglossie traversée. La littérature en tunisien et le tunisien dans la littérature. *International Journal of Sociology of Language* 278, 23-50.
DOI: <https://doi.org/10.1515/ijsl-2022-0018>
- Agudé, J. (2013). Des romans diglossiques : le cas de Youssef Fadel. En M. Benítez Fernández, C. Miller, J.J. de Ruiter y Y. Tamer (Eds.), *Évolution des pratiques et représentations langagières dans le Maroc du xx^e siècle*, vol. 1, pp. 57-149. París: L'Harmattan.
- Al-Wer, E. (2007). The formation of the dialect of Amman: from chaos to order. En C. Miller, E. Al-Wer, D. Caubet y J. C. E. Watson (Eds.), *Arabic in the city: Issues in dialect contact and language variation*, pp. 55-76. Londres: Routledge.

- Badawi, E. A. (1973). *Mustawayāt al-lughā al-‘arabiyya al-mu‘asira fi miṣr [Levels of Contemporary Arabic in Egypt]*. Cairo: Dar al-Ma‘arif.
- Bassiouney, R. (2010). Identity and code-choice in the speech of educated women and men in Egypt: Evidence from talk shows. En R. Bassiouney (Ed.), *Arabic and the media: linguistic analyses and applications*, pp. 97-121. Leiden-Boston: Brill. DOI: <https://doi.org/10.1163/ej.9789004182585.i-303.50>
- Behstedt, P. y Benabbou, M. (2005). Données nouvelles sur les parlers arabes du Nord-Est marocain. *Zeitschrift für Arabische Linguistik* 44, 17-70.
- Benítez Fernández, M. (2009). Árabe marroquí como proyecto editorial: ¿es una experiencia posible? En L. Abu Shams (Ed.), *Actas del Tercer Congreso Internacional de Árabe Marroquí: estudio, enseñanza y aprendizaje*. Vitoria-Gasteiz, 8 y 9 de mayo de 2008, pp. 37-54. Vitoria: Universidad del País Vasco.
- Benítez Fernández, M. (2010). *La política lingüística contemporánea de Marruecos: de la arabización a la aceptación del multilingüismo*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.
- Benítez Fernández, M. (2012). *Al-Amal*. Otro intento fallido de escribir en *dārīẓa* marroquí. En A. Barontini, Ch. Pereira, Á. Vicente y K. Ziamari (Eds.), *Dynamiques langagières en arabophonies: variations, contacts, migrations et créations artistiques. Hommage offert à Dominique Caubet par ses élèves et collègues*, pp. 379-391. Zaragoza: Pressas Universitarias de Zaragoza.
- Benítez Fernández, M. (2016a). Refranes marroquíes recopilados en Ouezzane (Norte de Marruecos). Un análisis desde un contexto de enunciación recreado. *Paremia* 25, 171-181.
- Benítez Fernández, M. (2016b). Notes sur le sociolecte des jeunes d'Ouezzane (Nord du Maroc). En G. Grigore y G. Bituna (Eds.), *Arabic varieties: Far and wide. Proceedings of the 11th International Conference of AIDA - Bucharest, 2015*, pp. 99-105. Bucarest: Editura Universitatii din București.
- Benítez Fernández, M. (2017). Variation linguistique dans le parler arabe d'Oujda: Nord-est marocain vs. communauté marocaine dans la diaspora. *Dialectologia* 18, 1-18.
- Benítez Fernández, M. (2019). Une approche sociolinguistique du parler arabe de Ouezzane depuis la perspective du genre. En C. Miller, A. Barontini, M.A. Germanos, J. Guerrero y Ch. Pereira (Dirs.), *Studies on Arabic dialectology and sociolinguistics - Proceedings of the 12th International Conference of AIDA, held in Marseille from 30th May-2nd June 2017* [en línea], pp. 370-379. Aix-en-Provence: IREMAM. DOI: <https://doi.org/10.4000/books.iremam.4252>
- Benítez Fernández, M. (2023). Linguistic variation, social meaning and covert prestige in a Northern Moroccan Arabic variety. *Languages* 8(1). DOI: <https://doi.org/10.3390/languages8010089>
- Benítez Fernández, M. y Guerrero, J. (2022). The Jebli speech between the media and the city: exploring linguistic stereotypes on a rural accent in Northern Morocco. *International Journal of the Sociology of Language* 278, 181-202. DOI: <https://doi.org/10.1515/ijsl-2022-0015>
- Bennis, S. (2003). Accent arabe et insécurité linguistique (cas de l'accent arabe du Tadla: Centre du Maroc). En I. Ferrando y J.J. Sánchez Sandoval (Eds.), *AIDA 5th proceedings*, pp. 349-362. Cádiz: Universidad de Cádiz.

- Benrabah, M. (2013). *Language conflict in Algeria: From colonialism to post-independence*. Berlín: De Gruyter.
- Blau, J. (1981). *The emergence and linguistic background of Judaeo-Arabic: A study of the origins of Middle Arabic*. Jerusalem: Ben-Zvi Institute for the Study of Jewish Communities in the East.
- Brustad, K. (2017). Diglossia as ideology. En J. Høigilt y G. Mejdell (Eds.), *The politics of written language in the Arab World. Writing change*, pp. 41-67. Leiden-Boston: Brill. doi: https://doi.org/10.1163/9789004346178_004
- Cantineau, J. (1960). *Cours de phonétique arabe*. París: C. Klincksieck.
- Caubet, D. (1993). *L'Arabe marocain: Syntaxe et catégories grammaticales, textes*. París-Lovaina: Peeters.
- Caubet, D. (2005). Generation *darija!* *Estudios de Dialectología Norteafricana y Andalusí* 9, 233-243.
- Caubet, D. (2017a). New elaborate written forms in Darija. Bloggin, posting and slamming in Morocco. En E. Benmamoun y R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics*, pp. 387-406. Nueva York: Routledge.
- Caubet, D. (2017b). Morocco: An informal passage to literacy in *dārija* (Moroccan Arabic). En J. Høigilt y G. Mejdell (Eds.), *The politics of written language in the Arab world. Writing change*, pp.116-141. Leiden-Boston: Brill.
doi: https://doi.org/10.1163/9789004346178_007
- Caubet, D. y Miller, C. (2013). Du rock au Maroc. Quelle place dans la nouvelle scène urbaine casablancaise? En L. Bonnefoy y M. Catusse (Eds.), *Jeunesses arabes. Du Maroc au Yémen: loisirs, culture et politique*, pp. 342-354. París: La Découverte
- Doss, M. y Davis, H. (2013). *al-ʿAmmiyya l-miṣriyya l-maktūba* [El dialecto egipcio escrito], Cairo: The General Egyptian Book Organization.
- Ech-Charfi, A. (2016). *The standardization of a diglossic low variety: The case of Moroccan Arabic*. Saarbrücken: Scholar's Press.
- El Khomsi, R. (2017). Étude des particularités linguistiques du parler *žabli* de la région d'Ouezzane. En A. Vicente, D. Caubet y A. Naciri-Azzouz (Eds.), *La région du Nord-Ouest marocain : Parlers et pratiques sociales et culturelles*, pp. 159-175, Zaragoza: Prensas de la Universidad de Zaragoza.
- Falchetta, J. (2019a). Better sound rural or criminal? Data from a case study: the city of Temara, Morocco. En C. Miller, A. Barontini, M. A. Germanos, J. Guerrero, Ch. Pereira (Dirs.), *Studies on Arabic dialectology and sociolinguistics. Proceedings of the 12th International conference of AIDA held in Marseille from May 30th to June 2nd 2017* [en línea], pp. 458-466. Aix-en-Provence: IREMAM.
<https://books.openedition.org/iremam/4395?lang=es> (consultado el 24 de junio de 2023). doi: <https://doi.org/10.4000/books.iremam.4395>
- Falchetta, J. (2019b). *The social connotations of linguistic variation in a Moroccan urban context: the case of Temara*. Tesis doctoral. Aix-en-Provence: Aix-Marseille Université-Institut des mondes arabes et musulmans.
- Ferguson, Ch. A. (1959). Diglossia, *Word* 15: 2, 325-340.
doi: <https://doi.org/10.1080/00437956.1959.11659702>
- Ferguson, Ch. A. (1996). Epilogue: Diglossia revisited. En A. Elgibali (Dir.) *Understanding Arabic: Essays in contemporary Arabic linguistics in honor of El-Said Badawi*, pp. 49-67. Cairo: American University of Cairo Press.

- Ferrando, I. (1998). On some parallels between Andalusí and Maghrebi Arabic. En J. Aguadé, P. Cresier y Á. Vicente (Eds.), *Peuplement et arabisation au Maghreb Occidental*, pp. 59-74. Madrid-Zaragoza: Casa de Velázquez-Universidad de Zaragoza.
- Ferrando, I. (2012). Apuntes sobre el uso del dialecto en la narrativa marroquí moderna. En M. Meouak, P. Sánchez y Á. Vicente (Eds.), *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*, pp. 349-358. Zaragoza: Universidad de Zaragoza.
- Fishman, J. (1967). Bilingualism with and without diglossia; Diglossia with and without bilingualism. *Journal of Social Issues* 23 (2), 29-38.
DOI: <https://doi.org/10.1111/j.1540-4560.1967.tb00573.x>
- Guerrero Parrado, J. (2015). *El dialecto árabe hablado en la ciudad marroquí de Larache*. Zaragoza: Prensa de la Universidad de Zaragoza.
- Guerrero, J. (2021). On interdental fricatives in the first-layer dialects of Magrebi Arabic. *Brill's Journal of Afroasiatic Languages and Linguistics* 13, 288-308.
- Guerrero, J. (2023). Sur le traitement des fricatives interdentes dans les parlers arabes contemporains : entre conservation et disparition. *Quaderni di Studi Arabi* 18 <https://brill.com/view/journals/qsaa/aop/article-10.1163-2667016x-18010217/article-10.1163-2667016x-18010217.xml> (consultado el 24/07/2023).
- Grandguillaume, G. (1983). *Arabisation et politique linguistique au Maghreb*. París: Maisonneuve et Larose.
- Hachimi, A. (2007). Becoming Casablančan: Fessis in Casablanca as a case study. En C. Miller, E. Al-Wer, D. Caubet y J. Watson (Eds.), *Arabic in the city: Issues in language variation and change*, pp. 97-122. Londres-Nueva York: Routledge-Taylor & Francis.
- Hachimi, A. (2011). Réinterprétation sociale d'un vieux parler citadin maghrébin à Casablanca. *Langage et Société* 138, 21-42.
- Hachimi, A. (2012). The urban and the urbane: Identities, language ideologies, and Arabic dialects in Morocco. *Language in Society* 41(3), 321-341.
- Hachimi, A. (2013). The Maghreb-Mashreq language ideology and the politics of identity in a globalized Arab world. *Journal of Sociolinguistics* 17(3), 269-296. DOI: <https://doi.org/10.1111/josl.12037>
- Hachimi, A. (2018). Arabic dialect contact and change in Casablanca: The role of complexity and salience in the adoption of a morphosyntactic variable. *Anthropological Linguistics* 60(1), 1-34.
- Haeri, N. (2003). *Sacred language, ordinary people: State, religion, and modernity in Egypt*. Nueva York: Palgrave Macmillan.
- Kaye, A. S. (1994). Formal vs. informal in Arabic: Diglossia, triglossia, tetraglossia, etc., polyglossia – multiglossia viewed as a continuum. *Zeitschrift für Arabische Linguistik* 27, 47-66.
- Khoukh, A. (1993). *El habla jebli de la ciudad de Wazzan en 1992*. Mémoire de licence d'espagnol non publié. Rabat: Université Mohamad V.
- Labov, W. (1972). *Sociolinguistic patterns*. Filadelfia: University of Pennsylvania Press.
- Lentin, J. (2008). Middle Arabic. En K. Versteegh (Ed.), *Encyclopedia of Arabic language and linguistics*, vol III, pp. 215-224. Leiden-Boston: Brill.
- Lippi-Green, R. (1997). *English with an accent. Language, ideology and discrimination in the United States*. Londres-Nueva York: Routledge.

- Marçais, W. (1930). La diglossie arabe. *L'Enseignement Public* 97, 401-409.
- Mesaoudi, L. (2003). Parler citadin, parler urbain. Quelles différences? En T. Bulot y L. Messaoudi (Dir.), *Sociolinguistique urbaine. Frontières et territoires*, pp. 105-135. Cortil-Wodon: Éditions Modulaires Européennes.
- Miller, C. (2006). Dialect koine. En K. Versteegh (Ed.), *Encyclopedia of Arabic language and linguistics*, vol I, pp. 593-597. Leiden-Boston: Brill.
- Miller, C. (2017). Contemporary *dārija* writings in Morocco: Ideology and practices. En J. Høigilt y G. Mejdell (Eds.), *The politics of written language in the Arab world. Writing change*, pp. 90-115. Leiden-Boston: Brill.
- Miller, C. y Falchetta, J. (2021). Standardization and new urban vernaculars. En W. Ayres-Bennett y J. Bellamy (Eds.), *Language standardization*, pp. 713-40. Cambridge: Cambridge University Press.
- Mitchell, T. F. (1986). What is educated spoken Arabic? *International Journal of the Sociology of Language* 61 (1), 7-32. DOI: <https://doi.org/10.1515/ijsl.1986.61.7>
- Moustaoui, A. (2005). La diglosia y la poliglosia como proceso de subordinación lingüística. Estudio del caso de Marruecos. *Estudios de Dialectología Norteafricana y Andalusí* 9, 139-149.
- Moustaoui, A. (2007). *Lenguas, identidades y discursos en Marruecos*. Tesis doctoral. Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona. <http://www.tesisenxarxa.net/TDX-0124108-150539/>
- Moustaoui, A. (2016). *Sociolinguistics of Moroccan Arabic*. Berlín: Peter Lang.
- Naciri-Azzouz, A. (2020). *Las variedades árabes de Ghomara (Norte de Marruecos). Un estudio comparativo*. Tesis doctoral. Zaragoza: Universidad de Zaragoza.
- Pennis, R. (2020). *Arabe mixte 2.0 : la variation syntaxique et stylistique dans les journaux numériques marocains (janvier-décembre 2016)*. Tesis doctoral. Ca' Foscari University y Aix-Marseille Université.
- Sánchez, P. y Vicente, Á. (2012). Variación dialectal en árabe marroquí: *al-hadra š-šāmālīya u la-hdṛa l-marṛākšīya*. En A. Barontini, Ch. Pereira, Á. Vicente y K. Ziamari (Eds.), *Dynamiques langagières en Arabophonies: variations, contacts, migrations et créations artistiques. Hommage offert à Dominique Caubet par ses élèves et collègues*, pp. 223-252. Zaragoza: Prensas Universitarias de Zaragoza.
- Trudgill, P. (1984). *Dialects in contact*. Oxford: Basil Blackwell Ltd.
- Vicente, Á. (2000). *El dialecto árabe de Anjra (Norte de Marruecos). Estudio lingüístico y textos*. Zaragoza: Universidad de Zaragoza.
- Vicente, Á. (2010). Andalusí influence on Northern Morocco following various centuries of linguistic interference. En J. P. Moferrer-Sala y N. Al Jallad (Eds.), *The Arabic language across the ages*, pp. 141-159. Wiesbaden: Dr Ludwig Reichert Verlag.
- Vicente, Á. (2021) *l-hadṛa b-əl-qāla* (ou la réalisation [ʔ] de /q/) dans les parlers arabes du Nord-ouest du Maroc. En N. Comolli, J. Dufour y M. A. Germanos (Eds.), *Libellules arabes, sémitiques, italiennes, berbères - études linguistiques et littéraires offertes à Jérôme Lentin par ses collègues, élèves et amis*, pp. 415-430. París: Geuthner.
- Vicente, Á. Caubet, D. y Naciri-Azzouz, A. (Eds.) (2017). *La région du Nord-Ouest marocain: Parlers et pratiques sociales et culturelles*. Zaragoza: Prensas de la Universidad de Zaragoza.

