

# SIGNOS LINGÜÍSTICOS

Revista semestral • Departamento de Filosofía • CSH/UAM/Iztapalapa

Variación en anáfora de frase nominal: objetos nulos y pronombres de objeto directo



Los títulos de las revistas científicas cubanas: estudio exploratorio



Léxico, gastronomía y cultura: Loja (Ecuador) y Cerro Azul, Veracruz (México)



Los pronombres personales yo y tú en la interacción oral de aprendientes sinohablantes de español



La Biblia como corpus para tecnologías del lenguaje

# SIGNOS LINGÜÍSTICOS

## **Directora fundadora**

Laura A. Hernández Martínez

## **Directora**

Elizabeth Santana Cepero

## **Consejo de redacción**

Milagros Alfonso Vega

Edgar A. Madrid Servín

Laura A. Hernández Martínez

Lucio Armando Mora Bustos

María del Refugio Pérez Paredes

Julio César Serrano Morales

María Virginia Mercau Appiani

Irma Munguía Zatarain

## **Comité editorial**

Raul Ávila (El Colegio de México) • Adriana Bolívar (Universidad Central de Venezuela) • Ignacio Bosque (Universidad Complutense de Madrid) • Teresa Carbó (Centro de Investigaciones y Estudios Superiores en Antropología Social-D.F.) • Concepción Company Company (Universidad Nacional Autónoma de México) • Gabriela Coronado (Western Sydney University) • Violeta Demonte (Universidad Autónoma de Madrid) • Teun Van Dijk (Universitat Pompeu Fabra) • Zarina Estrada (Universidad de Sonora) • Rafael Núñez Cedeño (University of Illinois) • Dora Pellicer (Escuela Nacional de Antropología e Historia) • Gemma Rigau (Universidad Autónoma de Barcelona) • María Luisa Rivero (University of Ottawa) • Lidia Rodríguez Alfano (Universidad Autónoma de Nuevo León) • Juan Uriagereka (University of Maryland) • Marisela del Carmen Pérez Rodríguez (Universidad de Oviedo) ••••••

•

•

•

•

•



Casa abierta al tiempo

**UNIVERSIDAD AUTÓNOMA METROPOLITANA**

**Dr. José Antonio de los Reyes Heredia**

RECTOR GENERAL

**Dra. Norma Rondero López**

SECRETARIA GENERAL

**UNIDAD IZTAPALAPA**

**Dra. Verónica Medina Bañuelos**

RECTORA

**Mtra. Elizabeth Santana Cepero**

DIRECTORA

**Dr. Javier Rodríguez Lagunas**

SECRETARIO

**Dra. Sonia Pérez Toledo**

DIRECTORA DE LA DIVISIÓN DE CIENCIAS SOCIALES Y HUMANIDADES

D. R © UAM-Iztapalapa

Departamento de Filosofía

División de Ciencias Sociales y Humanidades

Av. Ferrocarril San Rafael Atlixco, núm. 186

Col. Leyes de Reforma, 1ª. sección, Alcaldía

Iztapalapa, Ciudad de México, 09340, México

**Dra. Marta Ortega Soto**

JEFA DEL DEPARTAMENTO DE FILOSOFÍA

- 
- Índices y bases de datos en donde aparece la revista: Fuente académica-EBSCO, Latindex, CLASE (Citas Latinoamericanas en Ciencias Sociales y Humanidades-UNAM); Lingmex, Bibliografía Lingüística de México desde 1970.
  - *Signos Lingüísticos*: año 24, vol. XX, núm. 39, enero-junio de 2024, es una publicación semestral editada por la Universidad Autónoma Metropolitana, a través de la Unidad Iztapalapa, División de Ciencias Sociales y Humanidades, Departamento de Filosofía, Prolongación Canal de Miramontes, núm. 3855, Col. Ex-Hacienda San Juan de Dios, Alcaldía Tlalpan, C.P. 14387, Ciudad de México, y Avenida Ferrocarril San Rafael Atlixco, núm. 186, Col. Leyes de Reforma 1a. Sección, Alcaldía Iztapalapa, C.P. 09310, Ciudad de México, Teléfono 5804-4600, ext. 2786. Página electrónica de la revista: <https://signoslinguisticos.izt.uam.mx>, correo electrónico [sili@xanum.uam.mx](mailto:sili@xanum.uam.mx), [slingui@gmail.com](mailto:slingui@gmail.com). Editora Responsable: Elizabeth Santana Cepero. Certificado de Reserva de Derechos al Uso Exclusivo del Título número 04-2022-112312032000-102, ISSN: en trámite, ambos otorgados por el Instituto Nacional del Derecho de Autor. Responsable de la última actualización de este número: Elizabeth Santana Cepero, Departamento de Filosofía, División de Ciencias Sociales y Humanidades, Unidad Iztapalapa, Avenida Ferrocarril San Rafael Atlixco, número 186, Col. Leyes de Reforma 1 A Sección, Alcaldía Iztapalapa, C. P. 09310, Ciudad de México; fecha de última modificación: 16 de diciembre de 2024. Tamaño del archivo 4.40 MB.
  - Corrección de estilo: Sonia Elisa Morett Álvarez, Daliana del Carmen Rodríguez Campos
  - Formación: Daliana del Carmen Rodríguez Campos, Sonia Elisa Morett Álvarez
  - Edición de imágenes y Arte Final: Daliana del Carmen Rodríguez Campos
  - Las opiniones expresadas por los autores no necesariamente reflejan la postura del editor responsable de la revista.
  - Queda estrictamente prohibida la reproducción total o parcial de los contenidos e imágenes de la publicación sin previa autorización de la Universidad Autónoma Metropolitana.

# SIGNOS LINGÜÍSTICOS

.

.

.

.

.

39

enero-junio 2024

## CONTENIDO

<b>ARTÍCULOS</b>	8	Variación en anáfora de frase nominal: objetos nulos y pronombres de objeto directo LAURA MELANY FLORES MORALES
	38	Los títulos de las revistas científicas cubanas: estudio exploratorio DIANA ELENA PRIETO ACOSTA / EVELYN RUBIO FIGUERAS
	58	Léxico, gastronomía y cultura: Loja (Ecuador) y Cerro Azul, Veracruz (México) SONIA E. MORETT ÁLVAREZ / ISIS NATHALY ZAMBRANO OJEDA
	104	Los pronombres personales <i>yo</i> y <i>tú</i> en la interacción oral de aprendientes sinohablantes de español LIU, MUZI / LIDIA ESTER CUBA VEGA / MARITZA GUERRERO HIDALGO
<b>NOTA</b>	136	La Biblia como corpus para tecnologías del lenguaje GERARDO E. SIERRA MARTÍNEZ / GEMMA BEL ENGUIX

- RESEÑAS 170 Auza Benavides, Alejandra (2021), *Semillas del lenguaje: Desarrollo típico y atípico en pequeños hablantes del español*, Minden, (Nevada), Barker and Jules, 149 pp.  
SILVIA ROMERO-CONTRERAS
- 174 Araiza, Karina y Stephen A. Marlett (2021), *Conceptos de lingüística descriptiva: Un acercamiento a las lenguas del mundo*, México, Instituto de Lingüística de Verano, 164 pp  
GABRIELA GARCÍA SALIDO / MANUEL ANDRÉS PÉREZ-PÉREZ
- 181 **Normas editoriales**





## ARTÍCULOS

## Variation in noun phrase anaphora: null objects and direct object pronouns

LAURA MELANY FLORES MORALES 

Universidad Nacional Autónoma de México

[laura.melany.fm@gmail.com](mailto:laura.melany.fm@gmail.com)

**Abstract:** *This paper presents quantitative results on the variation between the presence of direct object clitics –lo(s) and la(s)– and their omission –null object or zero anaphora– when the antecedent is a noun phrase in the monolingual Spanish of Mexico City from a variationist perspective. The set of variables analyzed is related to the features of the nominal antecedent. As will be shown, this variety of Spanish shows a reduction in the restrictions for the use of the null object.*

**KEYWORDS:** MONOLINGUAL SPANISH; DIALECTAL VARIATION; ACCUSATIVE CLITICS; ZERO ANAPHORA; VARIATIONIST SOCIOLINGUISTICS

**RECEPTION:** 15/12/2022

**ACCEPTANCE:** 22/08/2024

# Variación en anáfora de frase nominal: objetos nulos y pronombres de objeto directo

LAURA MELANY FLORES MORALES 

Universidad Nacional Autónoma de México

[laura.melany.fm@gmail.com](mailto:laura.melany.fm@gmail.com)

**Resumen:** En este artículo se presentan los resultados cuantitativos sobre la variación existente entre la presencia de los clíticos de objeto directo *–lo(s)* y *la(s)*– y su omisión –objeto nulo o anáfora cero– cuando el correferente es una frase nominal en el español monolingüe de la Ciudad de México, desde una perspectiva variacionista. El conjunto de variables analizadas está relacionado con los rasgos del correferente nominal. Como se demostrará, esta variedad del español muestra una reducción en las restricciones para el uso del objeto nulo.

**PALABRAS CLAVE:** ESPAÑOL MONOLINGÜE; VARIACIÓN DIALECTAL; CLÍTICOS ACUSATIVOS; ANÁFORA CERO; SOCIOLINGÜÍSTICA VARIACIONISTA

RECEPCIÓN: 15/12/2022

ACEPTACIÓN: 22/08/2024

## INTRODUCCIÓN

En español normativo se señala que los nominales indefinidos e inespecíficos como los sustantivos de masa y genéricos –*pipián, dinero, mota*– o los plurales escuetos –*sillas y mesas, o pantalones*– muestran resistencia a ser sustituidos por los pronombres átonos de acusativo: *lo, la, los, las*. (Campos, 1986: 354; RAE y ASALE, 2010: 656-657).

De acuerdo con lo anterior, cuando en una oración aparece un nominal definido y específico como en (1), un/el libro, éste puede ser correferencial con un clítico acusativo como en (1a); en este caso, lo; por el contrario, se trata de una oración agramatical si se codifica con un objeto nulo como en (1b).

- (1) Compré **un/el libro**<sub>i</sub>  
 a. **Lo**<sub>i</sub> compré  
 b. \***Ø**<sub>i</sub> compré  
 (Campos, 1986: 354)

En cambio, cuando el correferente es indefinido y no específico como *café* en (2), la anáfora cero es posible como objeto directo del verbo *comprar*:

- (2) ¿Compraste **café**<sub>i</sub>?  
 Sí, **Ø**<sub>i</sub> compré  
 (Campos, 1986: 354)

Uno de los trabajos pioneros en el estudio del objeto nulo en español es el de Campos (1986). En dicho trabajo, además de los rasgos de definitud y especificidad, otras restricciones de orden sintáctico que se postularon para el objeto nulo fueron los entornos complejos, como las oraciones subordinadas de relativo; las oraciones subordinadas de sujeto; las oraciones interrogativas parciales en las que la incógnita no es el OD, y las oraciones adjuntas (Campos, 1986: 355-358).

Posteriormente, se han realizado investigaciones descriptivas sobre el objeto nulo para otras variedades monolingües del español: dialectos castellanos septentrionales (Clements, 2006); el español de Río de la Plata (Masullo, 2017); así como las variedades mexicana y peninsular (Reig Alamillo, 2009).

En esta línea, Clements (2006) analizó el objeto nulo conforme a la Hipótesis de transitividad propuesta por Hopper y Thompson (1980), la cual descompone la transitividad en el siguiente conjunto de valores: i) participantes, ii) kinesis, iii) aspecto, iv) puntualidad, v) volición, vi) polaridad, vii) modo, viii) agentividad, ix) afectación del objeto y, finalmente, x) individualización del objeto. De acuerdo con lo anterior, el español diferenciaría oraciones más o menos transitivas dependiendo de sus rasgos. Cuando alguno de los valores arriba mencionados está presente, se incrementa la transitividad y es cuando se emplean los pronombres acusativos; mientras que, si los valores están ausentes, la oración presenta una transitividad menor y dispara el objeto nulo como mecanismo anafórico; finalmente, si la transitividad es media y el correferente es indefinido e inespecífico, se produce una variación entre los clíticos y la marca cero (Clements, 2006).

