

Año 42, Número 78, enero de 2024

ISSN 0185-2647



ELA ESTUDIOS DE LINGÜÍSTICA APLICADA



Universidad Nacional Autónoma de México

Coordinación de Humanidades

Guadalupe Valencia García, Coordinadora

Escuela Nacional de Lenguas, Lingüística y Traducción

María del Carmen Contijoch Escontria, Directora

Estudios de Lingüística Aplicada

Comité asesor

Mark Amengual, Universidad de California,
Estados Unidos

Christian Fandrych, Universität Leipzig, Alemania

Ana Isabel García Tesoro, Universidad de
Antioquia, Colombia

Sara Gómez Seibane, Universidad de la Rioja,
España

Natalia Ignatieva Kosminina, Universidad
Nacional Autónoma de México, México

Luis Fernando Lara, El Colegio de México,
México

Olivier Le Guen, Centro de Investigación y
Estudios Superiores en Antropología Social,
México

Paulette Levy Brzezinska, Universidad Nacional
Autónoma de México, México

Fernanda López, Universidad Nacional Autóno-
ma de México, México

Pedro Martín Butragueño, El Colegio de México,
México

Stanislav Mulík, Universidad Nacional Autónoma
de México, México

Leonor Orozco Vaca, Universidad Nacional
Autónoma de México, México

Natalia Prunes, Universidad de Buenos Aires,
Argentina

Álvaro Hernán Quintero Polo, Universidad
Distrital Francisco José de Caldas, Colombia

Dietrich Rall, Universidad Nacional Autónoma de
México, México

Ígor Rodríguez Iglesias, Universidad Autónoma
de Madrid, España

Cecilia Rojas Nieto, Universidad Nacional
Autónoma de México, México

Virna Velázquez Vilchis, Universidad Autónoma
del Estado de México, México

Comité editorial

Graciela Fernández Ruiz, El Colegio
de México, CELL

Ittay Gil Carrillo, Universidad Nacional
Autónoma de México, ENALLT

Georgia M. K. Grondin, Universidad Autónoma
del Estado de México, Facultad de Lenguas
María del Refugio Pérez Paredes, Universidad
Autónoma Metropolitana-Iztapalapa
Nadieżdha Torres Sánchez, Universidad
Nacional Autónoma de México, IIFL
Victoria Zamudio Jasso, Universidad Nacional
Autónoma de México, ENALLT

Coordinador del Comité editorial

Daniel Rodríguez Vergara, Universidad Nacional
Autónoma de México, ENALLT

Equipo editorial

Mariela Elizabeth Pérez Castañeda,
editora responsable

Claudia Isabel Carrillo Sánchez, editora adjunta

Estudios de Lingüística Aplicada está incluida
en los siguientes índices:

CLASE (Citas Latinoamericanas en Ciencias
Sociales y Humanidades)

Dialnet

EBSCO

Infoling, Información global sobre lingüística
hispanica

Latindex, Sistema Regional de Información en
Línea para Revistas Científicas de América Latina,
el Caribe, España y Portugal

LLBA (Linguistics and Language Behavior Abstracts)

MLA (MLA International Bibliography)

Revistas UNAM

SERIUNAM

Ulrich's Periodicals Directory

WorldCat

D. R. © 1981

Año 42, número 78, enero de 2024

Universidad Nacional Autónoma de México

Escuela Nacional de Lenguas, Lingüística
y Traducción

Circuito Interior, Cd. Universitaria, Ciudad de
México, México, 04510

Dirección electrónica: ela@enallt.unam.mx

Página electrónica: <http://ela.enallt.unam.mx/>

Impreso y hecho en México

ISSN (impreso) 0185-2647

doi: 10.22201/enallt.01852647p.2024.78

ELA ESTUDIOS DE
LINGÜÍSTICA APLICADA

Año 42 ▶ Número 78 ▶ enero de 2024

Universidad Nacional Autónoma de México
Escuela Nacional de Lenguas, Lingüística y Traducción

ARTÍCULOS ►

Autorreparaciones en la afasia anómica
en español. Un estudio de caso 7
Self-repairs in Spanish anomic aphasia.
A case study
Olivia Castillo Castillo

Técnicas básicas de procesamiento del lenguaje
natural en el análisis de menciones a mujeres
en lecturas graduadas de ELE 45
Basic natural language processing
techniques for the analysis of mentions
of women in Spanish graded readers
Iban Mañas Navarrete
& María-Valle Sell

NOTAS ►

Productividad relativa de los esquemas de
formación de palabras en el español de México.
Actualización 2022 79
Relative productivity of word formation schemes
in Mexican Spanish. Update 2022
Ramón Zacarías Ponce de León

Diseño de tareas y materiales para recopilar un
corpus de aprendices mexicanos de lengua
inglesa 95
Design of tasks and materials to collect a corpus
of Mexican learners of English
Ana Abigahil Flores Hernández
& Pauline Moore

RESEÑA ►

Daniela Lauria. *Lengua y política. Historia crítica de los diccionarios del español de la Argentina* 127
Luis B. Quesada Nieto

Lineamientos para publicar en *Estudios de Lingüística Aplicada* 133



Artículos

Autorreparaciones en la afasia anómica en español. Un estudio de caso

Self-repairs in Spanish anomic aphasia.
A case study

Olivia Castillo Castillo

Benemérita Universidad Autónoma
de Puebla, Facultad de Ciencias
de la Comunicación
olivia.castillocas@correo.buap.mx



Recepción: 19 de febrero del 2023

Aceptación: 19 de septiembre del 2023

doi: 10.22201/enallt.01852647p.2024.78.1055

Resumen

La afasia anómica es un trastorno neurológico que se caracteriza por la presencia constante de problemas en el acceso léxico. Estos problemas pueden ir desde una leve dificultad en la producción de palabras deseadas durante una conversación hasta la virtual incapacidad de producir sustantivos bajo cualquier condición. Las descripciones de este síndrome señalan a los silencios prolongados, falsos inicios y repeticiones constantes, como elementos problemáticos en el acceso léxico. En este trabajo se discuten dichas descripciones a partir del análisis de cuatro historias elicítadas narradas por un hablante con afasia anómica. Se muestran operaciones de autorreparación (AR) como búsqueda, reciclaje, eliminación y pausas entre unidades entonacionales. A partir de estos datos se presenta un instrumento de categorización que permite realizar observaciones sobre las formas gramaticales más recurrentes en cada una de las operaciones de AR y que proporciona información acerca de los procesos elaborados por el hablante con afasia anómica para acceder a elementos léxicos. Se presenta a las AR como elementos que privilegian la continuidad discursiva mediante elementos lingüísticos específicos; entre ellos destaca la frase nominal.

Palabras clave: reinicios; operaciones de autorreparación; narración; pausas

Abstract

Anomic aphasia is a neurological disorder characterized by constant problems in lexical access. These can range from a slight difficulty producing desired words during conversation to the virtual inability to produce nouns under any condition. Descriptions of this syndrome point to long silences, false starts, and constant repetitions as problematic elements in lexical access. This paper discusses these descriptions from the analysis of four elicited stories narrated by a speaker with anomic aphasia. The speaker uses self-repair operations, including searching, recycling, deletion, and pauses between intonation units. Based on these data, a classification tool is presented that allows observations about the most recurrent grammatical forms in each of self-repair operations, which provides information about the processes incurred by the speaker with anomic aphasia to access lexical elements. Self-repairs are presented as elements that favor discursive continuity using specific linguistics elements, particularly the noun phrase.

Keywords: restarts; self-initiated repair; narration; pauses

1. Introducción

La afasia se ha descrito como un desorden del lenguaje que se adquiere como resultado de un daño en áreas del cerebro que contribuyen a la expresión y comprensión del lenguaje y sus componentes, por ejemplo, la comprensión o producción sintáctica, morfológica, fonológica y semántica (Geschwind, 1974; Goodglass, 1980; Luria, 1980; Caramazza, 1991; Christiansen, 1995; Cuetos Vega, 2003; Kambanaros & van Steenbrugge, 2006; Gertz, 2007; Jonkers & Bastiaanse, 2007; Armstrong & Ferguson, 2010; Holland, Halper & Cherney, 2010; Olness & Englebretson, 2011; Andretta, Cantagallo & Marini, 2012; Helm–Estabrooks, Albert & Nicholas, 2014; Bastiaanse, Wieling & Wolthuis, 2015; Tremblay & Dick, 2016).

La mayor parte de las afasias tiene su origen en un accidente cerebrovascular (ACV) o un ataque cerebral (AC). Este último se define como un daño en el cerebro, resultado de una condición patológica de los vasos sanguíneos, especialmente las arterias.

La afasia se manifiesta de varias formas según los aspectos del lenguaje que resulten más afectados. En el caso de la afasia anómica, el síntoma central consiste en la presencia de dificultades en la producción de palabras. Estos problemas de acceso al léxico pueden ir desde una leve dificultad en la producción de palabras deseadas durante una conversación hasta la virtual incapacidad de producir sustantivos bajo cualquier condición, por ejemplo, en pruebas básicas de nombramiento de objetos familiares.

Se ha señalado que la dificultad en el discurso de personas con afasia anómica (PAA) es mínima (Geschwind, 1974; Goodglass, 1980), sin embargo, la imposibilidad patológica de acceder a sustantivos evidencia un comportamiento lingüístico particular. Diversos estudios señalan que las PAA formulan repeticiones constantes, silencios extensos y falsos inicios y caracterizan estos elementos como problemáticos. No obstante, en este trabajo se presentan resultados que apuntan a patrones de regularidad en el discurso de una PAA y que contribuyen al estudio de la afasia anómica en es-

pañol, en particular, se observa un uso frecuente de autorreparaciones (AR) en pruebas de narración elicitada. Los datos que surgen a partir de estas narraciones muestran rutas de AR que privilegian la continuidad discursiva mediante elementos lingüísticos específicos, de manera concreta, se presenta a la frase nominal como un elemento central en la recuperación léxica.

La estructura de este trabajo se divide en tres partes: en la primera, se ofrece una breve caracterización de la afasia anómica y de las narraciones producidas por PAA, asimismo, se define a las AR, su unidad de análisis y otras operaciones presentes en los datos, además de señalarse el propósito de este estudio. En la segunda parte se expone la metodología, se presenta al participante y se desarrollan los procedimientos seguidos en la investigación. Finalmente, en la tercera parte se presenta el análisis, los resultados y las conclusiones de este trabajo.

2. Antecedentes

2.1. Afasia anómica

Por definición, “la principal característica de la afasia anómica es encontrar palabras en el contexto de un habla fluida y gramaticalmente bien formada. Se distingue de las otras afasias fluidas por la ausencia de parafasias literales y verbales. Al igual que en la afasia de conducción, la comprensión auditiva se encuentra relativamente intacta” (Goodglass, Barresi & Kaplan 2005: 82). Las PAA hablan con facilidad, pero presentan una importante ausencia de sustantivos en su discurso, sin que ello excluya la pérdida selectiva de otra clase de palabras, como verbos, adjetivos, adverbios, entre otras. Asimismo, se ha documentado el uso de frases no específicas, circunloquios y pausas prologadas (Goodglass, 1980; Goodglass *et al.*, 2005; Gertz, 2007; Helm-Estabrooks *et al.*, 2014; Tippett & Hillis, 2016).

En cuanto a los rasgos neuropsicológicos, se trata de un síndrome que ocurre con una frecuencia significativa en las lesiones

de la unión paritemporal, aunque también se ha descrito una forma particularmente pura de afasia anómica asociada con lesiones limitadas a la segunda circunvolución temporal. De modo general, la diversidad de lesiones asociadas con este síndrome exhibe puntos de vulnerabilidad en la amplia red que se activa en el proceso de denominación (Goodglass *et al.*, 2005: 82).

2.2. Deficiencias en las narraciones producidas por PAA

Para los propósitos de este trabajo se retoma la propuesta de Givón (1990), quien señala que la coherencia del discurso es multiproposicional, lo que significa la continuidad o recurrencia de algunos elementos, como la coherencia referencial, temporal, espacial y de eventos o acciones. En lo que toca a la relación entre las proposiciones, párrafos o turno en un diálogo, de acuerdo con Halliday y Hasan (1983), se trata de una serie de relaciones semánticas, de tal manera que, cuando dos proposiciones son parte del mismo discurso, ambas se encuentran en una relación de coherencia informativa vis a vis una de la otra (*cf.* Castillo, 2009).

Teun van Dijk (1996) ha propuesto dos categorías muy útiles para los estudios de narraciones elicidadas: macroestructura y microestructura. Según el autor, las macroestructuras son la forma básica de la coherencia global. Estas estructuras constituyen la base y determinan la coherencia de la superficie lineal de las subsecuentes proposiciones del texto. Los macrocomponentes posibilitan la derivación de las macroestructuras, mientras que los microcomponentes describen la estructura individual de la proposición de un texto y sus interrelaciones inmediatas al interior de una secuencia. El autor también señala que la acción significativa en una narración involucra a un agente o una persona que acompaña una acción. En este sentido, van Dijk encuentra en todas las narrativas que documenta tres macrocategorías: exposición, complicación y resolución.

La coherencia discursiva es un tema central en los estudios sobre PAA, en los cuales es posible distinguir dos tendencias. Por

un lado, hay estudios que señalan la deficiente elaboración de macroestructuras narrativas en PAA (Christiansen, 1995; Coelho & Flewellyn, 2003; Marini, Andretta, Del Tin & Carlomagno, 2011; Andretta *et al.*, 2012). Por el otro, en ciertas investigaciones se han identificado casos de PAA que mantienen relativamente intacta su capacidad de asignar el nivel de relevancia informativa de las narraciones que se les solicita. Según esta tendencia, las PAA mantienen la intención comunicativa pese a la falta de recursos léxicos (Olness & Englebretson, 2011).

Las evidencias presentadas por las investigaciones sobre la deficiencia en la macroestructura narrativa de los hablantes con afasia anómica señalan sus dificultades en el nivel del enunciado como un factor elemental en los problemas discursivos y la omisión de información central como reflejo del daño léxico (Ulatowska, Allard & Bond, 1990).

Christiansen (1995) señala que cuando la PAA no puede recuperar la palabra objeto de su lexicón simplemente salta la proposición e introduce una nueva. Coelho y Flewellyn (2003) sostienen que a pesar de que pueda existir una mejora a nivel del enunciado en los hablantes con afasia anómica, hay una afectación a nivel de la macroestructura. Aunado a ello, el trabajo de Andretta *et al.* (2012) indica que la amplia recurrencia de pausas largas, así como la repetición de palabras, son elementos que contribuyen a la falta de coherencia discursiva en dicho nivel. Por su parte, la investigación de Marini *et al.* (2011) ofrece resultados cuantitativos a partir de los errores gramaticales producidos por PAA en dos pruebas elicítadas (descripción de una ilustración y narración de una historia de dibujos animados). Para estos autores, la interrupción en el nivel de la cohesión también repercute de manera directa en la macroestructura narrativa.

2.3. Las autorreparaciones

El fenómeno de la reparación en el análisis conversacional se ha estudiado de manera multidisciplinaria, en este sentido, cada pers-

pectiva ha hecho énfasis en áreas puntuales. En el enfoque psicolingüístico (Levelt, 1983; Dingemanse, Blythe & Dirksmeyer, 2014), se describe el proceso de automonitoreo del hablante y el reconocimiento en línea de un problema en el discurso, asimismo, se expone su naturaleza no mecánica, esto es, se señala que dichos procesos requieren de la elección de recursos lingüísticos particulares para producir una reparación.

El trabajo desarrollado desde la sociolingüística (Schegloff, Jefferson & Sacks, 1977; Montes Miró, 1992; Schegloff, 2013) identificó los tres momentos que integran la reparación: reconocimiento de la fuente problemática (FP), marca explícita de lo que debe repararse (M) y reparación (R). Según esta perspectiva, la preferencia de la autorreparación (AR) por encima de las iniciaciones del interlocutor se debe a una decisión operativa, ya que las oportunidades para la autoiniciación emergen antes que las oportunidades para el interlocutor, lo cual hace plausible una alta frecuencia de inicios de AR en los espacios de transición.

Por otro lado, desde la lingüística clínica se han desarrollado propuestas de análisis en torno a las AR en población con alguna patología del lenguaje. En el caso de México, la investigación de Lara Galindo, Rojas Nieto, Jackson-Maldonado y Díaz Argüero (2015) da cuenta del monitoreo de FP en niños con trastorno primario del lenguaje y su incidencia en la producción de AR con respecto a niños con un desarrollo típico del lenguaje. A su vez, desde una perspectiva generativista del lenguaje, Ortíz Carballo (2021) describe en el estudio de caso de un hablante con afasia acústica-mnésica la relevancia de las estructuras nominales en la recuperación de referentes particulares.

En las distintas aproximaciones de estudio se observa que las AR mantienen los mismos movimientos de la reparación e implican consecuencias interaccionales, informativas e interpersonales en cada una de sus ocurrencias (Dingemanse *et al.*, 2014).

2.3.1. Organización de la unidad de análisis

El análisis de las AR tiene su origen en los estudios conversacionales (Levelt, 1983; Montes Miró, 1992; Fox, Maschler & Uhlmann, 2010; Schegloff, 2013; Dingemanse *et al.*, 2014; Kasper & Wagner, 2014), de ahí que su organización estructural incluya una *unidad de construcción del turno* (UCT). Esta unidad básica puede estar constituida por varias unidades entonacionales (UE). Las características prosódicas de la unidad entonacional propuestas por Chafe (1994) son útiles para establecer criterios de transcripción: las pausas (marcadas por secuencias de periodos breves y en algunas ocasiones, debido a su relevancia, por su extensión); los contornos terminales (señalados por periodos, marcadores interrogativos y comas); acentos (graves o agudos).

Por su parte, Schegloff (2013) propone dos unidades de análisis para las reparaciones: la *macroproyección* y la *microproyección*. La macroproyección puede incluir múltiples unidades de turno, ya sea en la forma o trayectoria de una conversación, mientras que en la microproyección los elementos más relevantes son las palabras o sonidos inmediatos, pues se trata de una organización local. No obstante, esta organización local no hace referencia exclusiva a aspectos gramaticales o fonológicos, sino que incluye aquellos aspectos contextuales que deben ser referidos, por ejemplo, lugares o personas relevantes para la historia y que de alguna manera están presentes en los turnos conversacionales. Además, Schegloff señala que la interacción directa contribuye al avance del discurso a través de unidades que permiten complementarlo y que componen los distintos órdenes de organización que ocurren en cualquier momento de la interacción discursiva. En este sentido, la granularidad en la observación de una interacción dependerá de los elementos que se tomen en cuenta, por ejemplo, el sonido, la sílaba o la palabra de una microproyección.

2.3.2. Operaciones de búsqueda, reciclaje, eliminación y aborto

Las operaciones de AR descritas por Schegloff (2013) son: búsqueda, reciclaje, eliminación, inserción, aborto, reemplazo, paréntesis, reformato y reformulación.¹ La operación de *búsqueda* hace referencia a una búsqueda léxica, la cual puede iniciarse una vez que emerge la FP y puede interrumpirse y reanudarse de manera inmediata o después de uno o varios turnos.

En su clasificación de operaciones de reparación, Schegloff (2013) denomina a la repetición como *reciclaje*, que refiere a la reiteración de algún tramo o extensión producido de manera previa, por lo regular menor a la unidad lingüística completa más pequeña (TCU), interaccionalmente relevante en un contexto dado (Selting, 2000: 477). Las repeticiones suelen marcar una sintaxis incompleta, así como una entonación continua y el alargamiento de la palabra previa a la AR, elementos que abren la posibilidad de completar la información (Rieger, 2003). Sin embargo, el éxito de la AR no siempre está ligado a la resolución léxica del formato, sino a las expectativas de conocimiento de los participantes en la conversación, es decir, no se trata únicamente de la clarificación de términos, sino de una comprensión pragmática.

La repetición realiza funciones genéricas y revela un amplio uso técnico que permite enlazar nuevamente la conversación (Schegloff, 2013). Los resultados de la investigación de Dingemanse *et al.* (2014) señalan que la principal función de la repetición parcial es enmarcar el ítem de la FP que es el objetivo de la iniciación de la reparación. Las repeticiones suelen considerarse como prácticas genéricas o recursos utilizados para realizar otras operaciones de reparación, por ejemplo, reemplazo, inserción, entre otras.

Schegloff (2013) observó que la operación de reciclaje también puede funcionar de manera independiente en “turnos de sobrevivencia”. Este tipo de reciclaje repite completamente un frag-

¹ En este trabajo únicamente se explicitan las AR observadas en el comportamiento lingüístico de la PAA que participó en el estudio.

mento para añadir información nueva. Se ha reportado un amplio uso de AR de reciclaje en estudios interlingüísticos de hablantes sin patología (Rieger, 2003).

Por otra parte, la operación de *eliminación* consiste en la omisión de uno o más elementos que el hablante ya ha articulado de manera parcial o total. Schegloff (2013) utiliza para esta operación el término *aborto*, como un símil de una misión espacial que se interrumpe antes de ser completada. En este sentido, incluye dos comportamientos lingüísticos: el abandono de la proposición iniciada en conjunto, al menos durante algunos turnos, y el abandono de la producción que se estaba iniciando en favor de una expresión distinta.

2.4. *Propósito de este estudio*

El presente trabajo se circunscribe a un estudio de caso que describe las AR realizadas por una PAA en cuatro narraciones elicidadas, dos de estas con estímulo presente y dos sin este. De acuerdo con investigaciones previas que dan cuenta de AR en narraciones realizadas por PAA, se esperaría una amplia dificultad para acceder a sustantivos —sin que ello excluya a otro tipo de palabras— así como una amplia posibilidad de que la PAA eluda la búsqueda de sustantivos específicos mediante un uso frecuente de circunloquios y AR de aborto. Para comprobar o descartar esta hipótesis, se elaboró un instrumento de categorización que registraba las siguientes variables en las cuatro narraciones: partículas utilizadas en las AR de búsqueda con más frecuencia, número de ocurrencias de reciclaje, número de unidades entonacionales involucradas en cada operación de AR de búsqueda, otras operaciones de AR involucradas en la operación de búsqueda y número de operaciones con resultados exitosos.

A partir de esta clasificación se presentará en primer lugar un panorama general del comportamiento lingüístico de la PAA con respecto a las AR. En segundo lugar, se identificarán las derivacio-

nes de las relaciones de las variables documentadas y en tercer lugar se abrirá la posibilidad de comparación con otros casos.

3. Método

3.1. Participante

Para preservar los datos personales del participante, en su codificación se utilizó únicamente tres letras (TAP). Los datos demográficos del participante en el momento en que se realizaron las pruebas eran: hombre de 61 años, lateralidad diestra, que reportaba estudios de nivel medio superior concluidos y declaraba no tener conocimientos de una segunda lengua. Antes del evento cerebro vascular (ECV), TAP realizaba charlas en la parroquia de su comunidad. La narración de historias era una de sus actividades frecuentes. Su diagnóstico médico indicaba un síndrome de afasia anómica, confirmado mediante el Test de Boston. La localización de la lesión de TAP era la arteria cerebral posterior en la región temporo-occipital izquierda. TAP reportaba problemas de presión arterial alta, así como un tratamiento médico de seguimiento previo al ECV.

3.2. Procedimiento

A TAP se le explicó el objetivo de este trabajo y se solicitó su autorización para ser videograbado; se acordó no mostrar las videograbaciones y únicamente utilizarlas para los propósitos del análisis, de manera que se preservara su anonimato (*cfr.* Knight & Adolphs, 2010). El diseño del corpus se realizó mediante tareas de narración; se grabaron las pruebas elicítadas correspondientes a las historias: *Frog, where are you?* y *Frog goes to dinner* (Mayer, 1969, 1974; Berman & Slobin, 1994) con estímulo y sin él. Con el objetivo de limitar la temática de las narraciones a temas compartidos por el participante y la investigadora, TAP narró cuatro historias, dos con estímulo presente y dos sin este. Las narraciones se realizaron primero con estímulo e inmediatamente después sin este.

En lo que toca al uso de tareas de narración, se hace hincapié en la necesidad de que el participante no conozca de antemano el estímulo visual para que pueda narrar la historia con el mayor número de detalles posible (Faitaki & Murphy, 2019). Las unidades entonacionales producidas por TAP fueron 820. El método de transcripción utilizado fue el Child Language Exchange System —CHILDES— (Montes Miró, 1992; MacWhinney, 2000, 2011; McCarthy y O’Keeffe, 2010).² Se eligió este método de transcripción debido a su capacidad de segmentar unidades entonacionales, así como las pausas entre cada una de ellas.

Las grabaciones de las pruebas elicidadas se llevaron a cabo en el Servicio de Audiología y Foniatría del Hospital General de México Dr. Eduardo Liceaga. El equipo técnico consistió en una cámara de video de alta definición y una grabadora portátil, de esta manera se garantizó el registro adecuado de las grabaciones.

Entre los datos codificados en la conformación del corpus, se han registrado las pausas oralizadas, así como los alargamientos seguidos de silencios. Para la codificación de cada una de las transcripciones se han contabilizado las pausas en el programa Audition,³ con el fin de integrar dicha información en el análisis de datos.

4. Análisis y resultados

4.1. Tipos de AR en cuatro pruebas elicidadas

En tres de las cuatro narraciones producidas por TAP, con y sin estímulo presente, más de 50% de las UE contiene una o más AR que repercuten en los diferentes procesos de acceso léxico (Cuadro 1).

² En el Anexo se puede consultar la codificación utilizada.

³ Adobe Audition es una estación de trabajo de audio profesional para mezclar, finalizar y editar audio con precisión. Este programa permitió medir las pausas de los hablantes y realizar una segmentación sistemática de las unidades entonacionales.

Este aspecto que incide de manera relevante en las secuencias discursivas, un ejemplo de ello son las continuas interrupciones que generan confusión en el oyente.

CUADRO 1. Porcentaje de autorreparaciones en pruebas elicítadas

Narración	Número de UE	Número de UE que presentan una o más AR
<i>Frog, where are you?</i> Con estímulo presente (TAP 1)	225	120 (53%)
<i>Frog, where are you?</i> Sin estímulo presente (TAP 2)	162	104 (64.19%)
<i>Frog goes to dinner.</i> Con estímulo presente (TAP 3)	245	96 (39.18%)
<i>Frog goes to dinner.</i> Sin estímulo presente (TAP 4)	188	101 (53.19%)

Este porcentaje proporciona evidencia de al menos tres aspectos: (i) las AR son prueba de un constante automonitoreo del hablante y por ende de ajustes en su comportamiento lingüístico; (ii) la utilización de las AR resulta un recurso discursivo eficiente en la recuperación léxica por parte de un hablante con afasia anómica; (iii) las AR pueden ocurrir y resolverse en una misma unidad entonacional.

En este orden de ideas podemos observar que en una misma UE se producen distintas operaciones de AR, o bien que la solución de una AR puede incluir varias UE. En (1)⁴ se observa un caso de búsqueda que involucra otra operación de AR:

- (1) (64) * TAP: lo sigue buscando # sube <la> # [/] **la** # **ventana** ##.
[0:01.120]
(TAP, *Frog, where are you?* con el estímulo presente)

⁴ En cada uno de los ejemplos se indica en negritas la AR que se describe.

Este ejemplo, que corresponde a la primera narración producida por el hablante con estímulo presente, muestra que la operación de búsqueda recurre al uso de la operación de reciclaje. Si bien cada una de las AR presenta particularidades en cada narración, hay recurrencias en cuanto al tipo de AR producidas por TAP, tal como se observa en el Cuadro 2.

CUADRO 2. Tipo de autorreparaciones

Autorreparación	<i>Frog where are you?</i>		<i>Frog goes to dinner</i>	
	Con estímulo	Sin estímulo	Con estímulo	Sin estímulo
Búsqueda	95	81	59	80
Eliminación	22	21	35	25
Aborto	2	2	2	0
Inserción	0	1	0	3
Otros	0	0	0	0

Los resultados expuestos en el Cuadro 2 indican que la AR de búsqueda es la más recurrente en las cuatro narraciones. Este resultado sugiere que TAP realiza búsquedas constantes para acceder al ítem léxico que considera necesario a fin de dar continuidad a la narración.

Es importante señalar que en la operación de AR de búsqueda en las cuatro narraciones se hace uso de otras operaciones, por ejemplo, la operación de reciclaje. Dado que dicha operación está ligada con las de búsqueda, no se contabiliza como una operación autónoma sino como parte de la misma AR.

La segunda operación más utilizada por TAP es la AR de eliminación. Este resultado expone una tendencia particular en este caso del español, que contrasta con lo reportado por Schegloff (2013), quien señala que este tipo de AR es el que menos se presenta en un corpus con datos de habla sin patología en inglés.

La mínima ocurrencia de la AR de aborto aporta evidencia en torno a dos comportamientos lingüísticos: el primero se refiere a la presencia de una planificación del discurso y el segundo a que se

privilegia la utilización de estrategias discursivas que solventan la dificultad para acceder a elementos léxicos determinados.

Entre las estrategias discursivas se pueden observar circunloquios o la eliminación de un incipiente elemento léxico por otro ítem, por ejemplo, el uso de sustantivos genéricos ante la dificultad de acceder a uno específico. A continuación, en (2) podemos observar una operación de búsqueda que involucra distintas operaciones de AR.

- (2) (12) * TAP: en # <el> # [/] **el vitral** se encuentra adentro <un> ##. [0:00.550]
 (13) * TAP: [//] **unaaa ##. alargamiento [0:00.504] pausa [0:00.354]**
 (14) * TAP: **rana ##.** [0:01.661]
 (TAP, *Frog, where are you?* con el estímulo presente)

En (2) surgen al menos dos FP en la misma UE; al principio se presenta una AR de búsqueda mediante el uso de artículos indefinidos y luego un desarrollo más complejo de la AR de búsqueda a lo largo de tres UE con pausas intermedias largas. Este comportamiento sugiere que la especificidad del artículo definido genera un acceso más eficiente con respecto al uso de un artículo indefinido, puesto que la inespecificidad de la entidad a la que se busca acceder en el universo discursivo hace más complejo el proceso.

A continuación, en (3) se presenta un ejemplo de una operación de eliminación.

- (3) (60) * TAP: y # **hay una # hay u # tenemos música ##.** [0:01.200]
 (TAP. *Frog goes to dinner* con el estímulo presente)

En (3) la operación de eliminación se produce en la misma UE en la que se encuentra la FP y se resuelve de manera inmediata. El hablante inicia con una operación de reciclaje que intercambia por una de eliminación con el fin de ajustar la búsqueda del sustantivo

de un instrumento musical por el uso genérico de la proposición <tenemos música>.

4.1.1. Pausas

Las pausas son un elemento central en la realización de las AR, su distribución a lo largo de las cuatro narraciones elicitadas permite observar una evidente planificación entre cada UE. En su mayoría se trata de silencios con mínimas ocurrencias de alargamientos silábicos. La concentración máxima de pausas se situó en el rango de los 301 a los 999 milisegundos. Este dato expone que el hablante planifica por encima del tiempo promedio, desde dos unidades entonacionales hasta el triple de una pausa estándar (*cf.* White, 1997).

Las pausas que van de 1 segundo a 1.999 segundos hacen evidente que la planificación discursiva realizada por el hablante va en detrimento de la propia narración ya que, al ralentizar el discurso, se propicia una plausible distracción del oyente. La ocurrencia de pausas que van de 2 a 2.999 segundos es menor; tiene presencia en tres de las cuatro narraciones con y sin estímulo presente. En (4) se exhibe la utilización de déicticos en una prueba sin estímulo presente.

- (4) (87) * TAP: y estee ##. [0:02.589]
 (88) * TAP: aquí ##. [0:04.215]
 (89) * TAP: aquí # corren ##. [0:01.456]
 (90) * TAP: eeh ##. [0:01.690]
 (91) * TAP: corren ##. [0:01.774]
 (92) * TAP: y # aquí ##. [0:02.561]
 (93) * TAP: corren ##. [0:03.940]
 (94) * TAP: y ##. [0:00.451]
 (95) * TAP: hay un hueco ##. [0:02.267]
 (TAP: *Frog, where are you?* sin el estímulo presente)

En (4) se observa el uso del déictico <aquí>. Este segmento presenta pausas de planificación particularmente extensas, incluso se

puede observar la de rango más extenso de las cuatro narraciones producidas por TAP. En (88), después de la producción del deíctico ocurre una pausa de [0:04.215]; las UE que aparecen enseguida muestran cómo TAP produce de manera reiterada operaciones de reciclaje, seguidas de pausas de planificación que decide ajustar en (94) y (95) al iniciar una descripción que interrumpe las operaciones de reciclaje y da pie al reinicio de la narración.

En los datos reportados se observa que TAP produjo 53% de operaciones de AR en tres de las cuatro pruebas. Este resultado sugiere que el hablante realiza búsquedas constantes para acceder al ítem léxico que considera necesario para continuar con su narración. En las cuatro pruebas se observa un uso de otras operaciones de AR que acompañan a la operación de búsqueda, así como la presencia constante de operaciones de reciclaje y eliminación.