- Vicente, Á. y Naciri-Azzouz, A. (2017). The social value of linguistic practices in Tetouan and Ghomara (Northwestern Morocco). En R. Bassiouney (Ed.), *Identity and dialect performance. A study of communities and dialects*, pp. 192-209. Londres/Nueva York: Routledge.
- Walters, K. (2003). Fergie's prescience: The changing nature of diglossia in Tunisia. *International Journal of the Sociology of Language* 163, 77-109.
DOI: <https://doi.org/10.1515/ijsl.2003.048>
- Youssi, A. (1992). *Grammaire et lexique de l'arabe marocain moderne. Préface de D. Cohen*. Casablanca: Wallada.
- Youssi, A. (1995). The Moroccan triglossia: Facts and implications. *International Journal of the Sociology of Language* 112, pp. 29-43.
DOI: <https://doi.org/10.1515/ijsl.1995.112.29>
- Zack, L. y Schippers, A. (Eds.). 2012. *Middle Arabic and mixed Arabic: Diachrony and synchrony*. Leiden: Brill.
- Ziamari, K., Caubet, D., Miller, C. y Vicente, Á. (2020). Matériaux d'enquêtes autour des «usages jeunes» dans quatre villes marocaines Casablanca, Meknès, Tétouan, Marrakech. En C. Trimaille, et al. (Dirs.), *Sociolinguistique des pratiques langagières de jeunes. Faire genre, faire style, faire groupe autour de la Méditerranée*, pp. 23-54. Grenoble: UGA Éditions.

Montserrat Benitez Fernández
Escuela de Estudios Árabes, CSIC
Cuesta del Chapiz, 22
Granada 18010

FROM MIXED ARABIC TO «EDUCATED WRITTEN *DĀRIĠA*»: DIGLOSSIC VARIATION IN MOROCCAN WRITTEN PRODUCTION

Rosa PENNISI¹

University of Catania, associated to IREMAM, Aix-Marseille Université

Abstract

Following developments in destandardisation of linguistic norms in non-standard Arabic writing practices (Mejdell, 2017), the present study aims to analyse formal intermediate registers emerging in Moroccan written productions. Using colloquial Arabic (*ʕĀmmiyya*) in writing is increasing, also in Morocco (Kebede & Hinds, 2016; Caubet, 2017a-b, 2018; Miller, 2017), where *Dāriġa* (Moroccan Arabic) combined with *Fuṣṣḥā* (Standard Arabic), both in oral and written interactions (Mixed Arabic) is also emerging in more or less formal productions. Recently, Moroccan publications, such as Elmedlaoui's (2019) monograph *al-ʕArabiyyatu al-Dāriġatu* (a comprehensive description of *Dāriġa* in *Dāriġa*, and in Arabic script), corroborate the interest on «middle/educated» Arabic (intermediate registers of Moroccan Arabic, mixing *Fuṣṣḥā* and *Dāriġa* in formal, but ordinary, interactions among educated speakers). Based on syntactic descriptions in Elmedlaoui (2019), the present study aims at analysing the stylistic variation of (middle/educated) *Dāriġa* in a corpus of 346 articles and its corresponding 2176 comments of *Goud* (a Moroccan online newspaper expressing in Moroccan Arabic). Quantitative and qualitative approaches in analysing declarative sentences and genitive constructions, allow to observe contrasting trends. As opposed to the syntactic (and artificial) regularities described in *al-ʕArabiyyatu al-Dāriġatu*, *Goud*'s written production shows syntactic and stylistic heterogeneous features of middle/educated (written) *Dāriġa*.

Keywords: written middle *Dāriġa*; diglossic variation; online newspapers; annexation; declarative sentences

1. rosa.pennisi@hotmail.it.  <https://orcid.org/0000-0001-8653-7877>

DESDE EL ÁRABE MIXTO HACIA LA «DĀRIĠA ESCRITA CULTA»:
LA VARIACIÓN DIGLÓSICA EN LA PRODUCCIÓN ESCRITA MARROQUÍ

Resumen

De acuerdo con la evolución de la desestandarización de las normas lingüísticas en las prácticas de escritura árabe no estándar (Mejdell, 2017), el presente estudio analiza los registros formales intermedios en las producciones escritas marroquíes. En Marruecos, el uso del árabe coloquial (*ʿāmmiyya*) en la escritura está aumentando (Kebede & Hinds, 2016; Caubet, 2017a-b, 2018; Miller, 2017). La *dāriġa* (árabe marroquí), combinada con la *fuṣṣḥā* (árabe estándar), emerge en las interacciones orales y escritas (árabe mixto). Además, publicaciones como *al-ʿArabiyyatu al-dāriġatu* de Elmedlaoui (2019), una descripción de la *dāriġa* «intermedia/culta» (registros intermedios que mezclan la *fuṣṣḥā* y la *dāriġa* en interacciones formales ordinarias entre hablantes instruidos), corroboran el interés en las producciones escritas en registros intermedios en Marruecos. El presente estudio compara las descripciones sintácticas de Elmedlaoui (2019) y la variación estilística de la *dāriġa* (intermedia/culta) en un corpus de 346 artículos y sus correspondientes 2176 comentarios en *Goud* (un periódico online escrito en árabe marroquí). Los enfoques cuantitativo y cualitativo en el análisis de las frases declarativas y las construcciones de genitivo permiten observar tendencias contrastadas. Si las regularidades sintácticas descritas en *al-ʿArabiyyatu al-dāriġatu* son artificiales, la producción escrita de *Goud* muestra rasgos sintácticos y estilísticos heterogéneos de la *dāriġa* intermedia/culta.

Palabras clave: *dāriġa* escrita intermedia; variación diglósica; periódicos en línea; anexión; frases declarativas

RECIBIDO: 03/11/2023

APROBADO: 30/05/2023

1. INTRODUCTION

As Boussofara-Omar (2006, p. 636) points out: «Diglossia need not necessarily be seen as a problem to be solved, denied, or contained; it is a richness that is often dramatically undervalued». Her words clearly contextualise the framework in which this study is positioned. Starting from the diglossic situation in the Arab world, and in particular the current linguistic situation in Morocco, this study aims to investigate diglossic variation in the formal written production of contemporary journalistic language. Since Ferguson's (1959) pioneering study on diglossia, from the 1960s to the present, research on linguistic variation in Arabic has been increasingly concerned with intermediate varieties and registers of Arabic,

through the perspective of the diglossic continuum. More recently, from the *Mixed Styles* of formal oral communication in Egypt (Mejdell, 2006)², there has been a shift to the observation of linguistic variation in written productions (Høigilt & Mejdell, 2017). The phenomenon of dialectal writing - not only in informal textual typologies -, and the developments in the destandardisation of linguistic practices in the written production (Kebede & Hinds, 2016, Mejdell, 2017), have contributed to elucidating linguistic and ideological dynamics. In Morocco, the use of extensive written Moroccan Arabic in artistic and cultural productions has been interpreted as issues related to identity expression of *Moroccanness* (Caubet, 2017a-b, 2018), fuelling reflections on changing linguistic practices and attitudes by institutional actors and civil society (Miller, 2012, 2016, 2017). In particular, the interest in written production in non-standard varieties of Arabic in Morocco is corroborated by the ongoing initiatives of various social actors. Among them, formally structured and organised initiatives continue their activities to promote and defend *Dāriġā*³ (Moroccan Arabic), such as the *Markaz tanmiyat al-Dāriġā*

2. The expression «Mixed Arabic» is usually used to identify the mixing of Standard Arabic and colloquial Arabic in oral interactions. From this perspective, Standard Arabic (in the present study generalised as *Fuṣḥā*) and *ʕĀmmiyya* (colloquial Arabic) would be considered as two distinct varieties or entities. Instead, approaching diglossia in terms of *Mixed Styles*, allows Mejdell (2006) to observe the codemixing between *Fuṣḥā* and *ʕĀmmiyya* in formal oral communications, not as a mechanism that responds to precise linguistic constraints, but rather as (free) choices or tendencies adopted by locutors when in a formal context they perform their communications according to the shared norms of the speech community (colloquial ordinary communications) while remaining more or less adherent to the way of speaking codified in the standard norm. With regard to written productions, Arabic texts with mixing of linguistic elements in *Fuṣḥā* and *ʕĀmmiyya* are traditionally associated with Middle Arabic, which for some scholars, this term only refers to mixed texts written before the *Nahḍa*. To the extent that «the nature of the linguistic mixing in Middle Arabic displays many similarities to what can currently be found in some oral mixed varieties» (Lentin, 2008, p. 216), and agreeing with Lentin, «nothing prevents us in theory, as far as the particular nature of oral and written language is taken into account, from regarding Middle Arabic written until precontemporary times as belonging to a large ensemble that could be labeled ‘Mixed Arabic’.» (Lentin, 2008, p. 216).

3. Specialists in North African Arabic dialectology, who generally investigate oral corpora, transcribe the word *دارجة* with *dariġa* (literally ‘circulating, vulgar, popular, common, habitual’, indicating Maghrebi Dialects; in the present study it refers to Moroccan Arabic). In the present study, however, the same word will instead be rendered with *Dāriġā* because the corpora come from contemporary written production and in order to avoid an overestimation of the (real) pronunciation (which varies due to the diatopic and diastatic variation of the native speakers), it is preferred to transcribe the metalinguistic terms using the ISO transcription system of Standard Arabic, in order to keep with their orthographic realisation. For the same reason, the examples and quotations given by Elmedlaoui (2019) and *Goud* are not transcribed, but only transliterated using common transliteration system of Standard Arabic. In the analysis, the discussed linguistic features will be underlined, both in the original Arabic and in translation. Finally, the capitalisation in *Dāriġā*, *ʕĀmmiyya*, *Fuṣḥā*, such as Arabic, English, French, etc., responds to the conventions of the English language.

Zakūra, 'Zakoura Centre for the Promotion of *Dāriġa*'. Following the publication in 2017 of *Qāmūs al-Dāriġa al-Maġribiyya* (Mgharfaoui *et al.*, 2017), the first monolingual dictionary of Moroccan *Dāriġa*, the Zakoura Centre published the first handbook of «*Dāriġa* Arabic» edited by Mohamed Elmedlaoui (2019), written in the middle/educated⁴ register of Moroccan Arabic.

The present study, therefore, intends to focus on the syntactic and stylistic features of the Moroccan middle/educated register, starting from the descriptions provided by the work of Elmedlaoui (2019). In particular, two morphosyntactic features will be explored in depth, namely the analysis of the subordination in declarative propositions, and the genitive annexation, in both synthetic and analytical constructions, through a comparative approach, using a corpus of articles and commentaries from a Moroccan online newspaper, *Goud*, which represents a model of contemporary written production mixing and alternating expressions in Moroccan Arabic.

The first part aims to present the research of Elmedlaoui (2019) and his handbook *al-Ṣarabiyyatu al-Dāriġatu* with his descriptions of Moroccan Middle Arabic; the second part, on the other hand, focuses on the analysis of morphosyntactic traits in comparison with the descriptions in Elmedlaoui (2019) and in *Goud's* journalistic corpus.

2. AL-ṢARABIYYATU AL-DĀRIĠATU:

AUTHOR, LANGUAGE, AND CONTENT

Mohamed Elmedlaoui is a Moroccan linguist and writer, originally from Igoudar Mnabha, a rural municipality in Morocco, in the province of Taroudant (Souss-Massa region). He was Head of the Department of Languages and Literature at the Faculty of Humanities of the University of Oujda, Professor at the Mohammed V University of Rabat, researcher at the University of Paris VIII and Research Director at the *Institut Royal de la Culture Amazigh*.