Por otra parte, Masullo (2017), dentro del generativismo como marco teórico, encontró que la caída de objeto de acusativo no identificado morfológicamente está restringida pragmáticamente —el correferente debe ser prominente en el contexto situacional—, por lo que el objeto nulo tendrá que ser recuperable del contexto de habla inmediato: “no podrían referirse a cosas que no están en el campo visual o perceptivo de los participantes” (2017: 64), lo cual se formaliza en el aspecto puntual y la selección de tiempo (presente, pretérito o futuro inmediato).

Desde otra perspectiva metodológica, el variacionismo, Reig Alamiello (2009) estudió los objetos directos nulos con antecedentes proposicionales, y los codificados con el clítico acusativo neutro *lo*, en las variedades mexicana y peninsular.

La autora concluyó que la frecuencia de uso del pronombre de objeto directo y del nulo es muy distinta entre las dos variedades: la peninsular marca canónicamente con *lo* un 70 % de los casos; por su parte, la mexicana utiliza el pronombre acusativo en un 17 %. Por lo tanto, la variante nula es usada el 30 % de las veces por los hablantes peninsulares, mientras que la variedad mexicana registra un 83 % de marcas nulas de objeto directo con antecedente proposicional.

En tal estudio, las variables estadísticamente significativas para el español peninsular resultaron ser las siguientes: i) pronombre dativo y ii) tipo de oración; mientras que para el español mexicano fueron: i) adverbio de manera,

ii) tipo de antecedente, iii) pronombre dativo, y iv) tipo de oración. Con estos resultados, la autora observó que las restricciones que condicionan la variación para los dos dialectos monolingües son las mismas, pero con diferente jerarquización, lo que indica que comparten una gramática subyacente en el uso anafórico del objeto directo proposicional.

Cabe resaltar que el grueso de las investigaciones sobre el objeto nulo en español parte de distintas situaciones de contacto lingüístico. Estos estudios han demostrado que las lenguas que no son parte de la familia indoeuropea actúan como aceleradores del cambio lingüístico, por lo que la marca cero del pronombre de objeto directo es más frecuente en estas variedades, debido a la reducción de fuertes restricciones como la definitud y especificidad o la animacidad del correferente.

El uso del objeto nulo en las situaciones de contacto, además, apunta hacia una reorganización del sistema pronominal del OD átono de tercera persona como parte de un proceso de gramaticalización que acerca a los pronombres referenciales a marcadores de concordancia de objeto directo (Palacios Alcaine, 2013).

Los estudios de contacto muestran que las variables que suelen favorecer la omisión de objetos directos con correferentes nominales son: i) la jerarquía de animacidad (García Tesoro y Fernández Mallat, 2015;<sup>1</sup> Palacios Alcaine, 2015;<sup>2</sup> Camus Bergareche y Gómez Seibane, 2015;<sup>3</sup> Torres Sánchez, 2018);<sup>4</sup> ii) la marca dativa relacionada con los entornos de tres participantes (Palacios Alcaine, 2015; Camus Bergareche y Gómez Seibane, 2015); iii) el verbo en infinitivo o gerundio (García Tesoro y Fernández Mallat, 2015; Palacios Alcaine, 2015; Camus Bergareche y Gómez Seibane, 2015; Torres Sánchez, 2018); iv) la clase semántica verbal (García Tesoro y Fernández Mallat, 2015; Torres Sánchez, 2018), y v) la definitud y especificidad (García Tesoro y Fernández Mallat, 2015).

<sup>1</sup> Los autores estudiaron la convergencia lingüística con la variedad andina quechua de Perú.

<sup>2</sup> La investigación describe el contacto del español con la variedad andina quichua de Ecuador.

<sup>3</sup> Los datos provienen de la situación de contacto lingüístico en el País Vasco entre el castellano y la lengua vasca.

<sup>4</sup> Este es un estudio sobre el sistema pronominal de bilingües tepehuano-español en Santa María de Ocotán y Durango.

En el presente artículo, se explora el efecto de un conjunto de variables lingüísticas y sociales respecto al objeto nulo en el español monolingüe de la Ciudad de México, como en (3), donde se observa la aparición de la anáfora cero como complemento del verbo transitivo *lavar*, así como la presencia del clítico de objeto directo masculino singular *lo* en la estructura argumental de los verbos *sacar* y *tender*. Los tres mecanismos anafóricos comparten el correferente nominal *su ropa*:

- (3) I: no tienen// o sea ella va/ mete **su ropa**<sub>i</sub> a la máquina y ya estubo  
 E: pura lavadora  
 I: sí ya/ [ $\emptyset$ ]<sub>i</sub> lava/ **lo**<sub>i</sub> saca/ y **lo**<sub>i</sub> tiende y ya// sí entonces <-ntos>/ pues  
 <-ps> ella se sorprendía aquí/ porque la gente tendía a sus [azoteas]  
 (CSCM, entrevista 4)

De esta forma, se podrá observar el estado del fenómeno en esta variante y se presentará el comportamiento porcentual de seis variables relacionadas con los rasgos del correferente nominal que son la jerarquía de animacidad, la definitud, la especificidad, el rasgo contable o no contable, el número y el género gramatical; asimismo, se mostrará la distribución porcentual de las variables sociales recuperadas del *Corpus Sociolingüístico de la Ciudad de México* (Martín Butragueño y Lastra, 2011-2015), de ahora en adelante CSCM. Dichas variables son el género, grupo etario y nivel de instrucción.

En cuanto a la organización del artículo, en la metodología se expone la procedencia de los contextos que conforman la base de datos, la manera en la que se compilaron y la justificación de su selección; igualmente, se incluyen las hipótesis y el esquema de codificación de cada una de las variables, tanto lingüísticas como sociales. Posteriormente, en el tercer apartado, se presentan los resultados porcentuales de las variables analizadas, y finalmente, se encuentra la conclusión.

Por último, cabe aclarar que esta investigación forma parte de un estudio más amplio que toma en cuenta otras variables lingüísticas, tanto de orden sintáctico como factores pragmático discursivos; sin embargo, presentar los resultados de estos dos grupos de variables queda fuera del objetivo del presente artículo.

## METODOLOGÍA

Como se mencionó, en esta sección, primeramente, se describe el corpus que se utilizó y la selección de la muestra de entrevistas. En segundo lugar, se exhiben las decisiones que se tomaron para incluir contextos en la base de datos. Y, además, se presenta el esquema de codificación de las variables independientes y las hipótesis de las que se parte.

### El corpus y la muestra de entrevistas

El corpus utilizado fue el *Corpus Sociolingüístico de la Ciudad de México* (Martín Butragueño y Lastra, 2011-2015), mismo que forma parte del Proyecto para el estudio sociolingüístico del español de España y de América (PRESEEA, 2014), cuyo objetivo general es la creación de un corpus de lengua española hablada, representativo del mundo hispánico, en sus variedades geográficas y sociales.

El contenido disponible para consultarse en línea forma parte del módulo nuclear; contiene los registros de 108 personas distribuidos en muestras establecidas por cuotas con afijación uniforme. Para el caso de la Ciudad de México, conforme a Martín Butragueño y Lastra (2011-2015), se optó por 18 cuotas o casillas, con seis informantes para cada una, estratificadas por tres variables sociales: i) género, con las variantes mujer y hombre; ii) edad, clasificada en tres rangos –jóvenes (20-34 años), adultos (35-54 años) y mayores (55 y más años)–, y iii) nivel de instrucción con tres clases –alto, medio y bajo–.

En cuanto a las entrevistas que se seleccionaron para esta investigación, se eligió una por cada casilla del CSCM; por lo tanto, la muestra quedó conformada por 18 colaboradores, nueve hombres y nueve mujeres, distribuidos equitativamente en el universo relativo estudiado, según la estratificación establecida por las variables sociales exigidas por el proyecto PRESEEA; es decir, un participante por grupo de edad<sup>5</sup> –de cada uno de los tres grados de instrucción–, tanto para los hombres como para las mujeres, tal como se muestra en la tabla 1.

<sup>5</sup> Se emplean las abreviaturas J para los jóvenes, A para los adultos y M para los mayores.



TABLA 1. MUESTRA DE ENTREVISTAS

Nivel de instrucción		Alto			Medio			Bajo		
		J	A	M	J	A	M	J	A	M
Edad										
Género	Hombre	1	1	1	1	1	1	1	1	1
	Mujer	1	1	1	1	1	1	1	1	1
<b>TOTAL</b>					<b>18</b>					

Una vez seleccionadas las entrevistas, se procedió a generar la muestra de verbos. Para ello, se retomó la metodología aplicada por Reig Alamillo (2009). Esto es, se buscaron los contextos en los que había un pronombre de objeto directo (*lo(s)* y *la(s)*), ya fuera como enclítico (*comerlo*) o como proclítico (*lo comeré*); después, se analizó si estos verbos aceptaban, además, un objeto directo nulo, de tal suerte que solo se incluyeron los verbos que presentaron tanto pronombres acusativos como objetos nulos. De este modo, se seleccionaron, para cada una de las entrevistas del corpus de esta investigación (tabla 1), los contextos de la variable dependiente: la presencia del clítico acusativo o el objeto nulo correferenciales con frases nominales.

Al recopilar los datos, se generaron cuatro criterios para definir el contexto variable. Primero, (i) que los correferentes del pronombre de objeto directo y del objeto nulo debían ser rastreables, es decir, mencionados previamente en el discurso. Por consiguiente, en la base de datos se incluyen casos como (4), en el que se encuentran dos ocurrencias del verbo transitivo *quitar* con la marca nula de objeto directo; asimismo, se observa que en turnos anteriores aparece la frase nominal *los animalitos* como último nominal correferencial con el objeto nulo en la estructura argumental de dichos verbos resaltados en (4).

- (4) I: [...] estaba pasando la visita// y de repente/ empiezo a ver que un niño se estaba ahogando// y/ no vas a creer/ que volteo a ver/ y estaba arrojando lombrices  
 E: [¿en serio?]  
 I: [por la boca]  
 E: ¡ahh!/ ¡qué barbaridad!  
 I: ¿y qué crees que pude hacer?  
 E: pues <-pus> no sé

- I: ¡pues <-pus> quitarle las lombrices!/ no lo voy a dejar que se muera/  
 [¡imagínate!]
- E: [pues sí]
- I: o sea/ imagínate l- el asco/ la náusea/ la cuestión que te puede dar que/  
 ves **los animalitos**, y dices/ [“guác-”/ así/ ¿no?]
- E: [uh/ sí]
- I: pues en ese momento
- E: te vale
- I: quita [Ø<sub>i</sub>]/ ¡pues sí mi reina!/ no es cuestión de que lo pienses/ [es cuestión]
- E: [claro]
- I: de que le [Ø<sub>i</sub>] quitas/ y lo ves/ y lo haces reaccionar/ porque lo están  
 ahogando este// el <-el::>/ el nudo de lombrices que le estaba subiendo  
 subiendo por la garganta  
 (CSCM, Entrevista 20)

Un segundo criterio para incluir datos fue (ii) que tanto el nominal como el pronombre y el objeto nulo fueran conmutables entre sí sin modificar el significado de la oración huésped. Esto se muestra en el ejemplo (5), donde originalmente hay un objeto nulo correferencial con *la calle de San Pablo* en la estructura argumental de *conocer*. Como se puede observar, al modificar el contexto agregando el pronombre acusativo *la* o repitiendo la frase nominal con función de objeto –ambos entre llaves–, el resultado es un enunciado adecuado y gramatical. De esta forma, se pretende demostrar que el proclítico *la*, la frase nominal en la posición prototípica del objeto directo (después del verbo), y el objeto nulo son conmutables entre sí.