En lo que respecta al número de UE producidas por TAP, la mayor producción se da en las pruebas en que cuenta con el estímulo presente, en este sentido, las operaciones de AR corresponden con dicho número. Cabe señalar un dato porcentualmente relevante en su comportamiento lingüístico. En *Frog, where are you?* sin estímulo presente, TAP presentó el porcentaje más alto de operaciones de autorreparación (64.19%) con respecto a las otras pruebas realizadas.

4.2. Elección de la forma gramatical con y sin estímulo en la historia *Frog, where are you?*

La elección de la forma gramatical en cada operación de AR proporciona información acerca de los procesos elaborados por TAP para acceder a elementos léxicos, principalmente sustantivos. La relevancia de esta caracterización radica no únicamente en la identificación de las clases gramaticales a las que corresponden las formas utilizadas por TAP, sino que también proporciona información con respecto a las rutas que sigue el hablante en el proceso de acceso léxico durante sus narraciones. Así, identificar los componentes más eficientes en las AR de búsqueda permite una mirada

panorámica a la dimensión discursiva de TAP en los procesos de narración elicitada.

Para el análisis de los datos elicitados en las cuatro pruebas, se desarrolló un instrumento de categorización que registraba las siguientes variables: partículas con más ocurrencias utilizadas en las autorreparaciones de búsqueda, número de ocurrencias de reciclaje, número de unidades entonacionales involucradas en cada operación de AR de búsqueda, otras operaciones de AR involucradas en la operación de búsqueda y número de operaciones con resultados exitosos.

CUADRO 3. Clase gramatical de las formas involucradas en las operaciones de búsqueda en TAP 1, *Frog where are you?* con el estímulo presente

Palabra	Número de ocurrencias		Otras operaciones	Número de ue involucradas					Exitoso
	Total	De reciclaje		-1	-2	-3	-4	(5 o más)	
Artículo definido	26	23	5	20	3	2	0	1	23
Artículo indefinido	16	10	5	7	5	0	0	1	8
Cuantificador	10	4	1	2	0	1	0	7	3
Posesivo	4	4	0	1	2	1	0	0	4
Verbo	8	7	2	7	1	0	0	0	8
Clítico	3	3	1	1	1	1	0	0	3
Preposición	7	7	1	5	2	0	0	0	3
Pronombre reflexivo se	1	1	0	0	1	0	0	0	1
Adjetivo	1	1	0	1	0	0	0	0	1

La categorización de los datos expuestos en el Cuadro 3 exhibe la alta frecuencia del artículo definido, que tiene 26 ocurrencias; de ellas, 23 casos se produjeron en una operación de reciclaje y en 20 casos la AR se realizó en una misma UE. Con relación a lo anterior se presenta el ejemplo (5).

(5) (118) * TAP: entons <el> # [/] **el perro** # lo <quiere seguir> ##.
[0:01.302]

(6) (173) * TAP: <el> # [/] **el venado corre** ## @g [0:01.203]

En (5) y (6) se muestran operaciones de AR de búsqueda que hacen uso de la operación de reciclaje. En este sentido, la repetición sistemática de artículos definidos, que concuerdan en género y número y que a su vez contribuyen al acceso del sustantivo, sugieren la relevancia de la frase nominal, que adquiere una dimensión destacada en el proceso de preservación del acceso a sustantivos en español.

La frecuencia de operaciones de reciclaje muestra que los artículos definidos e indefinidos de la frase nominal son una vía plausible en la recuperación de los sustantivos, es decir, de acuerdo con los datos recopilados durante esta prueba, se observa un acceso exitoso a través de dichas partículas a los sustantivos involucrados en las operaciones de búsqueda.

Esta evidencia se manifiesta a lo largo de todo el corpus analizado, tal como se expondrá más adelante. La frecuencia de ocurrencias de artículos definidos en las operaciones de búsqueda da cuenta de la relevancia de la especificidad en el universo discursivo que demanda esta prueba.

El segundo tipo de partícula con mayor recurrencia en esta prueba es el artículo indefinido (singular y plural), tal como se observa en (7).

(7) (97) * TAP: aquí # **de** # **del hoyo** sale # <una ar> # [//] **una ardilla** ##. [0:00.934]

En (7) surgen al menos dos FP en la misma UE que inicia la AR de búsqueda mediante el uso de artículos indefinidos. Las operaciones de búsqueda en (7) están precedidas por operaciones de eliminación. Este comportamiento sugiere que la especificidad del artículo definido puede proporcionar evidencia de un acceso más eficiente con respecto al uso de un artículo indefinido, ya que la inespecifici-

cidad de la entidad a la que se busca acceder en el universo discursivo hace más complejo el proceso.

Los artículos definidos se resuelven en 20 de 26 ocurrencias totales en la misma UE, lo que sugiere una búsqueda específica con resultados exitosos para el hablante. Las operaciones de búsqueda están acompañadas de operaciones de reciclaje en 23 de 26 ocurrencias, por ende, es plausible una combinación eficiente del tipo:

- a) Artículo definido > operación de reciclaje > sustantivo
- b) Artículo indefinido > operación de reciclaje > pausa de planificación > sustantivo

En (a) se enfatiza la relevancia del sustantivo unido a un artículo definido; en términos gramaticales, la frase nominal adquiere una dimensión relevante en el proceso de preservación del acceso a sustantivos en español. En el caso analizado, la frecuencia de operaciones de reciclaje señala una tendencia al acceso léxico exitoso en las operaciones de búsqueda. A su vez, los artículos indefinidos tienen una presencia porcentual importante, sin embargo, los casos de éxito se reducen a un 50%. Este resultado apunta a una mayor complejidad en el acceso léxico cuando la operación de búsqueda no es específica. Cabe señalar que, si bien no se trata de una ruta porcentualmente equivalente a la del artículo definido, sí es utilizada de manera sistemática por TAP en esta prueba.

En la segunda prueba, sin estímulo presente, los resultados presentan factores constantes y otros nuevos, como se discutirá a continuación.

CUADRO 4. Clase gramatical de las formas involucradas en las operaciones de búsqueda en TAP 2, *Frog, where are you?* sin estímulo

Palabra	Número de ocurrencias		Otras operaciones	Número de UE involucradas					Exitoso
	Total	De reciclaje		-1	-2	-3	-4	(5 o más)	
Artículo definido	15	13	2	9	5	1	0	0	13
Artículo indefinido	11	10	1	3	7	1	0	1	10

(continuación)

CUADRO 4. Clase gramatical de las formas involucradas en las operaciones de búsqueda en TAP 2, *Frog, where are you?* sin estímulo

Palabra	Número de ocurrencias		Otras operaciones	Número de UE involucradas					Exitoso
	Total	De reciclaje		-1	-2	-3	-4	(5 o más)	
Cuantificador	1	1	0	1	0	0	0	0	1
Posesivo	9	4	0	6	2	1	0	0	4
Verbo	5	5	0	3	1	0	0	1	
Clítico	1	1	0	1	0	0	0	0	1
Preposición	3	3	0	2	1	0	0	0	1
Pronombre reflexivo <i>se</i>	6	4	0	6	0	0	0	0	3
Adjetivo	0	0	0	0	0	0	0	0	0
Pronombre relativo <i>que</i>	1	1	0	1	0	0	0	0	1

En el Cuadro 4 observamos que el artículo definido corresponde a la partícula con mayor frecuencia en las operaciones de AR de búsqueda con 15 ocurrencias, 13 de ellas exitosas. Una de las diferencias en el comportamiento lingüístico de TAP con respecto a la prueba anterior se manifiesta en un incremento de UE involucradas en las AR de búsqueda. A continuación, en (8), (9) y (10) se presentan algunos ejemplos de esta prueba sin estímulo.

(8) (82) * TAP: <el> # **el** perro se echa a correr ##. [0:00.573]

(9) (84) * TAP: como que se cae # para que no <los> # [/] **los** # **las** abejas pasan ##. [0:00.400]

(10) (43) * TAP: entonces # este ##. [0:00.474]

(44) * TAP: él se sube <al> # [/] **al** ##. [0:00.760]

(45) * TAP: **a los** # **a las** ##. [0:00.446]

(46) * TAP: **piern** # **piedras** ##. [0:00.634]

Si bien, las operaciones en (8), (9) y (10) incluyen elementos de definitud, como los artículos definidos, las particularidades de cada caso permiten observar la complejidad del acceso léxico en la prueba sin estímulo presente. En (8), el hablante realiza una operación de búsqueda en la misma UE en la que emerge la FP y sigue el mismo comportamiento lingüístico reportado: reciclaje del artículo definido y acceso al sustantivo.

En (9) se observa un siguiente paso en el grado de complejidad. El hablante inicia su AR de búsqueda con el reciclaje de un artículo definido. No obstante, se trata de un artículo plural masculino que elimina para producir uno femenino. Este último posibilita la frase nominal <las abejas>.

En el caso de (10), la complejidad en el proceso de acceso léxico muestra distintas fases. En primer lugar, la emergencia de la FP inicia en una UE; en segundo lugar, el proceso muestra distintas operaciones: reciclaje, dos eliminaciones y un acceso léxico exitoso después de 4 UE. Este comportamiento sugiere que la definitud en la búsqueda de sustantivos sin el estímulo podría no implicar necesariamente un acceso rápido en al menos 6 de las 15 ocurrencias de búsquedas con artículo definido; en otros 6 casos se trata de soluciones que implican al menos 2 UE en esta prueba. En operaciones que involucran más de una UE se posibilita la coincidencia de más de una operación de AR. Es decir, a mayor número de UE, mayor probabilidad de utilizar distintas operaciones de AR y de pausas de planificación.

Respecto al número de ocurrencias del siguiente tipo de partículas, en esta prueba, los artículos indefinidos, se incluye (11).

- (11) (65) * TAP: encuentra <a un> ##. [0:00.483]
 (66) * TAP: [/] a un # este ##. [0:01.017]
 (67) * TAP: <a un> # [/] a <un enjambre> # [/] un enjambre ##.
 [0:01.488]

En (11) observamos una búsqueda que involucra tres UE y las operaciones de AR de búsqueda y reciclaje que permiten acceder al

sustantivo <enjambre>. A partir de este ejemplo podemos señalar que en una narración sin estímulo el hablante exhibe dificultades tanto en el uso de partículas de definitud como de indefinitud a fin de resolver la FP. Sin embargo, las operaciones de reciclaje de dichas partículas son recurrentes y señalan una mayor frecuencia de estas rutas de búsqueda de sustantivos que de otras. Aunado a ello, si bien las operaciones de AR contribuyen a la continuidad discursiva, también añaden elementos que pueden disminuir la atención del oyente.

Los resultados de la prueba *Frog, where are you?* sin estímulo muestran una disminución en la cantidad de UE producidas —162 frente a 224— con respecto a la prueba con estímulo presente. Esta disminución sugiere una mayor síntesis narrativa, lo que no necesariamente implica que la narración sea más clara para el oyente. Así, uno de los resultados expone que cuando TAP utiliza varias UE para acceder a un sustantivo utiliza más de una operación de autorreparación, generando fragmentos discursivos complejos para el oyente.

En lo que toca a los artículos definidos, si bien TAP resuelve la FP en 9 de 15 ocurrencias en la misma unidad entonacional en que emerge, hay 6 de esas ocurrencias que requieren de 2 o más UE. Esto sugiere que en casos en los que el hablante no cuenta con el estímulo presente, el uso de artículos definidos no es necesariamente una garantía de un acceso rápido, pero sí exitoso (13 de 15 ocurrencias). Este comportamiento muestra la siguiente ruta de acceso al sustantivo que se muestra en (c).

c) Artículo definido > operación de reciclaje > sustantivo

Las operaciones de búsqueda en esta prueba muestran un uso sistemático de operaciones de reciclaje: para el caso de los artículos definidos en 13 de 15 ocurrencias y en el caso de los indefinidos (d) en 10 de 11 ocurrencias.

d) Artículo indefinido > operación de reciclaje > sustantivo

En cuanto al uso de otras operaciones, podemos observar que las de eliminación emergen de manera sistemática para hacer ajustes de género y número a partículas como artículos definidos e indefinidos (e).

- e) Artículo definido o indefinido > operación de eliminación > ajuste de género o número > artículo definido o indefinido

4.3. Elección de la forma gramatical con y sin estímulo presente en la historia *Frog goes to dinner*

En el Cuadro 5 se muestra el comportamiento lingüístico de TAP en la narración *Frog goes to dinner* con estímulo presente.

Cuadro 5. Clase gramatical de las formas involucradas en las operaciones de búsqueda en TAP 3, *Frog goes to dinner* con el estímulo presente

Palabra	Número de ocurrencias		Otras operaciones	Número de UE involucradas					Exitoso
	Total	De reciclaje		-1	-2	-3	-4	(5 o más)	
Artículo definido	28	18	12	20	6	1	0	0	25
Artículo indefinido	5	3	1	3	1	1	0	0	5
Cuantificador	1	0	0	0	1	0	0	0	1
Posesivo	7	6	1	5	0	1	0	1	6
Verbo	14	10	2	12	1	1	0	0	13
Clítico	4								
Preposición	6	4	2	3	3	0	0	0	6
Pronombre reflexivo se	8	5	2	7	1	0	0	0	6
Adjetivo	0	0	0	0	0	0	0	0	0
Adverbio	3	3	0	2	1	0	0	0	3
Negación (no)	1	1	0	0	1	0	0	0	0

En el Cuadro 5 se observa que el número de apariciones de artículos definidos (28 ocurrencias) duplica el de la clase gramatical que le sigue, verbos (14 ocurrencias). A continuación, en (12), (13) y (14) se muestran ejemplos de operaciones en las que TAP utiliza artículos definidos.

(12) (93) * TAP: de tratar de sacar <la> # [/] **la rana** ## @g [0:01.067]

(13) (115) * TAP: <el> # [/] **el mesero** no se ha dado cuenta ##. [0:00.800]

(14) (159) * TAP: y # levantan <la> # [/] **la sopa** # <la> [/] **la copa** ##. [0:00.788]

En (12), (13) y (14) se observa un comportamiento lingüístico consistente en el caso de operaciones de búsqueda. En (12) y (13), el hablante reconoce la FP en una misma UE y produce operaciones de reciclaje, resolviendo estas operaciones con los sustantivos <rana> y <mesero>. En (14), el hablante produce una operación de eliminación al modificar <sopa> por <copa>, que es el elemento ilustrado en la historia. Este comportamiento sugiere que el uso de los artículos definidos es una ruta de búsqueda léxica más rápida para TAP en la realización de esta prueba.

(15) @ G [6]

(68) * TAP: la música **está** ##. [0:03.032]

(69) * TAP: <está> # [//] <están> # [/] **están** # **este** ##. [0:03.532]

(70) * TAP: <está> # [//] <están> ##. [0:01.057]

(71) * TAP: [/] **están contentos** # **están** ##. [0:00.699]

(72) * TAP: **hay música** ##. [0:00.871]

En (15) se observa el inicio de la búsqueda a través de la operación de reciclaje y pausas largas entre cada una de las UE presentes. La particularidad de este ejemplo es el verbo <estar>, pues muestra un uso atributivo <están contentos>, es decir, el propio verbo

requiere de un atributo para ser completado en la operación de búsqueda del hablante.

A continuación, en (16) y (17) se presentan ejemplos de otros verbos.

(16) (137) * TAP: este # la rana ##. [0:00.480]

(138) * TAP: <salta> # [/] salta de la comida ##. [0:00.354]

(17) (179) * TAP: van a comer ##. [0:00.450]

(180) * TAP: van a tomar ##. [0:00.800]

(181) * TAP: y # <ven a> # [/] ven a la rana ##. [0:00.742]

En (16) y (17) se observa un comportamiento similar. En ambas ocurrencias, ante la emergencia de una FP, el hablante inicia con operaciones de reciclaje de dos verbos que se resuelven en la misma UE. En los resultados de esta prueba se observa una variable distinta con respecto al tipo de palabras involucradas en las operaciones de AR, ya que se presentan ejemplos del verbo <estar> atributivo, así como de verbos que siguen una ruta de reciclaje y acceso a sustantivos en una misma UE. Este uso de verbos se produce en 14 ocurrencias, 13 de ellas exitosas.

f) Verbo > operación de reciclaje > atributo o complemento

El uso de artículos definidos es más frecuente que el de cualquier otra clase de palabras o partículas (28 ocurrencias, de las cuales 25 son exitosas). Asimismo, la ruta de acceso al sustantivo es eficiente —20 de estas ocurrencias se resuelven en la misma UE en que emerge la FP— lo cual sugiere la siguiente ruta (g) para resolver la autorreparación.

g) Artículo definido > operación de reciclaje > sustantivo

En el Cuadro 6 se observa el comportamiento de TAP en la segunda prueba sin estímulo presente.

Cuadro 6. Clase gramatical de las formas involucradas en las operaciones de búsqueda en TAP 4, *Frog goes to dinner* sin estímulo

Palabra	Número de ocurrencias		Otras operaciones	Número de ue involucradas					Exitoso
	Total	De reciclaje		-1	-2	-3	-4	(5 o más)	
Artículo definido	19	13	7	12	5	1	0	0	17
Artículo indefinido	2	0	2	2	0	0	0	0	2
Cuantificador	0	0	0	0	0	0	0	0	0
Posesivo	13	9	7	4	7	2	0	0	13
Verbo	11	6	5	10	1	0	0	0	11
Clítico	2	2	1	1	1	0	0	0	2
Preposición	5	3	2	3	2	0	0	0	5
Pronombre reflexivo <i>se</i>	9	3	6	8	1	0	0	0	7
Adjetivo	0	0	0	0	0	0	0	0	0
Adverbio	6	4	2	5	0	0	1	0	5
Pronombre relativo <i>que</i>	2	2	0	0	1	0	0	0	2

En el Cuadro 6 se observa que los artículos definidos ocupan el primer lugar entre las palabras involucradas en operaciones de autorreparación. A continuación, en (18) y (19) se presentan algunos ejemplos.

(18) (27) * TAP: [/]se va con su familia # **su su el señor** ##. [0:01.141]

(28) * TAP: <la> # [/] <la> # [/] **la su** ##. [0:00.428]

(29) * TAP: **la señora** ##. [0:00.856]

(19) (83) * TAP: <el> # [/] <el me> # [//] <el mes> # [/] <el mes> ##. [0:00.885]

(84) * TAP: [//] <el mesero> # **el mesero!** ##. [0:01.156]

En (18) observamos el uso de una operación de reciclaje y el éxito de la operación de búsqueda en la UE (29). Los dos inicios de búsqueda

queda a partir del posesivo llaman la atención; en (27) la operación de eliminación resuelve de manera inmediata la búsqueda, mientras que en (28) la operación de eliminación ocurre luego de una pausa prolongada. El ejemplo (19) expone la emergencia de la FP al inicio de la UE, en este caso la PAA inicia una operación de reciclaje que no involucra únicamente al artículo definido, sino la primera sílaba del sustantivo <mesero>, y que resuelve exitosamente en la siguiente UE después de una pausa que duplica el tiempo de una pausa estándar. En los ejemplos (20) y (21) las operaciones de búsqueda son más largas e involucran diversas operaciones.

- (20) (32) * TAP: ella # siempre ##. [0:00.601]
 (33) * TAP: <en su> # [/] <en su> # [/] en su ro # pa # en su ##. [0:00.659]
 (34) * TAP: **saco** ##. [0:00.391]
 (35) * TAP: siempre se lleva # siempre trae # en su casa ##. [0:00.508]
 (36) * TAP: a un # en su casa siempre lleva una rana ##. [0:00.407]
- (21) (114) * TAP: y ##. [0:01.011]
 (115) * TAP: **han # eh # levantan # las # <la> # [/]<la> # [/] la ##. [0:01.084]**
 (116) * TAP: levanta **su so** # <su> ##. [0:00.488]
 (117) * TAP: [/] **su** ##. [0:00.488]
 (118) * TAP: **levanta <su>** ##. [0:00.655]
 (119) * TAP: [/] **su copa** ##. [0:01.067]

En (20) se observa una búsqueda que el hablante decide extender: pese a que en la misma UE en que emerge la FP accede al sustantivo genérico <ropa>, continúa con la operación de búsqueda a través del posesivo. TAP requiere de una nueva UE para producir el sustantivo <saco>, que correspondía a esa parte de la narración, aun cuando no tenía el estímulo presente. El ejemplo (21) involucra diversas operaciones de AR, como eliminación y reciclaje. La partícula que utiliza TAP para acceder al sustantivo <copa> es el pose-

sivo que inicia 4 UE antes; este ejemplo muestra un comportamiento similar al de los artículos indefinidos en las otras pruebas, pues se trata de búsquedas más extensas que implican más de dos UE.

Los resultados que se muestran en el Cuadro 7 (véase Anexo) presentan un uso destacado de los artículos definidos (h): 19 casos en total, 17 de ellos exitosos, que siguen un comportamiento similar después de cuatro pruebas.

h) Artículo definido > operación reciclaje > sustantivo

Este resultado confirma la relevancia de la definitud en el acceso léxico, lo que se traduce en una respuesta más eficiente, ya que la mayor parte de las AR se solucionan en una misma UE. Durante esta prueba, el segundo tipo de partículas más utilizadas fueron los posesivos (i): 13 casos, todos exitosos, con un amplio uso de operaciones de reciclaje previo al acceso a los sustantivos, motivo de las operaciones de búsqueda.

i) Posesivo > operación de reciclaje > pausa larga > cambio de UE > acceso al sustantivo

El posesivo se encuentra presente en las cuatro pruebas y su uso se puede equiparar con el de los artículos indefinidos, sin embargo, no alcanza un porcentaje tan exitoso en el acceso léxico.

5. Conclusiones

Si bien existe una amplia gama de estudios en los que se discuten las narraciones de PAA de manera sistemática (Ulatowska *et al.*, 1990; Christiansen, 1995; Marini, 2011; Andretta *et al.*, 2012), en estos se considera a las pausas, los falsos inicios y las repeticiones como errores motivados por la dificultad al acceso léxico, principalmente de sustantivos, característica central de este tipo de afasia. En contraste, en este estudio de caso varias de estas ocurrencias se describen como AR, según la caracterización propuesta

por Schegloff *et al.* (1977), Levelt (1983), Schegloff (1990, 2011, 2013), Montes Miró (1992), Levelt, Roelofs y Meyer (1999) y Dingemanse *et al.* (2014).

Las AR se presentaron en las narraciones de una PAA a partir de dos historias elicítadas, *Frog, where are you?* (Mayer, 1969) y *Frog goes to dinner* (Mayer, 1974), con y sin estímulo presente. A partir de la macroestructura narrativa, y en conjunto con las AR, se registraron pausas amplias entre cada UE. Su cuantificación individual exhibió silencios mayores a los reportados en hablantes sin patologías del lenguaje (Chafe, 1994).

El instrumento de categorización de los datos obtenidos en las cuatro narraciones elicítadas representa una contribución al estudio de las AR en PAA (véase Cuadros 4, 5, 6 y 7). Como resultado de esta propuesta se derivan las siguientes observaciones:

- 1) La operación de búsqueda fue la más utilizada por TAP. Este resultado se aleja de la hipótesis de partida de este trabajo, pues se estimaba una alta frecuencia de circunloquios y una evitación sistemática de sustantivos específicos.
- 2) Se registró una sobresaliente participación de la AR de reciclaje como parte de la AR de búsqueda, de manera particular en las pruebas en que el estímulo estaba presente durante la narración. En estas pruebas los artículos definidos presentaron una frecuencia más alta con respecto a las narraciones sin estímulo presente.
- 3) Las operaciones de reciclaje y de eliminación ligadas a la operación de búsqueda tuvieron ocurrencias en las cuatro pruebas. La primera contribuyó temporalmente en el proceso de búsqueda léxica, en tanto que la operación de eliminación estuvo ligada a la AR de búsqueda y posibilitó el ajuste de género y número de falsos inicios o búsquedas prolongadas y permitió soluciones rápidas y eficientes en los procesos de AR de TAP.
- 4) El instrumento de categorización mostró que en las cuatro pruebas la clase gramatical de las formas utilizadas por TAP

en las AR de búsqueda se concentró en artículos definidos, artículos indefinidos y posesivos.

- 5) De entre las operaciones que aparecieron con una frecuencia mínima, hubo una cuya ausencia aportó información fundamental: la AR de aborto. Su casi nula presencia proporcionó evidencia de la planificación del discurso en línea del hablante en las cuatro narraciones, así como de la producción de otras estrategias discursivas que le permitieron acceder a elementos léxicos o solventar dicha dificultad a través del uso de pausas de planificación.
- 6) Se registró una mínima ocurrencia de sustantivos genéricos y aun menor de circunloquios, lo que contraviene la hipótesis de partida de la investigación, esto es, una mayor presencia de circunloquios en las AR.
- 7) La elección de la forma gramatical en cada operación de AR proporcionó información acerca de los procesos elaborados por TAP para acceder a elementos léxicos, principalmente sustantivos, en cada una de las cuatro pruebas. La relevancia de esta caracterización radica no únicamente en la identificación de las clases gramaticales a las que corresponden las formas utilizadas por el hablante, sino que proporcionan información con respecto al uso particular de estas formas, no únicamente en el acceso léxico sino en su dimensión discursiva.

La prominencia de los artículos definidos, seguida por la de los artículos indefinidos y los posesivos en la búsqueda de sustantivos confirma la relevancia de las frases nominales como parte central del discurso de una PAA y muestra que no se trata de la emergencia de sustantivos aislados, sino de sustantivos referenciales. Este aporte nos obliga a examinar más casos con el objetivo de contribuir al estudio de la afasia anómica en español.

6. Referencias

- Andreetta, Sara; Cantagallo, Anna, & Marini, Andrea (2012). Narrative discourse in anomic aphasia. *Neuropsychologia*, 50, 1787–1793.
- Armstrong, Elizabeth, & Ferguson, Alison (2010). Language, meaning, context and functional communication. *Aphasiology*, 24(4), 480–496.
- Bastiaanse, Roelien; Wieling, Martijn, & Wolthuis, Nienke (2015). The role of frequency in the retrieval of nouns and verbs in aphasia. *Aphasiology*, 30(11), 1221–1239. doi: 10.1080/02687038.2015.1100709
- Berman, Ruth, & Slobin, Dan Isaac (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale: Lawrence Erlbaum.
- Caramazza, Alfonso (1991). Some aspects of language processing revealed through the analysis of acquired aphasia: The lexical system. *Neuropsychology and Cognition*, 3, 15–44. doi: 10.1007/978-94-011-3740-9_1
- Castillo Castillo, Olivia (2009). *‘Entonces’ de adverbio temporal a marcador del discurso: elaboraciones desde la perspectiva de un esquema de temporalidad* (Tesis de maestría inédita). Benemérita Universidad Autónoma de Puebla, Puebla.
- Chafe, Wallace (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.
- Christiansen, Julie Ann (1995). Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language*, 51(2), 291–317.
- Coelho, Carl, & Flewellyn, Laura (2003). Longitudinal assessment of coherence in an adult with fluent aphasia: A follow-up study. *Aphasiology*, 17(2), 173–182. doi: 10.1080/729255216
- Cuetos Vega, Fernando (2003). *Anomia*. Madrid: TEA Ediciones.
- Dingemans, Mark; Blythe, Joe, & Dirksmeyer, Tyko (2014). Formats for other-initiation of repair across language. An exercise in pragmatic typology. *Studies in Language*, 38(1), 5–43. doi: 10.1075/sl.38.1.01din
- Faitaki, Faidra, & Murphy, Victoria (2019). Oral language elicitation tasks in applied linguistics research. En Jim McKinley & Heath Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 360–369). Londres: Routledge. doi: 10.4324/9780367824471

- Fox, Barbara; Maschler, Yael, & Uhlmann, Susanne (2010). A cross-linguistic study of self-repair: Evidence from English, German, and Hebrew. *Journal of Pragmatics*, 42(9), 2487–2505. doi: 10.1016/j.pragma.2010.02.006
- Gertz, David (2007). *Liebman's neuroanatomy made easy and understandable*. Austin: Pro-Ed.
- Geschwind, Norman (1974). *Selected papers on language and the brain*. Dordrecht: Reidel.
- Givón, Talmy (1990). *Syntax: A functional-typological introduction*. Amsterdam: John Benjamins.
- Goodglass, Harold (1980). Disorders of naming following brain injury: Observation of the effects of brain injury adds another dimension to our understanding of the relations between neurological and psychological factors in the naming process. *American Scientist*, 68(6), 647–655.
- Goodglass, Harold; Barresi, Barbara, & Kaplan, Edith (2005). *Evaluación de la afasia y trastornos relacionados*. Madrid: Médica Panamericana.
- Holliday, M. A. K., & Hasan, Ruqaiya (1983). *Cohesion in English*. Londres: Longman.
- Helm-Estabrooks, Nancy; Albert, Martin, & Nicholas, Marjorie (2014). *Manual of aphasia therapy*. Austin: Pro-Ed.
- Holland, Audrey; Halper, Anita, & Cherney, Leora (2010). Tell me your story: Analysis of script topics selected by persons with aphasia. *American Journal of Speech-Language Pathology*, 19(3), 198–203. doi: 10.1044/1058-0360(2010/09-0095)
- Jonkers, Roel, & Bastiaanse, Roelien (2007). Action naming in anomic aphasic speakers: Effects of instrumentality and name relation. *Brain and Language*, 102(3), 262–272. doi: 10.1016/j.bandl.2007.01.002
- Kambanaros, Maria, & van Steenbrugge, Willem (2006). Noun and verb processing in Greek-English bilingual individuals with anomic aphasia and the effect of instrumentality and verb-noun name relation. *Brain and Language*, 97(2), 162–177. doi: 10.1016/j.bandl.2005.10.001
- Kasper, Gabriele, & Wagner, Johannes (2014). Conversation analysis in applied linguistics. *Annual Review of Applied Linguistics*, 34, 1–42. doi: 10.1017/S0267190514000014
- Knight, Dawn, & Adolphs, Svenja (2010). Building a spoken corpus: What are the basics? En Anne O'Keeffe & Michael McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 38–51). Londres: Routledge.

- Lara Galindo, Wendy; Rojas Nieto, Cecilia; Jackson—Maldonado, Donna, & Díaz Argüero, Celia (2015). ¿Qué problemas—fuente autorreparan los niños con trastorno primario de lenguaje? El él, el súsape, el pícipe. *Estudios de Lingüística Aplicada*, 61, 165–198. doi: 10.22201/enallt.01852647p.2015.61.156
- Levelt, Willem J. M. (1983). Monitoring and self—repair in speech. *Cognition*, 14(1), 41–104. doi: 10.1016/0010-0277(83)90026-4
- Levelt, Willem; Roelofs, Ardi, & Meyer, Antje (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. doi: 10.1017/S0140525X99001776
- Luria, Aleksandr (1980). *Fundamentos de neurolingüística*. Barcelona: Toray—Masson.
- McCarthy, Michael, & O’Keeffe, Anne (2010). Historical perspective. What are corpora and how have they evolved? En Anne O’Keeffe & Michel McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 3–13). Oxfordshire: Taylor & Francis Group.
- MacWhinney, Brian (2000). *The CHILDES project: Tools for analyzing talk*. Hillsdale: Lawrence Erlbaum.
- MacWhinney, Brian; Fromm, Davida; Forbes, Margaret, & Holland, Audrey (2011). Aphasia—Bank: Methods for studying discourse. *Aphasiology*, 25(11), 1286–1307. doi: 10.1080/02687038.2011.589893
- Marini, Andrea; Andretta, Sara; Del Tin, Silvana, & Carlomagno, Sergio (2011). A multi—level approach to the analysis of narrative language in aphasia. *Aphasiology*, 25(11), 1372–1392.
- Mayer, Mercer (1969). *Frog, where are you?* Nueva York: Dial Books.
- Mayer, Mercer (1974). *Frog goes to dinner*. Nueva York: Dial Books.
- Montes Miró, Rosa (1992). *Achieving understanding: Repair mechanisms. Mother—child conversations* (Tesis doctoral inédita). Universidad de Georgetown, Georgetown.
- Olness, Gloria Streit, & Englebretson, Elise (2011). On the coherence of information highlighted by narrators with aphasia. *Aphasiology*, 25(6–7), 713–726. doi: 10.1080/02687038.2010.537346
- Ortiz Carballo, Cecilia (2021). *Manifestación de la concordancia en las categorías funcionales de los grupos nominales y sus reparaciones en un paciente con afasia acústico—mnésica* (Tesis de licenciatura inédita). Benemérita Universidad Autónoma de Puebla, Puebla.