As a linguist, he is a specialist in segmental and prosodic phonology, derivational morphology and lexicology. His research work ranges from the poetic

4. In Arabic linguistic studies, Middle Arabic corresponds to «the language of numerous Arabic texts, distinguished by its linguistically (and therefore stylistically) mixed nature, as it combines standard and colloquial features with others of a third type, neither standard nor colloquial» (Lentin, 2008, p. 216), whereas Educated Arabic refers to «the higher-register spoken variety of a particular country or region [...] to a type of spontaneous inter-regional, «mixed», or «koineized» discourse used among Educated Arabic speakers when they need to accommodate differing regional dialects.» (Ryding, 2008, pp. 666-667).

metrics of Moroccan folk songs to issues of reproducing the written orthography of languages with an oral tradition; and also, from the contrastive analysis of Semitic languages to the linguistic description of Amazigh, as well as linguistic variation between varieties of Berber. He is interested in social sciences and Arabic language, as his abundant scientific production testifies. In fact, his studies focus on the comparison between Hebrew, Arabic and Amazigh languages, especially from a phonetic point of view.⁵

His work, *al-ʕArabiyyatu al-Dāriġatu*, was published in 2019 by the *Markaz tanmiyat al-Dāriġa Zakūra*, ‘Zakoura Centre for the Promotion of *Dāriġa*’, i.e., the academic side of the *Zakoura Education* founded by Nouredine Ayouch in 1995⁶. The title of the work clearly sums up its content: *al-ʕArabiyyatu al-Dāriġatu* (literally ‘the *Dāriġa* Arabic’, i.e., ‘the circulating/current (Moroccan) Arabic’) is an old expression modelled on the expression *al-ʕArabiyyatu al-Fuṣṣḥā* (‘the most eloquent Arabic’, that is the codified Arabic, which is often referred to as (Modern) Standard Arabic (McCarus, 2008). The term ‘*al-Dāriġatu*’, used as a qualifier of ‘*al-ʕArabiyyatu*’ (as well as the qualifier ‘*al-Fuṣṣḥā*’ in the expression *al-ʕArabiyya al-Fuṣṣḥā*), can be interpreted symbolically as the intention to rehabilitate the dignity of Moroccan Arabic, considering it fully as a language, i.e. with its own phonological, morphological, syntactic and lexical structures and rules. Elmedlaoui, for instance, made it explicit in the subtitle:

«إملائيَّة ونحو: الأصوات، الصرْف، التَّرْكِيْب، المُعْجَم . (كِتَابٌ مُحَرَّرٌ بِعَرَبِيَّةٍ مَغْرِبِيَّةٍ وَسَطَى)»

[*?Imlāʔiyyatun wa-naḥwun : al-ʔaṣwātu, al-ṣarfu, al-tarkību, al-muṣṣġamu (kitābun muḥarrarun bi-ʕarabiyyatin maġribiyyatin wuṣṭā)*,] ‘Orthography and Grammar: Phonetics, Morphology, Syntax, Lexicon (book edited in Middle

5. Note, for example, the work on the phonological representation of Hamito-Semitic languages (Elmedlaoui, 1995), or the collaboration with François Dell, from which Dell-Elmedlaoui’s famous algorithm on syllabic enucleation based on the study of Tashelhit variety of Berber (Dell & Elmedlaoui, 2002); these are just a few of his many studies.

6. The *Zakoura Education* foundation was created in 1995, by the initiative of Nouredine Ayouch, a Moroccan advertising entrepreneur and member of the Higher Council of Education and Scientific Research in Morocco. In its early days, the *Zakoura Education* aimed to foster the educational and professional development of Moroccan citizens in rural and disadvantaged areas of the country. The foundation’s various projects include the opening of educational centres for pre-school education, but also initiatives aimed at promoting mother tongues, such as *Amāzīġ* and *Dāriġa*. Nouredine Ayouch is indeed a well-known personality in the dynamics of defending and promoting *Dāriġa* in Morocco. See the media and academic debate following the organisation of the international conference *The Language, the Languages* in 2010, corroborated by the organisation in 2013 of the conference *Le chemin de la réussite* on the formal teaching of mother tongues in the educational system in Morocco (Miller, 2015), and the controversy following the publication of *Qāmūs al-Dāriġa al-Maġribiyya* (Mgharfaoui et al., 2017), the first monolingual dictionary of the *Dāriġa* (Pennisi, 2020a).

Moroccan Arabic). The specificity of Elmedlaoui's study, therefore, lies in the fact that his work does not limit itself to describing Moroccan Arabic, but proposes a comprehensive manual of *Middle Moroccan Arabic*, written entirely in this linguistic register. In the introduction, in fact, Elmedlaoui makes explicit the main features of the linguistic variety described and the metalanguage used, which represents/reproduces the *wustā* 'middle' / *muhaddāba* 'educated' register of Moroccan Arabic, that is the language used by Moroccan intellectuals, researchers, professors, politicians, officials, and so on. In particular, Elmedlaoui (2019, p. 29) states:

اللغة الوصفية (métalangue) التي تحرر بيها هذ الكتاب هيأ عربية مغربية مهذبة وقريبة من السبجيل دبال الفوصحي، بدون ما تكون بعيدة على ما هوأ مألوف في الاستعمال العادي العام عند جميع المغاربة، وحثى عند باقية البلدان المغاربة في ماعدا بعض الموفرادات والتعبير الخاصة.

فأ هيأ نفس اللغة التي كا يستعملوها الموثقون فالماجالس الخاصة ديالهوم، والتي كا تستعملها كثير من الباراميج الحيوارية في الإذاعة والتيليفزيون، والتي كا يستعملها كثير من الأساتذة والأكاديميين في الأقسام المدرسية وفي المودراجات الجامعية، والتي كا تستعمل كذلك في المحاكم وفي البرلمان وفي الأكشاك الإدارية العمومية.

[*al-lūga al-wasfiyya (métalangue) llī tuḥarrir bī-hā hād al-kitāb hiyyā ṣarabiyya maġribiyya muḥaddāba wa-qarība min al-siġil dyāl al-fuṣḥā, bi-dūn mā takūn baʿīda ʿalā mā huwā maʿlūf fī al-istiʿmāl al-ʿādī al-ṣamm ʿinda ġamīʿ al-maġāriba, wa-ḥattā ʿinda baqīyya al-buldān al-maġāriba fī māʿadā baʿd al-mūfrādāt wa-al-taʿābīr al-ḥāṣṣa. fā hiyyā nafs al-lūga llī kā yastamīlū-hā al-mūtaqqāfūn f-al-māġālis al-ḥāṣṣa dyāl-hūm, wa-llī kā tastaʿmil-hā kaṭīr min al-bārāmīġ al-ḥīwāriyya fī al-ʿiḍāʿa wa-al-tīlīfīziyūn, wa-llī kā yastamīl-hā kaṭīr min al-ʿasātīza wa-al-ʿakādīmīyyīn fī al-ʿaqsām al-madrāsīyya wa-fī al-mūdarrāġāt al-ġāmīʿiyya, wa-llī kā tastaʿmil kaḍālik fī al-maḥākīm wa-fī al-barlamān wa-fī al-ʿakšāk al-idāriyya al-ʿūmūmiyya.*]⁷

'The descriptive language (metalanguage) in which this book was written is an educated Moroccan Arabic, close to the register of *Fuṣḥā*, not far from what is familiar in the ordinary general usage of all Moroccans, and also in the rest of the Maghreb countries, with the exception of a few specific words and expressions.

It is the same language used by intellectuals in their private circles, used in many talk shows on radio and television, used by many professors and

7. As mentioned in footnote 3, all transcriptions from Arabic only aim to transliterate; they are not intended to observe phonetic or phonological phenomena, or to represent correct spelling (in Moroccan Arabic). The following Arabic letters, when found in the original text, are transliterated as follows: > ? ; ج > ġ ; خ > ḫ ; د > ḍ ; ع > ʿ

academics in school departments and university lecture halls, also used in the courts, in parliament and in public administration’.⁸

Elmedlaoui’s assertions, in particular the reference to the language of the media, politics, and administration inevitably refers back to Abderrahim Youssi’s definition of *Modern Moroccan Arabic*. Youssi (1992, p. 23) stated that:

... les usagers ont élaboré cette variété d’AMM [Arabe Marocain Moderne] qui [...] non seulement instaure le continuum qui faisait défaut entre l’arabe dialectal et l’arabe littéral, mais en vient à incarner le modèle de la variété standard de l’oralité.

Moreover, Youssi (1992, p. 25) adds:

L’AMM est [...] la variété des situations formelles de l’échange orale telles que celles prévalant dans l’administration publique, entre étrangers [...]. L’AMM est enfin, et surtout, le medium utilisé à la radio et à la télévision pour la présentation spontanée et/ou la diffusion de programmes de vulgarisation technique et scientifique.

Comparing the two sources, what emerges is that Youssi (1992), through the perspective of the diglossic continuum, traditionally bounded by the high variety (*al-ʕArabiyya al-Fuṣḥā*) and low variety (*al-ʕĀmmiyya / al-Dāriġā*),⁹ argued for the increasing spread of the middle/educated register of Moroccan Arabic, essentially in oral communication. Although Elmedlaoui in no case cites Youssi’s founding and pioneering work in the field of Moroccan dialectology, through *al-ʕArabiyyatu al-Dāriġatu* he intends to demonstrate how Middle Moroccan Arabic can also be used in the practices of written production, and to a certain extent he dictates its norms by informally codifying spelling rules, morphosyntactic notions, and lexical features. The work, *al-ʕArabiyyatu al-Dāriġatu*, represents, in

8. Translations from Arabic are by the author. Note also the spelling choices used by Elmedlaoui to write in *Dāriġā*. One of the most evident is the use of *mater lectionis*, *alif*, *wāw* and *yāʕ* in place of vowel signs. This phenomenon is attested in some Middle Arabic texts (Lentin, 2012). From Elmedlaoui’s extract, see, for example, اللوغة * *al-lūġa* instead of اللُغة *al-lūġa*). The orthographic and phonetic rules of Middle Moroccan Arabic are made explicit by Elmedlaoui in the second chapter (Elmedlaoui, 2019, pp. 75-107).

9. One certainly recalls Ferguson’s pioneering study (1959) and the subsequent *Diglossia Revisited* (Ferguson 1991), but also and especially the variationist studies of Blanc (1960), Badawi (1973), and Meiseles (1980), or diglossia as stylistic variation in Hudson (1994) and Mejdell (2006 and 2017); in the latter study, it is particularly evident how, in Egypt, the mixed Arabic characteristics of oral (formal) communication are related to the change and destandardisation of written production practices.

line with the ideologies of the *Markaz tanmiyat al-Dāriġa Zakūra*, an initiative to promote Moroccan Arabic.

2.1. Middle Moroccan Arabic in *al-ʕArabiyyatu al-Dāriġatu*

Before focusing on the morphosyntactic features of declarative subordination and annexation (the subject of the present study), it remains important to mention some main aspects of Middle Moroccan Arabic, summarised in Elmedlaoui (2019).

On the phonetic level, Elmedlaoui points out, for example, that the phonemes /rāʔ/ and /zāy/ are emphatic and that interdental phonemes are stopped in Moroccan Arabic and in its middle register (Elmedlaoui, 2019, p. 17). Furthermore, on the morphological level, the difference between dual and plural is not produced and the external regular plural in *-īn* (referring in *Fuṣḥā* to both accusative and genitive cases), is generalised to the nominative case (Elmedlaoui, 2019, p. 17). Finally, on the syntactic level, the *iṣrāb* (case inflection) and the indefinite mark (*tanwīn*) are not realised and particles and analytic conjunctions (from the *Dāriġa* linguistic system) are preferred for complex syntactic constructions (Elmedlaoui, 2019, pp. 17–18).

Returning specifically to its contents, Elmedlaoui states that it is a book of applied contrastive descriptive linguistics that has been produced with both the oral and the written communications taken into account, with the aim of maintaining continuity between *al-ʕArabiyya al-Fuṣḥā* and *Dāriġa*. In other words, the book takes into account the combination of the two varieties, considered as a linguistic continuum, that is *al-ʕArabiyya al-Wuṣṭā*, ‘Middle Arabic’ (in Morocco) (Elmedlaoui, 2019, p. 21).

In the introduction, Elmedlaoui also sets out the pedagogical objectives of the book, namely the importance of training teachers to use this middle register of Moroccan Arabic with students. According to Elmedlaoui – in agreement with the *Zakoura Centre* –, the use of Moroccan Middle Arabic in classrooms would facilitate the acquisition and learning of *Fuṣḥā* as well. Finally, and as already mentioned, Elmedlaoui’s study would offer a coherent way to write in Middle Moroccan Arabic, that is conventionalising the writing in order to simplify the reading and, thus, facilitate the comprehension of what is read (Elmedlaoui, 2019, p. 25).

On this last point, it must be mentioned, however, that in recent years studies on written productions of dialect varieties (Mejdell, 2017), and in particular of written *Dāriġa* in Morocco (Caubet, 2017a–b, 2018), have shown that the

evolution of dialect writing practices responds more to sociological mechanisms, such as the concept of conventionalisation from below (Caubet, 2017a) – that is the widespread, shared, and repeated use of linguistic practices –, rather than from top-down standardisation dynamics, which, by the way, should coincide with institutional language policies in order to be eventually accepted and shared by society.¹⁰ In other words, the attempt of a non-institutional codification as implemented by the *Markaz tanmiyat al-Dāriġa Zakūra*, and, in this specific case, by Elmedlaoui's study, does not take into account the existing linguistic practices of the users, and would create a standardised form of a linguistic register (a standardised Middle Moroccan Arabic) that would overlap, on the one hand, with the already standardised variety of Arabic (*Fuṣḥā*) and, on the other hand, with the colloquial variety (*Dāriġa*), establishing a double diglossia that would impose itself on the current linguistic situation.

Last but not least, the public reception of these dynamics should not be underestimated. Already between 2010 and 2013 (but also more recently in 2017), i.e., at the time of Ayouch's initiatives of promoting *Dāriġa*, such as the (formal) integration of *Dāriġa* in the educational system mentioned above,¹¹ were strongly rejected by public opinion, since they still represent a sensitive topic, a practice considered unacceptable by Moroccan society (Miller, 2016; Pennisi, 2020a).

However, with regard to the use of non-standard varieties of Arabic and Moroccan *Dāriġa* in written production, this is an accepted tacit practice, especially in digital written production. Various publishing experiences have made, and still make, History in the Moroccan cultural and media scene. The present study will only examine the electronic newspaper *Goud*,¹² founded in 2011 by Ahmed Najim, the current editorial director, as a continuation of the publishing adventure of the well-known Moroccan weekly *Nichane*.¹³ *Goud* is a Moroccan, generalist, and liberal online newspaper. Ahmed Najim claimed that the newspaper's editorial lines leave journalists and contributors free to choose *Dāriġa* as the language of expression for all news (Pennisi, 2020b). In fact, most articles are also written with

10. See for example the traditional stages of the language standardisation process in Haugen (1966).

11. See footnote 5 in this study.

12. See the newspaper's website, <http://www.goud.ma>

13. *nīšān*, 'direct', in Moroccan Arabic – as well as the meaning of *Goud* –, was a Moroccan weekly magazine in Arabic script, originating from *TelQuel*, its French version. *Nichane* represented for the Moroccan media scene an editorial experience that was as innovative as turbulent, due to the controversy triggered by the taboo topics it dealt with, and using a language that was, effectively, 'direct', that is, with some expressions also in *Dāriġa*. See Aguadé (2012), Miller (2012, 2015, 2017); Brigi (2016); Hoogland (2018).

elements from *Dāriġa* linguistic system. *Goud's* journalistic corpus and descriptions of the morphosyntactic features of declarative propositions and annexation in Moroccan Middle Arabic will be analysed and compared in the following section.