- (5) I: todo Circunvalación// y justo en **la calle de San Pablo**<sub>i</sub>/ si [Ø<sub>i</sub>]/{**la**<sub>i</sub>}  
 conoces {**la calle de San Pablo**} o has oído hablar de la [la calle]
- E: [mh]
- I: de San Pablo/ [que es]
- E: [mh]
- I: la que sale a/ a/ a Calzada de Tlalpan  
 (CSCM, Entrevista 8)

El tercer criterio de selección que se siguió fue (iii) el principio de exhaustividad o responsabilidad (Silva Corvalán y Enrique Arias, 2017: 71;

Tagliamonte, 2006: 13), cuyo fin es poder analizar el fenómeno en su totalidad, mediante el registro de todos los contextos donde la variación es posible.

Como el propósito de este artículo es analizar los factores que propician el objeto nulo, los contextos que se recolectaron, de acuerdo con el principio de exhaustividad, fueron aquellos en los que se esperaba un pronombre de objeto directo y, en su lugar, se encontró la anáfora cero. En consecuencia, no se incluyeron los contextos en los que el elemento correferente con un objeto directo era un pronombre personal tónico (*él, ella, usted*), porque se observó que, en el corpus analizado, éstos siempre se recuperan mediante el pronombre átono de objeto directo. Del mismo modo, se excluyeron las construcciones llamadas relativas reasuntivas o pleonásticas (RAE y ASALE, 2010: 852; Silva Corvalán y Enrique Arias, 2017: 196) en las que tampoco se documenta el objeto nulo. Esta etiqueta se refiere a relativas especificativas o restrictivas que se pueden parafrasear con *el (lo(s)/la(s) cual(es)*, en donde el nominal del que se predica se duplica dentro de la subordinada, como en (6) donde *Chapultepec* <~*chapultepé*> se duplica dentro de la relativa con el pronombre *lo*.

- (6) I: [...] me decía mi mamá/ “ay/ ya nada más empiezas a ponerte tu tubería/ y es que ya/ te vas a ir al-/ allá afuera/ ¿verdad <~verdad>?”// “ay mamá/ ¿a dónde vamos?/ no vamos a ningún lado”/ y de veras ¿eh?/ no salíamos a ningún lado/ ni a **Chapultepec**, <~**chapultepé**>, que **lo**<sub>i</sub> teníamos tan cerca (CSCM, Entrevista 106)

El último criterio fue (iv) de exclusión: no se incluyeron en el análisis las construcciones verbales invariables o fijas, como las locuciones verbales (Romera Martín, 2017). Un ejemplo de ello se muestra en (7), donde aparece el nominal escueto *pecho* en la construcción *dar yo pecho* con la función de objeto directo. Por esto, cuando se repite la locución, se agrega un enclítico dativo singular *le* y se omite el objeto en la estructura argumental de *dar*, entendiéndose que *pecho* es el complemento invariable que, en compañía del verbo, forma un único significado: ‘amamantar’.

- (7) I: nada ma- a mi hijo <-mijo>/ a mi hijo <-mijo> no quiso el pecho/ él pura mamila/ pero a mi hija <-mija> sí/ tres meses nada más porque ya no/ ya no le quise **dar yo pecho** dije “ay no ya es mucho”/ como **darle** por

ejemplo/ ya ves que hay unas que hasta los ocho nueve meses ya como que/ ya no me gustó/ y ya le empecé a dar/ biberón/ mh/ pero el niño con puro biberón se/ se crió se puede decir porque no/ el pecho no lo quiso/ mh (CSCM, Entrevista 96)

Con base en estos cuatro criterios, se seleccionaron diez verbos que aceptaban objeto nulo. Estos verbos son: *conocer, hacer, llevar, poner, quitar, tener, terminar, vender, ver* y *usar*, los cuales se recogieron en un total de 349 contextos. Una vez conformada la base de datos, se procedió a etiquetar cada contexto. A continuación, se detalla el procedimiento conforme las variables lingüísticas y sociales que se analizaron.

## Hipótesis y esquema de codificación

En esta sección, primero, se presentan las variables lingüísticas, relacionadas con el correferente, que se consideraron en el etiquetado de la base de datos y se recuperaron de investigaciones previas, por lo que la incidencia de tales variables en la variación del conjunto de clíticos acusativos y el objeto nulo es algo comprobado.

Entonces, las variables independientes lingüísticas sobre los rasgos del correferente son: i) *jerarquía de animacidad*, ii) *definitud*, iii) *especificidad*, iv) *rasgo contable o no contable*, v) *género gramatical* y vii) *número gramatical*.

Posteriormente, puesto que esta investigación tiene una perspectiva variacionista que indaga por la relación entre las variables sociales y el fenómeno variable, se incluyen las variables sociales, que –como se mencionó al comienzo de la metodología– se recuperaron del CSCM (Martín Butragueño y Lastra, 2011-2015). Éstas son: i) *género*, ii) *grupo etario* y iii) *nivel de instrucción*.

## Variables lingüísticas

La exposición de las variables lingüísticas comienza por la jerarquía de animacidad, que se clasificó como a) animado y b) inanimado.

En cuanto a la jerarquía, se parte de la hipótesis de que la omisión se verá favorecida con referentes inanimados, debido a que no son tan altamente topicales ni referenciales como los humanos. En cambio, dentro de la clase animados, se espera el uso del pronombre acusativo como una marca de prominencia discursiva –paralelo al uso de la preposición *a* con objetos directos

humanos—, por lo cual se considera que la animacidad será una restricción fuerte para la anáfora cero.

Sobre la definitud, se decidió discriminar de manera binaria entre dos variantes: a) los definidos, ya sea con artículos definidos o indefinidos, y b) los indefinidos, es decir, sustantivos escuetos.

Dado que el objeto nulo, desde la normatividad, es gramatical en español cuando su correferente es indefinido e inespecífico, se espera que la clase indefinida contenga el grueso de los casos para la anáfora cero.

Otro rasgo recurrente en la descripción de la marca nula del pronombre acusativo es la especificidad, que —en este caso— se analizó desde un punto de vista pragmático discursivo, cuando el hablante “dé a entender que se refiere a un objeto o individuo determinado” (Leonetti, 1999: 860), con dos posibilidades: a) específico —en los casos con un correferente concreto— y b) inespecífico —cuando no permite asociarse con un referente particular—.

Así, en el rasgo específico caben todos los nombres propios, demostrativos, nominales determinados con artículos o con posesivos, y sustantivos con adjetivos, además de nominales con frases prepositivas y con relativas. Por el contrario, en el inespecífico se agruparon nominales escuetos, sustantivos genéricos y nominales con cuantificadores. Una prueba adicional que se les aplicó a los correferentes nominales del objeto directo consistió en agregar un modificador como *en concreto*, *en particular* o *determinado*, para localizar los específicos; mientras que, si el nominal tenía una lectura más natural o aceptable con el modificador *cualquiera*, se clasificó como parte de los inespecíficos (Rigau, 1999; 317).

A partir de las características morfosintácticas del correferente, se analizó el rasgo contable o no contable; a su vez, se tomó en cuenta la interpretación del nominal en cuestión, por lo que hubo casos con recategorizaciones. Es importante mencionar que, dentro de este rasgo, se excluyeron los nombres propios y entidades humanas.

De tal suerte que la clasificación morfosintáctica está basada en la presencia o ausencia de determinantes y en el número gramatical. Por esto, los nominales sin determinante y en forma singular se etiquetaron como no contables; en cambio, los mismos nominales en plural se marcaron como contables. Las frases nominales con artículos indefinidos se consideraron contables, sin importar el número gramatical. En su mayoría, los nominales determinados con artículo definido plural se clasificaron como contables, aunque hubo

casos de recategorización; es decir, se encontraron nominales definidos que en la forma plural se refieren a sustancias o materias que carecen de límites claros, por tanto, se etiquetaron como no contables. Por último, los nominales definidos en singular son de la clase contable; sin embargo, se encontraron algunos casos con una interpretación no contable, al referirse –de nuevo– a sustancias o masa.

Sobre esta variable, se espera que el mayor número de objetos nulos se encuentre dentro de las entidades no contables, dado que, cuando hay recategorizaciones en las que prevalece el sentido de sustancia o masa o los correferentes no presentan determinantes, en consecuencia, adquieren rasgos menos individualizados y más inespecíficos. Por último, se revisaron dos variables del nivel morfológico: el género y el número gramatical. Estas dos variables se operacionalizaron de acuerdo con las propiedades del español.

En primer lugar, el género gramatical tiene las variantes femenino y masculino. No se espera que éstas incidan en la variación entre el pronombre acusativo y el objeto nulo, pues sólo se ha documentado que repercuten en la reorganización del sistema pronominal de variantes del español de contacto, pero no en el uso del objeto nulo. Sin embargo, por ser una propiedad intrínseca de los nominales en español pareció necesario incluirlo en el análisis.

La última variable lingüística que se analizó es el número gramatical, que cuenta con las variantes singular y plural. Se espera que esta variable sí tenga un efecto en el uso del objeto nulo, pues –junto con los valores indefinido e inespecífico– la pluralidad está relacionada con una baja referencialidad o individualización.

Como se mencionó, se espera que algunas de las variables lingüísticas incidan en el uso del objeto nulo, como la inanimación, la indefinitud, la inespecificidad, el valor no contable y el número plural; en la misma dirección, se parte de la hipótesis de que la animación, la definitud y la especificidad son restricciones fuertes para la marca cero. Por otro lado, en el siguiente apartado se muestran las variables sociales estudiadas.

### **Variables sociales**

Uno de los principios básicos de la perspectiva sociolingüística variacionista es que no es posible comprender el desarrollo de un cambio lingüístico fuera

de la estructura social de la comunidad en la que ocurre. Es por ello que en esta investigación se estudiará también la relación del fenómeno en cuestión con las variables sociales. Cabe recordar que las variables de este apartado fueron retomadas del *CSCM*.

Así pues, para el género de los informantes se consideraron dos clases: hombres y mujeres. De esta variable, se espera que sean las mujeres las que más hagan uso del objeto nulo, al considerar que se trata de un fenómeno de abajo (Labov, 1990); de acuerdo con dicho principio, estos cambios son los que se producen por debajo del nivel de consciencia de los hablantes y, en esta instancia, las mujeres suelen ser las innovadoras.

Por su parte, la edad de los informantes se clasificó en tres generaciones: la primera, los jóvenes, se corresponde con las personas de entre 20 y 34 años; la segunda, los adultos, contiene a los informantes de 35 a 54 años, y la última, los mayores, incluye a aquellos de 55 o más. Se hipotetiza que, del mismo modo que observó Reig Alamillo (2009), el grupo etario más joven contenga el grueso de los contextos del objeto nulo.

El nivel de instrucción se clasificó en tres grupos. El primero, el bajo, contempla a las personas con enseñanza formal primaria hasta los 11 o 12 años como máximo; considera a quienes hayan atravesado por seis años de escolarización; se incluye también población analfabeta y con primaria inconclusa. Luego, sigue el grupo medio, que contiene a la población que cursó como máximo la enseñanza media, sea hasta la escuela secundaria o hasta el bachillerato, con un máximo de 12 años de escolarización. La escolaridad más alta involucra a los informantes con enseñanza universitaria o técnica superior, quienes tentativamente habrían estudiado hasta los 21 o 22 años, tras 16 años de escolaridad o incluso más.

Aquí “de abajo” tiene una doble interpretación en el sentido que se puede relacionar también con un nivel bajo en la jerarquía social; por lo tanto, en esta variable se espera que el nivel de instrucción bajo sea el grupo que lidere el uso de la variante innovadora, y en cambio, el nivel de instrucción alto sea el grupo menos favorecedor para la marca nula.