- Rieger, Caroline (2003). Repetitions as self–repair strategies in English and German conversations. *Journal of Pragmatics*, 35(1), 47–69. doi: 10.1016/S0378-2166(01)00060-1
- Schegloff, Emanuel (1990). On the organization of sequence as a source of ‘coherence’ in talk–in–interaction in conversational organization and its development. *Conversational Organization and its Development*, 38, 51–77.
- Schegloff, Emanuel (2011). Word repeats as units ends. *Discourse Studies*, 13(3), 367–380. doi: 10.1177/1461445611402749
- Schegloff, Emanuel (2013). Ten operations in self–initiated, same–turn repair. En Makoto Hayashi, Geoffrey Raymond & Jack Sidnell (Eds.), *Conversational repair and human understanding* (pp. 41–70). Cambridge: Cambridge University Press.
- Schegloff, Emanuel; Jefferson, Gail, & Sacks, Harvey (1977). The preference for self–correction in the organization of repair in conversation. *Language*, 53(2), 361–382.
- Selting, Margret (2000). The construction of units in conversational talk. *Language in society*, 29(4), 477–517.
- Tippett, Donna, & Hillis, Arguye (2016). Vascular aphasia syndromes. En Gregory Hickok & Steven L. Small (Eds.), *Neurobiology of language* (pp. 913–933). Oxford: Academic Press/Elsevier.
- Tremblay, Pascale, & Dick, Anthony (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, 162, 60–71. doi: 10.1016/j.bandl.2016.08.004
- Ulatowska, Hanna; Allard, Lee, & Bond, Sandra (1990). Narrative and procedural discourse in aphasia. En Yves Joannette & Hiram H. Brownell (Eds.), *Discourse ability and brain damage: Theoretical and empirical perspectives* (pp. 180–198). Nueva York: Springer.
- van Dijk, Teun (1996). *Estructura y funciones del discurso*. México: Siglo XXI.
- White, Ron (1997). Back channelling, repair, pausing, and private speech. *Applied Linguistics*, 18(3), 314–344. doi: 10.1093/applin/18.3.314

7. Anexo

CUADRO 7. Codificación, representación, características particulares en la transcripción del corpus de la presente investigación. Ejemplo

Codificación	Representación	Características particulares en la transcripción del corpus	Ejemplo
*	Símbolo que precede a la identificación del hablante		*TAP
(TRES LETRAS MAYÚSCULAS)	Identificación del hablante	RAM /TAP	*RAM
:	Posterior a los dos puntos se encuentra la transcripción de lo dicho por el hablante		*TAP: el niño
(ESPACIO DE UN TABULADOR)			
.	Punto final que señala el final de una unidad entonacional.		*TAP: coe ##.
#	Pausa	(-300 milisegundos)	*TAP: en # un
##	Pausa prolongada	(+300 milisegundos)	*TAP: en # un # panal ##. [0:01.142]
xxx	Material ininteligible tratado como palabra o secuencia		*TAP: XXX ##. [0:00.865]
www	Fragmento de la conversación que se ha decidido no transcribir		
()	Permite completar las palabras que el informante emite incompletas		
[>]	Sigue un solapamiento		
[<]	Precede un solapamiento		
[/]	Repetición sin corrección		
[//]	Repetición con corrección		
[/?]	Repetición indefinida (superior a tres veces)	En este caso, las repeticiones son parte del objeto de estudio, por lo que se transcriben de manera literal, independientemente de su uso.	

(continuación)

CUADRO 7. Codificación, representación, características particulares en la transcripción del corpus de la presente investigación. Ejemplo

Codificación	Representación	Características particulares en la transcripción del corpus	Ejemplo
[?]	Aproximación o conjetura		
[*]	Marca de uso peculiar codificado como error	En este estudio no se marcan errores debido a la naturaleza del discurso del corpus.	
@o	Onomatopeya		
@d	Forma dialectal		
@s	Segunda lengua		
+	Entre las partes de una onomatopeya		
+//	El emisor se interrumpe		
+/	Al emisor lo interrumpen		
++	Una persona completa la frase que ha iniciado otra		
+	Una persona completa la frase que ha iniciado él mismo en otro turno		
+"/	El siguiente turno es una cita		
+"	Se inicia una cita		
	El turno anterior era una cita		
[=!]	Risas		
&	No es una palabra		
[0:00.000]	Conteo en minutos:segundos.milisegundos		*TAP: ehhh ##. [0:00.504]

Fuente: Cuadro realizado a partir de los códigos de MacWhinney *et al.* (2000, 2011)

8. Abreviaturas

AR Autorreparación
 FP Fuente problemática

- PAA Persona con afasia anómica
M Marca explícita de lo que debe repararse
R Reparación
ECV Evento cerebro vascular
UE Unidad entonacional

9. Agradecimientos

Trabajo realizado con apoyo de la beca de doctorado otorgada por el Consejo Nacional de Humanidades, Ciencia y Tecnología (Doctorado en Lingüística, IIF UNAM). Agradezco toda la guía del Dr. Ricardo Maldonado Soto, así como el invaluable apoyo del Dr. José Marcos Ortega, su equipo de trabajo y, sobre todo, la generosidad de TAP para la realización de esta investigación.

Asimismo, agradezco la atenta lectura y los valiosos comentarios de los dictaminadores de este artículo.



Técnicas básicas de procesamiento del lenguaje natural en el análisis de menciones a mujeres en lecturas graduadas de ELE

Basic natural language processing techniques
for the analysis of mentions of women
in Spanish graded readers

Iban Mañas Navarrete

Universidad de las Islas Baleares,
Filología Española, Moderna y Clásica
iban.manas@uib.es

María-Valle Sell

Universidad Autónoma de Barcelona,
Departamento de Didáctica de
la Lengua y la Literatura, y de las
Ciencias Sociales
valle.sell@gmx.de



Recepción: 29 de junio del 2023

Aceptación: 19 de septiembre del 2023

doi: 10.22201/enallt.01852647p.2024.78.1084

Resumen

Los análisis de la representación de los géneros en materiales de ELE se han visto limitados hasta ahora por los métodos de recuento manual. El presente estudio abandona estos métodos y analiza la representación de los géneros en un corpus compuesto por 47 lecturas graduadas para estudiantes de ELE aplicando técnicas básicas de procesamiento del lenguaje natural basadas en el trabajo de Lucy, Demszky, Bromley y Jurafsky (2020). Se examinan las diferencias en el espacio dedicado a cada uno de los géneros y su evolución en el tiempo. Asimismo, se analiza el valor explicativo del sexo de los protagonistas sobre la distribución de menciones de hombres y mujeres. Los resultados muestran que, incluso cuando el protagonismo es compartido, las mujeres aparecen representadas en menor medida. Esta investigación aporta además un set de herramientas de análisis automático de textos que queda a disposición de la comunidad para el análisis de otros textos o para su adaptación al análisis de diferentes aspectos de la lengua con el fin de profundizar en la caracterización de los materiales de ELE.

Palabras clave: estudios de género; análisis de materiales; lecturas por niveles; procesamiento del lenguaje natural; Español como Lengua Extranjera; el texto como práctica social

Abstract

Analyses of gender representation in textbooks of Spanish as a Foreign Language (SFL) have been limited by manual-count methods. The present study adopts a different approach and investigates gender representation in a corpus of 47 graded readings for students of L2 Spanish by applying basic Natural Language Processing techniques based on the work by Lucy, Demszky, Bromley, and Jurafsky (2020). The differences in the space devoted to each gender and its evolution over time are examined. Likewise, the explanatory value of the gender of the main characters on the distribution of mentions of men and women is explored. The results show that women are under-represented, even in books where the main characters are both women and men. This work provides the results of the analysis and a set of tools for text analysis that is available for the community to apply it to the analysis of other texts or to adapt it to analyze different social aspects of language use and to deepen the characterization of the teaching materials used in SFL.

Keywords: gender studies; textbooks; graded readers; NLP; Spanish as Foreign Language; text as social practice

1. Introducción

El número de investigaciones centradas en el análisis de materiales de español como lengua extranjera (ELE) desde una perspectiva de género ha experimentado un crecimiento notable en las últimas décadas (Angulo Blanco, 2010; Filipović & Kuzmanović, 2020; Kuzmanović, 2022; Kuzmanović & Đuričić, 2016; Queiro Pérez, 2019; Robles Fernández, 2005; Taboada, Cejas, Sánchez, Spada, Taboada, Caire, Olivera, Pais, Ábrigo & Martín, 2021); hecho que refleja un importante paso en el proceso de toma de conciencia del contenido ideológico de cualquier material didáctico (Kramsch & Vinall, 2015). El presente estudio entronca con esta línea de trabajo y recurre a las ventajas de las técnicas de procesamiento del lenguaje natural y el tratamiento cuantitativo de datos para el análisis de materiales didácticos.

El procesamiento del lenguaje natural (NLP por sus siglas en inglés) puede resultar de gran ayuda en la obtención de datos relevantes para la investigación en segundas lenguas (Meurers & Dickinson, 2017) y, concretamente, en el análisis de materiales, puesto que ofrece la posibilidad de manejar corpus extensos de textos de manera automática y sistemática. En la investigación educativa, no obstante, la adopción de técnicas de NLP es todavía infrecuente (Fesler, Dee, Baker & Evans, 2019); especialmente escasos resultan los trabajos sobre su aplicación en la búsqueda de respuestas de corte sociológico en la educación (Lucy *et al.*, 2020: 2).

En este trabajo se analizan las menciones de mujeres en un corpus de 47 lecturas graduadas de ELE con ayuda de técnicas básicas de NLP para la identificación automática de menciones de hombres y mujeres. Las lecturas graduadas son materiales didácticos, concretamente, textos ficcionales diseñados para ser usados en el proceso de aprendizaje de una lengua (Acquaroni, 1996). Las herramientas de NLP desarrolladas durante el estudio están disponibles tanto para su libre uso, a fin de replicar el estudio con otros corpus, como para su libre modificación, ampliación o mejora. Además de esta contribución, los resultados obtenidos suponen

un aporte cuantitativo al cuerpo de trabajos sobre la cuestión del género y su representación en los manuales de ELE, que generalmente se desarrollan dentro de un marco cualitativo–descriptivo. Se profundiza, así, sobre un mismo fenómeno desde una perspectiva que desvela aspectos poco explorados.

2. El pensamiento crítico y la enseñanza de ELE

Existe un amplio consenso entre especialistas de la educación acerca de la importancia del pensamiento crítico en cualquier proceso educativo (Castillo Cuadra, 2020). La práctica del pensamiento crítico, entendido este como “un pensamiento que facilita el juicio porque se basa en criterios de razón, es autocorrectivo y sensible al contexto” (Lipman, 2014: 37), contribuye a la formación de ciudadanos más preparados para vivir en una democracia plena (Nussbaum, 2010). Esto supone, entre otros aspectos, aprender a relacionarse con los otros desde el respeto mutuo y ser menos vulnerable a la manipulación mediática y a la presión del pensamiento único.

El aula de lenguas extranjeras, como espacio de enseñanza/aprendizaje, resulta un contexto propicio para desarrollar estas habilidades (Ascaso & Atienza, 2018); es, de hecho, un marco idóneo para la reflexión sobre cómo el uso de la lengua influye en la manera de ver el mundo. Desde la tradición lingüística sociocultural, el discurso no solo modela la cultura en la que se inscribe, sino que es modelado por esta y, en tal sentido, es representante de los rasgos de dichas culturas y sus sociedades (Cameron, 1992; Fernández Martínez, 2012). En el Marco Común Europeo de Referencia para las Lenguas (MCER), que comparte en muchos aspectos la visión de la lengua como transmisor de cultura, se subraya la importancia de “propiciar métodos de enseñanza de lenguas modernas que fortalezcan la independencia de pensamiento, de juicio y de acción, combinados con las destrezas sociales y la responsabilidad social” (Consejo de Europa [CE], 2002: 4). En consonancia con esto, el MCER describe a los aprendientes de lenguas principalmente como agentes sociales (CE, 2002: 9) y el Plan Curricular del

Instituto Cervantes (PCIC), a su vez, define los objetivos generales de los niveles de referencia para el español partiendo de una concepción del aprendiente como “agente social, hablante intercultural y aprendiente autónomo” (Instituto Cervantes, 2006: 9). Es desde esta dimensión sociocultural del hablante que se puede fomentar especialmente el espíritu crítico con los estudiantes (Fuertes Gutiérrez, Soler Montes & Klee, 2021; Leeman, 2014). Aprovechando las facilidades que ofrece el contexto de aprendizaje de una lengua extranjera (LE) para reflexionar sobre las creencias y los prejuicios (Guerrero, 2003), Lacorte y Atienza (2019) recomiendan sobre el uso de propuestas metodológicas orientadas a la acción en el aula de ELE, por ejemplo, el aprendizaje por proyectos o el aprendizaje basado en problemas, como propuestas que facilitan la integración del pensamiento crítico en los contenidos curriculares.

Ahora bien, dado que un texto es un elemento clave de la práctica social, este es siempre susceptible de ser analizado críticamente en un entorno social como es el aula. Lo fundamental no es tanto “lo que el texto dice, sino cómo este se deconstruye en clase” (Lacorte & Atienza, 2019: 146). Para deconstruir un texto es necesario conocerlo a fondo y un análisis cuantitativo como el presente puede revelar propiedades inapreciables desde una inspección cualitativa de los contenidos. En este sentido, las lecturas graduadas permiten introducir dicha práctica desde los niveles iniciales, contribuyendo así a formar el espíritu crítico del aprendiente desde el comienzo. Además, para muchos estudiantes en contexto de no inmersión representan un producto cultural más que les permite acercarse a la sociedad de la lengua meta. Las herramientas que ofrece el análisis crítico del discurso (Van Dijk, 2016) capacitan a los estudiantes para posicionarse ante el material didáctico como miembros de la sociedad (Lacorte & Atienza, 2019), con el objetivo de analizar el lenguaje como práctica social para descifrar las ideologías y relaciones de poder subyacentes (Amari, 2015; Fairclough, 1989; Trujillo, 2009). Una de las relaciones de poder que pueden estudiarse en los materiales didácticos de ELE son las diferencias de género.

3. La cuestión de género y la enseñanza de ELE

Este análisis se encuadra en la educación con perspectiva de género (EPG), cuyo objetivo es contribuir a erradicar desigualdades, así como “al empoderamiento de las mujeres y las niñas y a fomentar el análisis crítico de los roles que se han concedido a los distintos sexos a lo largo de la historia” (Martínez Lirola, 2022: 85).

Desde los inicios de los estudios en lenguaje y género (Lakoff, 1973) se han realizado innumerables trabajos sobre la representación de los géneros en los libros de texto (Cameron, 1992; Chisholm, 2018), entendiéndose *género* como el constructo social que define cómo ha de ser un hombre y cómo una mujer.¹

En la perpetuación o impugnación de la desigualdad que genera la distinción de ciudadanos por género, el sistema educativo, en cualquiera de sus modalidades y etapas, tiene un papel central. De ahí que desde las instituciones públicas se impulsen iniciativas dirigidas a promover la educación en igualdad, un enfoque pedagógico basado en la “reivindicación de la igualdad [...] que propone una reformulación del modelo de transmisión del conocimiento desde una perspectiva de género en los espacios de socialización destinados a la formación y el aprendizaje” (Instituto de la Mujer, 2008: 17), para una educación libre de estereotipos sexistas. En la misma línea, la IV Conferencia Mundial sobre las Mujeres recoge la importancia de “elaborar planes de estudio, libros de texto y material didáctico libres de estereotipos sexistas para todos los niveles de enseñanza” (UN Women, 1996: 56).

En los primeros niveles educativos esto es especialmente importante ya que los niños empiezan desde muy temprano a construir las nociones de feminidad y masculinidad en su interacción con el entorno. Pero los sesgos de género en los materiales didác-

¹ Esta interpretación de género responde al concepto elaborado desde el pensamiento filosófico del feminismo de la igualdad, que denuncia el desequilibrio estructural que el binomio sexo/género produce en sociedades patriarcales (Lerner 1986).

ticos se encuentran de forma transversal en todas las etapas y todos los contextos (Blumberg, 2007), incluida la literatura que se trabaja en aula de ELE (Chover Lafarga, 2007).

El informe de Blumberg (2007) destaca la infrarrepresentación sistemática de las mujeres en los materiales didácticos, así como el grado altamente estereotipado de las actividades que realizan. Asimismo, se afirma que los avances que han tenido lugar para las mujeres en el mundo laboral en las últimas décadas no se ven reflejados en los materiales didácticos y que únicamente se detecta una cierta tendencia de reducción del sesgo de género. En el ámbito de la educación en España, según Vaíllo Rodríguez, los estudios comparativos “han alertado sobre la insuficiencia de los cambios acaecidos” (2016: 107), cuando no de una invisibilización sistemática y constante en los libros de texto de la educación obligatoria (López Navajas, 2014; Peñalver Pérez, 2003).

En el área de ELE específicamente, el análisis de seis manuales por parte de Kuzmanović y Đuričić indica tan solo “una intención clara de evitar (o, al menos, minimizar) la presencia de contenidos o representaciones estereotipados de géneros” (2016: 117). A conclusiones similares llegaron otros análisis de materiales elaborados desde la perspectiva de género (Robles Fernández, 2005; Queiro Pérez, 2019, entre otros). Molina Gil (2022), en su estudio sobre la representación de mujeres escritoras en los cursos de literatura de ELE, confirma diferencias de género tradicionalmente asentadas en la historiografía literaria. Por tanto, es necesario continuar reivindicando la necesidad de introducir la perspectiva de género también en el contexto ELE.

En cuanto a la metodología aplicada, para el análisis del contenido se suele emplear el recuento manual de ocurrencias. Galiano Sierra (1993) realiza un recuento de las profesiones de las mujeres en las imágenes de nueve manuales de ELE. En otros trabajos, como el de Morales Vidal y Cassany (2020), los autores examinan hasta siete parámetros de desigualdad social en cinco manuales de ELE, entre ellas la categoría de género; si bien los autores se refieren a estereotipos de género y lo incluyen en la categoría de

diversidad sexual e identidades de género. En todos estos estudios, el tiempo necesario para realizar un recuento manual limita las dimensiones de los corpus.

Esta observación coincide con Lucy *et al.* (2020: 2), cuando afirman que “large textbook coding efforts may reduce their tasks to counting or identifying simple indicators”.² También Vaíllo desarrolla este tema y apunta que “las investigaciones producidas en la última década no abandonan la cuantificación, aunque evidencian un claro esfuerzo por perfeccionar los instrumentos de análisis en relación con los aspectos de corte cualitativo” (2016: 105).

Aunque se ha estudiado la cuestión de género en manuales de ELE, las lecturas graduadas están mucho menos exploradas. No se ha encontrado ningún análisis centrado en un corpus de este tipo. En el ámbito del inglés como LE, Muto–Humphrey (2005) analiza el lenguaje utilizado en dos lecturas graduadas y confirma un desequilibrio en las representaciones de género. Hunter y Chick, luego de analizar 47 lecturas simplificadas para niños nativos de inglés encontraron que los personajes masculinos adultos, además de aparecer con una frecuencia mayor estadísticamente significativa, ejercen una profesión con mucha más frecuencia que los femeninos (2005: 1). La principal dificultad en estos estudios es la cantidad de recursos humanos necesaria: Hunter y Chick (2005) contaron con la ayuda de 42 asistentes para la recolección de los datos. Resulta evidente que las investigaciones se ven limitadas por el sistema de recuento manual.

Algunos estudios, tomando conciencia de esta limitación, han aplicado técnicas de la lingüística de corpus, automatizando el análisis de datos en un tiempo razonable (McEnery & Hardie, 2011). Entre el conjunto de estos trabajos, cabe destacar el estudio de Hunt (2015) sobre las representaciones de género en los libros de *Harry Potter*, en el que la autora pondera el grado de agencia

² “[L]os grandes esfuerzos de codificación de libros de texto pueden reducir sus tareas al recuento o la identificación de indicadores simples” (la traducción de las citas es nuestra).

de los diferentes personajes en las situaciones en las que aparecen representados. Hunt expone que su estudio analiza aspectos más sutiles del texto y el discurso que estudios anteriores similares en el ámbito de la literatura infantil y llega a la conclusión de que en las novelas de *Harry Potter* “females [...] are systematically excluded from acting on the world in powerful or meaningful ways and the message is clear: agency is not for girls”³ (Hunt, 2015: 266).

El estudio de Lucy *et al.* (2020) es una de las investigaciones de referencia en la aplicación de técnicas de NLP para el análisis de cuestiones sociales en materiales didácticos. Los autores apuntan que “los recientes desarrollos metodológicos [...] son muy prometedores a la hora de arrojar luz sobre cuestiones clave de los aspectos políticos y sociales de la educación” (Lucy *et al.*, 2020: 1). Ahora bien, advierten que los enfoques computacionales no sustituyen los métodos de investigación clásicos (Lucy *et al.*, 2020: 3) y que lo recomendable es combinar ambos enfoques para aprovechar las fortalezas y compensar las debilidades de cada uno. En su estudio recogen un corpus de libros de texto de historia en Texas, Estados Unidos, y realizan un análisis de contenido para describir la imagen de grupos sociales históricamente marginados por cuestiones de sexo, raza y etnia. En sus resultados destacan la prevalencia del hombre blanco en los manuales, además de una infrarrepresentación de mujeres y personas negras.

Teniendo en cuenta los resultados de las investigaciones previas, el presente estudio se propone analizar la presencia de hombres y mujeres en las lecturas graduadas de ELE. Entre los estudios sobre esta cuestión es habitual encontrar líneas de trabajo dirigidas a modificar la presencia de la mujer en los materiales didácticos. Por ejemplo, Salido Machado (2017) destaca la importancia de tratar la imagen de la mujer en el aula de ELE desde una perspectiva intercultural. Ahora bien, para que esto se pueda plantear es

³ “[S]e excluye sistemáticamente a los personajes femeninos [...] de actuar en el mundo de manera poderosa y relevante, y el mensaje en claro: la agencia no es para chicas.”

imprescindible que, previamente, los docentes tengan a disposición materiales didácticos con suficiente presencia de mujeres en sus contenidos. Por este motivo, en este estudio se plantean las siguientes preguntas de investigación:

- ¿Cómo ha evolucionado la presencia de las mujeres, en términos cuantitativos, en las lecturas graduadas de ELE durante las dos últimas décadas?
- ¿Cómo se distribuyen las frecuencias de menciones a mujeres y hombres en las lecturas graduadas de ELE en función del sexo de los protagonistas?

4. Metodología

4.1. Descripción del corpus

El corpus se compone de 47 lecturas graduadas distribuidas en once colecciones, con un total de 472 768 tokens, 341 856 excluyendo la puntuación y 3 477 types. El corpus incluye 80% de las lecturas de la editorial Difusión. Se eligió dicha editorial por varios motivos. En primer lugar, se trata de una editorial de referencia en el área de la enseñanza de ELE (Kuzmanović & Đuričić, 2016) y ha sido objeto de estudio en diferentes análisis de materiales relacionados con el tratamiento del género y la diversidad sexual (Morales Vidal & Cassany, 2020; Ramírez, 2020; Taboada *et al.*, 2021). Por otro lado, se cuenta entre las primeras que incluyó en su catálogo lecturas graduadas de ELE, por lo que permite analizar la evolución en el tiempo de los parámetros estudiados.

En estas lecturas graduadas se encuentra un número suficiente de personajes principales femeninos para el análisis comparativo propuesto; en concreto, el corpus se compone de 18 lecturas con mujer protagonista, 12 lecturas con hombre protagonista y 17 con hombre y mujer como protagonistas. La incorporación de esta variable en el estudio y análisis de su efecto en la representación de mujeres y hombres contribuye al carácter innovador de esta inves-

tigación, en tanto que no se ha encontrado hasta la fecha trabajos similares que tengan en cuenta el sexo del protagonista. Este estudio se centra en cuantificar la presencia de hombres y mujeres, y no las diversas expresiones de género posibles. La flexión binaria de género en español permite rastrear únicamente las menciones a mujeres y hombres.

Para la creación del corpus, se digitalizaron todas las lecturas con la aplicación Scanner Pro y se convirtieron en texto mediante un programa de reconocimiento óptico de caracteres, ABBYY FineReader. No se incluyeron las ilustraciones ni el texto correspondiente a ejercicios y glosarios. A continuación, se eliminaron notas al pie, parte de la puntuación y caracteres especiales innecesarios. Cada libro fue etiquetado con su título, colección a la que pertenece, año de publicación y sexo de los protagonistas (véase el Anexo para conocer la lista completa de títulos incluidos en el análisis).

4.2. Las herramientas de NLP

El código desarrollado para el análisis se basó en el trabajo de Lucy *et al.* (2020). Los *scripts* programados están disponibles en GitHub para su libre uso con una licencia GNU General Public License v3.0. Con el fin de cuantificar el espacio destinado a hablar de hombres y mujeres fue necesario identificar términos referidos a unos y otras. Para ello se crearon dos listas, una de nombres propios masculinos y femeninos y otra de sustantivos con moción de género en la flexión para hombres (*camarero*), para mujeres (*jefa*) y sustantivos comunes en cuanto al género (*periodista*).

El listado de nombres propios se creó combinando datos extraídos de Wikipedia y del Instituto Nacional de Estadística de España. Se añadieron manualmente algunos nombres y también algunos apellidos, ya que en España se utilizan con frecuencia para referirse a hombres, y tan solo de manera anecdótica a mujeres (Guerrero, 2001). El listado final no incluyó nombres compuestos,

ya que el programa fue diseñado para ignorar la segunda parte del nombre, y contiene 2622 entradas.

Para la elaboración de la segunda lista se recurrió a WordNet (Fellbaum, 1998), una base de datos léxica en inglés. Se creó una lista de hipónimos de *woman* y *man* y se realizó una traducción automática al español. La Figura 1 ilustra la lógica simplificada del *script*. El algoritmo parte de una palabra y crea una lista que incluye dicha palabra y todos sus hipónimos. Después busca los hipónimos de cada una de las palabras contenidas en la lista y los añade a la misma. Este ciclo se repite tantas veces como se haya especificado en los parámetros de la función. La comprobación de si se trata de un término referido a personas se realiza mediante restricción a la categoría lexicográfica de sustantivos referidos a personas. El número de niveles o profundidad de la búsqueda se puede modificar cambiando el número de iteraciones ($n = 4$ en el código, *default* = 1). Asimismo, es posible introducir la categoría lexicográfica o dejar el argumento en blanco (*default* = *none*).

A continuación, se realizó una traducción automática al español del listado resultante utilizando la interfaz de programación de aplicaciones (API) de Google Translate (Han, 2020). Se revisó el resultado, primero por ordenador, para eliminar duplicados, y después manualmente, borrando sustantivos no adecuados y añadiendo también algunos términos a mano tras comprobar que faltaban muchas formas femeninas de sustantivos masculinos, como *ingeniera*. Se modificó la categoría de algunas palabras, por ejemplo, *turista*, de la categoría *hombres* a *sustantivos comunes en cuanto a género*. También se realizaron algunos ajustes manuales, como eliminar la palabra *amazonas* tras comprobar que en los libros aparece en su sentido geográfico, o *puta* y *puto*, porque aparecen en numerosas ocasiones como sustantivos adjetivados que forman parte de expresiones malsonantes. El resultado fue una lista de 459 términos, 186 de la categoría *mujeres*, 194 de la categoría *hombres* y 79 de la categoría *sustantivos comunes en cuanto al género*.

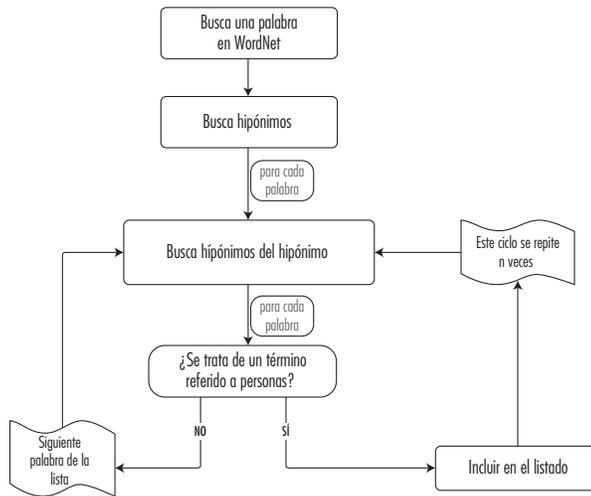


FIGURA 1. Lógica simplificada del algoritmo de búsqueda de hipónimos en WordNet

Por último, se añadieron los determinantes masculinos y femeninos y los pronombres y se combinó con el listado de nombres propios. Para comprobar que esta lista de términos era completa y, por tanto, que los resultados que se obtuvieran serían fiables, se extrajeron todos los núcleos nominales de las 47 lecturas que aparecen más de diez veces utilizando la biblioteca de Python spaCy (Honnibal & Montani, 2017), siguiendo a Lucy *et al.* (2020). Tras esta comprobación, se añadieron siete nombres propios (0.2% del total) y siete sustantivos (1.5% del total). Se creó un *script* modificando el de Lucy *et al.* (2020) para obtener el recuento de menciones de hombres y mujeres. La Figura 2 muestra la lógica simplificada del *script* modificado.

En el caso de los sustantivos comunes en cuanto al género, el algoritmo consultaba la palabra anterior y, si se trataba de un determinante incluido en el listado, añadía el sustantivo en la categoría correspondiente. Si la palabra anterior no era un determinante, confirmaba si se trataba de un adjetivo y, en caso afirmativo, comprobaba su flexión de género. Dado que muchos nombres propios

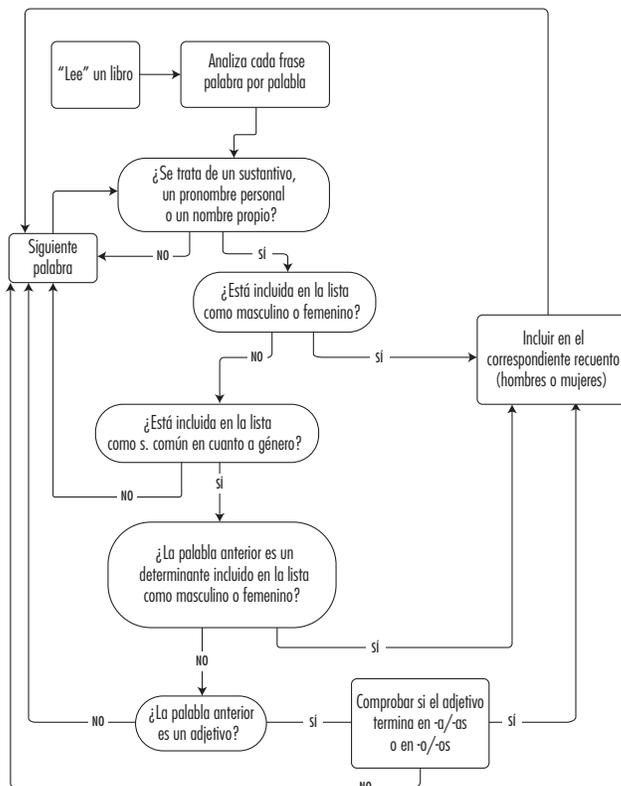


FIGURA 2. Lógica del recuento de menciones de hombres y mujeres

en español tienen carga semántica, algunas palabras, como *esperanza* entrarían en el recuento como nombre de mujer. Para evitar este resultado, se incluyó un listado de nombres propios con carga semántica, para que el algoritmo comprobara si los nombres estaban escritos con mayúscula inicial. Tras esta última modificación del código se llevó a cabo el proceso de recuento de menciones en el corpus.

4.3. Análisis

Se aplicaron dos análisis de correlaciones sobre el número de menciones a mujeres y a hombres. En el primero, se exploró la relación entre el número de menciones a mujeres y los años en que se publicaron las lecturas a través del índice de Spearman. En el segundo, se analizó la distribución de menciones a hombres y a mujeres dentro de cada subconjunto de lecturas según el tipo de protagonista (hombre, mujer o ambos sexos) a partir de la prueba de independencia χ^2 -cuadrado. Posteriormente, se exploró el alcance explicativo que tiene el sexo del protagonista sobre el número de menciones a hombres y mujeres. Tras confirmar sobredispersión de los datos en los dos casos (menciones a mujeres: $X^2 = 490,74$; $p < 0,001$; menciones a hombres: $X^2 = 589,24$; $p < 0,001$), se ajustaron sendos modelos de regresión binomial negativa con el paquete MASS en R. En los dos modelos se introdujo el número total de menciones como covariable centrada por la media y el tipo de protagonista como factor, ajustando como nivel de referencia la categoría de protagonistas de ambos sexos (MH).

5. Resultados

Una vez obtenido el recuento de menciones a mujeres y a hombres, se explora la relación entre el número de menciones a mujeres y el año de publicación de las obras. El intervalo temporal abarca aproximadamente las dos últimas décadas; desde 1999 (*El secreto de las flores* y *La chica del tren*) hasta 2020 (*Amor en línea*). Asumiendo que el número de menciones a mujeres en cada lectura está correlacionado con el número de menciones a personas, se computó un índice de correlación parcial entre las menciones a mujeres y el año de publicación, controlada por el total de menciones, calculado a partir de la suma de menciones a hombres y a mujeres en cada lectura (Figura 3).

Dado que no se cumplía la condición de distribución normal bivariada ($HZ = 0,89$; $p = 0,043$), se calculó el índice de correlación

de Spearman. El resultado apuntó a una ausencia de correlación entre el año de publicación y el número de menciones a mujeres ($r(45) = 0,19$; $p = 0,184$). Se confirma así una distribución uniforme de la presencia de mujeres en las lecturas graduadas durante los últimos 20 años.