3. DECLARATIVE SENTENCES AND ANNEXATION IN *AL-ʕARABIYYATU AL-DĀRIĠATU* AND *GOUD*

In this section, two morphosyntactic traits - the construction of declarative subordinations and the genitive constructions - will be compared through their descriptions in Elmedlaoui (2019) and from the samples of *Goud's* corpus (346 articles and its corresponding 2,176 comments). The aim is to observe regularities and variation of middle / educated (Moroccan) Arabic in written formal productions.

3.1. *Syntactic structure of complex sentences: declarative clauses*

In *al-ʕArabiyyatu al-Dāriġatu* declarative subordinations are described in the syntax section:¹⁴ الجومال وأنواعها ومنطيقها [al-qīsm al-ḥāmis. al-tarkīb. التركيب. القيسم الخامس.], [al-qīsm al-ḥāmis. al-tarkīb. al-ġūmāl wa-ʔanwāʕ-hā wa-manṭiq-hā], 'Part Five: Syntax. Phrases, Typology and Logic' (Elmedlaoui, 2019, pp. 227-286).

Elmedlaoui sketches declarative propositions according to two structures. The first is defined as الجملة الصوغرى الفاعلية [al-ġūmla al-ṣūġrā al-fāʕiliyya], 'subject subordinate' (Elmedlaoui, 2019, pp. 261), explained in the following two examples:

- (1) (نحفظ النصوص) عليا كما يصعب
 [(al-nūṣuṣ nḥfaḍ) ʕalayā kā yəṣʕāb]
 the texts (I) memorise on me it is difficult
 'It is difficult for me [that I] (memorise texts).'
- (2) (حفاظة النصوص) عليا كما يصعب
 [(al-nūṣuṣ ḥfāḍa) ʕalayā kā yəṣʕāb]
 the texts the memorisation on me it is difficult.
 'It is difficult for me (the memorisation of texts)'¹⁵

14. References to declarative sentences also appear in the third part devoted to morphology (Elmedlaoui, 2019, pp. 111-190 and pp. 150-151).

15. Translations tend to literally translate the original Arabic text to reflect the syntactic and grammatical categories of Arabic whenever possible. All brackets are in the original of all Elmedlaoui samples in Arabic, they are reported in translation too.

In (1) and (2), the same sentence, meaning ‘It is difficult to me to memorise texts’, follows an asyndetic construction (without the complementizer), and the subordinate clauses (in rounded brackets) replaces the subject of the verb of the main proposition (out of rounded brackets).¹⁶

In (1), the subordination is a verbal phrase (a VSO type, where the subject is implicit in the conjugation of the verb), while in (2), the subordination is composed of a nominal construct (the verbal noun of the verb *ḥafaḍa* ‘to memorise/ the memorisation’ + definite plural noun *al-nuṣūṣ* ‘the texts’, in construct state).

The second structure of subordination presented by Elmedlaoui is referred to as *الجوملة الصوغرى المفعولية*, [*al-ġūmla al-ṣūġrā al-maʿfūlīyya*], ‘object subordinate’ (Elmedlaoui, 2019, p. 262), in which the subordinate proposition replaces the direct verb complement of the main proposition, as indicated by the following examples:

- (3) {بغى الولد Ø (يمشي للموخيّام)}
- [[li-l-mūḥəyyām yəmsī Ø al-walad bgā]
- to-the camping (he) goes Ø the guy (he) wants
- ‘{The guy wants [to] (go to the camping)}’.
- (4) {تقدّ البنّت Ø (تنافس الولاد)}
- [al-walād tunāfis Ø al-bint taqadd]
- the boy (she) competes Ø the girl can
- ‘{The girl can (compete with the boy)}’.
- (5) {طلب القاضي من المتهم Ø (يدلي بالهويّة دياله)}
- [[dyāl-hu bi-l-hawwiyya yədlī min al-mathūm al-qāḍī ṭalab]
- of-him with-the identity (he) provides Ø the defendant to the judge (he) asked
- ‘{The judge asked the defendant [to] (provide his identity)}’.
- (6) {أمر القاضي الأعوان Ø (يحضرو المّمّاهم)}
- [[al-muttāhim yəḥḍḍirū Ø al-ʔaʿwān al-qāḍī amar]
- the defendant (they) bring Ø the officers the judge (he) ordered
- ‘{The judge ordered the officers [to] (bring the defendant)}’.

The symbol ‘Ø’ in (3)-(6) is used by Elmedlaoui to indicate the position of the complementizer where it could be inserted. Furthermore, Elmedlaoui points out

16. Morphological elements of Moroccan Arabic, such as in this case the preverb *kā* for prefix conjugation (كأ يصعب) [*kā yəṣʔāb*] ‘it is difficult’, are described in Elmedlaoui (2019). The morphemes of Moroccan Arabic are always clearly distinguishable in Elmedlaoui, such as the case of the preverb *kā*, which is not realised as a morpheme prefixed to the verb and is therefore visibly separated from the verb, unlike the common practices in *Goud* (although heterogeneous, as shown in the next section).

that the complementizer in *Fuṣḥā* corresponds to the particle *ʔan(n-)* and in Amazigh to *ad*, whereas in Moroccan Arabic there are two complementizers: بلي [*bəlli*] and باش [*bāš*], which are used depending on the modal verbs in the main proposition (Elmedlaoui, 2019, p. 262).

Elmedlaoui does not specify which modal verbs bear one or the other particle and never mentions whether or not the variation between these complementizers reflects a stylistic or register variation. It describes the two particles as belonging to the repertoire of Moroccan Arabic, without specifying whether the source of the examples or whether the examples represent the middle register of Moroccan Arabic. However, he gives another set of examples in which Elmedlaoui (2010, p. 263) indicates the construction with the obligatory or optional complementizer (in this case inserted between the symbols '><'), as follows:

- (7) {كا يظنّ الطماع <بلي> (البصقة درهم)}
 [[(dirham al-baṣqa) >bəlli< al-ṭammāf kā yəṭunn]]
 a dirham the spittle >that< the miser (he) thinks
 'The miser thinks >that< (spittle counts as a dirham)'.
 (8) {نكر المتهم بلي قام بالجريمة}
 [[(bi-l-ḡārīma qām) bəlli al-mathūm nakar]]
 with-the crime (he) committed that the defendant (he) denied
 'The defendant denied that (he committed the crime)'.
 (9) {اعترف المتهم بلي قام بالجريمة}
 [[(bi-l-ḡārīma qām) bəlli al-mathūm iṭāraf]]
 with-the crime (he) committed that the defendant (he) acknowledged
 'The defendant acknowledged that (he committed the crime)'.
 (10) {حاول الطبيب <باش> (ينقذ المرحوح)}
 [[(al-maḡrūḥ yənqad) bāš al-ṭabīb ḥāwal]]
 the wounded (he) saves >to< the doctor (he) tried
 'The doctor tried >to< (save the wounded)'.
 (11) {الكشافه (كا يعرفو بلي [الطريق واعرة])}
 [[([wāfira al-ṭarīq] bəlli kā yərfū) al-kaššāfa]]
 rough the road that (they) know the explores
 'Explorers (know that [the road is rough])'.

In this section devoted to declarative subordination, Elmedlaoui does not give examples using the complementizer *ʔan(n-)* (described as exclusive to Arabic *Fuṣḥā*). He thus suggests that in Moroccan Arabic the only complementizers used

for declarative subordinations are *bəllī* and *bāš*. However, in the section devoted to relative sentences (Elmedlaoui, 2019:265–267), an example appears, which is instead a declarative subordinate, where both *bəllī* and *ʔan(n)* appear, as given below:

- (12) {كا يفزح (كا يحبو التلاميذ) } بلبي يعرف أنه [الموعليم] }
 [al-talāmiḍ kā yəḥibbū] bəllī yəʔrif ʔann-hu al-mūʕallīm kā yəfarriḥ]
 the students (they) like that (it) knows that the teacher (he) is happy
 {((المدراسة ينظفون) }
 [([al-madrāsa yənaḍfū] }
 the school (they) clean }
 ‘{The teacher is happy [to know that (the students like to [clean the school])]}’.

Apart from example (7), which seems potentially referable to a proverbial expression, (thus closer to a familiar/popular register), all the other examples seem to represent ordinary communicative situations, more or less formal, but mostly all with the complementizer *bəllī* (with the exception of example 10).¹⁷

However, as stated in Caubet (1993) and Youssi (1992), both *bəllī* and *ʔan(n-)* are described as complementizers in Moroccan Arabic. In particular, Youssi points out that *ʔan(n-)*¹⁸ is the only particle used in Modern Moroccan Arabic, whereas *bəllī* is used in a more informal register of Moroccan Arabic. The stylistic variation that emerges from Youssi (1992) for formal oral communications is also reflected in the written journalistic production in *Goud*. Although the occurrences of *bəllī* in *Goud*'s articles are numerically lower than those of *ʔan(n-)*,¹⁹ such variation demonstrates the use of different stylistic strategies depending on the communicative functions of the news. See the following examples:

- (13) قالبيا (بلي) مزيان نساڤر و نشوف ناس اخرين و بلادات اخرى²⁰
 [ʔuḥrā blādāt wa aḥrīn nās nəšūf wa nəsāfr məzyān (bəllī) qāliyyā]
 other countries and other people (I) see and (I) travel good that (he)told-me

17. The use of *bāš* as a complementizer in declarative sentences is never attested in the corpus of journalistic Arabic used in the present study.

18. In particular, Youssi (1992, p. 279), gives the following example as a model of declarative subordination in Modern Moroccan Arabic: «ana ka-n-ṣtaqed ʔenna l-muneDDama dya-na lem t-məḥ bi-dalik. 'Personnellement je reste convaincu que notre Organisation ne permettra jamais cela.' The transcription from Moroccan Arabic corresponds to the original in Youssi (1992).

19. Out of a total of 1,173 occurrences of the complementizer, only 31 correspond to *bəllī*, while the remaining 1,142 comprise *ʔan(n-)*, including 40 occurrences of *ʔinna* (after the verb *qāla* 'to say').

20. Underlined words represent verbs introducing declarative sentences, words in (round brackets) are complementizers, and in italic words clearly written in *Dāriġa*.

(Nada Eljabli, *Goud, Opinion*, 04/10/2016)

He told me (that) it's good to travel and see other people and other countries'.

- (14) *بغات* *شئو* *مافهماتهاش* *ليها* *(باللي)* *شرحات*
 [bgāt šnū mā-fahmāt-hā-š (bəlli) lī-hā šarḥāt]
 (she) wanted what (she) did-understood-not that to-her (she) explained
 (Hajar Moutaouakil, *Goud, Opinion*, 19/08/2016)

'She explained (that) she had not understood what she wanted'.

- (15) *هنا* *كايين* *(أنه)* *للأسف* *ماعر فاش* *الناس*
 [hunā kāyn (?ann-hu) li-l-ʔasaf mā-ʔarfā-š al-nās]
 there there is that unfortunately (he) do-know-not the people
واحد *لا فونتناج* *ديال* *المستهلك*
 [al-musthaliq dyāl lāfūntāğ wāḥəd]
 the consumer of benefit a
 (Jawad Elansari, *Opinions, Goud*, 28/02/2016)

'People did not know, unfortunately, (that) *there is a consumer benefit*'.

- (16) *الخليجية* *المغربية* *القمة* *(ان)* *تخطيط* *بل* *دفة* *ماشى*
 [al-ḥaliğiyya al-mağribiyya al-qimma (an) taḥṭīṭ bal ṣadfā māšī]
 Gulf Morocco the summit that scheduled but a coincidence
 it-is-not

جات *قبل* *زيارة* *أوباما*
 [ʔŪbāmā ziyārat qəbla ġāt]
 Obama visit before (it) came

(Adil Meknessi et Hana Abu Ali, *Opinions, Goud*, 30/04/2016)

'It is not a coincidence, but [it is well] scheduled (that) the Morocco-Gulf summit came before Obama's visit'.

- (17) *الشركات* *عند بال* *هاد* *(أنه)* *للشك* *مجالا* *لا يدع*
 [l-šarikāt hād bāl ʔənd (?ann-hu) li-l-šakk mağālan lā yədaʔ]
 the enterprises those spirit in that for-doubt place there-is-not
(ان) *بنادم* *كيعيش* *بأقي* *فدار* *غفلون*
 [ğəflūn. f-dār bāqī ka-yəʔīš bənādəm (an)]
 ignorant in-house (he is) remaining (he) lives man that
 (Jawad Elansari, *Opinions, Goud*, 28/02/2016)

'No doubt (that) it is in the spirit of these enterprises [the fact] (that) man lives in ignorance'.

All the examples given by *Goud* are samples from articles originally published in the newspaper's Opinion Column. Example (13) is taken from a serial novel also published in *Goud's* Opinion Column, titled *Mudakkirāt mudallika sānk itwāl*, 'Memories of a five-star masseuse' by Nada Eljabli, in which the variation in register meets stylistic and literary needs. The other examples are taken from traditional opinion articles with various topical themes (from foreign policy to social and cultural issues within the country).

Note the syntactic variation between *bāllī* / *ʔan(n-)* in (13)-(17).²¹ While *bāllī* appears in syntactic contexts constructed more with elements in *Dāriġa* (examples 13 and 14), the *ʔan(n)* complementizer is integrated into mixed syntactic contexts in which elements in *Fuṣḥā* and *Dāriġa* are combined through diglossic code-switching (examples 15-17). While in (13) and (14) the communicative register is more informal (the examples reproduce in writing an informal oral communicative context), in (15)-(17) the communicative, formal register reproduces in writing a middle/educated style of (Moroccan) Arabic.

In short, *Goud* reproduces, without standardised rules, what Youssi (1992) called «Modern Moroccan Arabic», which is a middle, formal register composed of the combination of Modern Arabic and Moroccan Arabic.

3.2. Syntactic structure of complex sentences: annexation

In Elmedlaoui, annexation is treated both in the fourth chapter dealing to morphosyntax (Elmedlaoui, 2019, pp. 191-226) – in the paragraph devoted to definiteness (Elmedlaoui, 2019, pp. 196-197) – and in the fifth chapter devoted to syntax (Elmedlaoui, 2019, pp. 227-286), where he pays specific attention to the direct and indirect construction of annexation (Elmedlaoui, 2019, pp. 231-234), that is synthetic construction (the state construct) and analytic annexation through the use of genitive particles (of which he cites *dyāl*, *d-* and *mtāf* as genitive particles in Moroccan Arabic).²²

21. Also note the orthographic variation in the realisation of complementizers among *بالي* and *باللي* for instance. Such orthographic heterogeneity is a feature of written production in *Dāriġa* that does not lead to incomprehensibility.