Luego de haber explicado la forma en la que se operacionalizaron las variables lingüísticas y sociales, y de haber presentado sus correspondientes hipótesis, a continuación, se exponen los resultados en función del comportamiento porcentual de cada una.

## RESULTADOS

Primeramente, se presenta el panorama general del fenómeno variable en estudio. Así pues, de acuerdo con estos datos y como se muestra en la tabla 2, el porcentaje total del objeto nulo en el español de la Ciudad de México se presenta en un 23.5 % (N=82) de los posibles casos, por tanto, el pronombre acusativo alcanza el 76.5 % (N=267).

TABLA 2. DISTRIBUCIÓN GENERAL DEL CLÍTICO Y EL OBJETO NULO DE TERCERA PERSONA EN EL ESPAÑOL MONOLINGÜE SINCRÓNICO DE LA CIUDAD DE MÉXICO

<b>Pronombre acusativo de 3a. p.</b>	<b>Pronombre cero de acusativo de 3a. p.</b>
76.5 % (N=267)	23.5 % (N=82)
<b>TOTAL 349</b>	

Es importante recalcar que el 23.5 % (N=82) que alcanzó el objeto nulo con correferentes nominales es un hallazgo no previsto en la investigación. Tal porcentaje indica que el objeto nulo es un mecanismo anafórico productivo para recuperar correferentes nominales en el español monolingüe de la Ciudad de México.

### Variables lingüísticas

Se presenta los resultados de las variables lingüísticas relacionadas con los rasgos del correferente nominal.

#### Jerarquía de animacidad

El primer rasgo por analizar es la jerarquía de animacidad. Respecto a los animados, el 92.2 % (N=83) aparece con los pronombres *lo(s)* y *la(s)*, y el 7.8 % (N=7) con la marca nula. En cuanto a los inanimados, el 71 % (N=184) presenta el clítico y en el 29 % (N=75) de los casos fue usado el objeto cero. Lo anterior se observa en la tabla 3.

La distribución que muestra la tabla 3 concuerda con las tendencias de uso del objeto nulo en los trabajos consultados: la omisión se ve favorecida porcentualmente con referentes inanimados (Camus Bergareche y Gómez Seibane, 2015; García Tesoro y Fernández Mallat, 2015; Moreno Fernández, 2019; Palacios Alcaine, 2015; Torres Sánchez, 2018). En esta investigación,



los correferentes animados presentaron sólo el 7.8 % (N=7) de omisión del clítico con una fuerte preferencia por retomarse con el clítico, mientras que el uso del objeto cero se disparó hasta un 29 % (N=75) con correferentes inanimados.

TABLA 3. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO A LA JERARQUÍA DE ANIMACIDAD

	<b>Animado</b>	<b>Inanimado</b>
Pronombre acusativo de 3a. p.	92.2 % (N=83)	71 % (N=184)
Pronombre cero de acusativo de 3a. p.	7.8 % (N=7)	29 % (N=75)
<b>TOTAL</b>	<b>100 % (N=90)</b>	<b>100 % (N=259)</b>

Cabe mencionar que los correferentes animados usualmente cuentan con los rasgos definido y específico, y que, desde el prescriptivismo, están vinculados con la marca obligatoria del pronombre acusativo; aunque en esta investigación se encontraron algunos casos con el objeto nulo.

De este modo, dado que la animacidad supone una de las restricciones más fuertes para la marca cero (Palacios Alcaine, 2013), es interesante conocer las condiciones lingüísticas<sup>6</sup> que se presentaron para que tuviera lugar el 7.8 % (N=7) de objetos nulos cuando su correferente es una entidad animada, y éstas fueron: i) la ausencia del clítico dativo, ii) la atelicidad del verbo, iii) la distancia referencial corta, y iv) el número de menciones bajo. En cuanto a las variables sociales, todos los objetos nulos correferenciales con nominales animados fueron realizados por mujeres.

Por último, es importante mencionar que, desde una perspectiva de prototipos, los complementos directos tienden a ser entidades no animadas (Company, 2002), lo que se comprueba con los porcentajes globales: los correferentes animados representan sólo el 25.8 % (N=90) y los inanimados, el 74.2 % (N=259). Esto va de la mano con otro de los supuestos con los que se partió, el de que los ejemplares más prototípicos son menos marcados y más estables

<sup>6</sup> Si bien las variables lingüísticas de orden sintáctico y, sobre todo, los factores pragmático-discursivos quedan fuera de los alcances de este artículo, se optó por recuperar la interacción de aquellas variables con los rasgos del correferente nominal que aquí se discuten para un panorama más amplio del fenómeno estudiado.

diacrónicamente (Company, 2002: 207). De ello que la clase prototípica, los inanimados, suele codificarse con el objeto nulo, un mecanismo muy simple; mientras que la clase menos prototípica, los animados, hace uso de marcas más complejas como los clíticos acusativos.

## Definitud

Como se observa en la tabla 4, los pronombres acusativos con correferentes definidos alcanzaron el 78.4 % (N=232); por lo tanto, los mismos correferentes con el objeto nulo obtuvieron el 21.6 % (N=64). Respecto a los correferentes no definidos, el clítico reúne el 66 % (N=35) y el objeto nulo, el 34 % (N=18).

TABLA 4. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO A LA DEFINITUD

	<b>Definido</b>	<b>Indefinido</b>
Pronombre acusativo de 3a. p.	78.4 % (N=232)	66 % (N=35)
Pronombre cero de acusativo de 3a. p.	21.6 % (N=64)	34 % (N=18)
<b>TOTAL</b>	<b>100 % (N=296)</b>	<b>100 % (N=53)</b>

Los porcentajes obtenidos para esta variable revelan que los nominales escuetos son los más propensos a ser correferenciales con el objeto nulo, pero más interesante aún, que la definitud del nominal no restringe que la marca cero pueda ser el mecanismo anafórico utilizado para mantenerlo activo en el discurso.

El comportamiento de la variable definitud no era esperado, dado que, en español, la omisión de objeto directo, desde una visión normativa, sólo es gramatical cuando se trata de correferentes indefinidos e inespecíficos (Campos, 1986; Fernández Soriano, 1999: 1215; RAE y ASALE, 2010: 657).

Por esto, los resultados de la tabla 4 no corroboran lo que se ha señalado para el español normativo; en su lugar, son indicadores importantes del estado del objeto nulo en el español de la Ciudad de México porque apuntan a una disminución en la restricción que representa la variable definitud.

Finalmente, se exploró cuáles eran otras de las variables que daban lugar al cruce del objeto nulo y los nominales definidos, pero ninguna resultó ser una constante. Entre las que más se repitieron están el número de menciones bajo, con frecuencia de 60; la distancia corta, con 59 casos, y la inanimacidad, con 56 contextos.

## Especificidad

En cuanto a nominales específicos, se observó que el pronombre acusativo de tercera persona se usa alrededor de un 79.7 % (N=161) y el objeto cero en un 20.3 % (N=41) de los casos. Respecto a los no específicos, el porcentaje de los pronombres átonos es del 72.1 % (N=106); en cuanto al objeto cero, se registra un uso mayor al documentado para los no específicos, que asciende al 27.9 % (N=41). Estos porcentajes están vertidos en la tabla 5.

TABLA 5. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO A LA ESPECIFICIDAD

	<b>Específico</b>	<b>No específico</b>
Pronombre acusativo de 3a. p.	79.7 % (N=161)	72.1 % (N=106)
Pronombre cero de acusativo de 3a. p.	20.3 % (N=41)	27.9 % (N=41)
<b>TOTAL</b>	<b>100 % (N=202)</b>	<b>100 % (N=147)</b>

Sobre esta variable, se puede constatar una preferencia del objeto cero de aproximadamente siete puntos porcentuales por correferentes inespecíficos frente a correferentes específicos, con el 27.9 % (N=41) y con el 20.3 % (N=41), respectivamente. Sin embargo, estos porcentajes de especificidad en relación con los objetos nulos no indican que la especificidad sea todavía una restricción fuerte en el español monolingüe de la Ciudad de México, además de que todos los nominales específicos del 20.3 % (N=41) cuentan también con el rasgo definido. Por tanto, el español de la Ciudad de México utiliza la categoría del objeto nulo en contextos innovadores que se alejan de los usos normativos.

## Contable o no contable

Sobre el rasgo contable o no contable, como se puede ver en la tabla 6, la intersección del pronombre acusativo y los correferentes contables representó el 80.29 % (N=216), mientras que el objeto nulo con los mismos correferentes reunió el 19.7 % (N=53). Ahora, los pronombres de objeto directo con correferentes no contables alcanzaron el 71.73 % (N=33) y el objeto nulo obtuvo el 28.26 % (N=13) con los correferentes no contables.

TABLA 6. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO AL RASGO CONTABLE Y NO CONTABLE

	<b>Contable</b>	<b>No contable</b>
Pronombre acusativo de 3a. p.	80.29 % (N=216)	71.73 % (N=33)
Pronombre cero de acusativo de 3a. p.	19.7 % (N=53)	28.26 % (N=13)
<b>TOTAL</b>	<b>100 % (N=269)</b>	<b>100 % (N=46)</b>

El rasgo contable o no contable ha demostrado entrar en juego en distintos sistemas pronominales conocidos como referenciales –el sistema del español hablado en Asturias, Cantabria y Castilla (Fernández Ordoñez, 1999: 1355-1363)– y en la omisión del pronombre –el español hablado por bilingües tepehuanos (Torres Sánchez, 2018)–.

Así pues, la hipótesis inicial se comprobó, ya que, en el español monolingüe de la Ciudad de México, el objeto nulo respecto del rasgo contable o no contable muestra una preferencia porcentual por la segunda clase con el 28.26 % (N=13); por consiguiente, en relación con la anáfora cero se presenta una diferencia de alrededor de nueve puntos porcentuales frente a la clase de los contables, que registró el 19.7 % (N=53) de los casos.

Una vez más, se esperaba que la interacción entre el valor definido y el específico tuviera un efecto en las tendencias porcentuales, pues –como se mencionó– el etiquetado de los nominales se basó en sus propiedades morfosintácticas, según la presencia o ausencia de determinantes, y en el número gramatical.

Por tanto, al analizar a detalle las características de los correferentes contables y no contables correferenciales con el objeto nulo, se observa que entre los contables 44 son definidos y 24 son específicos, de los 53 casos totales, y de los no contables, con 12 ejemplares, 4 son definidos y sólo 1 específico.

Esto demuestra una relación entre los nominales contables y la definitud y, en menor medida, con la especificidad; en cambio, los correferentes no contables, en su mayoría, se componen de entidades indefinidas e inespecíficas. De este modo, se evidencia que el efecto de ambos valores persiste como restricción del objeto nulo, pero no de una forma categórica, sino que se trataría de un fenómeno incipiente: el objeto cero está extendiendo su ámbito y comienza a abarcar usos para los cuales los clíticos *lo(s)* y *la(s)* se especializaban.

## Género gramatical

En la tabla 7, se observa la distribución de los correferentes según el género gramatical. Para el género femenino, el pronombre de objeto directo fue utilizado en un 79.5 % (N=89) y el objeto cero se usó un 20.5 % (N=23); mientras que los correferentes masculinos se repartieron de la siguiente manera: 75.1 % (N=178) con el pronombre átono y 24.9 % (N=59) para el objeto cero.

TABLA 7. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO AL GÉNERO GRAMATICAL

	Femenino	Masculino
Pronombre acusativo de 3a. p.	79.5 % (N=89)	75.1 % (N=178)
Pronombre cero de acusativo de 3a. p.	20.5 % (N=23)	24.9 % (N=59)
<b>TOTAL</b>	<b>100 % (N=112)</b>	<b>100 % (N=237)</b>

La hipótesis inicial en la que se esperaba que el género de los nominales no incidiera en las tendencias del objeto nulo no se cumplió, como se puede observar en los porcentajes (*supra*), porque hay una preferencia de, al menos, cuatro puntos porcentuales por los correferentes con el género gramatical masculino. Posiblemente, la explicación a esto se encuentre en la distribución del número de nominales de acuerdo con el género gramatical. Del total de datos, 67.9 % (N= 237) corresponden a correferentes masculinos y 32.1 % (N=112) a los del género gramatical femenino. Por lo tanto, los nominales masculinos son el grueso de la muestra y esto repercute también en el comportamiento porcentual de la marca nula: los nominales masculinos con el objeto nulo son más del doble que los que son femeninos, lo que genera un porcentaje más alto.