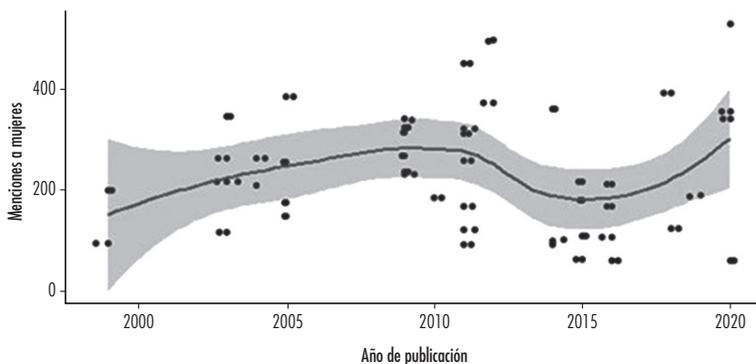


FIGURA 3. Correlación parcial del número de menciones a mujeres y año de publicación

La siguiente condición para el análisis de las menciones a mujeres fue el sexo del protagonista. Las lecturas se dividieron en tres grupos: protagonista hombre (H), mujer (M) y ambos (MH), y se observó la distribución de menciones a hombres y a mujeres en cada grupo a partir de una tabla de contingencia (véase Cuadro 1). En las lecturas con protagonista femenina las menciones a hombres representaron 47.4%, mientras que las menciones a mujeres supusieron 52.6% del total; a diferencia de los otros dos grupos, en los que se observó un desequilibrio pronunciado a favor de las menciones a hombres (61.5% en el caso de protagonistas de ambos sexos y 68.5% en las lecturas de protagonista masculino). El resultado confirmó la existencia de diferencias significativas en las distribuciones de menciones entre H y M según los protagonistas de las obras ($X^2 = 769,18$; $df = 2$; $p < 0.001$; $V = 0.17$, IC95% [0.14;0.18]).

CUADRO 1. Distribución de menciones a cada sexo en las lecturas graduadas según su protagonista

	Protagonista		
	Mujer y hombre	Hombre	Mujer
Menciones a hombres	6382 (61.5%)	3888 (68.6%)	4750 (47.4%)
Menciones a mujeres	3988 (38.5%)	1787 (31.4%)	5270 (52.6%)
Total menciones	10370 (100%)	5675 (100%)	10020 (100%)

La exploración de los residuos estandarizados permitió determinar dónde se encontraban las diferencias entre proporciones, después de ajustarse el nivel de significación al número de comparaciones múltiples ($p = 0,008$). Tal como se esperaba, las lecturas con protagonista H presentaron una proporción mayor de menciones a hombres ($Z_{res} = 18,763$), mientras que la proporción de menciones a mujeres fue superior en las lecturas con protagonista M ($Z_{res} = 26,386$). Más destacable resultó la proporción significativamente mayor de menciones a hombres en las lecturas con protagonistas MH ($Z_{res} = 10,404$). De estos resultados se desprende que, en las lecturas donde hombre y mujer comparten protagonismo, la presencia de mujeres fue significativamente inferior a la de los hombres.

Más allá de la relación entre el tipo de protagonista (H, M o MH) y las menciones a hombres y mujeres, en este estudio se examinó el alcance explicativo que tiene el sexo del protagonista de las lecturas sobre el número de menciones a hombres y mujeres. Los resultados que arrojaron los modelos de regresión creados para cada uno de los recuentos permitieron explorar esta cuestión. En el Cuadro 2 se reportan los resúmenes de los modelos finales (Modelo 2) comparados con sus respectivos modelos basales (Modelo 1), teniendo el número total de menciones como único parámetro.

Según los resultados del test de razón de verosimilitud, la inclusión del tipo de protagonista como parámetro resultó una contribución significativa a la capacidad de ajuste del modelo sobre las menciones a mujeres ($X^2 = 3,485$; $df = 2$; $p < 0,001$) y a hombres ($X^2 = 11,616$; $df = 2$; $p = 0,003$). Además, para ambos conjuntos

de datos los índices globales pseudoR² mejoraron del Modelo 1 al Modelo 2. En el caso de las menciones a mujeres, el modelo final explicó el 76% de la varianza de los datos; para las menciones a hombres, el 74%. La exploración de los residuos de ambos modelos (menciones a mujeres: $W = 0,956$; $p = 0,078$; menciones a hombres: $W = 0,980$; $p = 0,608$) confirmó el comportamiento óptimo del Modelo 2 en ambos casos.

CUADRO 2. Resultados de los modelos de regresión para las menciones a mujeres y a hombres

	Menciones a mujeres						Menciones a hombres					
	Modelo 2			Modelo 1			Modelo 2			Modelo 1		
	Est.	SE	sig.	b	SE	sig.	Est.	ES	sig.	b	ES	sig.
Constante												
Menciones total	5,327	,0630	**	5,389	,0534	**	5,718	,0578	**	5,695	,0388	**
Protagonista H	0,001	,0001	**	0,001	,0002	**	0,001	,0001	**	0,001	,0001	**
Protagonista M	-0,374	,1011	**				0,141	,0917	ns			
	0,318	,0867	**				-0,175	,0802	*			
pseudoR ²												
McFadden	0,119			0,059			0,107			0,088		
Nagelkerke	0,768			0,517			0,744			0,673		
AIC	516,94			547,43			540,74			548,36		

Nota: ** significativo a un nivel $p < 0,01$; * significativo a un nivel $p < 0,05$; ns = no significativo.

Los estimados del Modelo 2 para las menciones a mujeres indicaron que tener una protagonista M incrementaba significativamente la tasa de incidencia de mención a mujeres (IRR = 1,374; IC95% = 1,160; 1,628), lo cual supone un 37% más comparado con las menciones en lecturas con protagonista MH. Por su parte, las lecturas con protagonista H mostraron una reducción significativa de 31% en la tasa de menciones a mujeres (IRR = 0,687; IC95% = 0,566; 0,836) en comparación con las lecturas con protagonista MH. El cambio de protagonista MH a M conllevó una tasa superior

de menciones a mujeres, a la vez que el cambio de protagonista MH a H implicó una tasa inferior de menciones a mujeres. En la Figura 4 se recoge la media de menciones a mujeres esperada en función del número de menciones total para los diferentes tipos de protagonista.

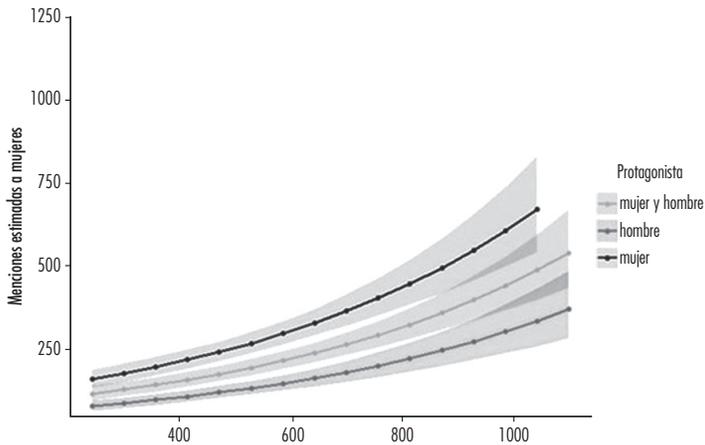


FIGURA 4. Media estimada de menciones a mujeres según el tipo de protagonista

En el Modelo 2 sobre las menciones a hombres, el estimado para la condición de protagonista M presentó el efecto opuesto al del Modelo 1 sobre menciones a mujeres. El protagonismo femenino conllevó una reducción significativa en la tasa de incidencia de mención a hombres ($IRR = 0,839$; $IC95\% = 0,716; 0,983$) en comparación con las lecturas con protagonista MH (Figura 5). En la tasa de menciones estimada para las lecturas con protagonista H, en cambio, no se presentó un aumento significativo comparado con la tasa de menciones de lecturas con protagonista MH ($IRR = 1,152$; $IC95\% = 0,958; 1,388$). Estos resultados indican que la tasa de menciones a hombres en los grupos de lecturas que tienen como protagonista MH y H son equivalentes. La ausencia de un cambio significativo entre estos dos grupos se puede explicar por la mayor presencia de hombres en las lecturas con protagonistas

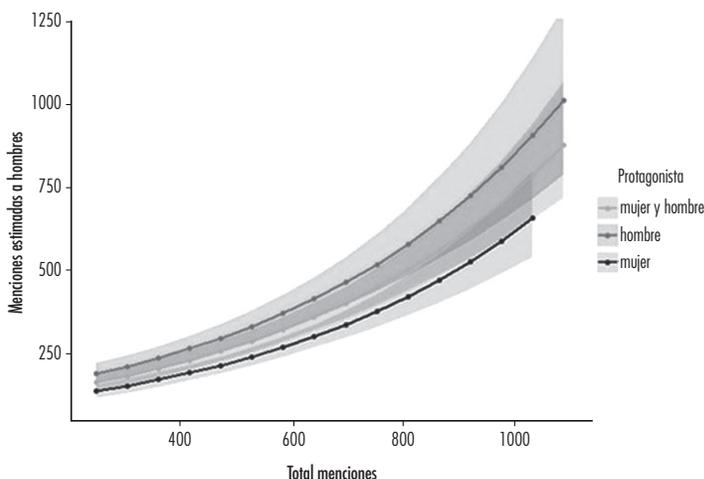


FIGURA 5. Media estimada de menciones a hombres según el tipo de protagonista

de ambos sexos, tal y como se mostró de manera descriptiva en el Cuadro 2.

Cuando las mujeres comparten protagonismo con los hombres, el espacio de las menciones a hombres fue equivalente al espacio que ocupan en obras con protagonista H. En el caso contrario, no se presentó este fenómeno. El número de menciones a mujeres aumentó significativamente de las lecturas con protagonista MH a las lecturas con protagonista M. Este efecto se debe a que las menciones a mujeres recuperan un espacio que, cuando comparten protagonismo con los hombres, poseen en una proporción inferior.

6. Discusión

En relación con la primera pregunta de investigación, sobre la presencia de las mujeres en las lecturas graduadas de ELE durante las dos últimas décadas, nuestros resultados han mostrado que el número de menciones a mujeres se mantiene a lo largo del tiempo. No se percibe, por tanto, ningún cambio en el espacio que ocupan las mujeres en el corpus. Estos resultados contrastan con las con-

clusiones que arrojan los análisis cualitativos de manuales de ELE, en los que generalmente se siguen detectando sesgos sexistas (Filipović & Kuzmanović, 2020; Kuzmanović, 2016; Kuzmanović & Đuričić, 2016), que se van reduciendo cuanto más actuales son los materiales.

La conjunción de estos resultados revela que, si bien se puede afirmar que se detecta un esfuerzo en el trabajo de editores y autores por reducir los sesgos sexistas a la hora de mostrar a las mujeres, en las últimas dos décadas no parece haberse superado el sesgo más elemental, el de la menor presencia de mujeres en los materiales didácticos. En veinte años el espacio que ocupan las mujeres en las lecturas graduadas no ha aumentado. Mientras, la imagen de las mujeres sí ha evolucionado en los manuales actuales de ELE, en concreto, se ha detectado una mayor presencia de mujeres y una reducción de lenguaje que las infravalore o subordine (Kuzmanović, 2016). Este resultado, relevante por sí mismo, se vuelve todavía más importante cuando se toma en consideración que, en el corpus, el número de lecturas con protagonista femenina ($n = 18$) es superior al de lecturas con protagonista masculino ($n = 12$). Entre los títulos de principio de siglo, donde se encuentran las lecturas de *Lola Lago*, cuya protagonista es una detective, se registra un número de menciones similar al de años posteriores, cuando las mujeres protagonistas no son tan frecuentes. Sabemos que en sociedades androcéntricas la producción cultural normalmente está copada por la presencia de hombres. Esto explicaría la inercia que opera por defecto y de manera tácita en la porción de espacio que se otorga a las mujeres, que persiste y resiste al margen del papel que ocupen las mujeres, protagonistas o no, en los relatos.

Respecto a la segunda pregunta de investigación, centrada en analizar las menciones a mujeres y hombres en función del sexo de los protagonistas, los resultados sobre la distribución de los totales señalaron una preponderancia ostensible de las menciones a hombres en las lecturas con protagonistas de ambos sexos. Una vez más, la disputa por el espacio en la narración se resuelve a favor de los hombres. Además, el análisis de la tabla de contin-

gencia desveló diferencias en las proporciones de menciones a hombres y mujeres según el tipo de protagonista. En lecturas con protagonista femenina, la presencia de hombres es notablemente superior (47.4%) a la presencia de mujeres en las lecturas protagonizadas por hombres, que no llegan a alcanzar un tercio del total de menciones (31.4%). Las lecturas con protagonista masculino prescinden en mayor medida de las mujeres que viceversa. Los resultados de los dos modelos ayudan a determinar el efecto de la variable sexo de los protagonistas. En el caso de las menciones a mujeres, el protagonismo femenino implica un incremento significativo del número de menciones, si bien tan solo supone poco más de la mitad, por lo que parecería que, incluso cuando las historias están protagonizadas por mujeres, la presencia de los hombres se antoja irrenunciable. Por otro lado, en las menciones a hombres, las lecturas protagonizadas por hombres y por ambos sexos muestran un número de menciones equivalente. Dado que el punto de referencia lo constituyen las lecturas con protagonistas de los dos sexos, en las que hay un desequilibrio patente a favor de los hombres, podemos afirmar que se observa un paso de la baja representación de mujeres en las lecturas coprotagonizadas con hombres a un cuasi-equilibrio de las menciones en las lecturas con mujeres protagonistas.

En cambio, al analizar las menciones a hombres, no se detectan diferencias en el espacio que estos ocupan entre las lecturas protagonizadas por hombres y las protagonizadas por ambos sexos. Estos resultados, leídos desde su reverso, podrían entenderse como la constatación de que las mujeres están infrarrepresentadas cuando comparten protagonismo con hombres. Estos resultados entroncan con las conclusiones que han alcanzado estudios previos sobre lecturas graduadas y simplificadas en inglés (Hunter & Chick, 2005; Muto–Humphrey, 2005).

La persistencia del patrón de dominación de los hombres sobre las mujeres se observa, por tanto, entre estudios de diferentes lenguas sobre el mismo género literario, lo cual justifica la importancia de este tipo de estudios con el propósito de alcanzar una des-

cripción más detallada de los materiales didácticos que se emplean en el aula de lenguas extranjeras. Un uso crítico de los materiales pasa inevitablemente por un conocimiento profundo de los mismos que capacite, tanto a profesores como a estudiantes, para desvelar la ideología subyacente de los discursos (Atienza Cerezo & Aznar Bertolín, 2020).

Además, los análisis de materiales desde una perspectiva crítica pueden estimular la toma de conciencia por parte de los creadores y despertar, como señalan Morales-Vidal y Cassany, un mayor “compromiso político de las editoriales del ámbito del español LE/L2 para asumir responsabilidades éticas más allá de sus legítimos intereses comerciales, con el propósito de reflejar mejor la diversidad de la comunidad y avanzar hacia un futuro con mayor equidad” (2020: 16). Para alcanzar una sociedad con mayor equidad, un paso ineludible es identificar y cuantificar la desigualdad existente derivada del sexismo.

La presencia del sexismo, como hemos podido comprobar en la revisión de estudios previos, se expresa no solo cualitativamente en la manera como se representa a las mujeres, sino también, como muestran los resultados de este estudio, cuantitativamente, limitando sistemáticamente el espacio que ocupan en relación con los hombres. En este sentido, cabe destacar el trabajo, desde el feminismo radical, humanista e ilustrado, por vindicar el derecho de las mujeres a ocupar el espacio público (de Miguel, 2004), del cual el espacio cultural y educativo forman parte. La lucha por la visibilización no tiene sentido si no se acompaña de la lucha por ocupar el espacio. Sin esta, la visibilización de la mujer corre el riesgo de convertirse en un cambio puramente estético, una estrategia de escaparate (Nogué, 2013). De esta postura se desprende una idea que no por ser evidente resulta menos reveladora: si las mujeres ocuparan más espacios, no necesitarían ser visibilizadas.

7. Conclusión

Los resultados de este trabajo han constatado que la presencia de hombres y mujeres no se distribuye de manera equilibrada en las lecturas graduadas de ELE. Análisis cuantitativos de este tipo son imprescindibles para proporcionar un escenario con datos robustos sobre el cual discutir los sesgos de género en los materiales didácticos. Al mismo tiempo, el profesor de ELE puede valerse de esta información para fomentar entre sus estudiantes una reflexión crítica acerca de las representaciones de hombres y mujeres en la sociedad. Si los sesgos de género son una de las “piedras mejor camufladas en el camino hacia la igualdad” (Blumberg, 2007: 4), información como la obtenida en esta investigación contribuye a la detección de características y patrones en los textos de ELE que, de otra manera, tal vez permanecerían camuflados.

En futuras investigaciones sería deseable aplicar un módulo de resolución de la anáfora, de manera que los sujetos elididos y los pronombres átonos sin flexión de género fueran asociados al personaje correspondiente. Si bien esto ayudaría a registrar un mayor número de menciones, la producción de sujetos nulos representa un mecanismo que afecta a ambos géneros por igual. Por ello, entendemos que los resultados del estudio no se desvirtúan por la ausencia de este módulo.

Por otro lado, en un siguiente paso, sería recomendable incorporar un análisis de contenido de las menciones a hombres y mujeres para desvelar qué imágenes de los géneros se promueven. Para ello se puede utilizar el código programado durante este trabajo, que permite explorar cuáles son los sustantivos más comunes utilizados para referirse a hombres y mujeres, las profesiones ejercidas y qué verbos y adjetivos les caracterizan con una frecuencia significativamente distinta. Asimismo, es importante investigar la evolución en el tiempo de dichas variables.

8. Referencias

- Acquaroni, Rosana (1996). Lecturas graduadas: algunas consideraciones para su incorporación a las actividades de clase. *Frecuencia L*, 1, 18–20.
- Amari, Fatma Zohra (2015). The role of critical discourse analysis in EFL teaching/learning. *Frontiers of Language and Teaching*, 6, 87–93.
- Angulo Blanco, María Esther (2010). *La evolución del papel de la mujer en dos manuales de ELE a través de las imágenes*. Las Lagunillas: Universidad de Jaén, Fundación Universitaria Iberoamericana.
- Ascaso, Samanta, & Atienza, Encarna (2018). El desarrollo del pensamiento crítico en ELE y su evaluación. *e-AESLA*, 4, 122–132.
- Atienza Cerezo, Encarna, & Aznar Bertolín, Joan (2020). Criterios para el desarrollo de las competencias interculturales en español LE/L2 desde una perspectiva crítica. *Journal of Spanish Language Teaching*, 7(2), 137–149. doi: 10.1080/23247797.2020.1844471
- Blumberg, Rae (2007). *Gender bias in textbooks: A hidden obstacle on the road to gender equality in education*. París: Unesco.
- Cameron, Deborah (1992). *Feminism and linguistic theory*. Londres: Palgrave Macmillan.
- Castillo Cuadra, Rodrigo (2020). El pensamiento crítico como competencia básica. Una propuesta de nuevos estándares pedagógicos. *Revista Latinoamericana de Filosofía de la Educación*, 7(14), 127–148.
- Chisholm, Linda (2018). Representations of class, race, and gender in textbooks. En Eckhardt Fuchs & Annekatriin Bock (Eds.), *The Palgrave handbook of textbook studies* (pp. 225–237). Londres: Palgrave Macmillan.
- Chover Lafarga, Ana (2007). El canon ‘otro’ de la literatura hispánica. Sobre la importancia de la perspectiva de género en la enseñanza de E/LE. En Jorge Martí Contreras (Coord.), *Actas del I Congreso Internacional de Lengua, Literatura y Cultura Española: la didáctica de la enseñanza para extranjeros* (pp. 161–168). Onda: Jorge Martí Contreras.
- Consejo de Europa (2002). *Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza y evaluación*. Madrid: Anaya.
- de Miguel, Ana (2004). La situación de las mujeres en el espacio público. En Adela García (Ed.), *Género y ciudadanía: un debate* (pp. 19–40). Barcelona: Icaria.

- Fairclough, Norman (1989). *Language and power*. Londres: Longman.
- Fellbaum, Christiane (Ed.) (1998). *WordNet: An electronic lexical database*. Cambridge: MIT Press.
- Fernández Martínez, Dolores (2012). Critical learning: Critical discourse analysis in EFL teaching. *Journal of Language Teaching and Research*, 3(2), 283–288. doi: 10.4304/jltr.3.2.283–288
- Fesler, Lily; Dee, Thomas; Baker, Rachel, & Evans, Brent (2019). Text as data methods for education research. *Journal of Research on Educational Effectiveness*, 12(4), 707–727. doi: 10.1080/19345747.2019.1634168
- Filipović, Jelena, & Kuzmanović, Ana (2020). Critical approach to intercultural education and foreign language learning and teaching: Neoliberal gender ideologies in textbooks of Spanish as a foreign language. *Porta Linguarum*, 34, 55–74. doi: 10.30827/portalin.v0i34.16733
- Fuertes Gutiérrez, Mara; Soler Montes, Carlos, & Klee, Carol (2021). Sociolingüística aplicada a la enseñanza del español. *Journal of Spanish Language Teaching*, 8(2), 105–113. doi: 10.1080/23247797.2021.2019448
- Galiano Sierra, Isabel María (1993). La mujer en los manuales de español para extranjeros. En Salvador Montesa y Antonio Garrido (Eds.), *Actas del Tercer Congreso Nacional de Asele. El español como lengua extranjera: De la teoría al aula* (pp. 119–136). Málaga: Asele.
- Guerrero, Susana (2001). Medios de comunicación y estrategias lingüísticas no sexistas. *Isla de Arriarán*, 17, 406–417.
- Guerrero, Susana (2003). Propuesta no sexista para favorecer la interculturalidad en ELE. En Manuel Pérez Gutiérrez & José Coloma Maestre (Eds.), *Actas del XIII Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera. El español, lengua del mestizaje y la interculturalidad* (pp. 393–403). Murcia: Asele.
- Han, SuHun (2020). *Googletrans: Free and unlimited Google translate API for Python* (Ver. 3.0.0). <https://py-googletrans.readthedocs.io/en/latest/index.html>
- Honnibal, Mathew, & Montani, Inés (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *Sentometrics Research*. <https://sentometrics-research.com/publication/72>

- Hunt, Sally (2015). Representations of Gender and Agency in the *Harry Potter* Series. En Jonathan Crichton (Ed.), *Corpora and Discourse Studies* (pp. 266–284). Londres: Palgrave Macmillan.
- Hunter, Maxwell, & Chick, Kay (2005). Treatment of gender in basal readers. *Reading Research and Instruction*, 44(3), 65–76. doi: 10.1080/1938807050955843
- Instituto Cervantes (2006). *Plan Curricular del Instituto Cervantes: niveles de referencia para el español*. Madrid: Biblioteca Nueva.
- Instituto de la Mujer (2008). *Guía de coeducación: síntesis sobre la educación para la igualdad de oportunidades entre mujeres y hombres*. Madrid: Instituto de la Mujer (Ministerio de Igualdad).
- Kramsch, Claire, & Vinall, Kimberly (2015). The cultural politics of language textbooks in the era of globalization. En Xiao Lan, Curdt-Christian & Csilla Weninger (Eds.), *Language, ideology, and education: The politics of textbooks in language education* (pp. 11–28). Nueva York: Routledge.
- Kuzmanović, Ana (2016). Los roles de género y los manuales de E/LE: ¿Perpetuación de estereotipos o cambio de ideología? En Anđelka Pejović, Ana Jovanović, Ivana Vučina Simović, Mirjana Sekulić, Vladimir Karanović & Aneta Trivić (Eds.), *Estudios hispánicos en la cultura y ciencia serbias* (pp. 339–348). Kragujevac: Universidad de Kragujevac.
- Kuzmanović, Ana (2022). ‘Haz realidad tus sueños’: las nuevas ideologías de género en los manuales de ELE. *Revista Internacional de Lenguas Extranjeras*, 17, 24–35. doi: 10.17345/rile17.3385
- Kuzmanović, Ana, & Đuričić, Milica (2016). ¿Manuales sexistas, aula sexista? El papel de docentes en la promoción de la igualdad de géneros en el aula de ELE. *Colindancias: Revista de la Red de Hispanistas de Europa Central*, 7, 209–224.
- Lacorte, Manuel, & Atienza, Encarna (2019). Dimensiones críticas en la enseñanza del español. En Javier Muñoz Basols, Elisa Gironzetti & Manel Lacorte (Eds.), *The Routledge handbook of Spanish language teaching* (pp. 628–640). Nueva York: Routledge.
- Lakoff, Robin (1973). Language and woman’s place. *Language in Society*, 2(1), 45–79.

- Leeman, Jennifer (2014). Critical approaches to the teaching Spanish as a local–foreign language. En Manuel Lacorte (Ed.), *The Routledge handbook of hispanic applied linguistics* (pp. 275–292). Nueva York: Routledge.
- Lerner, Gerda (1986). *The creation of patriarchy*. Nueva York: Oxford University Press.
- Lipman, Mathew (2014). *Pensamiento complejo y educación*. Madrid: Ediciones de la Torre.
- López Navajas, Ana (2014). Análisis de la ausencia de las mujeres en los manuales de la ESO: una genealogía de conocimiento ocultada. *Revista de Educación*, 363, 282–308.
- Lucy, Li; Demszky, Dorottya; Bromley, Patricia, & Jurafsky, Dan (2020). Content analysis of textbooks via natural language processing: Findings on gender, race, and ethnicity in Texas U.S. history textbooks. *AERA Open*, 6(3), 1–27. doi: 10.1177/2332858420940312
- Martínez Lirola, María (2022). *El análisis crítico del discurso y la pedagogía crítica: explorando sus relaciones y sus aplicaciones didácticas*. Granada: Comares.
- McEney, Tony, & Hardie, Andrew (2011). *Corpus linguistics*. Cambridge: Cambridge University Press.
- Meurers, Detmar, & Dickinson, Markus (2017). Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning*, 67(1), 66–95. doi: 10.1111/lang.12233
- Molina Gil, Raúl (2022). Las mujeres escritoras en los cursos de literatura de ELE: análisis de los manuales de Edelsa y Edinumen. *Lenguaje y Textos*, 55, 67–81. doi: 10.4995/lyt.2022.16485
- Morales Vidal, Elena, & Cassany, Daniel (2020). El mundo según los libros de texto: análisis crítico del discurso aplicado a materiales de español LE/L2. *Journal of Spanish Language Teaching*, 7(1), 1–19. doi: 0.1080/23247797.2020.1790161
- Muto-Humphrey, Keiko (2005). Gender balance in EFL textbooks: Graded readers. *Journal of School of Foreign Languages*, (35), 75–92. <https://core.ac.uk/download/pdf/235011991.pdf>

- Nogué, Neus (2013). Sexe, gènere i gramàtica. En Maria del Carme Junyent Figueras (Ed.), *Visibilitzar o marcar. Repensar el gènere en la llengua catalana* (pp. 47–66). Barcelona: Empúries.
- Nussbaum, Martha (2010). *Not for profit: Why democracy needs the humanities*. Princeton: Princeton University Press.
- Peñalver Pérez, Rosa (2003). *¿Qué quieres enseñar? Un libro sexista oculta la mitad*. Murcia: Instituto de la Mujer.
- Queiro Pérez, María Jesús (2019). *Análisis de imágenes de tres manuales de español para extranjeros bajo el enfoque de género* (Tesis de maestría inédita). Universidad de Barcelona, Barcelona. <https://diposit.ub.edu/dspace/handle/2445/135517>
- Ramírez, Guadalupe (2020). ¿Hablamos queer? La representación de la diversidad sexual en los manuales de ELE más vendidos en el cantón de Berna, Suiza. En Pilar Taboada de Zúñiga Romero & María del Rocío Barros Lorenzo (Eds.), *Perfiles, factores y contextos en la enseñanza y el aprendizaje de ELE/EL2* (pp. 903–916). Santiago de Compostela: Universidad de Santiago de Compostela.
- Robles Fernández, María Goretty (2005). *La mujer en los manuales de español lengua extranjera: del estereotipo al personaje real e histórico. Mujeres insignes, mujeres comunes, mujeres sin más* (Tesis de maestría inédita). Universidad Nacional de Educación a Distancia, Ciudad de México.
- Salido Machado, Elena (2017). Género e interculturalidad en la enseñanza del español como segunda lengua. En Marian Blanco Ruiz & Clara Sainz de Baranda Andújar (Eds.), *Investigación joven con perspectiva de género II* (pp. 285–300). Madrid: Universidad Carlos III de Madrid.
- Taboada, María Beatriz; Cejas, María Agustina; Sánchez, Claudia Carolina D.; Spada, María Fernanda; Taboada, Miriam Noemí; Caire, María Lucrecia; Olivera, Mariana Inés A.; Pais, Laura Gisela; Ábrigo, Carolina Araceli, & Martín, Ana Paula (2021). Una mirada crítica a la construcción de la mujer y la familia en materiales didácticos para ELE. En Carlos Valentini (Ed.), *Actas del IV Congreso Internacional SICELE: desafíos teóricos y nuevas perspectivas aplicadas*. Madrid: Instituto Cervantes. https://cvc.cervantes.es/ensenanza/biblioteca_ele/sicele/sicele04/indice.htm

- Trujillo, Juan Antonio (2009). Con todos: Using learning communities to promote intellectual and social engagement in the Spanish curriculum. En Manel Lacorte & Jennifer Leeman (Eds.), *Español en Estados Unidos y otros contextos de contacto: sociolingüística, ideología y pedagogía* (pp. 369–395). Madrid: Vervuert/Iberoamericana.
- UN Women (1996). *Informe de la IV Conferencia Mundial sobre la Mujer*. Nueva York: Naciones Unidas.
- Vaíllo Rodríguez, María (2016). La investigación sobre libros de texto desde la perspectiva de género: ¿hacia la renovación de los materiales didácticos? *Tendencias Pedagógicas*, 27, 97–124.
- Van Dijk, Teun A. (2016). Análisis crítico del discurso. *Revista Austral de Ciencias Sociales*, 30, 203–222. doi: 10.4206/rev.austral.cienc.soc.2016.n30-10

9. Anexo

Lista de lecturas graduadas analizadas

Serie: Lola Lago

Autoras: Lourdes Miquel y Neus Sans

Vacaciones al sol

Una nota falsa

Poderoso caballero

Por amor al arte

La llamada de La Habana

Lejos de casa

¿Eres tú, María?

Amor en línea

Sin noticias

Serie 24 horas en español. Cultura, convivencia, lengua, herencia.

Autores: Ernesto Rodríguez y Diana Chery

24 horas en español. Nueva York

24 horas en español. Miami

Serie: Un día en... Un día, una ciudad, una historia

Autor: Ernesto Rodríguez

Un día en Bogotá

Un día en Valencia

Un día en Sevilla

Un día en Barcelona

Un día en Madrid

Un día en Málaga

Un día en Salamanca

Un día en Ciudad de México

Un día en Buenos Aires

Un día en La Habana

Serie: Grandes personajes

Autoras: Aroa Moreno Durán (Frida Kahlo y Lorca), Laura Corpa (Dalí y Picasso), Daniel Cabrera (Che), Cecilia Bembibre (García Márquez)

Picasso. Las mujeres de un genio

Dalí. El pintor de sueños

García Márquez. Una realidad mágica

Che. Geografías del Che

Frida Kahlo. Viva la vida

Lorca. La valiente alegría

Serie: Aventura joven

Autores: Elvira Sancho y Jordi Surís

Aventura en La Habana

¿Dónde está Emiliano Fuentes?

Misterio en las Alpujarras

Objetivo: Barcelona

Perdidos en el Camino del Inca

Persecución en Madrid

El fantasma del instituto

El monstruo del rock

La chica de Mar del Plata

Trimestre maldito

Serie: Pepa Villa, taxista en Barcelona

Autoras: Alicia Estopiñá y Neus Sans

Fantasma en la escalera

Apartamento en la Costa Brava

Cenizas calientes

Una operación arriesgada

Serie: Hacerse mayor

Autora: Lourdes Miquel

Cosas del amor

Amor por sorpresa

Serie: El Mediterráneo

Autor: Jordí Surís

La chica del tren

El secreto de las flores

El puesto de fruta y otros relatos

Autor: Leo Thoma (Traductoras: María José Durán y Sonia Calvo)

El Camino de las Estrellas. Un viaje por el Camino de Santiago.

Autoras: Clara Villanueva y Josefina Fernández



Notas

Productividad relativa de los esquemas de formación de palabras en el español de México. Actualización 2022

Relative productivity of word formation schemes in Mexican Spanish. Update 2022

Ramón Zacarías Ponce de León
Universidad Nacional Autónoma
de México, Instituto de
Investigaciones Filológicas,
Centro de Lingüística Hispánica
rzacaria@hotmail.com



Recepción: 22 de junio del 2023
Aceptación: 19 de septiembre del 2023
doi: 10.22201/enallt.01852647p.2024.78.1074

Resumen

En este trabajo se presentan estadísticas sobre productividad de la morfología léxica en el español de México. Los datos se obtuvieron de la base neológica Morfolex y los cálculos se llevaron a cabo con el acumulado de neologismos a junio del 2022, el cual asciende a 21 605 palabras. A partir del corpus de neologismos se realizaron conteos de tipos, ocurrencias y *hápx legomena* para cada uno de los esquemas de formación de palabras representados en el corpus. La metodología seguida para estos cálculos es la utilizada en Morfolex desde 2010, presentada detalladamente en un trabajo anterior (Zacarías Ponce de León, 2016). En dicha investigación se mostraron los hallazgos correspondientes a diez mil neologismos. Con la actualización que se da a conocer ahora se reafirma que la base de datos Morfolex es representativa del léxico mexicano y ofrece un panorama de los procesos lexicogénicos más productivos en esta variedad del español. Por último, se hace una comparación entre los datos actuales y los correspondientes a 2016 para constatar la evolución que han experimentado los distintos procesos formativos de léxico.