22. Youssi (1992) describes the particles *mtāf*, *btāf*, and *tāf* as regional variants of *dyāl* (and *d-*, the shortened form of *dyāl*). In *Goud's* journalistic corpus, in fact, there are no occurrences of these particles in the articles. Only *tāf* is used very sporadically in some commentaries (a total of 129 analytical constructions of annexation and only 7 occurrences of *tāf* among them were found). The occurrences of *tāf* in the comments show an informal register tending towards rude and vulgar expressions.

He outlines the two structures schematically, indicating that both are used in Moroccan Arabic, without specifying the possible variation in style or register between the two. However, he indicates that the analytical construction is preferred where the rules of the state construct make the communicative faculty of language impractical. See the two (extreme) examples cited in Elmedlaoui (2019, p. 234):

- (18) كتاب وولاد بنت جارة نسيب مودير مدرسة
 madrāsa mūdīr nasīb ġāra bint walād kitāb]]
 school director brother-in-law neighbour daughter sons book
 حومة النجارين] كتاب مفيد.
 [mūfīd kitāb [Neġġārīn ḥūmā
 useful book Nejjarine district
 [‘The book of the nephew of the neighbour of the brother-in-law of the headmaster of the Nejjarine district school] is a useful book’.
- (19) [الكتاب ديال وولاد الجارة دنسيب
 d-nasīb al-ġāra walād dyāl al-kitāb]]
 of-the brother-in-law the neighbour sons of the book
 المودير ديال المدرسة اللي فحومة النجارين] هو كتاب مفيد.
 [mūfīd kitāb huwā [Neġġārī f-ḥūma illī al-madrāsa dyāl al-mūdīr
 useful book (it)is Nejjarine in-district that the school of the director
 [‘The book of the son of the neighbour of the brother-in-law of the school director who is in the Nejjarine district] is a useful book’.

The same nominal sentence (subject in square brackets + indefinite predicate), is reproduced in (18) with the subject composed of a «direct» genitive construct, that is the multiple state construct, and in (19) the genitive compound is instead constructed in an «indirect» manner, combining analytic and synthetic constructions and relative phrases.

Regarding the variation between synthetic and analytic constructions, Ech-Charfi (2014) describes the distribution of genitive constructions in Moroccan Arabic as *relations allowing variation*, that is semantic relations expressing the relationship of possession, form, substance, origin, time, space, etc. In fact, Ech-Charfi (2014, p. 180) states that «Although most of these relations can be expressed equally by the analytic genitive as well as the construct state, some are more natural with one than with the other».

In *Goud*, however, the stylistic variation between the two constructions emerges according to the columns, themes, and communicative functions of the texts. See in (20) below a sample from an Opinion article and in (21) a short report article:

- (20) الناس ما عرفاش للأسف أنه كاين هنا واحد لافونتاچ (ديال) المستهلك وهو انه عندو سند قانوني باش يقدر يرفع دعوى على هاد الشركات وكاينة احتمالية كبيرة جدا في كسب القضية، في حين انه الشركات لي قامت بهاد ماكننش عندها تا شمي سند قانوني. هي لعبت فقط على المعطى (ديال) انه أصلا تا شمي حد ما تسوق في الأول مني دارو الخطوة لي كان عبارة عن جس النض عبر الغاء خدمة VOIP من 3G من بعدا ما شافو أنه تا شمي ردة فعل قوية ما دارت إتجاهم إزاء هاد الخطوة ناضو حيدو تا 23.ADSL (ديال)

[al-nās mā ʕarfāš li-l-ʕasāf ʔann-hu kāyn hnā wāḥəd lāfuntāġ (dyāl) al-musthalik wa huwa ann-hu ʕəndū sanad qānūnī bāš yaqqar yərfaʕ daʕwā ʕalā ḥād al-šārikāt wa-kāyina ʔiḥtimāliyya kabīra ġiddan fī kasb al-qaḍiyya, fī ḥīn ʔann-hu al-šārikāt lli qāmat b-hāda mākānš ʕənd-ha tā-šī sanad qānūni.

hiyya laʕbat faqat ʕala l-muʕṭā (dyāl) ʔann-hu ʔašlan tā šī ḥad mā təswuq fī al-ʔawwal minnī dārū al-ḥuṭwa lli kān ʕibāra ʕan ġass al-nabḍ ʕabra ilġāʔ ḥədmət VOIP min 3G, min baʕdā mā šāfū ann-hu tā šī riddat fīʕl qawwiyya mā dārt tiġāham ʔizzāʔ ḥād al-ḥuṭwa nādū ḥīdū tā (dyāl) ADSL].

‘Unfortunately, people *don’t know* that *there is* a (consumer benefit), that is that *there is* a legal document *with which they can* pursue these companies with a very high probability of winning the claim; whereas the companies, *that did this, had no* legal document.

They just played a game *to hide the possibility (of) risks*; they took the first step to test the reaction of people after the withdrawal of the VOIP package from 3G, and *after finding that there was* no strong reaction against this measure, *they also withdrew (of) the ADSL*’.

(Jawad Elansari, Opinions, Goud, 28/02/2016)

- (21) هاد الموضة (ديال) تصوير الممارسة الجنسية غادية وتوسع علمت "كود" من مصادر خاصة، أن الغرفة الجنحية بالمحكمة الابتدائية الزجرية بالدار البيضاء، شرعت أمس في محاكمة متهم من أجل "حيازة ونشر وتوزيع عبر الهاتف المحمول صور منافية للأخلاق والآداب العامة والتهديد والسب والشتم والقذف". وأضافت المصادر ذاتها، أن المتهم غير دوز ليلة سخونة مع جورج عاملات جنس وصورهم بلا خبارهم وقلبيها من بعد تهديد وتوزيع صور إباحية... عبر القضية ما صدقاتش ليه من بعد ما دخل البوليس ودار خدمتو...

23. In (20) and (21), italic refers to *Dārīġa* expressions, words in (brackets) refers to the analytical annexation and underlined words refer to synthetic annexation, both in the original and respectively in translation.

[hād al-mūḍa (dyāl) taṣwīr al-mumārasi al-ġinsiyya ġādiyya wa-tatawassaʿ ʿalamat «Gūd» min maṣādir ḥāṣṣa, ʿann al-ġurfa al-ġanḥiyya, bi-l-maḥkama al-ibtidāʿiyya al-zaġariyya bi-l-Dār al-Bayḍāʿ, šaraʿat ʿamsi fi muḥākama muttahir min ʿaġl «ḥayāza wa-našr wa-tawziʿ ʿabr al-hātif al-maḥmūl šuwar munāfiyya li-l-aḥlāq wa-l-adāb al-fāmma wa-l-tahdīd wa-l-sabb wa-l-šatam wa-l-qadf». wa-ʿaḍāfat al-maṣādir dāt-hā, ʿann al-muttahir ġīr dūz layla *shūna* maʿa *ġūġ fāmlāt ġins* wa-šawar-hum blā ḥbār-him wa-qalab-ha *min baʿd* tahdīd wa-tawziʿ suwar ʿibāḥiyya ... *ġīr* al-qadiyya *mā šadqāt-š līh min baʿd mā daḥal al-būlīs wa-dār ḥadamtū...*] ‘*This trend (of) filming sexual relationships is increasingly expanding*’.

Goud has been informed by private sources that yesterday the correctional chamber of the court of first instance in Casablanca began the trial of an accused for ‘possession, publication and dissemination by mobile phone of images contrary to morality, threats, insults, and defamation’.

The same sources added that the accused *had only* spent a *hot night* with *two sex workers*, and photographed them *without* informing them, *and then* returned the photos threatening to disseminate pornographic images. *But* the thing *did not turn out well to him, especially when the police did their job...*

(Goud, Casablanca, *Casanigra*, 14/07/2016)

In (20), the author criticises the non-respect of legal rules by telecommunication companies. It is an Opinion article in which the author, in order to support his ideas, tries to explain and inform readers of their rights. The lexical choices and morphosyntactic structures used in the article refer to a neutral and formal register that conforms to Youssi’s (1992) descriptions of Modern Moroccan Arabic. Observing also the genitive constructions in the entire article, which is written mainly in *Dāriġa*, one notices that analytical constructions are preferred and synthetic constructions of annexation are fewer. Quantitative analysis of the set of occurrences found in Goud’s journalistic corpus indicates that this tendency is typical of articles published in the Opinion column written mainly in *Dāriġa* (that is, where the combination with Arabic *Fuṣṣḥā* is limited). However, the tendency to use the analytical construction in other columns is less evident, and the texts show a greater combination of expressions and syntactic constructions that mix *Fuṣṣḥā* and *Dāriġa*, as shown, for example, in (21).

The short article given in (21) in fact shows that only one occurrence of *dyāl* appears exclusively in the title (الموضة ديال تصوير الممارسة الجنسية), [*hād al-mūḍa*

dyāl taṣwīr al-mumārāsa al-ġinsiyya ġādiyya wa-tatawassaʿ], ‘This trend of filming sexual relationships is expanding’). This article, divided into two long paragraphs, shows in the first part the subject of the news (a court case for dissemination of pornographic material), while in the second part, some details and developments of the story are explained. In the incipit, which is constructed using the specific journalistic language (see for instance the typical formulas of introduction of sources, such as *أعلنت كود من مصادر خاصة، أن*, [*ʕalamat “Gūd” min maṣādir ḥāṣṣa, ʔann...*], ‘Goud has been informed by private sources that’), there are no linguistic elements in *Dāriġa*, while in the second part, where some details of the affair are made explicit, there is the insertion of expressions in *Dāriġa* that combine with the syntactic construction in *Fuṣṣḥā*. See, for example, *وأضافت المصادر ذاتها، أن المتهم غير دوز ليلة سخونة مع جوج عاملات جنس*, [*wa-ʔaḏāfat al-maṣādir ḏāt-hā, ʔann al-muttaḥim ġīr dūz layla ṣḥūna maʕa ġūġ ʕāmlāt ġins...*], ‘The same sources added that the accused had only spent a hot night with two sex workers’ in which the main proposition (*أن*) [*wa-ʔaḏāfat al-maṣādir ḏāt-hā, ʔann*], ‘The same sources added that’ is constructed in *Fuṣṣḥā*, while the declarative, which follows the structure of an SVO-type nominal sentence, is constructed according to the syntactic structure and lexical choices of *Dāriġa* (see, for example, the restrictive sentence with *غير*, [*ġīr*], ‘only’ and the numeral *جوج*, [*ġūġ*], ‘two’).

In conclusion, this article shows how the middle/educated Moroccan Arabic used in *Goud’s* written production corresponds to the (oral) Modern Moroccan Arabic described in Youssi (1992). Moreover, unlike the Opinion Column (in which stylistic variation is more evident due to the textual and communicative nature of opinion articles), in the report articles the journalistic style and register is clearer and consequently the journalistic language also tends to adhere more closely to standard language without, however, neglecting the mixed and middle register of Moroccan Arabic, which in this case emerges from the lexical choices, that characterises *Goud’s* written production.

4. CONCLUSION

Elmedlaoui’s descriptive handbook, *al-ʕArabiyyatu al-Dāriġatu*, is in line with the initiatives of the *Zakoura* Centre for the Promotion of *Dāriġa* and represents a fairly detailed description of Moroccan Arabic. However, the descriptions and examples cited in his work do not allow one to distinguish the stylistic variation

of the middle register of Moroccan Arabic in the actual (written and oral) practices of Moroccan users.

Goud's journalistic corpus represents a genuine model of written expressions and dissemination of the middle/educated register of Moroccan Arabic. As such, it allows us to observe the phenomena of stylistic variation according to different communicative contexts.

The focus on the syntactic variation of declarative sentences and genitive constructs reflects the dynamics of stylistic variation depending on the different columns of the newspaper and thus on the different communicative functions and strategies of texts. More generally, the journalistic language of written production in *Goud* reflects the middle/educated register of Moroccan Arabic of oral communication.

The stylistic variation – that is, the more or less formal register tending towards syntactic and lexical choices (more or less) adhering to the linguistic elements of *Dāriġa* or the combination of *Fuṣḥā* and *Dāriġa* – reproduces the register of ordinary formal communications, in which the linguistic and lexical repertoire of *Fuṣḥā* also plays an important role in communication. In other words, the middle/educated register of Moroccan Arabic integrates expressions and syntactic structures of *Fuṣḥā* into the repertoire of *Dāriġa*, which Moroccan users have acquired (see the example of the legal language in the report article, which has no equivalent in *Dāriġa*).

In Elmedlaoui (2019), diglossia is considered a problem that can be solved through the use, promotion, and writing of the middle/educated register of *Dāriġa*. However, if real user practices are not taken into account, the *Zakoura Centre's* attempt to informally codify this middle register runs the risk of creating an artificial diglossia instead.

In agreement with Boussofara-Omar (2006, p. 636), diglossia represents a «richness that is often dramatically undervalued». The complementarity between *Fuṣḥā* and *Dāriġa* represents the richness and communicative power of language in the many possible interactional contexts, both in oral and written communication.

REFERENCES

- Aguadé, J. (2012). Monarquía, dialecto e insolencia en Marruecos: el caso Nichane. In M. Meouak, P. Sánchez, Á. Vicente (Eds.), *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*, pp. 441-464. Zaragoza: Universidad de Zaragoza.
- Badawi, E. (1973). *Mustawayāt al-ʿarabiyya al-muʿaṣira fī miṣr*. Cairo: Dār al-maʿārif.