## Número gramatical

La última variable lingüística es el número gramatical. En la tabla 8 aparecen los resultados que se encontraron en la base de datos. El 80.2 % (N=210) de los nominales en singular se codificaron con el pronombre átono, mientras que el otro 19.8 % (N=52) se marcó con el objeto cero. Por su parte, con el número plural se observó una disminución respecto a la codificación por

medio del pronombre, pues ésta representa el 65.5 % (N=57), y con el objeto nulo hay un 34.5 % (N=30) de nominales en plural.

**TABLA 8. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO AL NÚMERO GRAMATICAL**

	<b>Singular</b>	<b>Plural</b>
Pronombre acusativo de 3a. p.	80.2 % (N=210)	65.5 % (N=57)
Pronombre cero de acusativo de 3a. p.	19.8 % (N=52)	34.5 % (N=30)
<b>TOTAL</b>	<b>100 % (N=262)</b>	<b>100 % (N=87)</b>

De acuerdo con los porcentajes en la tabla 8, se comprobó la hipótesis de la que se partió, puesto que se esperaba que la pluralidad, al estar relacionada con una baja referencialidad o individualización –como sucede con las frases escuetas–, generara las condiciones para utilizar el objeto nulo.

Cabe resaltar que la jerarquía de animacidad demostró un correlato con el número gramatical plural cuando es usado el objeto nulo, puesto que en 29 casos se trata de entidades inanimadas. Asimismo, tanto la definitud como la inespecificidad resultaron estar más relacionadas con este rasgo, pues 23 contextos del objeto nulo con correferentes plurales corresponden a entidades inespecíficas y 22, a definidas.

## **Variables sociales**

A continuación, se exponen los porcentajes e interpretaciones de los resultados de las variables sociales.

### **Género**

Como se puede observar en la tabla 9, el pronombre acusativo de tercera persona se registró en un 66.9 % (N=121) por las mujeres y en un 86.9 % (N=146) por hombres. Mientras que, el pronombre cero fue usado el 33.1 % (N=60) de las veces por las mujeres, más del doble que el 13.1 % (N=22) de casos contabilizados en el otro grupo.

TABLA 9. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO AL GÉNERO DE LOS INFORMANTES

	<b>Mujer</b>	<b>Hombre</b>
Pronombre acusativo de 3a. p.	66.9 % (N=121)	86.9 % (N=146)
Pronombre cero de acusativo de 3a. p.	33.1 % (N=60)	13.1 % (N=22)
<b>TOTAL</b>	<b>100 % (N=181)</b>	<b>100 % (N=168)</b>

Primeramente, es importante recordar que se considera que el objeto nulo sería un fenómeno de abajo, de acuerdo con dos principios propuestos por Labov (1990), los cuales describen dos situaciones en las que las mujeres lideran el uso en las formas no estándares: i) en cambios de arriba, esto es, cambios lingüísticos con cierto nivel de consciencia: las mujeres favorecen más la forma con más prestigio que los hombres, y ii) en cambios de abajo, es decir, por debajo del nivel de consciencia de los hablantes, donde las mujeres suelen ser las innovadoras.

Como el objeto nulo no es estigmatizado ni está en el nivel de consciencia de los hablantes, el principio que mejor describe la situación en el español de la Ciudad de México es el segundo, “[i]n change from below, women are most often the innovators” (Labov, 1990: 215).

De acuerdo con esto, se hipotetizó que las mujeres serían líderes en el uso de la categoría innovadora, lo que se comprueba con los porcentajes contenidos en la tabla 9, pues las mujeres utilizan el objeto nulo más del doble que los hombres.

Más aún, esta tendencia porcentual es recurrente, pues en ningún caso específico los hombres superan el uso del objeto nulo en comparación con las mujeres, ya sea en los tres grupos etarios o en los distintos niveles de instrucción.

### Grupo etario

Sobre el grupo etario, el pronombre de objeto directo obtuvo el 75.6 % (N=96) en el grupo jóvenes; el mismo pronombre alcanzó el porcentaje más bajo con los adultos, con el 68.6 % (N=72) de los casos; por último, los mayores emplearon los clíticos acusativos en un 84.6 % (N=99). Por su parte, el porcentaje de uso del objeto nulo dentro de los jóvenes fue de 24.4 % (N=31); en el grupo de los adultos el objeto cero se usó el 31.4 % (N=33) de las veces, y el porcentaje más bajo de la marca nula se encontró en el grupo

etario de los mayores, con el 15.4 % (N=18). La información anterior está vertida en la tabla 10.

TABLA 10. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO AL GRUPO ETARIO DE LOS INFORMANTES

	Jóvenes	Adultos	Mayores
Pronombre acusativo de 3a. p.	75.6 % (N=96)	68.6 % (N=72)	84.6 % (N=99)
Pronombre cero de acusativo de 3a. p.	24.4 % (N=31)	31.4 % (N=33)	15.4 % (N=18)
<b>TOTAL</b>	<b>100 % (N=127)</b>	<b>100 % (N=105)</b>	<b>100 % (N=117)</b>

En la tabla 10, se puede observar que los adultos son líderes del cambio en curso con una diferencia de siete puntos porcentuales respecto de los jóvenes, mientras que el grupo de mayores produjo casi la mitad de las marcas nulas en comparación con el grupo innovador: los adultos. Esto resultó contrario a la hipótesis inicial, pues se esperaba que la generación más joven fuera la que presentara mayor uso de la anáfora cero, es decir, que se replicaran los resultados obtenidos por Reig Alamillo (2009) en el español mexicano.

### Nivel de instrucción

La variable nivel de instrucción arrojó los siguientes porcentajes: la intersección del pronombre acusativo y el nivel bajo reunió el 81 % (N=136); en cambio, el grado medio y alto tuvieron un comportamiento porcentual similar con los clíticos *lo(s)* y *la(s)*, con el 72.6 % (N=85) y el 71.9 % (N=46) respectivamente. Ahora, el objeto cero, con el nivel de instrucción bajo, alcanzó el 19 % (N=32) de los ejemplares con pronombres acusativos. Por su parte, el uso pronominal obtuvo un porcentaje muy similar entre el nivel medio y alto: el 27.4 % (N=32) en el primer caso, y el 28.1 % (N=18) en el segundo. Los porcentajes aquí reportados se encuentran en la tabla 11.

Referirse al objeto nulo como un fenómeno de abajo, además de considerar la consciencia de los hablantes, también se puede relacionar con un nivel bajo en la jerarquía social (Labov, 1966: 128). Por lo anterior, la hipótesis esperaba que el nivel de instrucción bajo contuviera el grueso de los datos,



pero esto no fue así, sino que, por el contrario, el grupo con mayor nivel de instrucción fue el que más uso hizo de la marca nula, seguido por el nivel medio y, por último, con una diferencia porcentual importante, por el nivel bajo.

TABLA 11. DISTRIBUCIÓN DEL CLÍTICO Y EL OBJETO NULO RESPECTO AL NIVEL DE INSTRUCCIÓN DE LOS INFORMANTES

	Bajo	Medio	Alto
Pronombre acusativo de 3a. p.	81 % (N=136)	72.6 % (N=85)	71.9 % (N=46)
Pronombre cero de acusativo de 3a. p.	19 % (N=32)	27.4 % (N=32)	28.1 % (N=18)
<b>TOTAL</b>	<b>100 % (N=168)</b>	<b>100 % (N=117)</b>	<b>100 % (N=64)</b>

El que el porcentaje más alto se ubicara en el grupo con el mayor nivel de instrucción resultó un hallazgo inesperado, pues –de acuerdo con Labov (1990)– las innovaciones de arriba son características de formas de prestigio.

El porcentaje del nivel de instrucción alto, junto con el elevado uso del objeto nulo por parte de las mujeres (*supra*), podría interpretarse como el otro principio mencionado y que aquí también se recupera: “*In change from above, women favor the incoming prestige form more than men*” (Labov, 1990; 213), lo que podría indicar un valor positivo agregado a la variante innovadora, el objeto nulo. Además, si se trae a la discusión el porcentaje más alto encontrado en los grupos etarios correspondiente a la clase de los adultos, y se considera que los grupos de edades intermedias, con un papel activo en el mundo profesional, económico y social, suelen presentar el fenómeno de autocorrección (Blas Arroyo, 2004; 193, 202), con una inclinación por las variantes prestigiosas, el objeto nulo en el español de la Ciudad de México podría tratarse de una variante innovadora con un valor simbólico importante. Sin embargo, el hecho de que los hablantes no tengan un conocimiento consciente de la variante de la anáfora cero lleva a tomar este análisis con cuidado, y a adherirse al segundo principio.

## CONCLUSIÓN

A modo de conclusión, se ordenan de manera descendente las variables lingüísticas en función de las tendencias porcentuales de la marca cero.

De tal forma que el número gramatical plural mostró el porcentaje más alto para el objeto nulo en relación con la codificación del pronombre de acusativo, esto es, el 34.5 % (N=30). El segundo porcentaje más alto correspondió a las frases no definidas o escuetas dentro de la variable definitud con el 34 % (N=18).

Como se mencionó, existe una relación entre estas dos variables, dado que “[l]os grupos nominales escuetos pueden tener por núcleo sustantivos no contables en singular o contables en plural. Los sustantivos contables en singular están más restringidos en esta pauta” (RAE y ASALE, 2010: 295). Por ello, existe una preferencia por el plural en las frases escuetas.

La jerarquía de animacidad, con la clase inanimado, presentó el tercer porcentaje más alto para el objeto nulo con el 29 % (N=75). Cabe mencionar que el porcentaje más bajo del objeto nulo se encontró dentro de la clase animado, lo que indica que esta variante persiste como una restricción fuerte para el uso del objeto nulo.

La relación entre la jerarquía de animacidad y el objeto nulo también era esperada, pues –como se mencionó– las entidades animadas tienden a ser definidas y específicas, y los correferentes humanos a ser tópicos, y, por tanto, requieren el pronombre acusativo para marcar su prominencia discursiva. Esto se asemeja al uso de la preposición *a* que los acompaña cuando son referidos en frases nominales con la función de complemento directo.

Muy de cerca a la jerarquía de animacidad, la variable contable o no contable comprobó la hipótesis inicial: los correferentes no contables tienden al uso del objeto nulo, lo que se sostiene por haber reunido el 28.26 % (N=13) de los casos de acusativo para este valor, en comparación con el 19.7 % (N=53) de los contables, porcentaje que estuvo por debajo del promedio general.

En cuanto a la variable especificidad, se encontró que el uso del objeto cero se prefiere cuando sus correferentes son inespecíficos, en un 27.9 % (N=41) de los casos, frente a los específicos. Este comportamiento también era esperado, al ser un rasgo que propicia el uso del objeto nulo, incluso desde el prescriptivismo.







Sobre el género gramatical, en el español monolingüe de la Ciudad de México, el objeto nulo es usado más veces que su equivalente pronominal con correferentes masculinos, con el 24.9 % (N=59). Ahora bien, esta variable no parece incidir en ninguno de los trabajos consultados para el objeto nulo; sin embargo, se ha demostrado que entra en juego en los fenómenos de *leísmo*, *loísmo* y *laísmo* (Fernández Ordóñez 1993, 1999). Aquí se interpretó que la fuerza de esta tendencia responde al grueso de frecuencias que los nominales masculinos tienen en toda la base de datos, al contener el 67.9 % (N=237) de los casos. Queda pendiente encontrar si hay más variantes monolingües del español con la misma preferencia porcentual respecto a dicho género gramatical.