Palabras clave: productividad; neologismos; Morfolex; *hápx legomena*; morfología léxica

Abstract

This paper presents statistics on the productivity of lexical morphology in Mexican Spanish. The data were obtained from the Morfolex neologism database, and the calculations were carried out with the accumulated neologisms to June 2022, which amounted to 21,605 words. Counts of types, frequency of occurrence, and *hápx legomena* were calculated from the corpus of neologisms for each word formation scheme contained in it. These calculations followed the methodology used in Morfolex since 2010 and detailed in a previous work (Zacarías Ponce de León, 2016), which reported the findings corresponding to ten thousand neologisms. The update presented here confirms that the Morfolex database is representative of the Mexican lexicon and presents an overview of the most productive word-formation processes in this variety of Spanish. At the end of the work, a comparison is made between the current data and those corresponding to 2016 to explore the evolution of different lexical schemes..

Keywords: Productivity; neologisms; Morfolex; *hápx legomena*; lexical morphology

1. Introducción

La morfología léxica del español mexicano ha sido motivo de estudio del proyecto Morfolex desde 2008 (Zacarías Ponce de León, 2013). El principal producto de este proyecto es una base de datos que almacena palabras neológicas e incluye el análisis morfológico de su estructura. Esta base de datos permite agrupar las palabras de acuerdo con el proceso lexicogénico que las formó. Los pormenores de esta base de datos, tanto sus fundamentos como sus procedimientos, pueden consultarse en un trabajo anterior (Zacarías Ponce de León, 2016).¹ El objetivo de este artículo no es presentar la base de datos Morfolex, sino una actualización de la productividad morfológica para un total de 21 605 palabras.

Las palabras que aparecen registradas una sola vez en el corpus son llamadas *hápax legomena*. Este tipo de palabras tienen más probabilidad de haber sido creadas por un proceso productivo que aquellas que aparecen en más ocasiones. Por eso, en Morfolex se utilizan para el cálculo de la productividad relativa. En la siguiente sección se explicará brevemente el tipo de datos que se obtienen del corpus y el cálculo de productividad basado en el conteo de *hápax legomena*.

2. Metodología

Esta metodología es homogénea, verificable y replicable. Es homogénea en cuanto a la recolección de neologismos, el análisis morfológico y el etiquetado en la base de datos. Los cálculos de productividad, por su parte, son verificables y simples: el conteo de *hápax legomena* en la base de datos. Finalmente, la metodología es replicable porque puede reproducirse puntualmente en cualquier investigación similar. En publicaciones anteriores de Morfolex esta metodología es descrita con detalle (Zacarías Ponce de León, 2016;

¹ Para consultar las características del proyecto, su metodología y productos, remitimos a los interesados al artículo mencionado.

Lucio Tejeda & Zacarías Ponce de León, 2017). Asimismo, en el sitio morfolex.com puede encontrarse una descripción del proyecto, datos y estadísticas del corpus y publicaciones relacionadas.

2.1. *Medición de la productividad*

A fin de contextualizar los datos que se presentan en esta investigación, haremos una breve exposición sobre los fundamentos del cálculo de la productividad mediante el uso de neologismos.

Según lo que señala Vallès Botey (2002), la productividad relativa de los esquemas lexicogénicos de una lengua debe ser medida en el uso real de los hablantes. En tal sentido, hablar de productividad en la morfología léxica debe referirse siempre a palabras de nuevo cuño o, en palabras de esta investigadora, “debe basarse fundamentalmente en neologismos” (2002: 140). Esto significa que un esquema de formación de palabras será productivo si los hablantes lo utilizan constantemente en la creación de nuevo léxico. Vallès Botey afirma que: “ceñirse a los neologismos para medir la productividad actual es coherente desde el punto de vista conceptual [ya que] coincide con la noción básica de productividad más generalizada [...] la habilidad de los hablantes para producir nuevas palabras” (2002: 149).

El otro aspecto importante de la metodología es el uso de palabras que ocurren solamente una ocasión en el corpus; los llamados *hápax legomena*. El método estadístico basado en *hápax* es bastante conocido (Baayen, 1992; Aronoff & Fudeman, 2011). La importancia de estas piezas léxicas es que, evidentemente, no pertenecen a un léxico consolidado y, por lo tanto, su existencia es evidencia de que el esquema lexicogénico con el que se formaron es rendidor y, en función del número de nuevas palabras formadas, puede ser considerado productivo.

3. Productividad relativa de los esquemas lexicogénicos

En el Cuadro 1 se presenta la productividad morfológica del español mexicano actualizada a junio de 2022. Se indican los afijos más relevantes en esta variedad de lengua. Se muestran también los esquemas compositivos más productivos, tanto de la composición patrimonial como de la composición culta; así como otros procesos misceláneos. Los conteos se realizaron por Tipos, Ocurrencias y *Hápax legomena*. Esta última columna es la que se utiliza para ordenar los esquemas lexicogénicos según su productividad.

CUADRO 1. Productividad relativa con datos de junio 2022

Proceso morfológico	Ocurrencias		Tipos		<i>Hápax legomena</i>	
	21605		12803		8968	
	Total	%	Total	%	Total	%
Sufijación nombre	6516	30.2	3018	23.6	2023	22.558
-ista	848	3.9	341	2.7	208	2.319
-ero	1094	5.1	356	2.8	207	2.308
-azo	851	3.9	304	2.4	194	2.163
-ear	653	3.0	273	2.1	162	1.806
-ismo	478	2.2	245	1.9	154	1.717
-ense	179	0.8	126	1.0	98	1.093
-izar	256	1.2	127	1.0	91	1.015
-ano	128	0.6	102	0.8	86	0.959
-gate	102	0.5	70	0.5	58	0.647
-ar	189	0.9	82	0.6	57	0.636
-ico	114	0.5	75	0.6	53	0.591
-ería	163	0.8	80	0.6	52	0.580
-esco	76	0.4	58	0.5	50	0.558
Sufijación adjetivo	1136	5.3	679	5.3	491	5.475
-mente	114	0.5	103	0.8	92	1.026
-dad	142	0.7	111	0.9	84	0.937

(continuación)

CUADRO 1. Productividad relativa con datos de junio 2022

Proceso morfológico	Ocurrencias		Tipos		<i>Hápax legomena</i>	
	21605		12803		8968	
	Total	%	Total	%	Total	%
<i>-ismo</i>	216	1.0	84	0.7	48	0.535
<i>-izar</i>	158	0.7	67	0.5	44	0.491
<i>-ista</i>	116	0.5	50	0.4	31	0.346
<i>-ez</i>	76	0.4	40	0.3	29	0.323
<i>-ear</i>	60	0.3	38	0.3	23	0.256
<i>-ada</i>	18	0.1	16	0.1	12	0.134
<i>-ísimo</i>	10	0.05	10	0.1	10	0.112
Sufijación verbo	2613	12.1	1212	9.5	836	9.322
<i>-ción</i>	469	2.2	224	1.7	154	1.717
<i>-dor</i>	376	1.7	196	1.5	143	1.595
<i>-nte</i>	229	1.1	146	1.1	111	1.238
<i>-o</i>	151	0.7	120	0.9	77	0.859
<i>-miento</i>	189	0.9	99	0.8	65	0.725
<i>-ble</i>	122	0.6	83	0.6	59	0.658
<i>-ada</i>	121	0.6	64	0.5	48	0.535
<i>-e</i>	127	0.6	41	0.3	27	0.301
<i>-dero</i>	60	0.3	36	0.3	26	0.290
<i>-torio</i>	65	0.3	25	0.2	18	0.201
Prefijación nombre	2573	11.9	1651	12.9	1117	12.455
<i>anti-</i>	496	2.3	309	2.4	217	2.420
<i>narco-</i>	452	2.1	225	1.8	154	1.717
<i>na-</i>	113	0.5	92	0.7	75	0.836
<i>ex-</i>	154	0.7	103	0.8	70	0.781
<i>re-</i>	171	0.8	95	0.7	57	0.636
<i>des-</i>	89	0.4	59	0.5	43	0.479

(cont.)

CUADRO 1. Productividad relativa con datos de junio 2022

Proceso morfológico	Ocurrencias		Tipos		Hápax legomena	
	21605		12803		8968	
	Total	%	Total	%	Total	%
<i>pro-</i>	45	0.2	42	0.3	38	0.424
<i>pre-</i>	108	0.5	55	0.4	35	0.390
<i>sub-</i>	113	0.5	55	0.4	33	0.368
Prefijación adjetivo	896	4.147	645	5.038	450	5.018
<i>anti-</i>	187	0.9	128	1.0	86	0.959
<i>in-</i>	76	0.4	48	0.4	32	0.357
<i>pro-</i>	29	0.1	25	0.2	20	0.223
<i>des-</i>	23	0.1	21	0.2	18	0.201
<i>inter-</i>	29	0.1	23	0.2	18	0.201
<i>pre-</i>	41	0.2	26	0.2	18	0.201
<i>pos-</i>	36	0.2	22	0.2	16	0.178
<i>ex-</i>	28	0.1	21	0.2	15	0.167
<i>extra-</i>	25	0.1	18	0.1	14	0.156
Prefijación verbo	524	2.425	349	2.726	251	2.799
<i>re-</i>	216	1.0	123	1.0	84	0.937
<i>des-</i>	122	0.6	81	0.6	59	0.658
<i>pre-</i>	32	0.1	20	0.2	11	0.123
<i>sobre-</i>	14	0.1	13	0.1	11	0.123
<i>co-</i>	14	0.1	11	0.1	9	0.100
<i>in-</i>	7	0.03	6	0.05	5	0.056
<i>con-</i>	2	0.01	2	0.02	2	0.022
Parasíntesis	308	1.4	151	1.2	107	1.193
<i>en-</i> -ar	176	0.8	70	0.5	46	0.513
<i>a-</i> -ar	72	0.3	37	0.3	31	0.346
<i>des-</i> -ar	30	0.1	21	0.2	14	0.156

(cont.)

CUADRO 1. Productividad relativa con datos de junio 2022

Proceso morfológico	Ocurrencias		Tipos		<i>Hápax legomena</i>	
	21605		12803		8968	
	Total	%	Total	%	Total	%
<i>des-izar</i>	20	0.1	14	0.1	8	0.089
<i>en-ecer</i>	6	0.03	6	0.05	6	0.067
<i>en-izar</i>	1	0.005	1	0.01	1	0.011
<i>a-izar</i>	3	0.01	2	0.02	1	0.011
Composición culta	2503	11.6	2015	15.7	1322	14.741
<i>mega</i>	289	1.3	200	1.6	131	1.461
<i>super</i>	206	1.0	177	1.4	118	1.316
<i>multi</i>	160	0.7	122	1.0	65	0.725
<i>mini</i>	88	0.4	77	0.6	63	0.702
<i>micro</i>	129	0.6	101	0.8	58	0.647
<i>bio</i>	111	0.5	91	0.7	57	0.636
<i>ciber</i>	181	0.8	119	0.9	55	0.613
<i>semi</i>	84	0.4	77	0.6	51	0.569
<i>logo</i>	81	0.4	60	0.5	40	0.446
<i>metro</i>	63	0.3	45	0.4	34	0.379
Composición patrimonial	2510	11.6	1559	12.18	1187	13.236
N+N	830	3.8	632	4.9	489	5.453
V+N	947	4.4	472	3.7	366	4.081
N+Adj	183	0.8	129	1.0	102	1.137
Adj+Adj	148	0.7	116	0.9	90	1.004
N+de+N	63	0.3	46	0.4	36	0.401
Adj+N	77	0.4	56	0.4	34	0.379
Adj+i+Adj	81	0.4	31	0.2	17	0.190
Prep+N	64	0.3	21	0.2	14	0.156

(cont.)

CUADRO 1. Productividad relativa con datos de junio 2022

Proceso morfológico	Ocurrencias		Tipos		Hápax legomena	
	Total	%	Total	%	Total	%
	21605		12803		8968	
V+Adv	17	0.1	14	0.1	13	0.145
N++Adj	21	0.1	16	0.1	12	0.134
Traslapamiento*	954	4.4	781	6.1	611	6.813
Acortamiento	174	0.8	116	0.9	84	0.937
Acortamiento compuesto	747	3.5	486	3.8	363	4.048
Acronimia y siglación	143	0.7	133	1.0	120	1.338

* También conocidos como cruces o *blends*. Se refiere a palabras del tipo burrocracia, cenicebrijo, tuitaturata, suaperro, etcétera.

El cuadro anterior representa la productividad relativa de los procesos de formación léxica más relevantes en el español de México. Se desglosan los procesos afijales distinguiendo entre sufijos, prefijos y circunfijos. Cada una de estas agrupaciones se desglosa a su vez por afijo. Solo se presentan los afijos más productivos. De la misma manera, se desglosan los diferentes tipos de composición y también otros esquemas formativos menos relevantes.

3.1. Estadísticas por esquema lexicogenésico

Se presenta ahora los resúmenes de los esquemas lexicogenésicos más importantes, sufijación, prefijación y composición, con el objetivo de ofrecer una visión más general. En cuanto a la sufijación (Cuadro 2), la más productiva, y de manera considerable, es la de nombre; le sigue la de verbo y por último la de adjetivo. El panorama cambia con la prefijación (Cuadro 3). De nuevo, es la de sustantivo la más productiva, pero la de adjetivo se coloca en segundo

lugar, dejando la prefijación de verbo como menos productiva. En general, observamos que la sufijación total (37.3%) es mucho más productiva que la prefijación total (20.3%). Respecto a la composición patrimonial (13.2%) y a la composición culta (14.7%), como era de esperarse, se colocan en un nivel inferior al de la prefijación y al de la sufijación (Cuadro 4). Destacan en este caso, los dos tipos de composición que se utilizan más en español, el N+N y el de V+N, con niveles semejantes de productividad.

CUADRO 2. Productividad de la sufijación

	Ocurrencias		Tipos		<i>Hápax legomena</i>	
	Total	%	Total	%	Total	%
Sufijación nombre	6516	30.2	3018	23.6	2023	22.558
Sufijación adjetivo	1136	5.3	679	5.3	491	5.475
Sufijación verbo	2613	12.1	1212	9.5	836	9.322
Totales	10 265	47.5	4 909	38.3	3 350	37.3

CUADRO 3. Productividad de la prefijación

	Ocurrencias		Tipos		<i>Hápax legomena</i>	
	Total	%	Total	%	Total	%
Prefijación nombre	2573	11.9	1651	12.9	1117	12.455
Prefijación adjetivo	896	4.147	645	5.038	450	5.018
Prefijación verbo	524	2.425	349	2.726	251	2.799
Totales	3 993	18.5	2 645	20.6	1 818	20.3

CUADRO 4. Productividad de la composición patrimonial

	Ocurrencias		Tipos		<i>Hápax legomena</i>	
	Total	%	Total	%	Total	%
N+N	830	3.8	632	4.9	489	5.453
V+N	947	4.4	472	3.7	366	4.081
N+Adj	183	0.8	129	1.0	102	1.137
Adj+Adj	148	0.7	116	0.9	90	1.004
N+de+N	63	0.3	46	0.4	36	0.401
Composición	2510	11.6	1559	12.18	1187	13.24

En el Cuadro 5 se comparan todos los procesos de formación de palabras registrados en Morfolex; se incluye la composición culta, los cruces léxicos, la acronimia, los acortamientos y las siglas. Se presentan primero los casos de afijación, posteriormente los de composición, luego los de cruce léxico y, por último, el resto de los procesos.

CUADRO 5. Productividad global por esquema lexiconesic

	Ocurrencias		Tipos		<i>Hápax legomena</i>	
	Total	%	Total	%	Total	%
Sufijación	10265	47.5	4909	38.3	3350	37.3
Prefijación	3993	18.5	2645	20.6	1818	20.3
Circunfijación	308	1.4	151	1.2	107	1.2
Composición patrimonial	2510	11.6	1559	12.18	1187	13.2
Composición culta	2503	11.6	2015	15.7	1322	14.8
Traslapamiento	954	4.4	781	6.1	611	6.8
Acortamiento	174	0.8	116	0.9	84	0.94
Acortamiento compuesto	747	3.5	486	3.8	363	4.04
Acrónimos y siglas	143	0.7	133	1.0	120	1.34

El panorama de la morfología léxica puede apreciarse mejor de manera gráfica. En la Figura 1 se incluyen todos los procesos. Para mayor claridad, solo se grafica la columna correspondiente a la productividad calculada mediante las palabras *hápax legomena*.

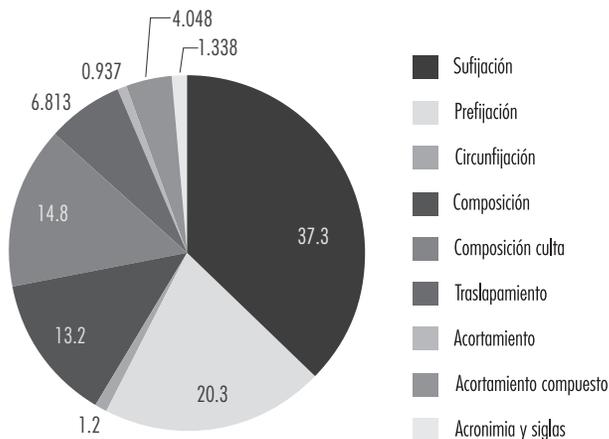


FIGURA 1. Productividad global por esquema lexicogenésico

4. Comparación de productividad 2016–2022

Ahora se presenta la comparación entre las estadísticas del 2022 y las estadísticas previas, que corresponden al 2016 (Cuadro 6). Se comparan los procesos generales a nivel de *hápax legomena*, es decir, el grado de productividad relativa.

CUADRO 6. Comparativo de productividad 2016–2022

	<i>Hápax legomena</i> 2016 (4618)		<i>Hápax legomena</i> 2022 (8968)	
	Total	%	Total	%
Sufijación	1738	37.6	3350	37.3
Prefijación	1198	25.9	1818	20.3
Circunfijación	67	1.45	107	1.2

(continuación)

CUADRO 6. Comparativo de productividad 2016–2022

	<i>Hápax legomena</i> 2016 (4618)		<i>Hápax legomena</i> 2022 (8968)	
	Total	%	Total	%
Composición	1047	22.7	2509	28.0
Traslapamiento	263	5.7	611	6.8
Acartamiento	50	1.1	84	0.94
Acartamiento compuesto	167	3.6	363	4.04
Acónimos y siglas	88	2.0	120	1.34

Al analizar el cuadro anterior se comprueba que los principales procesos de formación de palabras, la afijación y la composición, se mantienen como los más productivos. Sin embargo, se nota un cambio relevante: la composición en general, sumando la patrimonial y la culta, se ha robustecido, al pasar de un grado de productividad de 22.7 a 28. Esta alza en la productividad de la composición trae como consecuencia una disminución en la productividad relativa en el resto de los procesos. Esto es particularmente cierto para la prefijación, que pasó de ser el segundo proceso en orden de productividad a ser el tercero, cediendo su lugar a la composición. Una posible explicación al aumento de productividad de la composición es el contexto de la pandemia por COVID-19, ya que este acontecimiento generó gran cantidad de neologismos (Zacarías Ponce de León, 2022), muchos de ellos formados con los sustantivos *COVID* (*coviflu*, *covicumpleaños*) y *corona* (*coronabicho*, *coronabobo*). En cuanto al proceso más productivo de formación de palabras, la sufijación se mantiene claramente en el primer lugar, con una pequeña disminución en la estadística reciente.

5. Observaciones finales

Las estadísticas presentadas en esta nota nos proporcionan un panorama de la morfología léxica del español de México a través del estudio de la productividad relativa de los esquemas de formación de palabras. El conteo de palabras *hápx legomena* ha sido un parámetro importante en la configuración del corpus Morfolex, ya que mediante este es viable establecer un grado de productividad relativa. Uno de los objetivos de este proyecto ha sido la difusión de los cálculos de productividad relativa de nuestra variedad de lengua. La metodología del proyecto para el cálculo de la productividad (Zacarías Ponce de León, 2016), a partir de una base de datos neológica, es la única hasta el momento que ofrece la productividad de todos los esquemas léxicos de manera relativa. Aunado a esto, el acopio constante de datos permite comparar dos momentos diferentes en el corpus para observar los cambios en productividad. Esto es justamente lo que se ha presentado en este trabajo, la comparación entre los datos de 2016 con los de 2022. Consideramos que, como resultado de dicha comparación, es posible afirmar que la base de datos Morfolex ofrece un retrato adecuado del comportamiento de la morfología léxica del español mexicano.

6. Referencias

- Aronoff, Mark, & Fudeman, Kirsten (2011). *What is morphology?* Malden: Wiley-Blackwell.
- Baayen, R. Harald (1992). Quantitative aspects of morphological productivity. En Geert Booij & Jaap van Marle (Eds.), *Yearbook of morphology 1991* (pp. 109–149). Dordrecht: Kluwer Academic.
- Lucio Tejeda, Sofía, & Zacarías Ponce de León, Ramón (2017). Actualización del corpus Morfolex a partir de la 23ª edición del Diccionario de la Lengua Española. *Estudios de Lingüística Aplicada*, 66, 109–134.
- Vallès Botey, Teresa (2002). La productividad morfológica en un modelo dinámico, basado en el uso y en los usuarios. En M. Teresa Cabré, Judit Freixa & Elisabet Solé (Eds.), *Lèxic i neologia* (pp. 139–157). Bar-

celona: Observatori de Neologia, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Zacarías Ponce de León, Ramón (2013). El buzón buscapalabras. Procesos de formación de neologismos. *Anuario de Letras Hispánicas. Glosas Hispánicas* (Vol. 2, pp. 81–89). Ciudad de México: Facultad de Filosofía y Letras, Universidad Nacional Autónoma de México.

Zacarías Ponce de León, Ramón (2016). Morfología léxica en el español actual de México: neología y productividad. *Estudios de Lingüística Aplicada*, 64, 11–31.

Zacarías Ponce de León, Ramón (2022). Neologismos relacionados con la pandemia de COVID-19 en México. *Lingüística Mexicana. Nueva Época*, 4(1), 7–29.



Diseño de tareas y materiales para recopilar un corpus de aprendices mexicanos de lengua inglesa

Design of tasks and materials to collect
a corpus of Mexican learners of English

Ana Abigahil Flores Hernández
Universidad Autónoma del Estado de
México, Facultad de Lenguas
aafloresh@uaemex.mx

Pauline Moore
Universidad Autónoma del Estado de
México, Facultad de Lenguas
pmooreh@uaemex.mx



Recepción: 26 de abril del 2023
Aceptación: 22 de agosto del 2023
doi: 10.22201/enallt.01852647p.2024.78.1060

Resumen

Este trabajo describe el proceso de diseño de una entrevista oral para recopilar datos de producción lingüística de aprendices de inglés como segunda lengua en universidades públicas mexicanas. El objetivo de la aplicación de este instrumento es crear un corpus de aprendices variado y representativo que sirva como base para investigaciones en adquisición de segunda lengua, así como para el diseño de materiales y estrategias de enseñanza acordes a las necesidades de los estudiantes mexicanos. Para ello se realizó una rigurosa selección de tareas y diseño de materiales con base en dos conceptos clave para el diseño de un corpus de aprendices: la autenticidad y la representatividad. El resultado es una entrevista de cuatro secciones alineadas a las actividades productivas orales monológicas de acuerdo con el Marco Común Europeo de Referencia para las lenguas (monólogo descriptivo, informativo y argumentativo). Las tareas además representan tres de los géneros textuales de acuerdo con el modelo de análisis multidimensional para textos orales: textos informativos, narrativos y de posicionamiento.

Palabras clave: lingüística de corpus; corpus de aprendices; adquisición de L2; enseñanza de L2; lenguas extranjeras

Abstract

This paper describes the design process of an oral interview to be used in the collection of language-production data of university learners of English as a second language. The aim of applying this instrument is to build a relatively large and representative corpus of learners to be used in research about second-language acquisition in Mexico and for the creation of English language teaching strategies and programs consistent with the particular characteristics of Mexican students. To this end, we carried out rigorous task selection and materials design during this instrument design process to ensure two key aspects in the construction of the learner corpus: authenticity and representativeness. This process resulted in a four-part oral interview aligned to the monologic spoken productive activities according to the Common European Framework of Reference for Languages, namely descriptive, informative, and argumentative. These tasks also represent three text types according to the multidimensional analysis for oral texts, information-oriented, stance-oriented, and narrative.

Keywords: corpus linguistics; learner corpora; second language acquisition; language teaching; foreign languages

1. Introducción*

La lingüística de corpus se ha convertido en una fuente de evidencia para complementar datos obtenidos con técnicas de elicitación e introspección. Esta metodología permite visualizar lo que ocurre en el lenguaje y analizarlo desde el paradigma cuantitativo para entender los mecanismos de adquisición de una segunda lengua y mejorar, sobre esta base, la enseñanza de segundas lenguas (Granger, 2002; 2008). Hasta ahora, el diseño de programas y materiales de enseñanza de la lengua inglesa ha partido de un modelo que considera el análisis de necesidades desde dos aspectos: las características de los aprendices (estilos, necesidades, aptitudes, etc.) y las descripciones de la lengua (estructuras, funciones comunicativas, etc.). Ahora bien, los corpus de aprendices permiten agregar un tercer aspecto: la producción de quienes adquieren la segunda lengua. Esta producción brinda evidencia empírica del proceso de adquisición de la lengua y de diversos factores que influyen en él, al mismo tiempo que representa el producto de este proceso y de la influencia de dichos factores (Granger, 2002).

De acuerdo con el listado *Learner corpora around the world* (Centre for English Corpus Linguistics, 2020), que almacena las referencias y sitios web de los corpus de aprendices en el mundo, son pocas las colecciones dedicadas a recopilar la producción de aprendices de inglés que tienen como lengua materna el español (seis corpus en total) y ninguna de ellas se ha dado a la tarea de recopilar producción de aprendices mexicanos. Por otro lado, existen colecciones amplias, recopiladas en su mayor parte por grandes editoriales enfocadas en la enseñanza del inglés, que tienen como finalidad proveer material para obras de referencia y libros de texto para la enseñanza del inglés como lengua extranjera, las cuales contienen un porcentaje mínimo de representación de la variante

* Las autoras desean expresar su agradecimiento al Conahcyt por la beca otorgada a Ana Abigahil Flores Hernández (Convocatoria de Estancias Posdoctorales 2020, 2021 y 2022).

mexicana. Sin embargo, algunas de estas colecciones son de acceso restringido por tratarse de un recurso propiedad del sello editorial, otras dificultan la consulta directa de la producción de los aprendices mexicanos y otras solo incluyen producción escrita.

El corpus de aprendices mexicanos (Mexican Learner Corpus o MexLeC) constituye un proyecto que recolecta de manera longitudinal la producción oral de aprendices mexicanos de inglés como segunda lengua, utilizando una entrevista guiada conformada por cuatro diferentes tareas comunicativas. Los datos obtenidos con este instrumento pretenden conformar un robusto corpus de aprendices que permita, además de las metas inmediatas de monitorear el progreso de la producción oral, desarrollar estrategias y materiales específicos para los participantes y servir como base para un amplio rango de investigaciones sobre adquisición y enseñanza de segundas lenguas.

Para lograr estos objetivos, se ha diseñado un instrumento confiable y válido que permite recopilar suficientes ejemplares de producción lingüística, seleccionando con rigor cada una de las tareas y materiales incluidos para guiar tanto a los participantes como a los aplicadores y lograr una estandarización en la aplicación del instrumento, así como homogeneidad en los datos recabados con base en los conceptos de *autenticidad* y *representatividad*. La descripción amplia y detallada de este proceso de selección y diseño de tareas es el objetivo de este trabajo.

2. Corpus de aprendices: definición y tipología

Un área de auge en la lingüística actual es la investigación del uso de la lengua mediante los corpus electrónicos, que complementa la investigación basada en la intuición y juicios de los hablantes nativos (Granger, 2021). La lingüística de corpus es una metodología que implica el uso de corpus en investigaciones cualitativas o cuantitativas sobre la lengua por medio de su procesamiento y haciendo uso de *software* especializado, previa implementación de una normalización o etiquetado que posibilite el manejo de gran-

des volúmenes de datos y la replicabilidad de las investigaciones (McEnery & Gabriellatos, 2006; Lehmberg & Wörner, 2008; McEnery & Hardie, 2011).

Un corpus es una colección grande de textos auténticos y representativos de una cierta variante de lengua oral o escrita, que se encuentra almacenada en formato electrónico (Baker, Hardie & McEnery, 2006; Lüdeling & Kytö, 2008). Esta misma definición puede aplicarse para un corpus de aprendices, con la diferencia de que los textos recopilados son producidos por aprendientes de una segunda lengua (Granger, 2008). Así, un corpus de aprendices puede definirse como una colección sistemática de lengua continua, contextualizada y auténtica (escrita o hablada) que se almacena en formato electrónico y constituye una clase especial de datos empíricos que pueden usarse en diferentes disciplinas relacionadas con el aprendizaje de segundas lenguas, el diseño de materiales pedagógicos y el desarrollo de herramientas de entrenamiento y procesamiento de lenguaje natural (Granger, 2008; Callies & Paquot, 2015; Meunier, 2021).

Los corpus de aprendices pueden clasificarse según seis dimensiones: el modo, el género, la lengua meta, la lengua materna, el tiempo de recolección, su alcance y su propósito (Sinclair, 2005; Gilquin, 2015). El modo se refiere al contexto de recopilación y almacenaje de los datos, de acuerdo con este criterio un corpus puede ser oral o escrito, presencial o virtual y monológico o dialógico (Sinclair, 2005). Este criterio incluye el tipo de almacenaje, que se refiere a la clase de archivo en el que se guardan los datos. En específico, los corpus pueden ser textuales orales o callados, lo que implica, en el primer caso, el uso de archivos de texto, archivos de audio y video (con sus transcripciones), o bien, en el segundo caso, solo archivos de texto o transcripciones que representan la producción oral.

El género de un corpus hace referencia al contexto de uso de lengua o el propósito de las producciones que conforman la colección. Así, hay corpus de lengua con propósitos específicos, por ejemplo, de negocios o académicos, o cuyos textos representan

géneros discursivos, por ejemplo, cartas formales o informales, ensayos, narrativas, entre otros. Por último, a los corpus que reúnen varios géneros se les denomina corpus de lengua generales.

La lengua meta de un corpus es la utilizada por los participantes durante el proceso de recolección. Los corpus de aprendices se definen según dos lenguas. Por un lado, se considera la L2 que se está aprendiendo, en la que los participantes realizan su producción. Por otro lado, la lengua materna de los aprendices también juega un papel fundamental, debido a que influye en las expresiones y estructuras lingüísticas producidas.

Por el tiempo de recolección, un corpus puede ser sincrónico o diacrónico, es decir, puede ser recolectado en un solo punto en el tiempo (corpus transversal), o puede obtenerse mediante diferentes recolecciones con el fin de representar la evolución de la lengua de los aprendices a través del tiempo (corpus longitudinal). Según su alcance, un corpus puede ser local o global. Es global si contiene datos de muchos sujetos de diversas regiones o países del mundo y es local si los textos son recolectados para representar una población específica, por ejemplo, estudiantes de una universidad o pertenecientes a un curso o programa educativo específico. Además, los corpus pueden clasificarse en torno a su origen y propósito, es decir, el uso que se da a los datos recolectados. En este sentido pueden clasificarse en dos categorías, la primera, con fines de investigación y de acceso libre (o casi libre) para la comunidad científica y la segunda, de uso reservado, cuya finalidad es el diseño de materiales de enseñanza comercializables.

El corpus MexLeC, objeto de este trabajo, es una colección que puede considerarse en cuanto al modo como un corpus oral, virtual, monológico y callado, ya que representa la producción oral de aprendices que son entrevistados en modalidad virtual mediante tareas monológicas que después se transcriben y almacenan en archivos de texto (y no en archivos de audio). La lengua meta del corpus es inglés como segunda lengua y se trata de un corpus de aprendices general, pues no se circunscribe a algún género o variante lingüística específica. Respecto al tiempo de recolección,

constituye un corpus diacrónico, pues plantea un seguimiento longitudinal del desarrollo de la lengua de cada participante durante cuatro o cinco años. En cuanto a su cobertura, se considera de alcance nacional, ya que recopila producciones de estudiantes de diversas universidades en México que se especializan en lenguas modernas. Por último, MexLeC es un corpus de uso libre con fines de investigación o académicos.