- Blanc, H. (1960). Stylistic variation in spoken Arabic: A sample of interdialectal educated conversation. In C. A. Ferguson (Ed.), *Contributions to Arabic linguistics*, pp. 81-156. Cambridge, Mass: Harvard University.
- Bousofara-Omar, N. (2006). Diglossia. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski (Eds.), *Encyclopedia of Arabic language and linguistics*, Vol. I, pp. 629-637. Leiden: Brill.
- Brigui, F. (2016). De l'usage de l'arabe dialectal dans la presse écrite marocaine. In F.G. Moscoso & S. M. Adil (Eds.), *Identidad y conciencia lingüística*, pp. 249-264. Madrid: Universidad Autónoma de Madrid.
- Caubet, D. (1993). *L'Arabe marocain : Syntaxe et catégories grammaticales, textes*, Vol. I-II. Paris, Louvain: Peeters.
- Caubet, D. (2017a). Morocco: An Informal passage to literacy in *Dārija* (Moroccan Arabic). In J. Høigilt, G. Mejdell (Eds.), *The politics of written language in the Arab world - Writing change*, pp.116-141. Leiden: Brill, DOI: https://doi.org/10.1163/9789004346178_007.
- Caubet, D. (2017b). *Darija and the construction of 'Moroccanness'*. In R. Bassiouney (Ed.), *Identity and dialect performance. A study of communities and dialects*, pp. 99-124. New York: Routledge.
- Caubet, D. (2018). New elaborate written forms in *Darija*: Blogging, posting, and slamming in Morocco. In E. Benmamoun & R. Bassiouney (Eds.), *The Routledge handbook of Arabic linguistics*, pp. 387-406. London-New York: Routledge Taylor & Francis Group.
- Dell, F., Elmedlaoui M. (2003). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Springer Verlag. DOI: <https://doi.org/10.1007/978-94-010-0279-0>
- Ech-Charfi, A. (2014). The genitive in Moroccan Arabic. *International Journal of Linguistics* 6(1), 171-191.
- Elmedlaoui, M. (1995). *Aspects des représentations phonologiques dans certaines langues chamito-sémitiques*, (série Thèses et Mémoires, n°23). Rabat: Publication de la Faculté des Lettres et des Sciences Humaines de Rabat.
- Elmedlaoui, M. (2019). *Al-'arabiyyatu al-dāriġatu, 'imlā'īyyatun wa-naḥwun: al-'aṣwātu, al-ṣarfū, al-tarkību, al-mu'ġam*, al-Dār al-Bayḍā': Markaz tanmiyat al-dāriġa Zākūra.
- Ferguson, C. (1959). Diglossia. *Word* 15, 325-340.
- Ferguson, C. (1991). Epilogue: Diglossia Revisited. *Southwest Journal of Linguistics* 10(1), 214-234.
- Haugen, E. (1966). Dialect, language, nation. *American Anthropologist* 68(4), 922-35. DOI: <https://doi.org/10.1525/aa.1966.68.4.02a00040>
- Høigilt, J., Mejdell, G. (Eds.). (2017). *The politics of written language in the Arab World – Writing change*. Leiden: Brill. DOI: <https://doi.org/10.1163/9789004346178>
- Hoogland, J. (2018). *Darija in the Moroccan press. The case of the magazine Nichane*. *Sociolinguistic Studies* 12(2), 273-93. DOI: <https://doi.org/10.1558/sols.35567>
- Hudson, A. (1994). Diglossia as a special case of register variation. In B. Douglas & F. Edward (Eds.), *Sociolinguistic perspectives on register*, pp. 294-314. New York: Oxford University Press.

- Kebede, T. A., Kindt, K. T. (2016). *Language and social survey in Morocco, A tabulation report*, Fafo-report 2016:20.
- Lentin, J. (2008). Middle Arabic. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, (Eds.), *Encyclopedia of Arabic language and linguistics*, Vol. III, pp. 215-224. Leiden: Brill.
- Lentin, J. (2012). Normes orthographiques en moyen arabe : Sur la notation du vocalisme bref. In L. Zack, A. Schippers (Éds.), *Middle Arabic and mixed Arabic. Diachrony and synchrony*, pp. 209-234. Leiden-Boston: Brill.
- McCarus, E. N. (2008). Modern Standard Arabic. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski, (Eds.), *Encyclopedia of Arabic language and linguistics*, Vol. III, pp. 238-262. Leiden: Brill.
- Meiseles, G. (1980). Educated spoken Arabic and the Arabic language continuum. *Archivum Linguisticum 11*, 118-148.
- Mejdell, G. (2006). *Mixed styles in spoken Arabic in Egypt. Somewhere between order and chaos*. Leiden-Boston: Brill.
- Mejdell, G. (2017). Changing norms, concepts and practices of written Arabic. A 'long distance' perspective. In J. Høigilt & G. Mejdell (Eds.), *The politics of written language in the Arab world – Writing change*, pp. 68-89. Leiden: Brill.
DOI: https://doi.org/10.1163/9789004346178_005
- Mgharfaoui, K., Mabrou, A., Chekhayri, A. (2017). *Qāmūs al-dāriġa al-maġribiyya, al-Dār al-Bayḍā'*: markaz tanmiyat ad-dāriġa Zakūra.
- Miller, C. (2012). Observations concernant la présence de l'arabe marocain dans la presse marocaine arabophone des années 2009-2010. In M. Méouak, P. Sánchez, Á. Vicente (Eds), *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*, pp. 419-440. Zaragoza: Universidad de Zaragoza.
- Miller, C. (2016). Évolution des pratiques, évolutions des représentations ? Petit retour sur le débat autour de la valorisation de la darija au Maroc. *Revue des Études Berbères X*, .101-120.
DOI: <http://reb.centredechercheberbere.fr/revue-des-etudes-berberes.html?show=10>
- Miller, C. (2017). Contemporary dārija Writings in Morocco: Ideology and Practices. In J. Høigilt, G. Mejdell (Eds.), *The politics of written language in the Arab world - Writing change*, pp. 90-115. Leiden: Brill.
DOI: https://doi.org/10.1163/9789004346178_006
- Pennisi, R. (2020a). Written Dārija: “māšī mǝḳūl tǝktǝb-ha bi-ḥurūf al-luġa al-‘arabiyya!” It is not logical to write it with the Arabic letters! Media reception of the Zakoura dictionary project. *Annali di Ca' Foscari. Serie Orientale 56*, 129-154.
DOI: <https://doi.org/10.30687/AnnOr/2385-3042/2020/56/005>
- Pennisi, R. (2020b). *Arabe Mixte 2.0 : la variation syntaxique et stylistique dans les journaux numériques marocains (janvier-décembre 2016)*, PhD dissertation. Ca' Foscari University and Aix-Marseille Université.
- Ryding, K. C. (2008). Educated Arabic. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich, A. Zaborski (Eds.), *Encyclopedia of Arabic language and linguistics*, Vol III, pp. 666-671. Leiden: Brill.
- Youssi, A. (1992), *Grammaire et lexique de l'arabe marocain moderne*. Casablanca: Wallada.

NOTAS

LA PSICOLOGÍA COGNITIVA A EXAMEN: SOBRE EL LIBRO *CONVERSACIONES SOBRE LA MENTE COGNITIVA* DEL PROFESOR JOSÉ EUGENIO GARCÍA-ALBEA

José Manuel IGOA¹

Universidad Autónoma de Madrid

El 21 de enero del año 2020, en vísperas de la terrible pandemia que se cernía sobre nosotros, tuve la oportunidad de asistir a una mesa redonda celebrada en Tarragona en el marco del Simposio de la Sociedad Española de Lingüística, mesa que versaba sobre el problema del origen del lenguaje. En esa mesa redonda, el profesor José Eugenio García-Albea, autor del libro que motiva esta nota, oficiaba como moderador al lado de otros dos participantes, ambos destacados investigadores en distintos campos del saber, pero igualmente interesados y expertos en el tema que se debatía. Haciendo honor a su actitud siempre curiosa y entusiasta, además de bien informada, en cualquier asunto relacionado con el lenguaje y con la ciencia cognitiva, José Eugenio se embebió de tal modo en la discusión que acabó convirtiéndose en un participante más e hizo una encendida defensa de un punto de vista discrepante, en algunos aspectos, de los que mantenían los otros dos invitados, que por otra parte distaban bastante el uno del otro en sus planteamientos. La posición defendida por José Eugenio reflejaba una arraigada creencia que él siempre ha mantenido en relación con la psicología y otras disciplinas próximas a ella como la lingüística, y que se suele etiquetar como «realismo representacional». Aunque no me atrevería a afirmar rotundamente cómo reaccionó el público asistente en su totalidad ante la intervención del profesor García-Albea, sí me consta que una parte del mismo, formado en su mayoría por lingüistas, secundó con entusiasmo la postura que él defendía con determinación y que venía a sostener algo tan sencillo como que eso que llamamos «lenguaje», entendido como una facultad de la mente humana, realmente «existe» y, por tanto, es algo más que una etiqueta de

1. josemanuel.igoa@uam.es  <https://orcid.org/0000-0003-0628-5548>

conveniencia, el producto de una convención social o un agregado de elementos dispares, un mosaico de piezas que no le pertenecen en origen. El verbo *existe* tiene aquí un matiz ontológico que hace referencia a una forma de existencia empírica como categoría natural de fenómenos. El profesor García-Albea reaccionaba así contra quienes se afanan por disolver el lenguaje humano en un entramado de elementos no lingüísticos y están dispuestos a reducirlo a pura biología y también, y en sentido contrario, contra quienes pretenden elevarlo a la categoría de producto cultural sin una base natural específica, o sea, a «des-naturalizarlo».

Traigo a colación esta anécdota porque me parece un episodio que expresa de manera cabal el modo en que José Eugenio García-Albea ha transitado por las disciplinas en las que ha ocupado su vida académica, y no solo por sus compromisos teóricos, sino por una actitud que quienes lo han tratado reconocen como de profunda honestidad intelectual. De todo ello es reflejo el libro que es objeto de este comentario, un libro que puede provocar tanto adhesiones y coincidencias como desacuerdos y críticas, pero que refleja una trayectoria coherente de más de 40 años como teórico e investigador en las ciencias cognitivas.

El libro que nos ocupa es, por encima de todo, un libro de psicología cognitiva que trata de explorar los fundamentos teóricos de esta disciplina científica, partiendo del principio de que el comportamiento manifiesto de las personas se explica en función de los estados y procesos mentales que lo sustentan, de un sistema de representaciones simbólicas sobre las que opera un conjunto de procesos y mecanismos que activan, inhiben, acoplan y transforman esas representaciones y que tienen un valor causal en la génesis del comportamiento. Dicho de otro modo, la visión de la psicología humana que se defiende en el libro es la de un sistema de cálculo o cómputo inspirado en la llamada «metáfora del ordenador», pero tomándose muy en serio la idea de computación, que bajo este enfoque se contempla como una descripción literal, no metafórica, de los procesos cognitivos. Esto es, en esencia, lo que se conoce como «teoría computacional/representacional de la mente» (Pylyshyn, 1984; Horst, 2002).

La posición ontológica del realismo representacional a la que me he referido más arriba arranca de la corriente filosófica del racionalismo y ha aterrizado en las ciencias cognitivas de la mano de autores como Noam Chomsky, Jerry Fodor o Zenon Pylyshyn, tan queridos por el profesor García-Albea como discutidos y combatidos por otras corrientes actuales en esas ciencias, dentro de las cuales se sitúa también una parte importante de la lingüística. Estos tres autores configuran la espina dorsal del libro en forma de sendas entrevistas distribuidas a lo largo de

sus páginas. Cada entrevista, además, va precedida de un capítulo que conecta con el pensamiento del autor entrevistado a continuación. Así, a la entrevista con Jerry Fodor (capítulo 3) le antecede, en el capítulo 2, una revisión de la contribución de este autor a la ciencia cognitiva. Por su parte, la entrevista con Zenon Pylyshyn (capítulo 6) va precedida de un extenso tratamiento de la cuestión de la arquitectura mental (capítulo 5), que nuestro autor relaciona con (y a la vez distingue de) la noción de inteligencia enmarcada en la tradición de estudio de la psicología de las diferencias individuales. Finalmente, la entrevista con Noam Chomsky (capítulo 11) está ubicada a continuación de un capítulo dedicado al origen y desarrollo del lenguaje en los planos onto- y filogenético (capítulo 10), en el que se reconocen claramente los postulados del enfoque biolingüístico chomskiano en la explicación de estos procesos. A estas tres entrevistas se le añade una más que sirve de epílogo al libro (capítulo 12), en la que el propio autor expone, ante las preguntas de un avezado periodista, sus puntos de vista sobre las principales cuestiones que se abordan en el libro.

El compromiso del profesor García-Albea con una posición realista acerca de la existencia de las representaciones y los procesos de cómputo que conforman la arquitectura cognitiva humana y del poder causal de estos en la génesis del comportamiento lleva aparejada una concepción naturalista de la psicología cognitiva. Considerar la psicología como una ciencia natural, y los procesos cognitivos (incluido el lenguaje) como parte del mundo natural, impone severas restricciones sobre el tipo de explicaciones que pueden entrar en el dominio de lo psicológico, al tiempo que marca un contraste con otras concepciones de la psicología humana. Sin embargo, si damos por cierta la idea de que nuestro comportamiento está guiado por representaciones y que estas tienen un carácter simbólico -lo que equivale a decir que tienen significado-, nos encontramos con una suerte de paradoja, pues hemos de conciliar los principios explicativos de una ciencia natural que maneja entidades físicas (en el caso que nos ocupa, mecanismos y procesos neurobiológicos) con la explicación de comportamientos que se caracterizan como «intencionales», es decir, que se definen en función de su significado y, podría añadirse, de las metas y los objetivos que los agentes cognitivos tienen en mente al realizar conductas cognitivamente relevantes, es decir, basadas en creencias y deseos, como los que predica una psicología del sentido común (lo que en inglés se ha dado en llamar *folk psychology*). Así pues, el problema que afronta una psicología cognitiva de corte computacional como la que se defiende en este libro es el de *cómo naturalizar la intencionalidad* (Fodor, 1987). De esta espinosa cuestión se ocupa nuestro autor sobre todo en los capítulos 2, 5 y 7.

Este problema ciertamente no es nuevo en el panorama de la filosofía de la mente y las ciencias del comportamiento, y ha recibido distintas propuestas de solución a las que se alude en diversos lugares del libro. El profesor García-Albea, siguiendo sobre todo a Fodor y a Pylyshyn y basándose en otros notables antecesores suyos en filosofía de la mente, como Hilary Putnam, y en las ciencias de la computación, como Alan Turing, Allen Newell y Herbert Simon, apuesta decididamente por una forma de «funcionalismo computacional». En esencia, esta doctrina afirma que los estados mentales (computacionales) de un sistema, sea natural o artificial, se definen según el modo en que están causalmente relacionados entre sí y con las entradas sensoriales y las salidas motoras a las que están asociados, es decir, según la función que desempeñan en el sistema, y no por su constitución física. Sin embargo, para hacer compatible este postulado con el hecho incuestionable de que la causa material de los procesos de cómputo son los componentes físicos del sistema, los defensores del funcionalismo computacional recurren a la «hipótesis de la realización múltiple» (Putnam, 1967), según la cual la ejecución física de los procesos de cómputo se puede realizar (o «implementar», por usar un extendido neologismo) mediante diversos tipos de mecanismos, es decir, en distintos tipos de soporte físico.