El análisis de las variables lingüísticas se puede resumir diciendo que el objeto nulo muestra una tendencia porcentual a retomar correferentes plurales, escuetos, inanimados, no contables, no específicos y con el género gramatical masculino. Por su parte, los resultados de las variables sociales revelan que el objeto nulo se ve favorecido por el género femenino, el grupo etario de los adultos y el nivel de instrucción alto.


Por último, se discute la variable definitud. En esta investigación, las frases escuetas mostraron comportarse según las hipótesis iniciales, mientras que las frases definidas alcanzaron un porcentaje elevado y no esperado. Esto porque –como se ha repetido en distintos momentos– las variables definitud y especificidad son tradicionalmente las restricciones más recurrentes para la marca cero desde lo señalado en español normativo; no obstante, esto no se corrobora tras un estudio de corpus como el presente, pues, en la variante monolingüe del español de la Ciudad de México, la definitud muestra también ser una tendencia para el objeto nulo en comparación con la jerarquía de animacidad, donde la clase animado claramente restringe la aparición del objeto nulo.

De esto modo, se concluye que, con el análisis de las variables relacionadas con el correferente nominal, el objeto nulo parece ser una característica propia del sistema lingüístico del español monolingüe de la Ciudad de México; de ahí que resulte menester explorar otras variables de orden sintáctico y pragmático discursivo, que –por ahora– quedan fuera de los objetivos del presente artículo.

## BIBLIOGRAFÍA

- Blas Arroyo, José Luis (2004), *Sociolingüística del español. Desarrollos y perspectivas en el estudio de la lengua española en contexto social*, Madrid, Cátedra.
- Campos, Hector (1986), “Indefinite Object Drop”, en *Linguistic Inquiry*, vol. 17, núm. 2, pp. 354–359.
- Camus Bergareche, Bruno y Sara Gómez Seibane, (2015), “Nuevos datos sobre la omisión de objetos en el castellano del País Vasco”, en *Círculo de Lingüística Aplicada a la Comunicación*, vol. 61, pp. 211-236, consultado: 20 de agosto de 2024. 
- Clements, Joseph Clancy (2006), “Null direct objects in Spanish”, en *Functional Approaches to Spanish Syntax*, pp. 134-150, consultado: 20 de agosto de 2024. 
- Company Company, Concepción (2002), “Grammaticalization and category weakness” en *New Reflections on Grammaticalization*, vol.49, pp. 201-216, consultado: 20 agosto de 2024. 
- Fernández Ordóñez, Inés (1993), “Leísmo, laísmo y loísmo: estado de la cuestión” en Olga Fernández Soriano, *Los pronombres átonos*, Madrid, Taurus, pp. 63-96.
- Fernández Ordóñez, Inés (1999), “Leísmo, laísmo y loísmo”, en *Gramática Descriptiva de la Lengua Española*, vol.1, pp. 1317-1398.
- Fernández Soriano, Olga (1999), “El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos”, en *Gramática Descriptiva de la Lengua Español*, vol. 1, pp. 1209-1274.
- García Tesoro, Ana Isabel y Víctor Fernández-Mallat (2015), “Cero vs. lo en español andino (Chincheró, Cuzco)” en *Círculo de Lingüística Aplicada a la Comunicación*, vol. 61, pp. 131-157, consultado: 20 de agosto de 2024. 
- Hopper, Paul y Sandra Annear Thompson (1980), “Transitivity in Grammar and Discourse” en *Language*, vol. 56, núm. 2, pp. 251–299, consultado: 20 de agosto de 2024. 
- Labov, William (1990), “The intersection of sex and social class in the course of linguistic change” en *Language Variation and Change*, vol. 2, núm. 2, pp. 205-254, consultado: 21 de agosto de 2024. 
- Leonetti, Manuel (1999), “El Artículo” en *Gramática Descriptiva de la Lengua Española*, vol. 1, pp. 787-890.

- Martín Butragueño, Pedro y Yolanda Lastra (2011-2015). “Corpus sociolingüístico de la Ciudad de México (CSCM)”, *Laboratorio de Estudios Fónicos*, consultado: 21 de agosto 2024. 
- Masullo, Pascual José (2017), “La interfaz sintaxis-pragmática: caída de objeto acusativo definido sin clítico en el español rioplatense” en *Saga. Revista De Letras*, vol. 7, pp. 53-72, consultado: 21 de agosto 2024. 
- Palacios Alcaine, Azucena (2013), “Contact-induced change and internal evolution: Spanish in contact with Amerindian languages” en *The Interplay of Variation and Change in Contact Settings*, vol. 12, pp. 165-198, consultado: 21 de agosto de 2024. 
- Palacios Alcaine, Azucena (2015), “De nuevo sobre la omisión de objeto directo en el español andino ecuatoriano” en *Círculo de Lingüística Aplicada a la Comunicación*, vol. 61, pp. 104-130, consultado: 21 de agosto 2024. 
- PRESEEA (2014): “Consulta del Corpus del Proyecto para el estudio sociolingüístico del español de España y de América”, *Proyecto para el estudio sociolingüístico del español de España y de América*, consultado: 21 de agosto de 2024. 
- Real Academia Española (RAE) y Asociación de Academias de la Lengua Española (ASALE) (2010), *Nueva gramática de la lengua española: Manual*. Madrid, Espasa Calpe.
- Reig Alamillo, Asela (2009), “Cross-dialectal variation in propositional anaphora: Null objects and propositional lo in Mexican and Peninsular Spanish”, en *Language Variation and Change*, vol. 21, núm. 3, pp. 381-412, consultado: 22 de agosto 2024. 
- Rigau, Gemma (1999), “La estructura del sintagma nominal: los modificadores del nombre” en *Gramática Descriptiva de la Lengua Española*, vol.1, pp. 311-361.
- Romera Martín, Celia (2017), Construcciones con verbo de apoyo. Caracterización en español y comparativa con el catalán, tesis de licenciatura en Grau en Estudis de Català i Espanyol, Barcelona, Universidad Autónoma de Barcelona, consultado: 21 de agosto 2024. 
- Silva Corvalán, Carmen y Andrés Enrique-Arias (2017), *Sociolingüística y pragmática del español*, Washington D.C., Georgetown University Press.
- Tagliamonte, Sali (2006), *Analysing Sociolinguistic Variation*. Nueva York, Cambridge University Press, Key Topics in Sociolinguistics.

Torres Sánchez, Nadiezdha (2018), Aquí hablamos tepehuano y allá español. Un estudio de la situación de bilingüismo incipiente entre español y tepehuano del sureste (o'dam) en Santa María Ocotán y Durango, tesis de doctorado en Centro de Estudios Lingüísticos y Literarios, Ciudad de México, El Colegio de México, consultado: 21 de agosto de 2024. 

**LAURA MELANY FLORES** Morales: Maestra en Lingüística Hispánica por la Universidad Nacional Autónoma de México. Actualmente es traductora de inglés a español y profesora de español como lengua extranjera. Sus principales líneas de investigación son sintaxis y cambio lingüístico del español.

D. R. © Laura Melany Flores Morales, Ciudad de México, enero-junio, 2024.



## ***Titles Used In Cuban Scientific Journals: An Exploratory Study***

**DIANA ELENA PRIETO ACOSTA** 

Instituto de Literatura y Lingüística José A. Portuondo Valdor  
Facultad de Artes y Letras, Universidad de La Habana  
[dianaelenap@gmail.com](mailto:dianaelenap@gmail.com)

**EVELYN RUBIO FIGUERAS** 

Dirección de Gestión Documental y Archivos. CITMA (Cuba)  
[erubio2506@gmail.com](mailto:erubio2506@gmail.com)

**Abstract:** *The objective of this research is to describe the most prominent linguistic characteristics of scientific titles written today in Cuba. We analyzed 200 scientific articles, written in Spanish by Cubans, that were recently published in Anales de la Academia de Ciencias de Cuba, a centenary and multidisciplinary journal. We obtained global data that corroborate findings described by other authors for Spanish from other regions, both in average length and in the general formal structure of the titles, without the need to discriminate between disciplines, genres, or fields. This will allow for the subsequent diachronic description of the evolution of scientific style in Cuba.*

**KEYWORDS:** SCIENTIFIC ARTICLES; TITLING PRACTICES; SYNCHRONY; SCIENTIFIC STYLE; CUBA.

**RECEPTION:** 08/06/2024

**ACCEPTANCE:** 02/10/2024



# Los títulos de las revistas científicas cubanas: estudio exploratorio

**DIANA ELENA PRIETO ACOSTA** 

Instituto de Literatura y Lingüística José A. Portuondo Valdor,  
Facultad de Artes y Letras, Universidad de La Habana  
[dianaelenap@gmail.com](mailto:dianaelenap@gmail.com)

**EVELYN RUBIO FIGUERAS** 

Dirección de Gestión Documental y Archivo. CITMA (Cuba)  
[erubio2506@gmail.com](mailto:erubio2506@gmail.com)

**Resumen:** El objetivo de esta investigación es describir las características lingüísticas más prominentes de los títulos científicos que se escriben hoy en Cuba. Analizamos 200 títulos de artículos científicos, escritos en español por cubanos, que fueron publicados recientemente en *Anales de la Academia de Ciencias de Cuba*, revista centenaria y multidisciplinaria. Obtuvimos datos globales que corroboran los descritos por otros autores para el español de otras latitudes, tanto en longitud promedio como en estructura formal general del aparato titular, sin necesidad de discriminar entre disciplinas, géneros o campos disciplinares. Estos datos permitirán la ulterior descripción diacrónica de la evolución del estilo científico en Cuba.

**PALABRAS CLAVE:** ARTÍCULOS CIENTÍFICOS; PRÁCTICAS DE TITULACIÓN; SINCRONÍA; ESTILO CIENTÍFICO; CUBA

**RECEPCIÓN:** 08/06/2024

**ACEPTACIÓN:** 02/10/2024

## INTRODUCCIÓN

Los títulos de los textos intervienen en tres importantes momentos de la lectura: antes de ésta, desempeñan una función denominativa (identifican) y expresiva o retórica (despiertan interés); durante la lectura, cumplen un rol orientador (cómo leer) y jerarquizador (en qué orden leer) y, finalizada la lectura, actúan como *locus* en la memoria –según entendían los rétores clásicos–, donde el título se fusiona como un todo con el contenido (De la Fuente González, 1997: 186). No en vano afirma Genette (2001: 270) que “uno de los avales de la eficacia de este paratexto es su transparencia, su transitividad. El mejor intertítulo, el mejor título en general, es posiblemente aquél que también sabe hacerse olvidar”.

En los últimos decenios, numerosos trabajos procuran dar cuenta de las características lingüísticas de los títulos de las publicaciones científicas, dada la primordial función que estos desempeñan en la comunicación de la ciencia. En efecto, los títulos aportan datos básicos para la organización de la información y para su posterior búsqueda; esto es, desde la indexación en los catálogos de las bibliotecas y las bases de datos bibliográficos hasta la recuperación de la información por sus usuarios finales (Soler, 2009: 50).

En la actualidad, posiblemente por la altísima circulación de información científica, muchos investigadores seleccionan los artículos que desean consultar orientándose sobre todo por los títulos y no por el contenido de los resúmenes, como fuera a finales del siglo pasado (Salager-Meyer y Alcaraz Ariza, 2013: 258). Más aún, los actuales sistemas de publicación continua, donde los artículos son difundidos sin esperar por la conformación de números o volúmenes de revista completos (*article base-publishing*), colocan a los títulos de los artículos en una posición todavía más central en la comunicación científica, pues la recuperación de los artículos se basa generalmente en búsquedas por palabras clave en metabuscadores más que en la revisión de los índices de contenido de las revistas, como fuera otrora (Hyland y Zou, 2022: 1).