Los corpus de aprendices más notables por su tamaño y sus innumerables aplicaciones en la investigación sobre adquisición y enseñanza de segunda lengua son cuatro, dos comerciales y dos académicos. Por un lado, los corpus comerciales incluyen el Cambridge Learner Corpus, que contiene más de 1800 millones de palabras producidas durante las certificaciones de Cambridge ESOL, y el Longman Learners' Corpus, integrado por 10 millones de palabras representativas de aprendices de 20 lenguas maternas distintas (Meyer, 2002). Ambos fueron creados con el objetivo de sustentar materiales comercializables de enseñanza y aprendizaje de inglés, y han sido utilizados con frecuencia en la creación de libros de preparación para certificaciones y diccionarios bilingües para hablantes no nativos.

Por otro lado, entre los corpus académicos se encuentran el International Corpus of Learner English (Granger, Dupont, Meunier, Naets & Paquot, 2020) y el Trinity Lancaster Corpus (Gablasova, Brezina & McEnery, 2019), ambos con un promedio de 4.5 millones de palabras. Estas colecciones tienen como objetivo común conformar una base para la investigación sobre adquisición y enseñanza del inglés como segunda lengua. Estos corpus son de acceso libre y se han empleado en estudios sobre adquisición de L2, por ejemplo, Gilquin (2019), que realiza una exploración sobre el uso de verbos ligeros como *take* o *make* en el Trinity Lancaster Corpus.

3. Utilidad de un corpus de aprendices

En el campo de la enseñanza de lenguas extranjeras, los datos de un corpus de aprendices representan una aportación valiosa, sobre

Corpus spot

Be careful with modal verbs – the Cambridge Learner Corpus shows exam candidates often make mistakes with these.

I **must fill in** an application form for a visa.
NOT I ~~must to fill in~~ an application form for a visa.

I **don't have to show** my passport at the border any more.
NOT I ~~haven't to show~~ my passport at the border any more

Correct the mistakes that the candidates have made with obligation and necessity.

- You needn't much space to park your car.
- Another thing, should I take my camera with me?
- You needn't smoke in this part of the restaurant; it's a no smoking area.
- It is better when you go by car because you must not get up early.
- We have get to the exhibition early or we won't get a ticket.
- You don't have to swim off the rocks because it's dangerous.
- My doctor says I need give up smoking.
- Lisa must to buy a ticket before getting on the bus.
- I don't have to be late or I'll miss my plane.

FIGURA 1. Ejercicio basado en el análisis de errores según un corpus de aprendices (Chapelle & Sharp, 2014)

todo para la creación de materiales pedagógicos y el diseño de programas e instrumentos de evaluación (Granger, 2008). El uso de un corpus auténtico que muestra los errores comunes y los conocimientos adquiridos por los aprendientes permite que los diseñadores puedan enfocarse en necesidades reales de aprendizaje. Entre los resultados de estas aportaciones se encuentran los diccionarios para aprendices, como el *Longman Wordwise Dictionary* o el *Longman Dictionary of Contemporary English*, ambos basados en el Longman Learners' Corpus. El análisis de los corpus de aprendices permite que el lexicógrafo pueda comentar en las entradas léxicas problemas de uso que presentan los aprendices. Por ejemplo, en la entrada para *clothes* del *Longman Wordwise Dictionary* se encuentra la acotación de uso mostrada en (1).

- (1) *Do not use cloth or cloths to mean “the things that people wear”. Instead use clothes, a clothes shop. | The guests all wore casual clothes.*

‘no utilice *cloth* o *cloths* para referirse a lo que la gente viste. En lugar de ello use *clothes, a clothes shop*. | Los invitados usaban ropa casual’

Otras aplicaciones incluyen el diseño de materiales para enseñanza que abordan errores usuales de los aprendices (Granger, 2002). La editorial Cambridge University Press incluye contenidos de práctica en sus libros de preparación para exámenes de certificación derivados de su corpus de aprendices, el Cambridge Learner Corpus. En la Figura 1 se muestra un ejercicio tomado de Capel y Sharp (2014), que presenta los errores más comunes en el uso de los verbos modales *must* y *have to*, según el Cambridge Learner Corpus.

4. Consideraciones en el diseño de un corpus

El diseño de un corpus tiene una relación directa con el objetivo de uso de la colección. En el caso de un corpus general es necesario definir con cuidado el contenido con respecto a las personas que participan y sus características, los materiales, el medio de recolección y el método de transcripción elegido (Hunston, 2008). Meunier (2021) considera que los criterios principales para seleccionar los textos a incluir en un corpus de aprendices son: 1) los aprendices y sus características, su lengua materna y sus niveles de dominio de la lengua meta; 2) los tipos de tareas seleccionadas y su propósito, ya sean orales, escritas, formales, informales, expositivas, narrativas, entre otras, y 3) las condiciones en las que se producen estas tareas: interactivas, en ambientes virtuales o como parte de una evaluación. Estos criterios pretenden orientar las decisiones iniciales del proceso de diseño y obedecen a las características fundamentales que debe cumplir un corpus lingüístico: autenticidad y representatividad (McCarthy & O’Keeffe, 2009), así como practicidad y conveniencia para el investigador (McEneaney & Hardie, 2011).

4.1. Autenticidad

La *autenticidad* en un corpus se refiere a que las muestras de la lengua contenida hayan sido producidas por los hablantes en contextos naturales (McEnery & Hardie, 2011), es decir, que se trate de una comunicación genuina generada en contextos reales y situaciones cotidianas, no artificiales o forzadas (Sinclair, 1996). En el caso de un corpus de aprendices, la producción incluida no podría ser considerada auténtica en el sentido estricto del concepto. El uso de una L2 como lengua extranjera se reduce al ambiente escolar y a situaciones de aprendizaje en el salón de clases —tareas específicas que deben realizar los aprendices—, de tal suerte que la producción lingüística casi siempre es el resultado de una solicitud de un hablante más competente, sea este el maestro o el entrevistador.

A pesar de lo anterior, la lengua incluida en un corpus de aprendices, si bien no se considera del todo auténtica o natural, puede ser vista como “quasi-natural” o “quasi-auténtica” (Meunier, 2021). De acuerdo con Gilquin (2015) y Granger (2008), la lengua producida por aprendices puede clasificarse según diferentes grados de naturalidad, dependiendo de la tarea en que participan los aprendices y el grado de edición o espontaneidad del discurso producido. En este sentido, un corpus oral, a diferencia de un corpus escrito, podría considerarse como una producción más auténtica o más representativa del interlenguaje de los aprendices, debido a que es más espontánea, transcurre en tiempo real y no da tiempo a los procesos de monitoreo o edición que se observan en la producción escrita.

Asimismo, existen niveles de naturalidad que reflejan el tipo de datos producidos en relación con las tareas utilizadas para estimular dicha producción. Así, las tareas “más naturales” son aquellas que se enfocan en la comunicación de significados, en las cuales el aprendiz elige las palabras que usará para expresarse. En estas el participante no es obligado a producir estructuras o vocabulario específicos, ya que generan cierta clase de comunicación cuyas estructuras y vocabulario son impredecibles (Granger, 2008). Wich-

mann (2008) define el grado de naturalidad del discurso producido a partir de una tarea de acuerdo con el grado de restricciones en las estructuras requeridas en la respuesta. Considerando lo anterior, se puede concluir que un tipo de tarea muy poco natural podría ser la lectura en voz alta o la repetición, mientras que las más naturales parten de la solicitud de opiniones o recuentos de experiencias personales.

En resumen, la autenticidad de un corpus de aprendices dependerá de la selección de tareas que privilegien la comunicación y no la generación de estructuras específicas, y que se realicen dentro de un contexto que no permita la preparación previa, el uso de materiales de apoyo o la edición, con el fin de que se produzca un discurso con un mayor grado de espontaneidad que permita el acceso al repertorio léxico real del aprendiz.

4.2. *Representatividad*

La *representatividad* se define a partir de que el contenido de una colección determinada sea una muestra en menor escala de la variedad de lengua que se pretende documentar. Esto a fin de que su análisis y resultados puedan generalizarse para la totalidad de la variedad lingüística representada. Se puede afirmar que un corpus general y robusto es representativo en la medida en que captura la variabilidad de la lengua (Baker *et al.*, 2006; McEnery, Xiao & Tono, 2006). En este caso, la lengua que se desea representar es la producida por aprendices de inglés, la representatividad, entonces, estará determinada por la selección de los bloques de textos incluidos en la colección (corpus), que deberá de haber sido conformada de modo equilibrado.

Según Biber (1993), son dos los criterios para determinar la representatividad: el criterio externo o situacional, definido por el contexto o el tipo de discurso de la lengua producida y que da lugar a una clasificación de géneros textuales (novela, artículo periodístico o noticia); y el criterio interno, determinado por los rasgos lingüísticos específicos de la lengua producida (verbos en infiniti-

vo, tercera persona, modales), que resulta en una clasificación de tipos textuales (narrativos, informativos, o persuasivos, entre otros).

En un corpus de aprendices los géneros textuales corresponden a la variedad de tareas, ya que no se cuenta con una gran cantidad de producciones que concuerden con los géneros textuales de un corpus de hablantes nativos, como textos periodísticos, poemas o entradas de blog. De la misma manera, la variabilidad de la lengua es determinada por las tareas que realizan los aprendices en el salón de clases, así como los materiales diseñados para su aplicación, que influyen de manera decisiva sobre el tipo de discurso generado.

McCarthy y O’Keeffe (2009) señalan que las tareas orales pueden clasificarse en dos tipos: dialógicas o monológicas, dependiendo si hay una interacción con uno o varios interlocutores o si el participante genera un discurso sin intervención de un interlocutor. Wichmann (2008) agrega entre estos dos tipos un nivel intermedio, el diálogo asimétrico o monólogo dirigido, en el cual existe un interlocutor que solo funciona como guía; en general se espera que la mayor parte del discurso sea producido por el participante.

De acuerdo con el listado de corpus de aprendientes en el mundo (Centre for English Corpus Linguistics, 2020), existen alrededor de 40 corpus de aprendientes de lengua inglesa como L2 que contienen textos orales. Considerando las tres categorías de tareas orales antes mencionadas, las de tipo monológico incluyen la repetición de palabras o frases, narraciones, descripciones, presentaciones, lectura en voz alta y preguntas de opinión que no contemplan una interacción con el entrevistador o un par. Las tareas dialógicas constituyen conversaciones, tareas de discusión, juegos de roles o interacciones áulicas que se desarrollan entre participantes. Por último, las tareas que representan un diálogo asimétrico incluyen ciertas preguntas de opinión, conversaciones y entrevistas en las que el participante interactúa con un examinador.

Se puede decir, entonces, que el criterio externo en un corpus de aprendices está determinado por el tipo de tareas que realizan

los aprendices, pues dependiendo de ello se producen distintos géneros textuales, como la narrativa, la descripción, la argumentación, así como diversas formas de discurso: espontáneo, ensayado o repetitivo. Todos estos tipos están centrados en el intercambio comunicativo, la entrega de información o bien, la simple producción oral de figuras acústicas. La representatividad entonces estará delimitada por el tipo de tarea que se seleccione y las condiciones en las que se aplique.

Con respecto al criterio interno de la representatividad, en un corpus oral existen cuatro tipos textuales definidos con base en los rasgos lingüísticos característicos del discurso producido. De acuerdo con el análisis multidimensional (MDA) para los textos orales (Biber, 2004), estos tipos son: discurso orientado a la interacción, discurso de posicionamiento, discurso orientado a la información y discurso narrativo.

El primer tipo textual está orientado a la interacción, y se caracteriza por intercambios con turnos cortos cuya temática refiere aspectos cotidianos sociales o relacionales. En términos semánticos, en este tipo de textos se espera que la mayoría de los verbos sean de actividad, como *walk*, *work*, *create*, y que estén conjugados en el tiempo presente y en primera y segunda persona por enfocarse en los interlocutores.

El segundo tipo es el texto con orientación al posicionamiento, cuya interacción se enfoca en la expresión o intercambio de opiniones. En este caso, se espera que los verbos utilizados pertenezcan a la categoría semántica de verbos mentales, como *think* y *believe*, o factivos, como *be* y *seem*, además, se espera toda la gama de expresiones de modalización, tanto el uso de verbos modales como de procesos de atenuación e intensificación mediante expresiones evasivas, como *probably*, *I am not sure that...*, *it is very true that...* Debido a que la modalización es prosódica (Hood, 2010), también pueden observarse mezclas de expresiones modalizadas con verbos factivos, esto para generar significados más o menos atenuados como en (2).

- (2) *Actually, I don't think that could be true*
 FACTIVO MENTAL MODAL FACTIVO
 'De hecho, no creo que eso pueda ser verdad'

Dentro de la clasificación de textos orientados hacia la información existen dos variantes de acuerdo con el tipo de tema: académico informativo o cotidiano descriptivo (Biber, 1993). Si el tema es académico, los elementos léxicos empleados podrán ser más largos, como en una presentación escolar. En este caso, se espera el uso de nominalizaciones para representar pensamiento abstracto. Por el contrario, los elementos léxicos serán cortos si la tarea solicitada es, por ejemplo, la descripción de una imagen en que se busca representar puntos más concretos. En ambos casos serán frecuentes las frases preposicionales, en el caso del académico, para subordinaciones, como en (3); en lo cotidiano, para referenciar la ubicación de elementos, como en (4). Los procesos de personalización en el lenguaje académico harán más frecuente el uso de la voz pasiva en este tipo de texto, mientras que en el plano cotidiano se espera un mayor empleo de la voz activa.

- (3) *In spite of their centrality to Lakoff's theory these marginal figures have been frequently if not entirely overlooked in subsequent discussions of her work*
 'A pesar de su centralidad en la teoría de Lakoff, estas cifras marginales se han ignorado con frecuencia, aunque no por completo en las discusiones posteriores sobre su trabajo'
- (4) *The cat is under the table*
 'El gato está debajo de la mesa'

La cuarta clase textual es la narrativa, que engloba el contar datos biográficos, experiencias de vida, o historias ficticias que pueden recurrir a una serie de imágenes como soporte. Las tareas narrativas requieren del tiempo pasado, pero los verbos no están restringidos en términos semánticos, y pueden ser materiales, mentales

o relacionales de acuerdo con las necesidades de la historia, por ejemplo, *happened, lived, thought, been, became*. Asimismo, predomina el uso de la tercera persona, como en *Snow White lived* o *Judith wrote*, y el reporte de diálogo con verbos de comunicación en cláusulas relativas con o sin *that*, como se presenta en (5) y (6).

(5) *Angela said that there were too many*
‘Angela dijo que eran demasiados’

(6) *the man talked about fixing the window*
‘el hombre habló sobre las reparaciones de la ventana’

La representatividad en un corpus también está determinada en términos sociolingüísticos, pues la población define el tipo de producción de lengua contenida en los textos de la colección (Baker *et al.*, 2006). Para Egbert, Biber y Gray (2022), la población y el criterio de selección definen la parte de la realidad que se desea representar, un cierto dominio de uso de la lengua. Los autores comentan que en el caso de los corpus de aprendices el criterio de selección con frecuencia obedece a una autoselección de los participantes, lo que podría poner en tela de juicio su representatividad. Es decir, los participantes en el corpus son aquellos que se sienten más cómodos con su competencia en la lengua, por lo que se sesga la muestra hacia alumnos más competentes. Esta tendencia es aún más marcada en el caso de las recolecciones longitudinales que requieren un seguimiento de al menos dos años por cada participante, ya que solo siguen participando los aprendientes más motivados.

Además de las variables sociolingüísticas que deben reportarse y considerarse para el balance de la muestra (como edad, sexo, nivel educativo, etc.), existe una manera específica y crucial de clasificar las características de los aprendices a incluir, y que incide sustancialmente en las producciones que derivan de las tareas solicitadas: el nivel de dominio de la lengua. Los diferentes niveles de dominio de la lengua inglesa, así como las funciones comunicativas orales características de estos niveles, están determinados por

el Marco Común Europeo de Referencia para las lenguas (MCER). Este marco describe la habilidad lingüística en general como un conjunto de competencias, actividades y estrategias. Así, considera esta habilidad o eficiencia como un sistema basado en tres dimensiones: competencias generales, competencias comunicativas y actividades / estrategias comunicativas. Las competencias generales de acuerdo con el MCER incluyen el saber ser, saber hacer y saber aprender; las competencias comunicativas abarcan la sociolingüística, lingüística y pragmática; y las actividades / estrategias comunicativas engloban la recepción, producción, interacción y mediación (Council of Europe, 2018).

Ahora bien, la producción oral se encuentra dentro de las actividades comunicativas junto con la producción escrita. La producción oral consta de cinco modalidades o tipos de discurso oral: monólogo descriptivo, monólogo informativo, monólogo argumentativo, discurso público y dirigirse a una audiencia. Dentro del monólogo descriptivo se encuentran las tareas que implican la narración y la descripción de información cotidiana, experiencias pasadas y expectativas a futuro, además de descripciones de situaciones y temas complejos dentro del campo de acción y/o interés del aprendiz. En este caso, el discurso utilizado puede variar en su extensión, desde frases y fórmulas memorizadas hasta el desarrollo de descripciones detalladas de subtemas relacionados. El monólogo informativo implica un flujo unidireccional de la información, aunque el hablante puede ser interrumpido ocasionalmente o para solicitarle que aclare, aporte mayor detalle o repita una respuesta inaudible. Este tipo de actividad incluye la descripción de un objeto, indicaciones e información personal sobre hechos comunes y la descripción de procesos especializados académicos o profesionales del hablante. El monólogo argumentativo abarca la descripción de intereses, opiniones y preferencias sobre diversos temas, desde los más comunes a los más especializados. Así, también puede englobar la fundamentación y defensa de la información mediante el uso de estrategias de énfasis, contraste y expansión de las posturas personales.

Como recomendaciones finales en la construcción de un corpus de aprendices, Granger (2002, 2008) y Gilquin (2015) sugieren documentar y reportar diversos aspectos relacionados con los participantes, su contexto de aprendizaje y los procedimientos de recolección de la producción de los aprendices, con la finalidad de ofrecer la información del modo más completo y transparente posible. Estos datos representan las diversas variables que afectan la producción de una L2: la lengua materna, el contexto de adquisición de la L2 (segunda lengua o lengua extranjera), el nivel de dominio de la L2, el tiempo y modo de contacto con la L2 objeto de análisis y el contacto con otras segundas lenguas, así como el contexto de recolección de los datos (educativo o natural, por ejemplo, en el trabajo o durante las actividades cotidianas). Finalmente, en el caso de que no sea una recolección natural, es necesario registrar datos sobre las tareas, como tipos, temas, condiciones (tiempo y preparación), materiales utilizados para estimular la producción y detalles sobre su aplicación, además de incluir información sobre los aplicadores de dichas tareas (sexo, lengua materna, conocimiento de otras lenguas y grado de familiaridad con el participante).

5. Metodología

El objetivo del corpus MexLeC es crear una base de datos extensa y representativa de la producción oral de aprendices universitarios de inglés como segunda lengua. Para lograr este objetivo las tareas para la estimulación de la producción oral deben ser diseñadas de acuerdo con los dos criterios principales en la construcción de un corpus de aprendices: la autenticidad y la representatividad.

Con el objetivo de capturar diversos “grados de naturalidad” de los datos se han seleccionado cuatro diferentes tareas, desde aquellas que permiten un final abierto, la libre selección del vocabulario y las estructuras con que el participante desee expresarse hasta aquellas basadas en imágenes o temáticas muy específicas, que resultan más restrictivas en relación con el vocabulario y las

estructuras requeridas para responder de manera adecuada a la tarea. Al mismo tiempo, se ha privilegiado el registro de datos monológicos no interactivos, que facilitan la producción del aprendiz en turnos extendidos, así como la mínima interrupción o participación del entrevistador o guía en la tarea, con la finalidad de obtener la muestra más extensa posible del habla de los aprendices. No obstante, los aplicadores tienen como instrucción conducir la aplicación de las tareas a modo de conversación o charla informal, y participar frecuentemente mediante interjecciones que muestren interés y eviten que el aprendiz se sienta juzgado o evaluado.

En consideración al criterio externo de representatividad de un corpus, se ha considerado el MCER (Council of Europe, 2018) para decidir las temáticas y el tipo de tarea, ya que dicho marco señala, por medio de los descriptores de la producción oral monológica, las funciones comunicativas y los temas específicos que el aprendiz es capaz de abordar en cada nivel. Como resultado, la variedad de tareas descritas en las secciones de monólogo descriptivo y monólogo argumentativo dieron lugar a las cuatro partes del instrumento aplicado que detallaremos más adelante. Aunado a ello, los descriptores del MCER sirvieron de apoyo para determinar cuáles tareas eran aplicables según el nivel de lengua del aprendiz. Esto permitió que los participantes hablaran con mayor espontaneidad y facilidad en cada una de las tareas asignadas, puesto que no se sienten rebasados o forzados al requerírseles esfuerzos más allá de sus capacidades de acuerdo con el momento en el que se encuentran de su adquisición o desarrollo de la lengua inglesa. Los descriptores para cada tarea seleccionada pueden observarse en el Cuadro 1.

CUADRO 1. Secciones en la entrevista MexLeC y descriptores correspondientes según el MCER (Council of Europe, 2018)

TAREA 1. Preguntas sobre temas cotidianos	
Nivel de lengua	Función comunicativa: monólogo sostenido. Descripción de experiencias
A1	Puede describirse a sí mismo/a, a qué se dedica y dónde vive.
A1	Puede describir aspectos simples de su vida diaria.

(continuación)

CUADRO 1. Secciones en la entrevista MexLeC y descriptores correspondientes según el MCER (Council of Europe, 2018)

TAREA 1. Preguntas sobre temas cotidianos	
Nivel de lengua	Función comunicativa: monólogo sostenido. Descripción de experiencias
A2	Puede describir aspectos diarios de su ambiente, como gente, lugares y experiencias de trabajo o estudio.
A2	Puede realizar descripciones breves y básicas de eventos y actividades.
A2	Puede describir planes y preparativos, hábitos y rutinas, actividades y experiencias personales pasadas.
A2	Puede describir a su familia, sus condiciones de vida, experiencia educativa y su trabajo actual o más reciente.
A2	Puede describir gente, lugares y posesiones.
B1	Puede brindar descripciones precisas sobre varias temáticas conocidas dentro de su campo de interés.
B1	Puede brindar descripciones detalladas sobre experiencias, sentimientos y reacciones.
B1	Puede describir sueños, expectativas y ambiciones.
B2	Puede brindar descripciones claras y detalladas sobre una gran variedad de temáticas relacionadas con su campo de interés.
B2	Puede describir detalladamente la importancia personal de eventos y experiencias.
C1	Puede dar descripciones claras y detalladas sobre temas complejos.
TAREA 2. Preguntas de opción sobre preferencias e intereses	
Nivel de lengua	Función comunicativa: monólogo sostenido. Descripción de experiencias.
A2	Puede explicar lo que le gusta o disgusta sobre algo.
A2	Puede usar lenguaje simple y descriptivo para comparar objetos y posesiones usando enunciados breves.
B1	Puede expresar claramente sentimientos acerca de una experiencia y expresar las razones para explicar estos sentimientos.
Nivel de lengua	Función comunicativa: monólogo sostenido. Argumentación.
A2	Puede explicar lo que le gusta y no le gusta acerca de algo. Puede explicar las razones de sus preferencias de una cosa sobre otra.
A2	Puede brindar su opinión en términos simples.
B1	Puede brindar razones simples para justificar su punto de vista sobre un tema conocido o común.
B1	Puede expresar opiniones sobre temáticas relacionadas con la vida cotidiana.
B1	Puede brindar brevemente razones y explicaciones de sus opiniones.

(cont.)

CUADRO 1. Secciones en la entrevista MexLeC y descriptores correspondientes según el MCER (Council of Europe, 2018)

TAREA 3. Narrativa con base en imágenes	
Nivel de lengua	Función comunicativa: Monólogo sostenido. Descripción de experiencias.
A2	Puede contar una historia o describir algo de manera general, utilizando un listado simple de hechos.
B1	Puede relatar una narrativa o descripción simple, utilizando una secuencia lineal de hechos o aspectos.
B1	Puede describir eventos reales o imaginarios.
B1	Puede narrar una historia.
C1	Puede brindar descripciones y narrativas muy elaboradas, integrando subtemas y desarrollando a profundidad hechos y detalles particulares, cerrando con una conclusión apropiada.
TAREA 4. Preguntas de opinión	
Nivel de lengua	Función comunicativa: Monólogo sostenido. Argumentación.
B1	Puede desarrollar un argumento adecuado y comprensible.
B2	Puede desarrollar un argumento de manera sistemática, resaltar puntos importantes y significativos brindando detalles relevantes que apoyen su argumento.
B2	Puede desarrollar un argumento claro y expandir y apoyar sus puntos de vista con cierta amplitud, así como brindar argumentos adicionales y ejemplos relevantes.
B2	Puede construir un argumento razonable y coherente.
B2	Puede explicar un punto de vista sobre una temática determinada presentando las ventajas y desventajas entre varias opciones.
C1	Puede sostener su postura sobre un tema complejo formulando puntos precisos y empleando efectivamente estrategias para enfatizar sus argumentos.

Con el objetivo de considerar la representatividad interna se han tratado de representar tres de los tipos textuales en Biber (2004), el discurso informativo, el discurso orientado al posicionamiento y la narrativa. Dado que el discurso orientado a la interacción constituye una tarea de tipo monológico no se ha incluido en el instrumento de recopilación. Para asegurar el balance en la producción, se han determinado tiempos mínimos y máximos para cada una de las tareas, de esta manera todos los participantes tendrían los mismos intervalos para responder, cuestión que homogeneiza la aplicación y permite la comparabilidad de los datos en la investigación.

Ahora bien, una vez determinadas la variedad de tareas (géneros textuales), las temáticas adecuadas para cada tarea y nivel de lengua (población), así como los tipos textuales por abarcar en cada una de ellas (tipos textuales) y el balance de los textos que se incluirán (tiempo para cada tarea), se pudo finalizar el guion de la entrevista para la recolección de datos. Se presenta a continuación la estructura del instrumento y los detalles de cada tarea.

6. Resultados: estructura de la entrevista MexLeC

El instrumento diseñado para la construcción del corpus MexLeC consiste en una entrevista que contiene cuatro tareas o secciones. La entrevista tiene una duración total de aplicación de 12–16 minutos dependiendo del nivel del participante. Cada una de las secciones tiene un nivel de lengua recomendado según las funciones comunicativas que el aprendiz es capaz de cumplir de acuerdo con su nivel de lengua. Además, cada tarea corresponde a una tipología textual. En el Cuadro 2 se puede observar la estructura de cada una de las partes de esta entrevista y sus detalles.

CUADRO 2. Tareas de la entrevista MexLeC

Género textual (MCEC)	Tipología textual (Biber, 2014)	Descripción de la tarea	Tiempo	Nivel de lengua
Monólogo descriptivo	Informativo	Batería de preguntas relacionadas con temas como familia, amigos o tiempo libre.	3–5 minutos	Pre A1
Monólogo argumentativo	Posicionamiento	Batería de preguntas sobre preferencias o elecciones de la vida cotidiana (gastar o ahorrar el dinero, ropa cómoda o ropa de moda).	2–3 minutos	Pre A1
Monólogo descriptivo	Narrativo	Narrativa con base en una secuencia de imágenes (tira cómica a color).	3–5 minutos	A2
Monólogo argumentativo	Posicionamiento	Batería de preguntas de opinión sobre temas relacionados con el uso de la tecnología, la educación y los problemas sociales (por ejemplo, la tecnología contribuye o interfiere en la comunicación).	2–3 minutos	B1

La primera sección “Questions on familiar topics” (preguntas sobre temas conocidos) consiste en una serie de preguntas descriptivas relacionadas con la cotidianidad del participante. Estas giran en torno a dos temáticas seleccionadas de una lista de cuatro temas centrales: familia, amigos, ocupación (escuela/trabajo) y tiempo libre. Cada una de las temáticas contiene preguntas guía que no necesariamente se repiten palabra por palabra. El criterio para juzgar el empleo correcto de las preguntas es que el participante produzca discurso según la temática general seleccionada. Se espera que en esta sección la producción lingüística del participante incluya descripciones de sus intereses, actividades, experiencias, expectativas y planes futuros. Esta es la sección más básica de la entrevista, por lo que se sugiere su aplicación para todos los niveles de lengua a partir del nivel pre-A1. El tiempo de aplicación es de 3 a 5 minutos, algunos ejemplos de estas preguntas pueden observarse en (7–9).

- (7) *What do you do? Do you work or study?*
‘¿A qué te dedicas? ¿Trabajas o estudias?’
- (8) *What are your plans when you graduate?*
‘¿Cuáles son tus planes una vez que te gradúes?’
- (9) *Can you share a good memory about school?*
‘¿Puedes compartirme un recuerdo bonito sobre tu escuela?’

La segunda sección, “Choice questions” (preguntas de elección), incluye tres preguntas que representan una elección de la vida cotidiana. En esta se espera que el participante exprese y explique sus preferencias e intereses. La sección puede ser aplicada a partir del nivel A1 y su tiempo esperado de duración es de 2 a 3 minutos. Algunos ejemplos de las preguntas de esta sección se muestran en (10–12).

- (10) *Staying at home or going out*
‘¿Quedarse en casa o salir?’
- (11) *Saving or spending money?*
‘¿Ahorrar o gastar el dinero?’
- (12) *To study or to work?*
‘¿Estudiar o trabajar?’

La sección tres lleva por título “Picture-based narrative” (narrativa basada en imágenes). En esta se muestra a los participantes una tira cómica que apareció originalmente en *Ferdinand*, obra del autor danés Mikkelsen (Ortíz, 2021). Los participantes deben crear una historia con base en la secuencia de imágenes presentada. El tiempo para esta tarea es de 2 a 3 minutos y es aplicable a partir del nivel A2 de lengua. La secuencia utilizada se presenta en la Figura 2.

La última sección de la entrevista, la sección cuatro, se titula “Opinion questions” (preguntas de opinión) y consiste en un par de preguntas argumentativas seleccionadas de cuatro opciones posibles. En esta sección se espera que los participantes expresen su opinión y establezcan argumentos que apoyen puntos de vista con relación a temáticas sobre educación, uso de la tecnología y sociedad. El tiempo esperado para esta sección es de 3 a 5 minutos. Algunos ejemplos de estas preguntas se presentan en (13) y (14).

- (13) *Do you think mobile devices have destroyed communication?*
‘¿Crees que los dispositivos móviles han destruido la comunicación?’
- (14) *Which do you think is more practical online or face-to-face activities?*
‘¿Crees que es más práctico trabajar en línea o de manera presencial?’

Es importante mencionar que para asegurar la confiabilidad en la aplicación de la entrevista MexLeC, se ha creado una guía para el entrevistador, que contiene instrucciones detalladas y las preguntas



FIGURA 2. Secuencia de imágenes utilizada en la tercera sección de la entrevista MexLeC (Ortiz, 2021)

específicas, así como los tiempos y objetivos (funciones comunicativas) de cada tarea. Esta guía incluye, además, una presentación (diapositivas) con los apoyos visuales necesarios. Con el objetivo de preparar a los participantes para esta entrevista, se ha diseñado también una guía rápida, que tiene como finalidad dar a conocer la estructura general de la entrevista, y que esto permita a los aprendices reducir la ansiedad y el sentido de evaluación o juicio.

Como parte de las buenas prácticas de documentación en la construcción de un corpus, a fin de complementar los materiales creados para la aplicación de la entrevista se ha elaborado un formato con el perfil de los participantes. Este formato facilita recabar datos claves para el uso de un corpus de aprendientes: nivel de dominio de la lengua, exposición a la lengua meta, lengua materna del aprendiente y de sus padres, exposición a otras lenguas extranjeras y datos personales de identificación y contacto. Todos los datos recabados se anonimizan antes de ofrecerlos al público en general.

Por cuestiones prácticas, esta entrevista se aplica de manera remota, mediante videollamada y de forma voluntaria. Los participantes se autoseleccionan, sin embargo, se invita a los estudiantes de los primeros semestres, pues el objetivo del corpus es hacer un seguimiento longitudinal durante cuatro años, tiempo promedio de estancia en la facultad. Cada uno de los participantes firma

un consentimiento para el uso de sus datos y grabaciones en cada recolección anual.

7. Conclusiones

El objetivo de este trabajo ha sido presentar el diseño de los materiales que se utilizarán para la construcción de un corpus de aprendices que contiene ejemplos de la producción oral de aprendices universitarios de inglés. La recolección pretende un seguimiento longitudinal de estos estudiantes, desde que ingresan a la facultad hasta que se gradúan, durante un tiempo aproximado de cuatro años y mediante recolecciones anuales. Este tipo de colecciones puede ser de utilidad para estudiar el proceso de adquisición de una segunda lengua, así como para diseñar materiales más efectivos con base en las necesidades específicas de los aprendices.