Otro de los postulados fundamentales de la teoría computacional/representacional de la mente es la defensa de una arquitectura cognitiva de tipo simbólico, en la cual las unidades de procesamiento están dotadas de estructura y significado, pero los procesos operan sobre la estructura de las representaciones mentales y no directamente sobre su significado, con independencia de cómo se defina este último. Dado el papel primordial que se atribuye a la estructura de las representaciones, en una arquitectura simbólica cobra especial relevancia la noción de *regla*, como principio operativo en la manipulación de las representaciones mentales, en contraste con otro principio fundamental muy extendido en las teorías psicológicas como es el de *asociación* (Pinker, 1998). El contraste entre reglas y asociaciones no es patrimonio exclusivo de la psicología, sino que, como es sabido, se encuentra presente también en las teorías sobre la naturaleza del conocimiento del lenguaje y de las operaciones que subyacen a la adquisición y al uso de dicho conocimiento. Así, por ejemplo, un principio fundamental de la lingüística generativa es el carácter jerárquico de las expresiones lingüísticas y la sensibilidad a la estructura que exhiben los procesos de formación e interpretación de dichas expresiones. En esta teoría se tiende a ubicar los mecanismos de asociación lineal y secuenciación de unidades como parte de los procesos de externalización de cadenas de símbolos lingüísticos,

ya sea en la recepción sensorial o en la expresión motora de las mismas. En relación con ello, en el ámbito de la ciencia cognitiva, la dicotomía reglas-asociaciones marca la diferencia entre las arquitecturas simbólica (o clásica) y conexionista en diversos dominios y tareas cognitivas, asunto que también es objeto de reflexión en varios capítulos del libro (véase los capítulos 2, 3 y 5) y ha generado incontables debates entre los expertos (Fodor y Pylyshyn, 1988).

En lo que atañe a los dominios de indagación teórica e investigación empírica que son objeto de atención en el libro, el lenguaje ocupa, sin lugar a duda, un lugar primordial, seguido a distancia por la percepción. No en vano estos han sido, en ese mismo orden de importancia, los temas que han marcado la trayectoria del profesor García-Albea en su actividad docente e investigadora. En lo que respecta al lenguaje, y teniendo en cuenta la publicación para la que esta nota ha sido elaborada y sus potenciales lectores, es preciso aclarar dónde y cómo encaja el lenguaje, en tanto que facultad de la mente, en el entramado teórico que despliega nuestro autor en el libro. Como es bien sabido, el interés por el lenguaje como objeto de estudio científico ha sido fundamental en la génesis y el desarrollo de la ciencia cognitiva, y las investigaciones sobre el lenguaje en lingüística teórica, psicolingüística experimental y, posteriormente, en neurociencia han ocupado un lugar prominente en las ciencias del conocimiento. De hecho, uno de los factores influyentes al respecto es la irrupción de la teoría generativa del lenguaje en los albores de la moderna ciencia cognitiva, uno de cuyos hitos más recordados es la crítica que hizo Chomsky del conductismo skinneriano en la explicación de la naturaleza y la adquisición del lenguaje (Chomsky, 1959, y capítulo 11 del libro de García-Albea). Otro destacado factor que entró en juego años después fue la hipótesis de la modularidad de la mente, una propuesta sobre la arquitectura cognitiva que debemos al filósofo Jerry Fodor y que es objeto de especial atención en varios capítulos del libro (especialmente los capítulos 2, 3, 4 y 5). Tirando de estos hilos surgen varias cuestiones que se discuten en el libro relacionadas con el estudio del lenguaje como facultad de la mente. Todas ellas se exponen a partir de los testimonios recogidos en las entrevistas con Fodor y con Chomsky, en una suerte de diálogo implícito entre estos dos autores en el que ambos expresan su respeto y consideración por el otro, aun reconociendo las diferencias que los separan.

Una primera cuestión de carácter general atañe a las relaciones entre la lingüística teórica y la psicolingüística como disciplinas de referencia en el estudio del lenguaje, y las diferencias que se observan entre ellas en el modo de entender la arquitectura modular de esta facultad de la mente. El origen de estas diferencias remite a la

conocida distinción chomskiana entre competencia y actuación. Reconociendo la diferencia de perspectiva que caracteriza a cada ámbito de estudio, la competencia como foco de interés para los lingüistas y la actuación para los psicolingüistas, ambos autores coinciden en el carácter complementario de ambas perspectivas, aunque reconocen que cada una de ellas conlleva una concepción diferente de la arquitectura de la facultad del lenguaje: así, mientras que Chomsky propone una arquitectura modular para la competencia lingüística (o el Lenguaje-I, según la terminología más reciente), basada en la especificidad de los dominios de conocimiento que conforman la facultad del lenguaje (Chomsky, 1980), Fodor concibe la modularidad como una propiedad del lenguaje entendido como sistema de entrada (y posiblemente de salida) que media entre el pensamiento (los sistemas centrales de su teoría) y las actividades perceptiva y motora en la comprensión y producción del lenguaje, destacando el carácter autónomo, o informativamente encapsulado, de estos sistemas de procesamiento periféricos (Fodor, 1983). Fodor añade además un matiz interesante en su entrevista, al considerar que el estudio de la competencia, en cualquier dominio de interés científico, corresponde a un nivel de abstracción mayor, mientras que la investigación sobre la actuación es una indagación sobre los efectos de interacción que se dan cuando un hipotético sistema abstracto de conocimiento entra en contacto con otras variables que afectan a su funcionamiento. A mi juicio, esta es una forma de resaltar el interés común de la lingüística y la psicología en el estudio del lenguaje y de reivindicar la dependencia mutua entre ambas disciplinas.

Una segunda cuestión se refiere a las relaciones entre el lenguaje y el pensamiento y al papel que el lenguaje desempeña en los procesos de conceptualización y razonamiento, materia que ha sido y sigue siendo objeto de debates entre filósofos, lingüistas y psicólogos. En este asunto, las posturas de Chomsky y Fodor son más distantes y difíciles de conciliar. Así, Fodor defiende que el lenguaje, entendido como módulo de entrada-salida, es un sistema de propósito específico diseñado para expresar e interpretar intenciones comunicativas y, por tanto, es subsidiario y dependiente del pensamiento, es decir, de los «sistemas centrales» de la arquitectura cognitiva. Estos operan de acuerdo con un «lenguaje del pensamiento», un sistema combinatorio y conceptual amodal y prelingüístico con una fuerte base innata, necesario para el aprendizaje de cualquier facultad cognitiva, entre ellas la del lenguaje, y por tanto anterior a todas ellas (Fodor, 1975). En cambio, Chomsky asume que al menos una parte del pensamiento, la que se puede formular en términos de expresiones lingüísticas estructuradas, está vinculada al mecanismo generativo del

lenguaje mediante reglas de semántica-pragmática formal. Esto es lo que Chomsky denomina «pensamientos lingüísticamente articulados» en su entrevista, en la que subraya las diferencias de este planteamiento con respecto a las ideas de Fodor. Sin embargo, por encima de estas diferencias, ambos autores parecen coincidir en la idea de que lo que llamamos *pensamiento* es un dominio de estudio difícil de definir, y expresan serias reservas de que podamos llegar a alcanzar un conocimiento fiable del mismo. Así, Chomsky lo considera un fenómeno mal definido y, por tanto, científicamente inabordable, y Fodor asevera que adolece de un problema irresoluble que ya advirtieron los expertos en inteligencia artificial y se ha dado en llamar «problema del marco» (o de la relevancia), consistente en la imposibilidad de fijar un límite en las variables que pueden intervenir en procesos cognitivos complejos como los del razonamiento no demostrativo. Este problema, omnipresente en la psicología, se puede ilustrar con un aforismo atribuido al profesor Mariano Yela, uno de los pioneros de la psicología académica en España, que además fue maestro de profesor García-Albea en sus años de formación y a quien nuestro autor dedica uno de los capítulos del libro, aforismo que reza así: en psicología «todo tiene que ver con todo, pero no del todo».

Derivadas de los anteriores problemas, hay dos cuestiones más que marcan importantes diferencias entre Chomsky y Fodor en sus concepciones acerca del lenguaje. La primera de ellas es la cuestión del significado de las expresiones lingüísticas. En esta materia, Chomsky aboga por una aproximación internista, según la cual los símbolos lingüísticos son entidades mentales cuya conexión con objetos y eventos extralingüísticos, esto es, su significado referencial, es incierto y, por tanto, difícil de establecer. Desde este punto de vista, la dimensión referencial del significado queda más allá de las posibilidades de una explicación científica satisfactoria. Este escepticismo hacia las teorías referencialistas del significado le lleva a sostener que una semántica mentalista es la aproximación más fecunda, o incluso la única posible por ahora, al análisis del significado, que debe quedar circunscrito a los límites expresivos que marca el lenguaje (Chomsky, 2000). En cambio, Fodor trata de conciliar las perspectivas internista y externista del significado con su teoría del «atomismo informacional» (Fodor, 1998). Esta teoría defiende una relación de dependencia asimétrica entre las representaciones verídicas del significado de una expresión lingüística (p. ej., las instancias de caballos en nuestra experiencia como causa del concepto CABALLO) y las expresiones erróneas del mismo (así, instancias de vacas como causa del concepto CABALLO), de suerte que la formación de estas últimas depende del vínculo causal creado por las primeras, pero nunca a la inversa.

Esta teoría es, según Fodor, compatible con una explicación mentalista (internista) de las representaciones del significado en el formato del lenguaje del pensamiento.

La última cuestión sobre las diferencias de opinión de Fodor y Chomsky que merece la pena señalar es la que compete al papel de la biología y su relación con la psicología en la explicación de la facultad del lenguaje humano. Este asunto se aborda en la entrevista a Fodor a propósito de la defensa del funcionalismo computacional (*vid supra*), que este autor expresa con claridad en afirmaciones como la siguiente: «desde el punto de vista del nivel natural de explicación, es una especie de accidente que los sistemas psicológicos resulten estar encarnados en sistemas biológicos» (p. 49; en cursiva en el original). En este sentido, Fodor no oculta sus diferencias con la postura biologicista, y posiblemente reduccionista, que Chomsky asume en la versión más reciente de la teoría generativa, el programa minimista adscrito al enfoque biolingüístico. Con todo, Chomsky también ha manifestado en otras ocasiones que la lingüística es una rama de la psicología cognitiva (Chomsky, 1972), lo cual, considerando el énfasis que pone en el carácter biológico de la facultad del lenguaje, invita a pensar que la psicología cognitiva debe ser, en último término, una rama de la biología. Sea como fuere, la matriz biológica o psicológica del lenguaje como facultad de la mente no parece haber suscitado debates en la lingüística o la psicolingüística comparables al que ha despertado en la filosofía de la mente, aunque no por ello carece de interés científico.

Antes de finalizar esta reseña sobre el libro del profesor García-Albea, he de referirme a tres capítulos del libro no mencionados hasta ahora, que inciden en cuestiones epistemológicas, en la medida en que hacen alusión a las fuentes de adquisición de conocimiento sobre la naturaleza y el funcionamiento de los procesos cognitivos. Así, el capítulo 1 es una respuesta a una crítica de la práctica científica en psicología vertida en un artículo de un experto en metodología de las ciencias del comportamiento, crítica que incide en la situación de crisis permanente que afecta a esta disciplina y subraya el dogmatismo metodológico en que se halla sumida. Aun compartiendo este sombrío diagnóstico en algunos aspectos, el profesor García-Albea rompe una lanza a favor del pluralismo metodológico de esta ciencia y defiende la unidad de la psicología en virtud de su objeto de estudio. Por otra parte, los capítulos 8 y 9 están dedicados a enjuiciar la aportación de las neurociencias a la empresa de explicar los procesos cognitivos. García-Albea emprende aquí una labor pedagógica dirigida a situar las teorías y los métodos de investigación en neurociencia en el lugar que les corresponde dentro del estudio de los procesos cognitivos. Por una parte, se afana en aclarar algunas confusiones terminológicas

en las que a veces se incurre en estas disciplinas, y por otra, critica de forma bien argumentada su pretensión de erigirse en la solución definitiva a los problemas de la ciencia cognitiva y de dar una respuesta concluyente a los interrogantes que plantean, si bien reconoce los avances teóricos y metodológicos que han aportado en aras de un mejor conocimiento de su objeto de estudio.

Para concluir esta nota, y retomando las palabras con las que la iniciaba, quisiera subrayar la coherencia temática de esta obra del profesor García-Albea y recalcar la original articulación del libro en torno a las entrevistas que le sirven de hilo conductor. Estas entrevistas tienen la virtud de haber extraído, a través de sus bien planificadas preguntas y sus oportunos comentarios, la información pertinente que cada entrevistado ha podido aportar a la ciencia cognitiva desde su campo particular de especialización. El motivo central del libro que aparece en el título, la conversación como forma de interacción humana, se puede contemplar como una guía de lo que debe ser la actividad de indagación científica, una actividad que solo es posible en colaboración con otros y que ha de estar basada en el intercambio de ideas y el sometimiento de estas a la consideración y al juicio crítico de los demás. Este no es ciertamente un libro neutral o ecléctico en sus planteamientos, sino franco y transparente en sus intenciones y en sus argumentos, lo que no le resta un ápice de interés. Como bien señala el autor en su presentación del libro, esta es una obra que puede atraer a cualquier persona interesada en la psicología humana, sobre todo en aspectos relacionados con la naturaleza, el desarrollo y el funcionamiento de los procesos cognitivos, con especial énfasis en el lenguaje, por lo que considero que es recomendable para estudiantes y estudiosos de la filosofía, la lingüística, la psicología, la neurobiología y las ciencias de la computación, así como del resto de las ciencias sociales y las humanidades.

REFERENCIAS

- Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language* 35, 26-57. [Crítica de *Verbal Behavior* de B.F. Skinner. En *¿Chomsky o Skinner? La génesis del lenguaje*. Barcelona: Fontanella, 1977.]
- Chomsky, N. (1972). *Language and mind*. Nueva York: Harcourt Brace Jovanovich. [*El lenguaje y el entendimiento*, Barcelona: Planeta-Agostini, 1992.]
- Chomsky, N. (1980). *Rules and representations*. Nueva York: Columbia University Press. [*Reglas y representaciones*, México: Fondo de Cultura Económica, 2001.]
- Chomsky, N. (2000). *New horizons in the study of language and mind*, Nueva York: Cambridge University Press.