Por lo general, los textos inéditos, tanto manuscritos como mecanografiados o redactados en dispositivos informáticos, son resultado solamente de un proceso de producción (Moya Méndez, 2013: 34). A diferencia de éstos, los textos científicos son obra de autores que se han visto obligados a una mayor elaboración: la corrección editorial, la cual consiste en velar por el cumplimiento de normas ortográficas, gramaticales y estándares de prestigio

de la época, así como otras relativas a las convenciones de las comunidades disciplinares y las normas particulares de las casas editoriales. Más tarde, el texto es modificado por un editor,<sup>1</sup> a partir de sus conocimientos, competencias y las instrucciones editoriales de quien lo contrata.<sup>2</sup> En síntesis, en el texto especializado, tal vez más que en otros documentos, se puede rastrear la ideología y la cosmovisión de toda una época, posiblemente por su naturaleza institucional.

Los títulos de los textos científicos han sido investigados con fines utilitarios inmediatos (por ejemplo, didácticos), y también por interés teórico, en estudios de discurso, estilísticos, retóricos. Han sido conducidas investigaciones sobre el aparato titular en un género textual específico, así como de varios géneros, disciplinas e incluso lenguas. De estas últimas, predominan las comparaciones con el inglés, que continúa siendo la *lingua franca* de la ciencia. También hay estudios diacrónicos (Salager-Meyer y Alcaraz Ariza, 2013).

De la titulación científica en español hay relativamente menos investigaciones. Entre ellas se citan estudios sobre los títulos de varios géneros discursivos en disciplinas específicas, como la odontología (Morales, Perdomo, Cassany, *et al.*, 2020b; Morales, Perdomo, Cassany, *et al.*, 2022; Morales, Perdomo, Cassany, *et al.*, 2020a); otros que indagan en géneros específicos, como las tesis doctorales (Soler, 2018), y estudios comparativos de varias disciplinas y géneros (Soler, 2009).

Hasta donde tenemos conocimiento, no está documentado el comportamiento de los títulos en las publicaciones científicas cubanas contemporáneas, aspecto al que atendemos en este trabajo. Éste se inscribe en una investigación más amplia, diacrónica, con miras a describir la evolución del estilo científico en Cuba y su retórica,<sup>3</sup> la cual se enmarca en la confluencia de los estudios

<sup>1</sup> Llegado el siglo xx se fortalece el control del editor sobre los textos, figura que antes era prácticamente invisible (Posteguillo-Gómez y Piqué-Angordans, 2007). Es posible que este fortalecimiento de la intervención editorial sea reflejo de la consolidación de la institucionalización de la ciencia y de la vigorosa entrada al mercado que la ciencia experimentó en el siglo xx, la centuria que nos enseñó a “publicar o perecer” (*publish or perish*).

<sup>2</sup> Según Moya Méndez, en el caso de los textos publicados, sería más apropiado calificar las investigaciones de lingüístico-editoriales que de lingüísticas (Moya Méndez, 2013: 33-34).

<sup>3</sup> En el sentido aristotélico del término; esto es, la observación teórica de lo persuasivo en el discurso, diferente del sentido de la palabra en la tradición pedagógica de una metodología educativa en la

históricos del discurso (pragmática histórica), la historia de la lengua y la historia de la propia ciencia. En definitiva, para explicar el discurso científico en todas sus posibilidades es necesario integrar diferentes disciplinas: a) la filosofía de la ciencia, que plantea la correspondencia entre lengua y ciencia; b) la sociología de la ciencia, que propone la correlación entre ciencia, cultura y sociedad; y c) el enfoque sociocognitivo del análisis del discurso, que proporciona los elementos necesarios para demostrar que la construcción del discurso científico no es ajena a su contexto (Prieto Castro, 2018: 14).

Para analizar las características lingüísticas de los títulos científicos que se producen hoy en Cuba hemos seleccionado la revista *Anales de la Academia de Ciencias de Cuba*,<sup>4</sup> publicación especializada de la Academia de Ciencias de Cuba. Según describen varios autores (McClellan III, 1979; Gross, Harmon, Reidy, 2002; Swales, 2008 (1990); Kotwica, 2018), los géneros científicos nacieron en las instituciones científicas que fueron creadas entre los siglos XVII y XIX. Estas fundaciones, sociedades científicas y academias de ciencias, se procuraron vehículos formales para la comunicación de sus resultados, las revistas científicas, las cuales, en consecuencia, devinieron en fuente documental de la historia de las propias corporaciones y de la ciencia de sus países.

La actual *Anales de la Academia de Ciencias de Cuba* fue creada en 1864, pero con el título *Anales de la Real Academia de Ciencias Médicas, Físicas y Naturales de La Habana*, nombre de la corporación por aquella época.<sup>5</sup> Desde

---

competencia discursiva, que comienza con los sicilianos e Isócrates (Vidal Ramírez, 2022: 74). Ramírez Vidal explica que en el siglo II a. C., con Hermágoras de Temnos, los maestros empiezan a utilizar el término “retórica” de manera indistinta hasta que se legitima de esa forma en época tardía y todavía actualmente se emplea para designar ambas orientaciones (2022: 75).

<sup>4</sup> En línea en <https://revistaccuba.sld.cu/index.php/revacc/index>

<sup>5</sup> La Academia de Ciencias de Cuba, fundada en 1861, nace con el calificativo de *Real* bajo el mandato de Isabel II, el cual mantiene hasta 1899 como reflejo de su vínculo con el contexto colonial de la época. En palabras de Pruna Goodgall (2011), se satisface de ser “la única Real Academia que se creó fuera de la metrópoli española, cuando el antiguo imperio se había reducido ya a su expresión última” (2011: 19-20). Iniciado el siglo XX conserva el nombre de *Academia de Ciencias Médicas, Físicas y Naturales de La Habana*. Sin embargo, ya triunfada la Revolución de 1959, el Gobierno decide dotarla de un carácter nacional y pasa a llamarse *Academia de Ciencias de Cuba*. Aunque desde su inicio hasta la actualidad, la historia de la Academia está plagada de hitos, su período dorado es el siglo XIX, pues en él se enmarca la conformación de la primera comunidad científica del país, que defendió el reconocimiento de la ciencia como una faceta de la cultura nacional. Encargados

el 2011, año del 150 aniversario de la Real Academia de Ciencias, *Anales* inicia una nueva etapa de vida, en formato digital.

En sus inicios en el siglo XIX, a la vez que revista científica, *Anales* funcionaba como periódico oficial de la institución y se destinaba a la difusión de conocimientos sobre los temas que eran discutidos en el seno académico, especialmente los relacionados con los fenómenos comunes del suelo patrio, como medio de anticipación de los adelantos y reformas (Mestre y Muñoz, 1864). Eso sí, también eran tratadas en la revista cuestiones relativas al funcionamiento y la administración de la entidad.

*Anales* tuvo la gracia de presentar en sus páginas todos los estudios vinculados con el descubrimiento del agente transmisor de la fiebre amarilla, la edificación del primer acueducto cubano, el tratamiento de epidemias como el cólera o la viruela, el establecimiento del primer alumbrado público en Cuba, la conformación del catálogo de plantas cubanas, entre otros. Figuras cimeras de la ciencia en Cuba y el mundo, como Carlos J. Finlay, Felipe Poey, José de la Luz y Caballero, Francisco de Albear, Juan Bruno Zayas, Joaquín Albarrán y muchos más, honraron la publicación con sus trabajos.

Factores como éstos dejan clara la marcada incidencia de la obra académica en el despertar de la ciencia en Cuba. No en balde Enrique José Varona expresó que “la academia habanera representaba la mayor suma de cultura, la mayor suma de saber que hasta entonces había producido nuestra patria” (citado en Pruna Goodgall, 2011: 21). Los historiadores de la ciencia en Cuba han llegado a afirmar rotundamente que “cualquier estudio futuro de la ciencia cubana en su conjunto requerirá la evaluación con técnicas modernas del corpus completo publicado de *Anales*” (Valero González y Jiménez Pérez, 2022).

El objetivo en la presente investigación es explorar las características lingüísticas más prominentes de los títulos científicos que se escriben en la actualidad en Cuba. Buscamos responder la siguiente pregunta de investigación: ¿cómo se comporta lingüísticamente la titulación de los textos científicos cubanos en la actualidad? La información resultante permitiría ilustrar una tendencia general, sin establecer distinciones entre disciplinas ni géneros, que serviría

---

de exponer sus conocimientos y más actuales investigaciones, históricamente los académicos han buscado dar respuesta a las problemáticas sociales que ha enfrentado el país. Por ello, para referirse a la historia nacional y al desarrollo de la ciencia en Cuba, sin duda es necesario mencionar a la Academia.

para documentar el panorama global de la titulación en las publicaciones científicas de Cuba. Por esta razón, hemos seleccionado una de las revistas científicas más antiguas del país, que tiene la ventaja adicional de ser multidisciplinaria. *Anales* abarca todas las temáticas que ocupan a la Academia cubana: ciencias naturales y exactas, sociales y humanísticas, agrarias y de la pesca, técnicas y biomédicas. Por tratarse de una publicación centenaria, nos permitirá establecer comparaciones y trazar tendencias a través del análisis de todo el arco temporal de su historia, cuyo período digital revisamos ahora.

## METODOLOGÍA

En esta comunicación adoptamos una aproximación textual sincrónica para explorar las características lingüísticas de los títulos especializados que se producen en la actualidad en Cuba. Para la muestra actual, seleccionamos una revista multidisciplinaria, *Anales de la Academia de Ciencias de Cuba*, tomada en su periodo digital. Tal segmentación obedece a la hipótesis de que la introducción de las tecnologías de la información y las comunicaciones (TIC) ha constituido un cambio de contexto importante en la evolución de los géneros especializados (Kress, 2003), por lo que, en una segunda fase del estudio, los resultados actuales serán comparados con las épocas de salida impresa, en una aproximación diacrónica.<sup>6</sup> Para esta última, consideraremos otras etapas importantes en el desarrollo interno de la revista (González Navarro y Arduengo García, 2022), de la ciencia y de las publicaciones científicas, en Cuba y en el mundo.

La muestra actual quedó integrada por 200 títulos (2639 palabras), similar en extensión a las de otros autores (Morales, Perdomo, Cassany, *et al.*, 2022; Morales, Perdomo, Cassany, *et al.*, 2020b; Morales, Perdomo, Cassany, *et al.*, 2020a; Jasrial, Syahrul, Mukhaiyar *et al.*, 2024). Los textos fueron tomados de los tres números más recientes publicados por la revista, correspondientes a los años 2022 a 2024. De ellos seleccionamos sólo textos escritos originalmente en español y por cubanos, y descartamos los textos de opinión, como

<sup>6</sup> Algunos autores ya indagan sobre el posible efecto de las TICs en la estructura retórica de los géneros especializados (*cf.*: Trebucq, Strieder y Ruzo, 2012).

los editoriales y algunas contribuciones especiales, así como otros textos de la vida interna de la Academia, no propiamente científicos (por ejemplo, discursos, memorias).

En el análisis, de tipo textual, examinamos manualmente la muestra en varias ocasiones, primero, por separado, cada investigador y, en un segundo tiempo, de manera conjunta, para beneficiar la fiabilidad de los resultados. Para que estos sean comparables con otros estudios, adoptamos la propuesta que Salager-Meyer y Alcaraz Ariza (2013) hicieron en su estudio diacrónico de los títulos científicos en inglés, pues es la que siguen los autores cuyos resultados compararemos con los nuestros. Así, además de la longitud de los títulos, calculada en cantidad de palabras, consideramos la siguiente estructura del aparato titular:

**Títulos proposicionales** (llamados por otros “oraciones completas” para describir títulos que contienen un verbo conjugado). Aquí se incluyen los títulos escritos tanto en modalidad aseverativa como interrogativa, ya sea presentados con signos de interrogación o sin ellos. Ejemplo 1:

- (1) La unión y formación de poros de las actinoporinas están determinadas por las propiedades fisicoquímicas de la membrana

Con todo, el contenido de los títulos interrogativos no es oracional, sino que al usarlos se interpreta la información nominal que corresponde a la palabra interrogativa con que se forman: “modo o modos de hacer algo” (Real Academia de la Lengua Española y Asociación de Academias de la Lengua Española (ASALE), 2009: 3156). Con este tipo de título no se solicita al lector que satisfaga una incógnita, sino que se le anticipa que en el texto se presentará tal información.