El corpus MexLeC sigue creciendo. Debido a su naturaleza longitudinal tendrán que pasar alrededor de 5 años antes de que alcance su tamaño final. Sin embargo, podemos reportar que en este momento contiene 142 transcripciones ortográficas que han pasado por un proceso riguroso de revisión. El corpus de producción oral resultante tiene un tamaño aproximado de 110 mil tokens. Estos datos son el resultado de las dos primeras recolecciones realizadas durante los años 2021 y 2022 en la Licenciatura en Lenguas de la Universidad Autónoma del Estado de México. A principios del año 2022, se inició la aplicación de los instrumentos y el levantamiento de datos a los estudiantes de la Licenciatura en Enseñanza de la Lengua Inglesa de la Universidad Autónoma de Hidalgo, con una segunda aplicación en 2023. En ese mismo año, el proceso también se inició en la Licenciatura en Lenguas Modernas Inglés de la Universidad Autónoma de Querétaro. Actualmente, se están afinando los detalles para la primera aplicación de los instrumentos en la Licenciatura en Lengua Inglesa de la Universidad Autónoma del Estado de Quintana Roo.

Uno de los retos más grandes relacionados con el instrumento diseñado es lograr la capacitación de los entrevistadores de mane-

ra que permita conservar la homogeneidad en la aplicación de la entrevista, y asegurar mínimas interrupciones para conservar los turnos largos del entrevistado (pero sin que el entrevistador parezca desinteresado en lo que tiene que decir el participante), así como el cuidado de los tiempos para cada tarea. Los datos recabados son de acceso libre y se encuentran disponibles para su descarga en la página web de MexLeC (Flores & Moore, 2023).

Además de la planeación de recolecciones para incluir nuevos aprendices, es necesario garantizar el seguimiento longitudinal de los participantes actuales. Otra meta a mediano plazo es comenzar la aplicación de los datos en el diseño de materiales e intervenciones didácticas, así como la mejora y reestructuración de programas que consideren las características de la lengua producida por los estudiantes.

El uso de datos de libre acceso, extensos, auténticos, representativos y estandarizados de los aprendices mexicanos de inglés permitirá contar con un modelo más amplio para el diseño de materiales, mejora de programas y planeación de intervenciones, que incluya no solo factores individuales o lingüísticos, sino también análisis empíricos cuantitativos que validen nuestras intuiciones prácticas y teóricas.

8. Referencias

- Baker, Paul; Hardie, Andrew, & McEney, Tony (2006). *A glossary of corpus linguistics*. Edimburgo: Edinburgh University Press.
- Biber, Douglas (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257.
- Biber, Douglas (2004). Conversation text types: A multi-dimensional analysis. En Gérald Purnelle, Cédric Fairon & Anne Dister (Eds.), *Le poids des mots. Actes des 7es Journées internationales d'analyse statistique des données textuelles* (pp. 15–34). Lovaina la Nueva: Presses universitaires de Louvain.

- Collies, Marcus, & Paquot, Magali (2015). Learner corpus research: An interdisciplinary field on the move. *International Journal of Learner Corpus Research*, 1(1), 1–6. <https://benjamins.com/catalog/ijlcr.1.1.00edi>
- Capel, Annette & Sharp, Wendy (2014). *Objective First. Student's book with answers*. 3a. ed. Edimburgo: Cambridge University Press.
- Centre for English Corpus Linguistics (2020). Learner corpora around the world. *Université Catholique de Louvain (UCLouvain)*. <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>
- Council of Europe (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume*. Estrasburgo: Council of Europe.
- Egbert, Jesse; Biber, Douglas, & Gray, Bethany (2022). *Designing and evaluating language corpora. A practical framework for corpus representativeness*. Cambridge: Cambridge University Press.
- Flores, Ana, & Moore, Pauline (2023). *Mexican Learner Corpus*. <https://sites.google.com/view/mexlec/intro>
- Gablasova, Dana; Brezina, Vaclav, & McEnery, Tony (2019). The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research*, 5(2), 126–158.
- Granger, Sylviane (2002). A bird's-eye view of learner corpus research. En Sylviane Granger, Joseph Hung & Stephanie Petch Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). Ámsterdam: John Benjamins.
- Granger, Sylviane (2008). Learner corpora. En Anke Lüdeling & Merja Kytö (Eds.), *Corpus linguistics. An international handbook* (Vol. 1, pp. 259–275). Berlín: Walter de Gruyter.
- Granger, Sylviane (2021). Phraseology, corpora and L2 research. En Sylviane Granger (Ed.), *Perspectives on the L2 phrasicon* (pp. 3–24). Ámsterdam: Multilingual Matters.
- Granger, Sylviane; Dupont, Maïté; Meunier, Fanny; Naets, Hubert, & Paquot, Magali (2020). *The International Corpus of Learner English. Version 3*. Lovaina la Nueva: Presses universitaires de Louvain. <https://dial.uclouvain.be/pr/boreal/object/boreal:229877>
- Gilquin, Gaëtanelle (2015). From design to collection of learner corpora. En Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (Eds.), *The*

- Cambridge handbook of learner corpus research* (pp. 9–34). Cambridge: Cambridge University Press.
- Gilquin, Gaëtanelle (2019). Light verb constructions in spoken L2 English: An exploratory cross-sectional study. *International Journal of Learner Corpus Research*, 5(2), 181–206. <https://benjamins.com/catalog/ijlcr.18003.gil>
- Hood, Susan (2010). *Appraising research: Evaluation in academic writing*. Londres: Palgrave Macmillan.
- Hunston, Susan (2008). Collection strategies and design decisions. En Anke Lüdeling & Merja Kytö (Eds.), *Corpus linguistics. An international handbook* (Vol. 1, pp. 154–167). Berlín: Walter de Gruyter.
- Lehmborg, Timm, & Wörner, Kai (2008). Annotation standards. En Anke Lüdeling & Merja Kytö (Eds.), *Corpus linguistics. An international handbook* (Vol. 1, pp. 484–500). Berlín: Walter de Gruyter.
- Lüdeling, Anke, & Kytö, Merja (Eds.) (2008). *Corpus linguistics. An international handbook* (Vol. 1). Berlín: Walter de Gruyter.
- Meyer, Charles F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.
- McCarthy, Michael, & O’Keeffe, Anne (2009). Corpora and spoken language. En Anke Lüdeling & Merja Kytö (Eds.), *Corpus linguistics. An international handbook* (Vol. 2, pp. 1008–1023). Berlín: Walter de Gruyter.
- McEnery, Tony, & Gabrielatos, Costas (2006). English corpus linguistics. En Bas Aarts & April McMahon (Eds.), *The handbook of English linguistics* (pp. 33–71). Oxford: Blackwell.
- McEnery, Tony, & Hardie, Andrew (2011). *Corpus linguistics: Method, theory, and practice*. Cambridge: Cambridge University Press.
- McEnery, Tony; Xiao, Richard, & Tono, Yukio (2006). *Corpus-based language studies: An advanced resource book*. Londres: Routledge.
- Meunier, Fanny (2021). Introduction to learner corpus research. En Nicole Tracy Ventura & Magali Paquot (Eds.), *The Routledge handbook of second language acquisition and corpora* (pp. 23–36). Londres: Routledge.
- Ortiz, Federico (2021). *Ferdinand. Ven a mi mundo*. <http://www.venamimundo.com/DeAquiyaAlla/TirasComicas/Ferdinand.html>

- Sinclair, John (1996). *Preliminary recommendations on corpus typology* (Reporte técnico). Expert Advisory Group on Language Engineering Standards. <http://www.ilc.cnr.it/EAGLES96/corpusyp/corpusyp.html>
- Sinclair, John (2005). Corpus and text: Basic principles. En Martin Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxbow Books.
- Wichmann, Anne (2008). Speech corpora and spoken corpora. En Anke Lüdeling & Merja Kytö (Eds.), *Corpus linguistics. An international handbook* (Vol. 1, pp. 187–206). Berlín: Walter de Gruyter.





Reseña

Daniela Lauria. *Lengua y política. Historia crítica de los diccionarios del español de la Argentina*. Buenos Aires: Eudeba, 2022. 374 pp.

Luis B. Quesada Nieto
The Graduate Center,
The City University of New York

¿A quién le pertenece la lengua? ¿Qué fuerzas deciden cómo unificar en un estándar una multiplicidad de variedades? ¿De dónde emana la autoridad para establecer lo que cuenta como arcaísmo, neologismo, extranjerismo, vulgarismo? ¿En función de qué ideas y poderes se desplazan, manipulan o adaptan estas definiciones en el eje diacrónico? Estas y otras interrogantes subyacen en *Lengua y política. Historia crítica de los diccionarios del español de la Argentina* (2022), de Daniela Lauria, publicación que expone los resultados de la investigación doctoral de la autora, consistente en el análisis de los discursos lexicográficos producidos en Argentina, a través de un corpus de diccionarios monolingües que se organiza en dos grandes periodos: el que va de la Revolución de mayo a su conmemoración centenaria en 1910 y el que transcurre de 1910 al Bicentenario de 2010.

Como han señalado diversos estudios sobre política lingüística e historiografía lingüística, los largos procesos de estandarización por los que pasaron los objetos-lengua, constructos actualmente reconocidos como *español*, *inglés* o *guaraní*, han requerido gestionar la heterogeneidad de las prácticas lingüísticas por medio de su eliminación, invisibilización y denostación, según la forma en que estas son referidas o no en los instrumentos lingüísticos que dichos procesos producen (gramáticas, ortografías, diccionarios, manuales, tratados). Con base en la idea de la unidad lingüística, que por lo general acompaña a la formación y consolidación de proyectos políticos tan amplios como el Estado nación, mediante estas herramientas de gestión se va formando una lengua estándar que es

reflejo del uso de una élite culta, a la que han quedado adheridas, además, las cualidades de ser apropiada y gramatical. En otras palabras, las expresiones ideológicas asociadas a un proyecto político que requiere de la instrumentalización de la lengua inevitablemente impregnan el trabajo de los individuos y las instituciones que participan en la creación de dichos instrumentos y lo que en ellos se presenta como “error gramatical”, barbarismo, extranjerismo, indigenismo, neologismo, etcétera. Es, en resumen, un acto político encubierto, al menos, de científicidad y buenas intenciones.

Es este tipo de alineaciones político-ideológicas lo que a Lauria le interesa evidenciar, más con un propósito hermenéutico que descriptivo. Para lograrlo la autora hace un recorrido detallado de la producción lexicográfica de la nación argentina analizándola desde una perspectiva glotopolítica, definida como el estudio de “las ideologías lingüísticas y las intervenciones en el espacio público del lenguaje, asociándolas con posiciones y espacios institucionales e indagando en los modos en que aquellas participan en la instauración, reproducción o transformación de entidades políticas, relaciones sociales y estructuras de poder” (Narvaja de Arnoux, 2008, en Lauria: 29).

A partir del análisis discursivo de los textos que aparecen, tanto en la macroestructura (elementos paratextuales de los diccionarios: prólogos, títulos, notas, comentarios, entre otros) como en la microestructura —entendida como “la forma de lematización (también llamada entrada); y todo el cuerpo del segundo enunciado: la formulación de las definiciones [...]; la categorización gramatical e información morfosintáctica, la forma gráfica; la presencia o no de un sistema de marcas [...] entre otras indicaciones (44)—, la obra expone importantes tensiones que contribuyen a comprender por qué fenómenos léxicos como los indigenismos reciben determinado tipo de tratamiento y cómo este tratamiento se desplaza según el momento histórico y las ideas de progreso asociadas a él. Por ejemplo, el condenar los usos léxicos de la vida del campo mediante diversas estrategias, como la clasificación diastrática (rural) o pragmática (coloquial, vulgar), el uso de tiempos verba-

les (pasado), la selección de ejemplos de uso (literarios o *ad hoc*), en un momento en que el discurso nacional buscaba desterrar todo aquello que entendía relacionado al subdesarrollo y la barbarie; un gesto propio de los diccionarios producidos en torno al Centenario de la Independencia. El mismo objeto lingüístico, no obstante, atraviesa por una revalorización y cien años más tarde es abordado de formas más positivas en los diccionarios producidos alrededor del Bicentenario, cuando la nación estaba más preocupada por hacer resaltar su esencia, su singularidad, y figuras como la del gaucho son apreciadas con otros ojos, sobre todo a partir de la publicación del *Idioma nacional de los argentinos* de Lucien Abeille (1900). La aparición de esta obra propició un autonomismo lingüístico producto del antihispanismo y de la idea de que la independencia política debía venir acompañada de una independencia lingüística y cultural, “propia de la etapa inicial del proceso de formación de la nación” (163). En este marco de ideas, recuerda Lauria, Abeille, de origen francés, se configura como un lexicógrafo externo, objetivo y científico, para quien el cambio lingüístico no es una corrupción de la lengua, sino por el contrario, la evolución de la lengua es un síntoma de la evolución social, de su progreso; razón por la que, en lo léxico, su obra muestra gran amplitud para incluir voces de diversos orígenes.

Algo similar sucede con las voces extranjeras. La élite intelectual, los “guardianes de la lengua”, se acercaban a normarla con una batería de aspiraciones y justificaciones sobre las que tampoco se aprecia un criterio lingüístico estable. Alrededor de los años 70 del siglo XIX, cuando la inmigración masiva era apreciada como positiva y necesaria para el desarrollo civilizatorio de la nación, aparecen en la producción diccionarística intervenciones que interpretan las palabras propias del contacto lingüístico como muestras de modernidad y progreso. No obstante, en otro momento, son vistas con recelo por las élites criollas, como índices del rechazo a la integración o de la descomposición y fragmentación lingüística de la lengua castellana y, por extensión, de la descomposición y el retraso nacional que debía superarse: “la élite letrada observaba

con recelo la presencia inquietante del inmigrante como nuevo sujeto social. Estas circunstancias particulares generaron que el proceso de diccionarización del español de la Argentina se viera afectado [...] por la aparición de una nueva modalidad lexicográfica: los diccionarios de barbarismos” (131), acontecido entre 1890 y 1903. De fondo, la pregunta sobre qué hacer con los extranjerismos en los diccionarios, cómo definirlos y cómo tratarlos, es un eco de la política hacia los inmigrantes y la idea que se tiene sobre su contribución a la vida nacional, sujeta por supuesto a los cambios sociales y políticos. Esto propicia el encuentro con distintos criterios normativos. Cabe mencionar la obra de Turdera, *Diccionario de barbarismos argentinos y otros excesos de la misma nacionalidad* (1896), que presenta la peculiaridad de emplear ejemplos *ad hoc* creados por el mismo autor y se pronuncia en contra del uso de galicismos, anglicismos y, en general, la oralidad de los extranjeros (italianos y españoles, sobre todo gallegos, catalanes, aragoneses, vascos y andaluces). En contraste, la de Monner Sanz, *Notas al castellano de la Argentina* (1903), observa un criterio más o menos flexible, en especial hacia los neologismos.

Lauria explora los distintos criterios normativos en torno a algún aspecto o fenómeno lingüístico en concreto, los cuales determinan que para algunos el cambio lingüístico representado por las innovaciones léxicas o los neologismos sea algo positivo, mientras que para otros constituya algo negativo y deba por tanto erradicarse, dado que se consideran usos “incorrectos” o desvíos de la norma, establecida por el *Diccionario de la Real Academia Española*. La autora asegura que “se entrevé un continuum que va de posturas rigurosas a posturas más moderadas [...] [vinculadas] con la concepción de la lengua que subyace a cada instrumento lingüístico [...] [y] a que los autores están inmersos en distintas discursividades, producto de las distintas condiciones de producción de los textos” (160).

Con claridad expositiva, Lauria muestra las diversas tensiones constitutivas de la labor lexicográfica de modalidad complementaria y contrastiva, propias del primer periodo que estudia, entre

estas: norma-uso, prescripción-descripción, propio-ajeno, letrado-popular, escritura-oralidad, centro-periferia, de entre las cuales identifica el eje prescripción-descripción como el más sensible a los desplazamientos discursivos. Asimismo, a partir del juego de inclusiones y exclusiones, la autora señala cómo las obras convocan un imaginario sobre la dicotomía civilización-barbarie, en las que además se aprecian las “regularidades y los desplazamientos ocurridos debido a las transformaciones operadas en el espacio social” (173), tales como la modernización, la urbanización, la industrialización, la inmigración y la consolidación nacional, acontecimientos que “dejaron sus huellas en la memoria lexicográfica en términos de ruralismos, indigenismos, regionalismos y barbarismos” (173).

La segunda parte del libro, dedicada al análisis de diccionarios del periodo del Bicentenario, continúa el mismo método de análisis de las macro y microestructuras. Esta parte concluye con los efectos de los macroprocesos de globalización en las políticas lingüísticas hispánicas y la producción de diccionarios monolingües. La autora identifica en este contexto la coexistencia de dos modalidades diccionarísticas, por un lado, los diccionarios complementarios y contrastivos y, por el otro, los diccionarios integrales; cada uno reproduce diversas ideologías lingüísticas, en las que han entrado nuevos agentes reguladores de la lengua. Previo a esto, Lauria revisa diccionarios de argentinismos de la primera mitad del siglo xx que permiten identificar una valoración de las formas lingüísticas asociadas típicamente a la cultura rural criolla y el gaucho, como se dijo antes, en una etapa en la que se busca reforzar la identidad nacional. Sobre estos instrumentos (escritos por Garzón, Segovia y Díaz Salazar), a partir del análisis y mediante el uso de ejemplos concretos, la investigadora plantea que se constituyen como el lugar ideal para observar la tensión política e ideológica entre la defensa de la unidad de la lengua y la conformación de la lengua nacional como un sello distintivo de cara al mundo hispánico. Otras tensiones también se observan en el eje de las representaciones de la naturaleza, como opuestas a las de la

urbanidad, a través de enunciados polarizados en que “lo rural se afilia con lo ordinario, lo rústico y el pasado; la ciudad se coliga con la civilización y el progreso” (248).

La obra de Lauria es extensa, rigurosa y organizada y constituye una gran aportación a la historia de las ideas sobre la lengua en Argentina y Latinoamérica realizada desde la perspectiva glotopolítica. En efecto, la historia crítica de los diccionarios monolingües y de la actividad lexicográfica, como la autora plantea hacia el final del libro, van de la mano con el desarrollo del capitalismo, lo que permite interpretar las intervenciones acerca de la lengua en función de procesos como la construcción de la nación o su integración en el sistema económico regional y mundial. De gran valor es el hecho de que este estudio posibilita la reconstrucción de una historia lexicográfica argentina que, al apoyarse en una perspectiva de análisis crítico como lo es la glotopolítica, ofrece un acercamiento a las formas de la lengua, su conocimiento y contextualización, además de que hace explícitas las condiciones de su producción. Por esto, visitar esta historia lexicográfica crítica significa a la vez visitar la historia argentina reciente, es decir, un ejercicio en que lo social y lo político son indisociables de lo lingüístico.

Referencias

Narvaja de Arnoux, Elvira (2008). *Los discursos sobre la nación y el lenguaje en la formación del Estado (Chile, 1842-1862): estudio glotopolítico*. Buenos Aires: Santiago Arcos.

Lineamientos para publicar en *Estudios de Lingüística Aplicada**

Todos los manuscritos propuestos para publicación en *Estudios de Lingüística Aplicada* (ELA) deben ser enviados a través de su sitio en internet (ela.enallt.unam.mx). Estos lineamientos, así como las instrucciones para hacer los envíos en línea, pueden consultarse en dicho sitio.

De acuerdo con los principios del doble dictamen ciego, en los manuscritos que se envíen por primera vez deberá omitirse el nombre, la adscripción y la dirección de contacto del autor. Además, deberá eliminarse la identificación del autor/a de las propiedades del archivo (véase “Garantizar una evaluación por pares anónima” en el sitio web de la revista). Esta información se solicitará a los autores en el proceso de envío en línea. Posteriormente, y solo si el manuscrito ha sido aceptado, estos datos se incorporarán en el archivo en Word de la versión final del manuscrito.

Los artículos de investigación deberán tener una extensión máxima de 30 cuartillas, las notas de 20 cuartillas y las reseñas de 10 cuartillas, sin incluir referencias bibliográficas y anexos, con interlineado doble, en letra Times New Roman de 12 puntos, en formato carta (21.59 cm x 27.94 cm) y respetando los siguientes márgenes: izquierdo 3 cm, derecho 3 cm, superior 2.5 cm, inferior 2.5 cm. Los cuadros y figuras deberán incluirse dentro del texto en el sitio que les corresponde y presentarse por separado en el formato del programa original en el que fueron creados, así como

* *Estudios de Lingüística Aplicada* sigue la hoja de estilo de *The generic style rules for Linguistics* (Haspelmath, 2014) y, para el formato de las referencias, el manual de publicaciones de la American Psychological Association (APA, 2010), con algunas modificaciones. La elaboración de los lineamientos para publicar en *Estudios de Lingüística Aplicada* estuvo a cargo de Sergio Ibáñez Cerda, Zazil Sobrevilla Moreno y Mariela Elizabeth Pérez Castañeda, con la colaboración de Blanca Estela Solís Valdespino, Armando Rodríguez Briseño y Sandra Contreras Macías.

en archivo de imagen (véase *infra* §11). La versión del artículo en formato PDF y los cuadros y figuras en archivo de imagen son indispensables para controlar los cambios de procesadores y los caracteres fonéticos u otros detalles en el proceso editorial.

Se sugiere a los autores cuya lengua materna no sea el español o el inglés (según la lengua en que esté escrita la contribución) que un hablante nativo revise su manuscrito antes de que sea enviado. Las contribuciones en español deberán seguir las innovaciones ortográficas aprobadas por la Real Academia Española y la Asociación de Academias de la Lengua Española en 2010 (RAE & Asale, 2010).¹

Se recomienda minimizar el uso de marcas tipográficas tales como negritas, versalitas, cursivas, comillas, etc., y conservarlas solo cuando sea indispensable. Cualquier necesidad editorial que esté fuera de las especificaciones de estos lineamientos deberá ser comunicada y puesta a consideración de los editores.

1. Partes del texto

El texto del artículo va precedido por el título y por un resumen, de 1000 a 1200 caracteres con espacios, seguido por cinco palabras clave que no coincidan con las que aparecen en el título del artículo. El título, el resumen o *abstract* y las palabras clave o *keywords* deberán proporcionarse en la lengua principal del texto y en su traducción al español o al inglés, según sea el caso. Tanto las palabras clave como las *keywords* deben ir separadas por punto y coma y no llevar punto final.

¹ Se seguirá, por ejemplo, la eliminación de la tilde diacrítica en el adverbio *solo* y en los pronombres demostrativos (*este, ese, aquel*, etcétera), incluso en casos de posible ambigüedad. Un resumen sobre las principales novedades de la última edición de la *Ortografía* se ofrece en Real Academia Española (RAE, 2010) y puede recuperarse en http://www.rae.es/sites/default/files/Principales_novedades_de_la_Ortografia_de_la_lengua_espanola.pdf

Los artículos se subdividen en secciones numeradas y, de ser necesario, en subsecciones, cada una de las cuales lleva un encabezado. La numeración empieza siempre en 1 (Sección 1: 1.1, 1.2, Sección 2: 2.1, 2.2, etc.). Solo en casos excepcionales se empleará un tercer nivel de división, que llevará también un encabezado. La última sección numerada puede ir seguida de varias secciones opcionales (Corpus o Fuentes, Abreviaturas, etc.), y uno o más Anexos (A, B, etc.). Al final del documento aparece la sección de Referencias. Las normas de estilo de los distintos tipos de referencias bibliográficas se detallan en §15.

Los encabezados de las secciones no llevan punto final ni más mayúsculas que la inicial (véase §2). Si una subsección tiene (sub-)subsecciones, deberá haber al menos dos de ellas. Todo el texto deberá pertenecer a alguna sección, aunque, excepcionalmente, podrá haber un párrafo introductorio.

2. Uso de mayúsculas

Las oraciones, nombres propios, títulos y encabezados inician con mayúscula. Obsérvese que en los manuscritos escritos en inglés no se emplean mayúsculas en el resto de las palabras que componen títulos y encabezados de secciones, figuras, cuadros, etc. (es decir, no se emplea *title case*). Con respecto a las referencias bibliográficas, en los títulos de artículos y libros en inglés también se seguirá este principio, si bien los títulos de las revistas y series respetarán el uso de las mayúsculas de cada publicación, pues se tratan como nombres propios. Si el título de una referencia es en inglés y contiene un subtítulo, el subtítulo iniciará con letra mayúscula solo si se trata de una oración completa; si es en español, siempre iniciará con minúscula. Se presentan a continuación algunos ejemplos:

1.1. Overview of the issues (NO: Overview of the Issues)

Figure 3. A schematic representation of the workflow (NO: A Schematic Representation of the Workflow)

Anderson, Gregory (2006). *Auxiliary verb constructions*. Oxford: Oxford University Press. (NO: *Auxiliary Verb Constructions*)

Clyne, Michael (Ed.) (1991). *Pluricentric languages: different norms in different nations*. Berlín: De Gruyter Mouton.

Cuando se haga referencia a una parte del artículo numerada, esta iniciará con mayúscula (Figura, Cuadro, Apéndice), *e. g.*

como se muestra en el Cuadro 5

En ningún caso se usarán mayúsculas para hacer énfasis o introducir conceptos o como sistema de marcación tipográfica. Tampoco se usarán para iniciar los nombres de leyes, teorías, modelos, procedimientos estadísticos o hipótesis.

3. **Cursivas**

Las cursivas se emplean para los siguientes propósitos:

- Para todas las formas de la lengua objeto (letras, palabras, frases, oraciones) citadas dentro del texto, a menos que se trate de transcripciones fonéticas o representaciones fonológicas en Alfabeto Fonético Internacional (AFI).
- Cuando los ejemplos en párrafo aparte sean de lenguas que no sean el español o el inglés, y por lo tanto vayan glosados. Los ejemplos en español y en inglés no glosados irán en redondas.
- Para llamar la atención del lector en algún aspecto particular de un ejemplo lingüístico en un ejemplo numerado que aparezca en redondas.
- Para palabras en una lengua distinta a la del texto principal.
- Cuando se hace referencia a un término técnico como forma de la lengua objeto, por ejemplo:

el término *citativo* no es apropiado para esto.

- Cuando se hace referencia a un término técnico en primera mención. En tales contextos, los términos técnicos se manejan como formas de la lengua objeto, por ejemplo:

la noción de *implicatura convencional*, que durante algún tiempo había mantenido un bajo perfil

- Para enfatizar alguna palabra que no es un término técnico, *e. g.*

Esto es posible en este contexto, pero *solo* en este contexto.

- Para enfatizar algo en una cita, siempre con la indicación al final de la misma:

[*énfasis añadido*]

- Para títulos de libros, revistas, tesis y otras fuentes como periódicos, videos y películas.
- Para los símbolos estadísticos (*N, n, ds, etc.*).²

4. Versalitas

Las versalitas se emplean:

- Cuando se considere necesario llamar la atención sobre un término en su primera mención o cuando se ofrece su definición, *e. g.*

Se distinguen dos tipos de alineamiento principales, a saber,
NOMINATIVO-ACUSATIVO Y ERGATIVO-ABSOLUTIVO.

En la medida de lo posible, evite el uso de las versalitas (o de cualquier marcación tipográfica) para introducir términos que sean de

² Excepto las letras griegas, subíndices y exponentes que sirven como identificadores (es decir que no son variables) y las abreviaturas que no son variables.

uso general en la teoría lingüística, por ejemplo, sujeto, objeto directo, pronombre, pretérito, cuando se usen en su sentido convencional.

- También se emplean en las abreviaturas de categorías en las glosas interlineales (véase §10.1) y se pueden usar para indicar foco o prominencia acentual en las oraciones de ejemplo:

(1) PINTURA salía de las tuberías.

- Finalmente, en las contribuciones en español, los siglos se marcarán en versalitas: siglo XVI, XIII, XX, etcétera.

5. Negritas y otras formas de énfasis

Las negritas se utilizan:

- Para llamar la atención del lector en algún aspecto particular de un ejemplo lingüístico, sea en un ejemplo numerado o en un ejemplo que aparezca en el texto cuando estos estén en cursivas:

(2) *das Kind*, ***dem*** *du* *geholfen hast*
El niño.NOM **quien.DAT** tú.NOM ayudado has
'el niño a quien ayudaste'

- En general, no se emplean mayúsculas ni subrayado con el fin de destacar algún elemento. De manera excepcional, se podrá emplear el subrayado para resaltar una letra en una palabra de ejemplo y en otros casos donde no se puedan emplear otras formas de énfasis.

6. Uso de comillas

Las comillas dobles se usan en las siguientes situaciones:

- Cuando en el texto se cita un pasaje de otro texto, por ejemplo:³

³ Las citas en párrafo aparte no llevan comillas.

De acuerdo con Takahashi (2009: 33), “los citativos nunca se empleaban en oraciones subordinadas en japonés antiguo”.

- Cuando se menciona un término técnico o alguna expresión que el autor no quiere adoptar, como en el ejemplo de abajo. Alternativamente, en estos casos se puede también usar cursivas:

A esto se le llama comúnmente “pseudo-conservadurismo”, pero no emplearé ese término, pues se presta a confusión.

Las comillas simples se emplean:

- Para significados lingüísticos, por ejemplo:

El latín *habere* ‘tener’ no es cognado con *hafian* ‘tener’ del inglés antiguo.

- Para marcar citas dentro de citas.

7. Empleo de otros signos ortográficos

- En las acotaciones parentéticas —como en este ejemplo— se emplea el guion largo o raya, sin que sea precedido ni seguido de espacios.
- En caso de que el artículo esté escrito en inglés, se emplea mismo signo, antecedido y seguido de espacios, por ejemplo:

The starting hypothesis is that the prosodic expressiveness is divided into two major types — encouragement and discouragement — described through several restrictions within the framework of the Optimality Theory.

- Para indicar rangos numéricos se emplea el guion medio, por ejemplo, 1995–1996.
- La elipsis dentro de una cita se indica con [...], *e. g.*

“Barcelona aceptaba la noche sobre el mar [...] Las ciudades se aceptan porque abrigan.”

- Los corchetes angulares < > se emplean para hacer referencia a grafemas y demás símbolos escritos específicos, por ejemplo, <q>.
- Los millares, millones, etc., no se separan con coma ni con punto (por ejemplo: 35 542 y no 35,542, ni 34.542). La separación se hará con espacios en blanco en los números de más de cuatro cifras, mientras que los números de cuatro cifras se escribirán sin espacios de separación: 1000, 10 000, 100 000, 1 000 000, 100 000 000, etcétera. Los decimales van precedidos por un punto, no por coma (por ejemplo 2.35 y no 2,35).
- En los textos escritos en español no se usará la tilde diacrítica en el adverbio *solo* ni en los pronombres demostrativos (*este, ese, aquel*) incluso en casos de posible ambigüedad (por ejemplo: *solo llevaba un par de monedas en el bolsillo, este es tonto, quiero aquella* y no *sólo llevaba un par de monedas en el bolsillo, éste es tonto, quiero aquélla*). Tampoco se utilizará tilde diacrítica en la conjunción disyuntiva *o* escrita entre cifras (*1 o 2* y no *1 ó 2*).

8. Abreviaturas

- Se deben evitar las abreviaturas de expresiones no comunes.
- Los nombres de lenguas no deberán ser abreviados.
- Las abreviaturas se deben usar para etiquetas de categorías gramaticales en las traducciones morfema por morfema. Las reglas de glosado de Leipzig incluyen una lista de abreviaturas estándar para las categorías gramaticales más usadas (véase §10.1).
- Cuando se hace referencia repetida a un término complejo que no es ampliamente conocido, se podrá abreviar (por ejemplo, CDO para “construcción de doble objeto”). En tal caso, la abreviatura se deberá explicar la primera vez que aparece en el texto y también al final del artículo en la sección Abreviaturas. Las abreviaturas de expresiones no comunes no se usarán en los encabezados ni en los pies de figuras y cuadros y se deben evitar al inicio de las secciones.

9. Citas dentro del texto

Las citas textuales, cuando son menores de 5 líneas se incorporan en el texto entre comillas; cuando son de 5 líneas o más, se despliegan en un bloque independiente (a bando y separado con espacios antes y después), sin comillas y con tipo menor. En cualquier caso debe señalarse la fuente. Las citas en lenguas distintas a las de la lengua principal del texto deberán traducirse, dentro del texto entre corchetes si son cortas y en nota al pie si son largas.

Las fuentes se citan señalando el autor, el año del trabajo y la página específica de la cita. El autor puede aparecer como un elemento dentro del texto principal (como en el primer ejemplo que se muestra abajo) o entre paréntesis (como se muestra en el segundo ejemplo). Los números de página van después del año, separados por dos puntos y un espacio, sin omitir dígitos. Cuando aparece el nombre del autor dentro del paréntesis, este se separa del año con coma.

Thomason y Kaufman (1988: 276–280) señalan que los dialectos del inglés del norte tienen más innovaciones morfológicas que los dialectos de inglés del sur.

La notación que empleamos para representar esto se toma de las teorías que afirman que los rasgos- ϕ aparecen en una geometría de rasgos (Gazdar & Pullum, 1982: 34).