- Fodor, J.A. (1975). *The language of thought*. Nueva York: Crowell. [*El lenguaje del pensamiento*, Madrid: Alianza, 1985.]
- Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press. [*La modularidad de la mente*. Madrid: Morata, 1986].
- Fodor, J.A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press. [*Psicosemántica*, Madrid: Tecnos, 1994.]
- Fodor, J.A. (1998) *Concepts: Where cognitive science went wrong*. New York: Oxford University Press. [*Conceptos: donde la ciencia cognitiva se equivocó*, Barcelona: Gedisa, 2014.]
- Fodor, J.A. y Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition* 28(1-2), 3-71.
- Horst, S. (2002). Teoría computacional de la mente, en R.A. Wilson y F. Keil (Eds.), *Enciclopedia MIT de ciencias cognitivas*. Madrid: Síntesis.
- Pinker, S. (1998). Words and rules. *Lingua* 106, 219-242.
- Putnam, H. (1967). The nature of mental states, en W. H. Capitan y D. D. Merrill (Eds.), *Art, mind and religion*. Pittsburgh: University of Pittsburgh Press. Reimpreso en Ned Block, Ed., *Readings in Philosophy of Psychology*. Cambridge, MA: Harvard University Press, 1980, pp. 223-231.
- Pylyshyn, Z.W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press. [*Computación y conocimiento: hacia una fundamentación de la ciencia cognitiva*, Madrid: Debate, 1988.]

José Manuel Igoa González
Universidad Autónoma de Madrid
Facultad de Psicología
C/ Iván Pavlov, 6
Ciudad universitaria de Cantoblanco
Madrid 28049

RESEÑAS

RESEÑA

GALLEGO, Ángel J. y Edita GUTIÉRREZ RODRÍGUEZ. *Trabajando el Glosario de términos gramaticales. Ejercicios reflexivos y competenciales*. Madrid: Arco/Libros (Cuadernos de Lengua Española, 148), 2022, 89 pp., ISBN 978-84-7133-875-4.

El contenido fundamental de este libro se circunscribe a la presentación de una serie de ejercicios sobre gramática española cuya finalidad es aprender a «observar, pensar, argumentar, hacer deducciones y generalizaciones, para elaborar hipótesis y elegir entre distintas opciones de análisis» (p. 12). Todos estos ejercicios tienen la particularidad de originarse, al menos en parte, en los conocimientos teóricos que se pueden extraer del reciente *Glosario de términos gramaticales* (2019) (*GTG*, de ahora en adelante), concebido *grosso modo* como un pequeño diccionario de gramática con unas 500 entradas y confeccionado por la Real Academia Española (RAE) en alianza con la Asociación de Academias de la Lengua Española (ASALE). A zaga de cada una de ellas, encontramos los contenidos terminológicos de la *Nueva gramática de la lengua española* (*NGLE*), un extenso tratado gramatical descriptivo y normativo a la vez que atiende, asimismo, a cuestiones de índole variacional en el ámbito hispánico (*vid.* prólogo del *GTG*; Bosque (2020) o Bosque y Brucart (2021) para una actualización de la futura edición de esta gramática panhispánica). A pesar de que estos ejercicios bien podrían usarse por los diferentes grupos de discentes de Educación Secundaria y Bachillerato, están más bien pensados como posible modelo para los instructores en esas etapas de la Enseñanza Media.

Como bien se advierte en el introito de la publicación, el germen de este cuaderno está concebido a partir de los manuales *Problemas de morfosintaxis* (Bosque, 1980) y *Repaso de sintaxis tradicional* (Bosque, 2022, originalmente publicado en 1999), aunque estos últimos quizá están más orientados, en los tiempos que corren, hacia estudiantes de las carreras de filología y lingüística. En términos generales, el volumen está organizado de la siguiente manera: el prefacio (pp. 5-9) lo firma

Ignacio Bosque, destacado gramático del español, y sirve como desiderátum de la obra en su conjunto (pp. 11-16); una relación de 60 ejercicios, con opciones de respuesta múltiple y, también, de respuesta abierta (pp. 19-47); respuestas razonadas a los mismos (pp. 49-84); e índice temático que permite la identificación de los conceptos gramaticales utilizados (pp. 85-89). La selección de los ejercicios, ciertamente el grueso del libro, no es en absoluto azarosa, pues responde a la necesidad (no perentoria) de identificar «las unidades básicas de la sintaxis, como las clases de palabras, las funciones sintácticas y semánticas, o los diferentes tipos de sintagmas y oraciones» (p. 11), objetivo principal que se marcan al alimón los profesores Gallego y Gutiérrez Rodríguez en los prolegómenos del volumen.

El compendio está organizado en torno a seis tipos de ejercicios, a saber: de análisis directo, de secuencias agramaticales, de análisis inverso, de pares mínimos, de análisis de secuencias ambiguas y de elección de análisis (*vid.* Bosque y Gallego, 2016, para más información sobre esta tipología de ejercicios gramaticales; o el reciente estudio en Bosque y Gallego, 2023). Esta clasificación nos recuerda *mutatis mutandis* a la que ya se empleó en Bosque (2022) entre *ejercicios identificativos* –identificar categorías y funciones, y, como variante de estos, pares mínimos– y *ejercicios constructivos* –determinar si una serie de secuencias es sintácticamente válida o no–. Cada uno de estos seis bloques temáticos contiene una selección de 10 actividades de diferentes niveles de dificultad, que podrían considerarse como *elemental* (*), *intermedio* (**), *avanzado* (***) y *superior* (****), tal y como se realizó en Bosque (2022). Todos los bloques, además, contienen al menos un ejercicio de cada nivel de dificultad, lo que facilita su manejo en diferentes escenarios educativos.

Los *ejercicios de análisis directo* –posiblemente los de corte más tradicional del libro– exigen el análisis sintáctico de algunos de los elementos constitutivos de una secuencia en español. Los *ejercicios de secuencias agramaticales*, como bien indica su nombre, solicitan al estudiante que «formule un principio positivo que permita explicar la agramaticalidad; es decir, una generalización que incluya todos los elementos que funcionen igual» (p. 14). Este tipo de actividades permiten al alumno reflexionar sobre los principios subyacentes del sistema gramatical del español, como la transitividad en **El barco fue llegado*. Los *ejercicios de análisis inverso*, por su parte, exigen la producción de secuencias que respondan a una serie de condiciones gramaticales, como sería la creación de una oración subordinada que funcione sintácticamente como sujeto y, paralelamente, esté encabezada por un adverbio de modo. Los *ejercicios de pares mínimos* contienen dos tipos de tareas: a) actividades que incorporan dos secuencias posibles en la lengua en las

que se debe determinar en qué radica la diferencia (de significado) (e.g., *Escribí a Ana* vs. *Dibujé a Ana*), y b) actividades que amalgaman construcciones gramaticales y agramaticales donde se debe deducir el principio que viola la secuencia que no es posible (e.g., *Sale gas* vs. **Molesta gas*). Los *ejercicios de análisis de secuencias ambiguas* presentan una serie de oraciones ambiguas en español que los alumnos, por un lado, deben parafrasear para dar cuenta de los distintos sentidos posibles y, por el otro, tienen que proponer análisis alternativos que expliquen las posibles interpretaciones dadas, como ocurre con el sintagma nominal *La elección del presidente*. Por último, los *ejercicios de elección de análisis* muestran una gama de diferentes análisis para una secuencia con el fin de que los estudiantes puedan determinar la opción más apropiada y desechar las representaciones que son erróneas, fundamentándose en argumentos gramaticales (de tipo empírico) como, por ejemplo, la adscripción categorial de la lexía *excepto* como posible preposición, adverbio, conjunción coordinante o conjunción subordinante.

Es conveniente mencionar en este punto también que cada ejercicio de este libro está acompañado de una serie de referencias que remiten tanto al *GTG* como a la *NGLE*. Esto facilita sobremanera la utilización de una fuente de recursos didácticos para el aprendizaje (o refuerzo) de los contenidos gramaticales que el estudiante debería manejar con soltura para resolver las tareas propuestas y que podemos encontrar en dichas obras académicas. Como ya se ha mencionado anteriormente, la inclusión de respuestas razonadas permite, tanto al estudiante como al docente, tomar como referencia el tipo de contestación que se podría esperar en cada caso. Hay que recordar que «la renovación pedagógica en el ámbito de la gramática básica no se consigue cambiando la terminología [...], sino más bien cambiando la rutina por la reflexión y sustituyendo los hábitos memorísticos por los que fomentan el desarrollo de la capacidad argumentativa» (Bosque, 2022: 9-10).

Sirva igualmente esta impresión como modelo para la elaboración de recursos didácticos sobre la gramática de otras lenguas, como el inglés, algo que, hasta donde sabemos, no se ha publicado todavía, a pesar de la existencia de interesantes publicaciones como la monografía de Berry (2021), donde se intercala magistralmente lo lingüístico y lo pedagógico, o las siempre interesantes observaciones realizadas por Huddleston cuando analizaba la enseñanza de la gramática inglesa en Australia (*vid.* Huddleston, 1989). La situación de la enseñanza de la gramática inglesa es, no obstante lo anterior, un tanto diferente (*vid.* Crystal, 2017, particularmente el documento «English grammar in the UK: A political history» suplementario al estudio, o Aarts *et al.*, 2019).

Esta publicación supone una importante incorporación más al ya cuantioso grupo de estudios sobre la enseñanza de la gramática española, en todas sus manifestaciones y vertientes metodológicas, que, desde hace algunos años, se han ido publicando en diferentes editoriales y revistas especializadas (*vid.* Bosque y Gallego, 2016, 2018, 2021; Bosque, 2018; Marqueta Gracia, 2023; Gallego y Gutiérrez Rodríguez, 2022; o Gil Laforga y Gutiérrez Rodríguez, 2022; o el exitoso volumen, que va ya por la novena edición, de Lozano Jaén, 2020; por mencionar algunos estudios relevantes), así como algunas otras colecciones de ejercicios donde, igualmente, se persiguen fines pedagógicos (e.g., Borrego Nieto y Gómez Asencio, 2010, o Alonso Pascua, 2023, por mencionar dos de ellas). Nos encontramos, pues, ante una excelente adición a las numerosas monografías que la colección «Cuadernos de lengua española» ha venido publicando desde su origen sobre diferentes aspectos de la sintaxis y la morfología del español. Además, se puede concebir como un valioso recurso para ir construyendo poco a poco los hábitos y las bases metodológicas en pro de una *gramática reflexiva* cimentada sobre sólidos posicionamientos científicos. Este logro, en efecto, conseguiría que se introdujesen de forma paulatina las premisas de la conocida como *grammar-as-science conjecture* (Bosque y Gallego, 2023) en el currículo educativo actual.

REFERENCIAS

- Aarts, B., I. Cushing y R. Hudson. (2019). *How to teach grammar*. Oxford: Oxford University Press.
- Alonso Pascua, B. (2023). *Prácticas de dialectología sincrónica española*. Salamanca: Ediciones Universidad de Salamanca.
- Berry, R. (2021). *Doing English grammar: Theory, description and practice*. Cambridge: Cambridge University Press.
- Borrego Nieto, J. y J. J. Gómez Asencio. (2010). *Prácticas de fonética y fonología*. 3ª ed. Salamanca: Ediciones Universidad de Salamanca.
- Bosque, I. (1980). *Problemas de morfosintaxis*. Madrid: Editorial de la Universidad Complutense.
- Bosque, I. (2018). Qué debemos cambiar en la enseñanza de la gramática. *ReGroC. Revista de Gramática Orientada a las Competencias* 1, 11-36.
- Bosque, I. (2020). La segunda edición de la *Nueva gramática de la lengua española* y el nuevo *Glosario de términos gramaticales*, en S. Muñoz Machado (dir.), *Crónica de la lengua española 2020*, pp. 201-217. Barcelona: Espasa-Planeta.
- Bosque, I. (2022). *Repaso de sintaxis tradicional: ejercicios de autocomprobación*. 9ª ed. Madrid: Arco/Libros.

- Bosque, I. y J. M. Brucart. (2021). Los trabajos para la nueva edición de la *Nueva gramática de la lengua española*, en S. Muñoz Machado (dir.), *Crónica de la lengua española 2021*, pp. 558-564. Barcelona: Espasa-Planeta.
- Bosque, I. y Á. J. Gallego. (2016). La aplicación de la gramática en el aula. Recursos didácticos clásicos y modernos para la enseñanza de la gramática. *Revista de Lingüística Teórica y Aplicada* 54(2), 63-83.
- Bosque, I. y Á. J. Gallego. (2018). La gramática en la Enseñanza Media. Competencias oficiales y competencias necesarias. *ReGroC. Revista de Gramática Orientada a las Competencias* 1, 141-202.
- Bosque, I. y Á. J. Gallego. (2021). La terminología gramatical en Secundaria y Bachillerato. Ventajas e inconvenientes didácticos de algunos conceptos gramaticales antiguos y modernos. *Revista Española de Lingüística* 51(2), 51-78.
- Bosque, I. y Á. J. Gallego. (2023). Addressing language in secondary education: Advantages of a scientific approach to grammar teaching, en A. Corr y A. Pineda (eds.), *Theoretical linguistics in the pre-university classroom*, pp. 207-230. Oxford: Oxford University Press.
- Crystal, D. (2017). *Making sense: The glamorous story of English grammar*. Londres: Profile Books.
- Gallego, Á. J. y E. Gutiérrez Rodríguez. (2022). La subordinación en el *Glosario de términos gramaticales*. Teoría, didáctica y su aplicación a la escritura. *Tejuelo: Didáctica de la Lengua y la Literatura. Educación* 35(2), 173-204.
- Gil Laforga, I. y E. Gutiérrez Rodríguez. (2022). Terminología gramatical, en B. Marqueta Gracia, N. López Cortés y A. Ariño Bizarro (eds.), *Avances de la Lingüística y su aplicación didáctica*, pp. 129-147. Madrid: Akal.
- Huddleston, R. (1989). *English grammar in school textbooks: Toward a consistent linguistic alternative*. Mt. Gravatt, Queensland: Applied Linguistics Association of Australia.
- Lozano Jaén, G. (2020). *Cómo enseñar y aprender sintaxis. Modelos, teorías y prácticas según el grado de dificultad*. 9ª ed. Madrid: Cátedra.
- Marqueta Gracia, B. (2023). *La reflexión sintáctica: teoría y práctica para la docencia*. Zaragoza: Prensas de la Universidad de Zaragoza.
- RAE y ASALE (2009). *Nueva gramática de la lengua española*. 2 vols. Madrid: Espasa.
- RAE y ASALE. (2019). *Glosario de términos gramaticales*. Salamanca: Ediciones Universidad de Salamanca.

PABLO M. TAGARRO

Universidad del País Vasco/Euskal Herriko Unibertsitatea

MIGUEL CUEVAS-ALONSO

Universidade de Vigo