Las referencias bibliográficas completas correspondientes a cada trabajo citado se enlistan alfabéticamente al final del texto. A continuación se hacen algunas especificaciones sobre el formato de las citas dentro del texto:

- En caso de que se cite una obra de dos autores, ambos nombres se coordinarán con *y* cuando aparecen como un elemento del texto principal —como en el primer ejemplo arriba—, y con *&* cuando apa-

recen entre paréntesis —como en el segundo ejemplo. En el primer caso la referencia es a los autores y en el segundo a la obra.

- Cuando hay más de dos autores, la referencia, en la primera mención, incluye a todos y, a partir de la segunda, solo al primero, seguido por la abreviatura *et al.* (véase el ejemplo abajo). En la lista de referencias al final del texto se dará la lista completa de los autores.

López *et al.* (1992) = López, García & Silva (1992)

- Cuando se citan múltiples trabajos de un mismo autor, el nombre no deberá repetirse y los años se separarán por coma.

Mientras que Hawkins (2004, 2014) ha argumentado en favor de un principio de minimización de dominio, otros autores han tratado de explicar los efectos observados en términos puramente gramaticales.

Estudios empíricos previos reportan que la anteposición de objeto en estas lenguas ocurre bajo las mismas condiciones contextuales para los verbos transitivos canónicos y para los verbos de objeto-experimentante (véase Verhoeven, 2008b, 2010a para el turco y el chino).

- Cuando se enlistan varias referencias en paréntesis, estas se separan por punto y coma y normalmente se ordenan cronológicamente (aunque si el autor lo considera adecuado para la exposición, pueden seguirse otros criterios como el alfabético o por orden de importancia).

Los hablantes se basan mucho en trozos formulaicos o “pre-fabricados” durante la producción y comprensión del habla (Pawley & Syder, 1983; Sinclair, 1991; Erman & Warren, 2000; Bybee, 2006; véase Wray, 2002 para un panorama histórico más amplio).

- En algunos casos, cuando se cita una reimpresión de un texto antiguo o una referencia clásica, es conveniente dar el año de la publicación citada, seguida por el año del texto original entre corchetes, por ejemplo:

Correas (1984 [1626])

Lapesa (2000 [1973])

- Las comunicaciones personales (cartas privadas, memorandos, mensajes electrónicos, etc.) no se incluyen en la lista de referencias, pero sí deben aparecer citadas en el texto. Para ello, se debe proporcionar los datos del autor, así como una fecha tan exacta como sea posible:

Thomas Lutes (comunicación personal, 18 de abril, 2001)

- Los números de página deben escribirse completos, sin omitir dígitos; tampoco se usará la abreviatura “y ss.” para ‘y siguientes’.
- Los números de página se podrán omitir solo si la referencia concierne al trabajo entero. Se pueden dar números de capítulo o de sección en lugar de números de página, si se considera pertinente (por ejemplo: *Auer, 2007: capítulo 7*; *Matras, 2009: §6.2.2*). Asimismo, para la citación de material sin paginación, como en algunas fuentes en línea, puede hacerse referencia a otras secciones como el párrafo, por ejemplo: *Reynolds, 2016: párr. 5*.

Es importante asegurarse de que las referencias bibliográficas correspondientes a cada trabajo citado se encuentren enlistadas al final del texto en la sección de Referencias (véase §15) y viceversa, y que la ortografía, nombres de los autores, fechas y paginado correspondan en ambos lugares.

10. Ejemplos en párrafo aparte

- Todos los ejemplos en párrafo aparte deben ir numerados. Esta numeración debe aparecer entre paréntesis.
- Cuando hay múltiples ejemplos dentro de un mismo número (“sub-ejemplos”), se distinguen por las letras *a*, *b*, etc., seguidas de un punto.

- (3) a. Él la vio a ella.
b. Ella lo vio a él.

- En las referencias cruzadas a ejemplos también se emplearán paréntesis, incluso cuando la referencia está dentro de un comentario en paréntesis:

Como se muestra en (6) y (8–11), esta generalización se extiende a las construcciones transitivas, pero (29b) es una excepción.

En todos los demás contextos, el acento recae en la segunda sílaba (véase (15a)).

- A menos que los ejemplos sean del español o el inglés, deberán ser glosados y traducidos (véase abajo §10.1.). El texto de los ejemplos numerados irá en redondas cuando se trate de ejemplos que no requieran glosas. Si el texto principal está escrito en español y el ejemplo está en inglés, el autor debe ofrecer una traducción general, en otra línea entre comillas simples (véase el siguiente ejemplo), y puede omitir la glosa interlineal si no lo considera relevante para la discusión. Este mismo criterio es pertinente para los textos escritos en inglés en los que se ofrezcan ejemplos en español.

- (4) John: Condi is smart but conservative.
‘John: Condi es listo pero conservador.’

- En los ejemplos lingüísticos que no requieren glosas, las fuentes se presentan después del ejemplo; en los ejemplos glosados, directamente después de la traducción idiomática, como se muestra en (5):

(5) Luganda

Maama a-wa-dde taata ssente.
 madre ella.PRS-dar-PRF padre dinero
 ‘La madre le ha dado dinero al padre.’ (Ssekiryango,
 2006: 67)

- Si la fuente no es una referencia bibliográfica, sino el título de un texto (probablemente inédito), como en (6), o el nombre o clave de identificación de un informante o del material elicitado, como en (7), se cita la fuente en corchetes y el artículo deberá incluir una sección especial antes de las Referencias donde se especifique más información sobre dichas fuentes. Esta sección, por lo general, se titula *Corpus* o *Fuentes*.

(6) Jalonke

I sig-aa xon-ee ma.
 2SG ir-IPFV extraño-DEF en
 ‘Vas con el extraño.’ [Mburee 097]

- (7) No entendía la sensación de que era un chiste. Desde el principio pensé que no era y al final no le entendí la gracia. [Alejandra, 5to, chiste 6]

10.1. *Ejemplos glosados*

Cuando un ejemplo proviene de una lengua distinta a la del texto principal, se proporciona una glosa interlineal, alineada palabra por palabra, en la segunda línea. Enseguida se proporciona la traducción idiomática. Véase (8).

(8) Islandés

Storm-ur-inn rak bát-inn.
tormenta-NOM-DEF conducir barca.ACC-DEF
á land.
en tierra
'La tormenta condujo la barca a tierra.'

Las convenciones para el glosado interlineal se pueden encontrar en las Reglas de Glosado de Leipzig,⁴ que son el estándar mundial. El principio básico es que cada elemento del texto primario corresponda a un elemento de la línea de glosa, y los símbolos que marcan los límites entre elementos (especialmente el interno a la palabra (-) y el de límite de clítico (=)) deben estar presentes en el texto primario y la glosa. Las etiquetas de categoría gramatical abreviadas se escriben en versalitas y la traducción idiomática con comillas simples. Se debe proporcionar una lista de abreviaturas al final del trabajo. La correspondencia y alineación entre las palabras del texto primario y las glosas deben ser claras; las palabras y sus glosas deben separarse utilizando la tecla Tab (␣).

En los ejemplos glosados:

- El texto primario del ejemplo va en cursivas, con uso regular de mayúsculas y signos de puntuación: se inician con mayúscula y se cierran con punto final, como en (8) arriba, excepto cuando el ejemplo no es una oración completa, como se muestra en (9).

(9) *das Kind, dem du geholfen hast*
El niño.NOM quien.DAT tú.NOM ayudado has
'el niño a quien ayudaste'

⁴ Las Reglas de Glosado de Leipzig (Comrie, Haspelmath & Bickel, 2008–2015) están disponibles en www.eva.mpg.de/lingua/resources/glossing-rules.php

- Cuando la lengua del ejemplo no suele escribirse, el texto primario puede carecer de mayúsculas iniciales y puntuación normal, por ejemplo:

(10) Hatam

a-yai bi-dani mem di-ngat i
 2SG-tener a-mí para 1SG-ver INT
 ‘¿Me lo darías para que pueda verlo?’ (Reesink, 1999: 69)

- La línea de glosa va en redondas y no lleva mayúsculas ni puntuación. La traducción idiomática tiene nuevamente el uso normal de las mayúsculas y de puntuación, como se ve arriba en (8).
- Cuando se mencionan varias lenguas en un mismo trabajo, se dará el nombre de la lengua en la línea del número del ejemplo, como en (10) y (11). Solo en estos casos el nombre de la lengua comenzará con mayúscula por ser el único elemento en esa línea.

(11) Sakha

a. *En bytaan buol-uoq-uy*
 tú lento ser-FUT-2SG
 ‘Serás lento’ (Baker, 2012: 7)

b. **En bytaan-yaq-yy*
 tú lento-FUT-2SG
 (‘Serás lento’) (Baker, 2012: 7)

- Para los ejemplos agramaticales se proporcionará la traducción idiomática buscada entre paréntesis, como en (11b). También se podrá proporcionar una traducción literal entre paréntesis después de una traducción idiomática, por ejemplo:

(12) Japonés

Tsukue no ue ni hon ga aru.
 mesa GEN alto en libro SUJ estar
 ‘Hay un libro sobre la mesa.’ (Lit. ‘En lo alto de la mesa está un libro.’)

- El texto de la lengua objeto puede aparecer en dos líneas, una no analizada (“superficial”) y una analizada, en caracteres redondos, que puede contener una representación más abstracta, por ejemplo:

(13) Karbi

amatlo la kroikrelo

amāt=lo là krōi-Cē-lò

y.luego=FOC este aceptar-NEG-RL

‘Y luego, ella disintió’ (Konnerth, 2014: 286)

- Los corchetes angulares o diples no van nunca en cursivas, incluso si el texto que aparece dentro de ellos va en cursivas.

11. Cuadros y figuras

- En caso de que se presenten cuadros y/o figuras, estos deberán integrarse en el cuerpo del texto y, además, presentarse por separado en el formato del programa original en el que fueron creadas (por ejemplo, Word, Excel, Praat, ELAN), así como en archivo de imagen (en formato tif, png o jpg). Procure que la resolución de las figuras sea lo más alta posible (no menor a 600 dpi) para que las imágenes sean nítidas en la versión impresa. Asegúrese de que todo elemento gráfico sea pertinente y que contenga los datos necesarios para su lectura.
- El texto debe indicar al lector cuando convenga la consulta de estos objetos gráficos, de aquí que sea necesario darles una numeración. La numeración deberá ser consecutiva y la palabra *figura* o *cuadro* deberá llevar inicial mayúscula (Cuadro 1, Cuadro 2; Figura 1, Figura 2). Los objetos deben insertarse tan cerca como se pueda del punto en el que se los menciona (véase ejemplo *infra*).
- Cada cuadro y figura debe tener un título, idealmente no mayor a una línea, con la palabra *figura* o *cuadro* en versal-versalitas (CUADRO 1, CUADRO 2; FIGURA 1, FIGURA 2) seguida de punto. El título no se cierra con un punto y en él no se hace un uso especial de las

mayúsculas. Este va como encabezado en el caso de los cuadros y al pie en el caso de las figuras.

Ejemplo de una figura y de su mención en el texto principal:

La mayor amplitud la muestra el vocativo de duda, lo que se refleja en un final marcadamente ascendente (Figura 18).

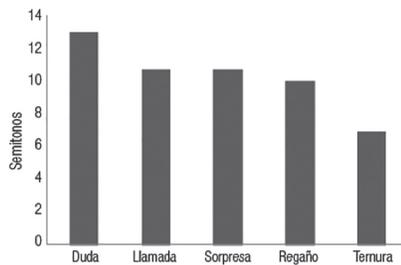


FIGURA 18. Campo tonal de los vocativos en todos los dialectos

FIGURA 18. Campo tonal de los vocativos en todos los dialectos

- Los cuadros generalmente están delimitados por una línea superior y una línea inferior, además de una debajo de los encabezados de columna. Se recomienda editar el cuadro de tal manera que solo sean visibles las líneas necesarias.
- Los cuadros y figuras pueden tener notas y llamadas internas. Estas irán al pie del cuadro y se considerarán independientes a las notas al pie de página en el resto del texto. Bajo el rótulo de *Nota* irán las explicaciones o informaciones generales del cuadro; las llamadas servirán para brindar información específica sobre alguna columna, fila o celda en particular. Estas últimas se indicarán con asterisco(s) (*) o con letras minúsculas (*a*, *b*, *c*) a manera de superíndices (véase ejemplo *infra*). Ejemplo de un cuadro con notas:

CUADRO 1. Porcentaje de la media de respuestas correctas para cada unidad según grado

UNIDAD	POSICIÓN	1ER GRADO ^a		3ER GRADO ^b	
		Media	DE	Media	DE
Sílabas	Inicial	71.49	29.16	94.51	8.6
	Final	57.59	33.56	87.35	12.63
	Rima	62.76	36.14	84.71	22.59

Nota: Los niños de 1^{er} grado son lectores iniciales y los de 3^{er} grado, lectores entrenados.

^a Los niños de 1^{er} grado son lectores iniciales.

^b Los niños de 3^{er} grado son lectores entrenados.

- Como figuras se considerarán las gráficas en eje coordenado o de pastel, diagramas (incluyendo estructuras arbóreas), mapas, dibujos, fotografías e imágenes. Señale, cuando sea necesario, la fuente de las figuras dentro del texto de la nota. Considere que estas irán en blanco y negro en la versión impresa y tome esto en cuenta para su elaboración, sobre todo cuando tenga que mostrar contrastes.

12. Referencias cruzadas en el texto

Para las referencias cruzadas a cuadros, figuras o notas se usan los nombres de estos elementos en mayúsculas (por ejemplo: Figura 3, Cuadro 2, Nota 17). No se emplearán abreviaturas, como “Fig. 3”, “Cap. 4”, o “n. 17”. Las referencias cruzadas a secciones emplearán el carácter § (por ejemplo, §2.3).

13. Notas al pie

El número de referencia para nota al pie debe ir inmediatamente después de un punto o coma, aunque en casos excepcionales irá inmediatamente después de una palabra. Su numeración comenzará con 1. Los agradecimientos se darán en una nota al pie cuya llamada será un asterisco (*) al final del título del trabajo. Las abreviaturas y otras convenciones notacionales, como los sistemas de transcripción, se incluirán en una sección aparte.

Los ejemplos numerados en las notas al pie llevan los números (i), (ii), etc. Si hubiera sub-ejemplos, estos seguirán la numeración (i.a), (i.b), etcétera.

14. Alfabetos no latinos

Todas las formas en lenguas que normalmente no se escriben en alfabeto latino (como japonés o armenio) deben acompañarse de una transcripción o de una transliteración.

Cuando el artículo versa en su totalidad sobre una de estas lenguas, no se omitirá la escritura original, al menos en los ejemplos numerados. Las formas no-latinas no necesitan ir en cursiva.

15. Lista de referencias

La lista de referencias al final del artículo lleva el encabezado Referencias. Las entradas se enlistan alfabéticamente. Se incluirán solo las obras referenciadas explícitamente en el texto.

15.1. Puntos generales

Para el formato de las referencias, *ELA* adopta en casi todos sus aspectos el modelo propuesto por la American Psychological Association (APA, 2010) con algunas excepciones.

Nótese que:

- Se deben incluir los nombres completos de todos los autores y editores, a menos que el autor use habitualmente nombres abreviados (por ejemplo, R. M. W. Dixon). Los nombres se considerarán completos según las convenciones de cada lengua, por ejemplo, los de autores anglosajones irán con un solo apellido y los de autores hispanos con los dos apellidos, salvo cuando sea convención de un autor utilizar uno solo.

- Los números de página son obligatorios tanto en las referencias de artículos de revista como en las referencias de artículos o capítulos en volumen editado. Los números de volumen de las revistas y los títulos de series son opcionales (aunque se recomienda incluirlos).
- Los títulos de revistas, editoriales y asociaciones no se abrevian.
- En el caso de las referencias que hayan sido consultadas en la red, es indispensable incluir el doi o la URL que permita su localización (véase §15.7).

Es responsabilidad del autor corroborar que toda la información de su lista de referencias sea precisa y completa. Se recomienda verificar cada entrada contra la publicación original. Las referencias precisas contribuyen a la credibilidad del autor como investigador cuidadoso y son signo de una ética profesional que valora el trabajo de los demás investigadores.

15.2. Partes estándar y tipos de referencia estándar

Hay cuatro tipos de referencia estándar: artículo de revista, libro, artículo o capítulo en libro editado, tesis. Los distintos tipos de referencia se componen de diferentes partes, como se muestra en el Cuadro 1.

CUADRO 1. Partes estándar de las referencias bibliográficas

	Lista de autores	Año	Título del artículo o capítulo	Lista de editores	Título de la publicación	Número de volumen	Número de páginas	Ciudad	Editorial
Artículo de revista	*	*	*		*	*	*		
Libro	*	*			*			*	*
Capítulo o artículo en libro editado	*	*	*	*	*		*	*	*
Tesis	*	*			*			*	*

Una referencia bibliográfica se conforma por las partes explicadas en el Cuadro 1 (aunque algunas de ellas son opcionales): lista de autores, año, título del artículo o capítulo, lista de editores, título de la publicación, número del volumen, número de páginas, ciudad y editorial. Las partes no estándar pueden añadirse entre paréntesis.

15.3. *Reglas generales de formato*

- Los títulos de artículo o capítulo irán en redondas, sin comillas.
- Los títulos de publicación (tanto de libros como de revistas) aparecerán en cursivas.
- Los títulos inician con mayúscula. En los títulos de artículos y libros en inglés no se emplean mayúsculas en el resto de las palabras (es decir, no se emplea *title case*), salvo en los nombres de revistas y series, y cuando la ortografía de la lengua lo exija. Se sugiere seguir el mismo criterio en títulos en otras lenguas.
- El título principal y el subtítulo se separan con dos puntos. Si es en español, el subtítulo iniciará con letra minúscula; si es en inglés, también iniciará con minúscula a menos que se trate de una oración completa.
- La lista de editores, o compiladores, coordinadores, etc., va seguida de (Ed.) o (Eds.), dependiendo del número de editores, o (Comp.) o (Comps.), etc.
- La ciudad de publicación va seguida de dos puntos.
- Las partes no estándar se pueden añadir al final de una referencia entre paréntesis.
- En las referencias, las abreviaturas, las preposiciones para introducir el volumen editado en el que aparece un capítulo, los nombres de la ciudad de publicación y cualquier información adicional deben ir en la lengua principal del texto. Por ejemplo, en una contribución en español, se usará la preposición “En”, la abreviatura “Ed.” y el nombre “París” en la siguiente referencia:

Erdal, Marcel (2007). Group inflection, morphological ellipsis, affix suspension, clitic sharing. En M. M. Jocelyne Fernandez-Vest (Ed.), *Combat pour les langues du monde: Hommage à Claude Hagège* (pp. 177–189). París: L'Harmattan.

En cambio, cuando la lengua principal del artículo es el inglés, las convenciones de abreviación y especificaciones irán en esta lengua:

Yu, Alan C. L. (2003). *The morphology and phonology of infixation* (Unpublished doctoral dissertation). University of California, Berkeley.

15.4. Tipos estándar de referencia

Estos son algunos ejemplos de los cuatro tipos estándar de referencia: artículos de revistas, libros, artículos en volúmenes editados y tesis:

- Artículo de revista:

▶ Milewski, Tadeusz (1951). The conception of the word in languages of North American natives. *Lingua Posnaniensis*, 3, 248–268.

En los artículos consultados en línea debe agregarse el doi o el URL:

Ender, Andrea (2016). Implicit and explicit cognitive processes in incidental vocabulary acquisition. *Applied Linguistics*, 37(4), 536–560. 10.1093/applin/amu051

Cid Uribe, Miriam Elizabeth, & Ortiz-Lira, Héctor (1998). La conducta prosódica del vocativo en el español culto de Santiago de Chile. *Onomazein*, 3, 143–162. http://www.onomazein.net/Articulos/3/8_Cid.pdf

- Libro. El título va seguido de un punto, ya sea que el libro sea de autor o de editor:

Matthews, Peter (1974). *Morphology*. Cambridge: Cambridge University Press.

Lightfoot, David W. (Ed.) (2002). *Syntactic effects of morphological change*. Oxford: Oxford University Press.

- Artículo o capítulo en volumen editado. La lista de editores va precedida de “En” y seguida de “(Ed.)” o “(Eds.)” y coma. Los nombres de los mismos irán en el orden regular, es decir, nombre-apellido. Siempre se incluirá la información completa sobre el volumen, incluso si se enlistan en las referencias otros artículos del mismo libro. No se requiere enlistar el volumen por separado, a menos que así se le cite en el texto.

Erdal, Marcel (2007). Group inflection, morphological ellipsis, affix suspension, clitic sharing. En Marie-Madeleine Jocelyne Fernandez-Vest (Ed.), *Combat pour les langues du monde: Hommage à Claude Hagège* (pp. 177–189). París: L’Harmattan.

Las presentaciones de congresos publicadas en memorias se pueden tratar como artículos en volúmenes editados:

Fridman, Boris (1996). Gestos y oraciones de la lengua de señas de México. En *Memorias de la xxiv Mesa Redonda de la Sociedad Mexicana de Antropología*. México: Sociedad Mexicana de Antropología/Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México.

- Tesis (el tipo de tesis se menciona en paréntesis y se señala el nombre de la institución, seguido del lugar) de acuerdo al siguiente modelo:

Autor, Nombre (año). *Título de la tesis* (Tesis de licenciatura, maestría o doctoral inédita). Nombre de la institución, Lugar.

Yu, Alan C. L. (2003). *The morphology and phonology of infixation* (Tesis doctoral inédita). University of California, Berkeley.

En el caso de las tesis en línea debe agregarse el URL como en el siguiente ejemplo:

Sanz Martín, Blanca Elena (2008). *Construcciones complejas con el verbo tener: caracterización sintáctica y semántica* (Tesis doctoral). Universidad Nacional Autónoma de México, México. <http://132.248.9.195/ptd2009/enero/0638276/Index.html>

15.5. Partes opcionales

- El volumen de la revista puede ir seguido del número, el cual debe aparecer entre paréntesis en redondas:

Coseriu, Eugenio (1964). Pour une sémantique diachronique structurale. *Travaux de linguistique et de littérature*, 2(1), 139–186.

Puesto que los números de revista son muy útiles para rastrear artículos, se recomienda incluirlos.

- El título de un libro puede ir seguido de información sobre la serie (título de la serie y número del volumen en la serie), las ediciones o el número de volumen, proporcionada entre paréntesis:

Lahiri, Aditi (Ed.) (2000). *Analogy, leveling, markedness: Principles of change in phonology and morphology* (Trends in Linguistics, 127). Berlín: De Gruyter Mouton.

En los títulos de las series se usan las mayúsculas de manera especial, como sucede en los títulos de las revistas.

- En el caso de una reimpresión de un texto antiguo o una referencia clásica se usan los datos de la reimpresión o la edición citada y se incluye el año de la publicación original entre corchetes después del año de la reimpresión o edición citada:

Alonso, Amado (1951 [1933]). *Estilística y gramática del artículo en español*. En *Estudios lingüísticos: temas españoles* (pp. 126–160). Madrid: Gredos.

15.6. *Nombres y apellidos de los autores*

- Los nombres de los autores siempre deberán aparecer en el orden “apellido, nombre”, de manera que quede claro cuáles elementos pertenecen al apellido y cuáles al primer nombre. Si el segundo nombre en el ejemplo siguiente estuviera dado en el orden “nombre apellido” (Francisco José Ruiz de Mendoza), la segmentación de los apellidos no sería clara para un lector no hispanohablante.

Pérez Hernández, Lorena, & Ruiz de Mendoza, Francisco José (2002). Grounding, semantic motivation, and conceptual interaction in indirect directive speech acts. *Journal of Pragmatics*, 34(3), 259–284.

En el caso de los traductores, editores, coordinadores, etc., cuyo nombre aparece no al inicio de la referencia, si no en el cuerpo de la misma, el orden corresponde al normal, es decir, nombre-apellido.

- Cuando hay más de dos autores o editores, cada uno se separa por punto y coma, con excepción del último nombre, que es precedido por coma seguida de *et* (&). No se omite el nombre de ningún autor, es decir, en la sección de referencias no se usa *et al.*

Chelliah, Shobhana, & de Reuse, Willem (2010). *Handbook of descriptive linguistic fieldwork*. Dordrecht: Springer.

Johnson, Kyle; Baker, Mark, & Roberts, Ian (1989). Passive arguments raised. *Linguistic Inquiry*, 20, 219–251.

- Los apellidos compuestos o complejos no se tratan de manera especial y se alfabetizan bajo la primera parte, aunque comiencen con letra minúscula. Así, los apellidos alemanes o neerlandeses que empiezan con *von* o *van* (por ejemplo *van Riemsdijk*) o los apellidos neerlandeses, franceses o españoles que empiezan con *de* (por ejemplo, *de Groot* o *de la Garza*) se tratan como los apellidos belgas (por ejemplo, *De Schutter*) y los apellidos italianos (por ejemplo, *Da Milano*). Los siguientes nombres están ordenados alfabéticamente como se indica (es decir, de manera mecánica):

Da Milano, Federica > de Groot, Casper > de Saussure, Ferdinand > De Schutter, Georges > van der Auwera, Johan > Van Langendonck, Willy > van Riemsdijk, Henk > von Humboldt, Wilhelm

- Los nombres chinos y coreanos se podrán tratar de manera especial. Dado que los apellidos no son muy distintivos, se puede dar el nombre entero dentro del texto, por ejemplo:

la negación neutral *bù* es compatible con verbos estativos y de actividad (cf. Hsieh Miao-Ling, 2001; Lin Jo-wang, 2003; Teng Shou-hsin, 1973)

15.7. Referencias electrónicas (publicaciones en internet)

- Las publicaciones regulares que son accesibles en línea no se tratan de manera especial, excepto que deberán llevar al final de la referencia el número de Digital Object Identifier (doi) o, en caso de no tenerlo, el URL (es decir, la dirección electrónica del recurso referido).

Ender, Andrea (2016). Implicit and explicit cognitive processes in incidental vocabulary acquisition. *Applied Linguistics*, 37(4), 536–560. 10.1093/applin/amu051

Cid Uribe, Miriam Elizabeth, & Ortiz-Lira, Héctor (1998). La conducta prosódica del vocativo en el español culto de Santiago de Chile. *Onomazein*, 3, 143–162. http://www.onomazein.net/Articulos/3/8_Cid.pdf

- Cuando se cita un recurso de la red que no es una publicación científica regular, por ejemplo, una página de internet, esta se deberá tratar como una referencia impresa, en la medida de lo posible:

Redish, Laura, & Lewis, Orrin (1998–2015). Vocabulary in Native American languages: Salish words. http://www.native-languages.org/salish_words.htm

- Previo al envío y/o publicación del artículo es necesario corroborar que los URL de las referencias electrónicas indiquen la ubicación correcta. Cuando esto no ocurra, es necesario actualizar el URL o sustituirlo por otra fuente.

15.8. *Miscelánea*

- Los libros pueden incluir entre paréntesis información sobre ediciones, números de volúmenes y números de páginas inmediatamente después del título y con el punto después del paréntesis:

Croft, William (2003). *Typology and universals* (2a. ed.). Cambridge: Cambridge University Press.

Rissanen, Matti (1999). Syntax. En Lass, Roger (Ed.), *Cambridge history of the English language* (Vol. 3, pp. 187–331). Cambridge: Cambridge University Press.

- Si una editorial está asociada a varias ciudades, solo es necesario dar la primera, por ejemplo: “Berlín: De Gruyter Mouton” o “Ámsterdam: Benjamins”.
- Los títulos de los trabajos escritos en lenguas distintas al español, inglés o francés pueden ir acompañados de una traducción, que debe aparecer entre corchetes cuadrados después del título en la lengua original:

Li, Rulong (1999). *Minnan fangyan de daici* [Demostrativos y pronombres personales en min del sur]. En Rulong Li & Song-Hing Chang (Eds.), *Daici* [Demostrativos y pronombres personales] (pp. 263–287). Guangzhou: Ji'nan University Press.

- Si el título no solo está en una lengua distinta, sino también en una escritura diferente, se puede proporcionar la escritura original entre paréntesis, además de la transliteración. Del mismo modo, el nombre del autor se podrá dar entre paréntesis después del nombre transliterado, como en el siguiente ejemplo:

Plungian, Vladimir A. (Плунгян, Владимир А.) (2000). *Obščaja morfologija: Vvedenie v problematiku* (Общая морфология: Введение в проблематику) [Morfología general: introducción a los problemas]. Moscú: Editorial URSS.

- En el caso de las ponencias no publicadas, se incluirá el mes y año de la reunión (aunque se recomienda que solo se citen ponencias de congresos recientes, si se puede suponer que el material será publicado eventualmente):

Filppula, Markku (septiembre, 2013). Areal and typological distributions of features as evidence for language contacts in Western Europe. Ponencia presentada en el congreso de la Societas Linguistica Europaea, Split, Croacia.

Para resolver otros casos específicos no contemplados en estos lineamientos, se recomienda a los autores consultar el manual de publicaciones de la American Psychological Association (APA) en su última versión o su sitio web (www.apastyle.org/index.aspx), donde encontrarán un extenso inventario de modelos de fichas para casos especiales (*software*, medios audiovisuales, conjuntos de datos, referencias virtuales, etcétera).

16. Referencias

- American Psychological Association (apa) (2010). *Manual de publicaciones de la American Psychological Association* (3a. ed.). México: El Manual Moderno.
- Comrie, Bernard; Haspelmath, Martin, & Bickel, Balthasar (2008–2015). Leipzig glossing rules. <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>
- Haspelmath, Martin (2014). *The generic style rules for Linguistics*. <http://www.eva.mpg.de/lingua/pdf/GenericStyleRules.pdf>
- Real Academia Española (RAE) (2010). Principales novedades de la última edición de la *Ortografía de la lengua española* (2010). http://www.rae.es/sites/default/files/Principales_novedades_de_la_Ortografia_de_la_lengua_espanola.pdf
- Real Academia Española (RAE) & Asociación de Academias de la Lengua Española (Asale) (2010). *Ortografía de la lengua española*. Madrid: Espasa.

Distribución, venta y suscripciones

Katsumi Yoshida Hernández
Escuela Nacional de Lenguas, Lingüística
y Traducción (ENALLT), Circuito Interior s/n,
Ciudad Universitaria, Alcaldía Coyoacán,
c. p. 04510, Ciudad de México, México
ela@enallt.unam.mx

Números sueltos y suscripciones	Precio unitario	Precio con envío incluido			
		República Mexicana	Zona 1	Zona 2	Zona 3
Ejemplar regular Cualquier número, excepto <i>ELA</i> número 50	\$100 <small>MXN</small>	\$130 <small>MXN</small>	USD \$22	USD \$25	USD \$29
Ejemplar especial <i>ELA</i> número 50	\$200 <small>MXN</small>	\$230 <small>MXN</small>	USD \$35	USD \$38	USD \$41
Suscripción regular 2 ejemplares regulares	\$160 <small>MXN</small>	\$220 <small>MXN</small>	USD \$40	USD \$45	USD \$53
Suscripción especial 1 ejemplar regular más <i>ELA</i> número 50	\$240 <small>MXN</small>	\$300 <small>MXN</small>	USD \$50	USD \$55	USD \$63
Zona 1: Norteamérica y Centroamérica Zona 2: Sudamérica y Europa Zona 3: Resto del mundo					

Índice de números anteriores
<http://ela.enallt.unam.mx>





se terminó de imprimir en febrero de 2024 en los talleres de Solar, Servicios Editoriales, s. a. de c. v., ubicados en Calle 2 número 21, San Pedro de los Pinos, Alcaldía Benito Juárez, c. p. 03800, Ciudad de México, con un tiraje de 200 ejemplares. Se utilizó papel bond blanco de 90 g en los interiores y couché de 250 g para los forros. En su composición se utilizaron los tipos Times LT Std y Futura Std.

Coordinación editorial

Enio Ramírez Campos

Edición y cuidado editorial

Mariela Elizabeth Pérez Castañeda

Claudia Isabel Carrillo Sánchez

Corrección de estilo

Armando Rodríguez Briseño

Corrección de textos en inglés

María Elena Sánchez Salazar

Diseño de portada

Soledad Martínez Piña

Diseño de interiores y formación

Héctor Sandoval Sandoval

ARTÍCULOS ►

Autorreparaciones en la afasia anómica en español.
Un estudio de caso 7
Self-repairs in Spanish anomic aphasia. A case study
Olivia Castillo Castillo

Técnicas básicas de procesamiento del lenguaje natural en
el análisis de menciones a mujeres en lecturas graduadas de ELE 45
Basic natural language processing techniques for the analysis
of mentions of women in Spanish graded readers
Iban Mañas Navarrete & María-Valle Sell

NOTAS ►

Productividad relativa de los esquemas de formación
de palabras en el español de México. Actualización 2022 79
Relative productivity of word formation schemes
in Mexican Spanish. Update 2022
Ramón Zacarías Ponce de León

Diseño de tareas y materiales para recopilar un corpus
de aprendices mexicanos de lengua inglesa 95
Design of tasks and materials to collect a corpus
of Mexican learners of English
Ana Abigahil Flores Hernández & Pauline Moore

RESEÑA ►

Daniela Lauria. *Lengua y política. Historia crítica de
los diccionarios del español de la Argentina* 127
Luis B. Quesada Nieto