**Títulos de formato complejo** (llamados por otros “títulos compuestos”), pues incluyen dos o más elementos, separados formalmente al modo de títulos secundarios (subtítulos o “segundos títulos”). Ejemplo 2:

- (2) Gestión de Gobierno basada en ciencia e innovación: avances y desafíos

Los segundos o terceros componentes son opcionales y suelen introducir paráfrasis –infrecuentes en los textos altamente especializados– o indicaciones de tipo especificativo o explicativo del título principal, que sí es obligatorio

en nuestra cultura (Genette, 2001: 52-53). Se separan del título principal por algún signo de puntuación: punto y seguido, coma, dos puntos.

En sus estudios, los autores de habla inglesa los han denominado *colon titles* (Dillon, 1982), *hanging titles* (Anthony, 2001), *colonic titles* (Hartley, 2005), *compound titles* (Hartley, 2007), *multiple-unit titles* (Moattarian 2015) o *series titles* (citados en Salager-Meyer, Alcaraz Ariza, Luzardo Briceño, 2013; Morales, Perdomo, Cassany, et al., 2020b). Según Salager-Meyer, Alcaraz Ariza y Luzardo Briceño (2013), la función semántica de estos segundos y terceros componentes ha variado con el tiempo, asunto al que atenderá nuestra indagación diacrónica.

**Títulos simples.** Integrados por un sólo componente no proposicional. Ejemplo3:

(3) Aplicaciones del ozono en la agricultura cubana

Estos títulos suelen ser frases nominales o frases preposicionales que contienen una preposición tematizadora (por ejemplo, *de, sobre, acerca de*), a veces precedida de alguna información superflua, del tipo *consideraciones sobre, observaciones acerca de* (Gross, Harmon, Reidy, 2002). Finalmente, también se documentan títulos simples introducidos por frases verbales no personales, en particular gerundios, habituales en lengua inglesa.

Otros rasgos morfosintácticos que consideramos sintomáticos y buscamos documentar fueron la presencia de artículo (determinado/indeterminado) y el uso de acortamientos gráficos, fueran siglas o abreviaturas. En la configuración semántica, buscamos la presencia de información que, a modo de etiquetas, identificara de manera explícita el género textual o la metodología seguida en el estudio, factores a los cuales atienden otras investigaciones (Yang, 2019; Kotwica, 2018) y será preciso explorar en diacronía.

## RESULTADOS Y ANÁLISIS

En relación con la extensión de los títulos, encontramos un promedio de 13.19 palabras por título, longitud similar al promedio de las extensiones documentadas en otros estudios sincrónicos que han sido conducidos en períodos históricos similares (13.29 palabras), pero considerando un solo



género discursivo: 12.7 palabras en casos clínicos (Morales, Perdomo, Cassany, Burdiles, *et al.*, 2022); 13.68 en revisiones sistemáticas (Morales, Perdomo, Cassany, Izarra, 2020b); 15.8 en artículos originales de investigación (Morales, Perdomo, Cassany, *et al.*, 2020a) e incluso en estudios comparativos entre disciplinas: media de 11 palabras (Soler, 2009). Según Genette (2001: 269), la extensión de los títulos ha disminuido diacrónicamente, y ya desde principios del siglo XIX, tras un largo período de preferencia por títulos extensos, el estilo de los títulos pasa de locuaz a discreto, como fuera en la Antigüedad clásica. Sin embargo, pareciera que en el ámbito especializado el comportamiento ha sido basculante entre menores y mayores extensiones, dependiendo de las convenciones retóricas al uso por la comunidad discursiva en diferentes períodos de desarrollo científico y tecnológico. Actualmente, *Anales* pide a sus autores una extensión límite de 15 palabras.<sup>7</sup>

En la actualidad, si bien algunos autores han recomendado proporcionar en el título información suficiente como para indicar el tópico o temática de estudio y mostrar también el alcance de la investigación (Swales y Feak, 2004), otros han recomendado la menor cantidad posible de palabras (Day y Gastel, 2008). Es posible que, tras una etapa de acortamiento de los títulos especializados, seamos testigos más recientemente de cierto aumento de la extensión debido al interés por explicitar información metodológica e incluso de género textual que en ciertas etapas no solía figurar en el aparato titular. Así sugiere Yang (2019) y, de estar sucediendo, es lógico que se documente primero en inglés, que continúa siendo la *lingua franca* de la ciencia, y luego en las otras lenguas nacionales.

Con todo, especulamos que ha de estar aconteciendo un proceso global de normalización y unificación de todos los aspectos formales del estilo científico, que afecta a todas las lenguas, géneros y disciplinas, como en su momento aconteció con el esquema superestructural IMRyD (*introducción, métodos, resultados y discusión*).<sup>8</sup> En definitiva, las condiciones actuales de

<sup>7</sup> Consúltense sus instrucciones editoriales en <https://revistaccuba.sld.cu/index.php/revacc/about/submissions#authorGuidelines>

<sup>8</sup> Sollaci y Pereira (2004) refieren que, desde el surgimiento de los documentos públicos de comunicación científica (*scientific papers*) en 1665 y hasta finales del siglo XIX, la carta y el reporte de experimentos coexistieron de manera hegemónica. La superestructura IMRyD comenzó a aparecer a partir de 1940, y su frecuencia creció aceleradamente después de la Segunda Guerra Mundial. Para 1950 llegó a estar

internacionalización, no sólo de las relaciones económicas y políticas, sino también de la ciencia, y en general de la información como activo económico de tanto valor agregado como la transferencia misma de bienes y servicios, refuerzan la necesidad de normalizar los elementos que canalizan el intercambio: las unidades básicas de transferencia (los términos) y los sistemas de almacenamiento y recuperación de la información (Prieto Acosta, 2015).

Aquí nos interesa apenas documentar valores promedio para más adelante describir la ruta diacrónica en Cuba.<sup>9</sup> Siguiendo este enfoque, documentamos que sólo en tres trabajos (1.5 %) se reflejaba de manera explícita el género textual, pero advertimos que, en la mayoría de las revistas científicas cubanas contemporáneas, la información genérica suele estar contenida en antetítulos con que identifican la sección de la revista donde el artículo ha sido incluido. Suele ocurrir que los títulos de esas secciones sean designativos del género textual y no temáticos. Por tanto, de figurar la información del género en el título del artículo, sería remática, casi redundante. No es de extrañar, pues, que fuera mayor (31.5 %) la cantidad de títulos que contuviera información de corte metodológico (*research procedure titles* en Salager-Meyer, Alcaraz Ariza y Luzardo Briceño, 2013), ausente en los antetítulos. Véanse algunos ejemplos (4-5):

- (4) Producción científica cubana. Estudio bibliométrico comparativo en SciELO Citation Index y en el Directorio de Revistas en Acceso Abierto
- (5) Evaluación preclínica y clínica de SOBERANA® 01. Candidato vacunal contra la COVID-19

La estructura formal del aparato titular se comportó como sigue: un 2.5 % de títulos proposicionales (ninguno interrogativo), un 29.5 % de títulos complejos (todos integrados por dos componentes), y un rotundamente mayoritario 68 %

---

presente en poco más del 10 % de los textos científicos, y durante la década de 1970 se convirtió en la más frecuente forma de publicación de resultados científicos (80 % de las publicaciones especializadas en medicina).

<sup>9</sup> Por ejemplo, según Yang (2019: 98), los títulos están ganando en longitud y cada día los autores emplean más palabras significativas para ampliar la visibilidad del artículo y mejorar las posibilidades de citación.

de títulos simples formados por frases nominales con frases preposicionales como posmodificadores. El reducido porcentaje de títulos proposicionales (“oraciones completas”), la comparativamente baja frecuencia de títulos complejos y la notoria mayoría de títulos nominales simples coinciden con lo que documentan otras series sobre el estilo científico (Soler, 2009; Morales, Perdomo, Cassany, *et al.*, 2020a), pero con algunas diferencias en cuanto a la frecuencia de títulos complejos (Morales, Perdomo, Cassany, Izarra, 2020b).

Según Morales Perdomo, Cassany *et al.* (2022), la declaración PRISMA ha establecido recientemente que el título de las revisiones sistemáticas y los metaanálisis debe incluir, al menos, el tema y el método; es decir, además de la información temática, debe contener la especificación de que se trata de revisiones sistemáticas o de metaanálisis, lo cual ha ido condicionando el uso de títulos complejos en ese género (Moher et al., citados en Morales, Perdomo, Cassany *et al.* 2020a). La puntuación de estos últimos se caracterizó por el empleo de dos puntos y punto y seguido como separadores mayoritarios (cada uno con el 41.3 %), y en mucho menor medida por el uso de la coma (15.5 %), comportamiento que también coincide con el explicado por los autores de las series aquí citadas. Véanse algunos ejemplos (6-8):

- (6) Primer brote de COVID-19 en Cuba. Factores de riesgo de morbimortalidad y respuesta específica de anticuerpos en convalecientes
- (7) Pasado, presente y futuro de las vacunas veterinarias clásicas cubanas: aportes a la salud animal
- (8) Gestión de seguridad psicológica, su efectividad durante la emergencia sanitaria por COVID-19

En general, pasada la Edad Media, se documenta una gradual preferencia por los títulos nominales frente a los proposicionales (narrativos), y predominan los títulos temáticos, introductores de información nueva, frente a los remáticos o puramente designativos (Genette, 2001: 264). Ésta habría de ser la tendencia general de comportamiento de los títulos especializados, si consideramos el modo como se gestó el estilo científico desde los comienzos de la institucionalización de la ciencia en el siglo xvii. Impulsado primero por Francis Bacon y “fabricado” luego por la Royal Society inglesa en rechazo

al estilo recargado del Barroco, el estilo expositivo idóneo para la ciencia habría de reunir tres importantes cualidades: brevedad, perspicuidad (claridad, precisión) y decoro (adecuación), además de sobriedad expresiva (Hernández Guerrero y García Tejera, 1994: 119-20). Como se dijo antes, aspiramos a documentar –y explicar– las variaciones de longitud de los títulos y, sobre todo, el probable aumento de la extensión, en el análisis diacrónico.

Con respecto al uso de artículos, determinados o indeterminados, hallamos sólo 26 ocurrencias (13 % de los títulos), 18 de las cuales correspondieron al artículo determinado. Documentamos la utilización del indeterminado en construcciones explicativas en aposición a un sustantivo, lo cual es considerado un anglicismo sintáctico (Segura, 2001: 54). Es posible que en el análisis diacrónico podamos establecer el momento de aumento de la frecuencia de uso de artículos indeterminados en los textos cubanos y, en general, del comportamiento del uso de artículos. En un par de títulos complejos entramos la confluencia de ambos (9-10):

- (9) *Las ataxias en Cuba: Un proyecto inclusivo desde la ciencia, la sociedad y las familias*
- (10) *La guarda y cuidado y el régimen de comunicación de los menores de edad en familias ensambladas. Una propuesta para Cuba*

También hallamos una baja frecuencia (6.5 %) de empleo de otros premodificadores del núcleo nominal en los títulos simples (11-13), que ha sido documentada (6.2 %) por autores como Morales, Perdomo, Cassany *et al.* (2022). Su evitación en los textos actuales podría ser un recurso para acortar los títulos, al modo de la omisión de artículos (Soler, 2009: 54).

- (11) *Aportes al estudio de la entonación del español de Cuba. Patrones entonativos y prosodia del discurso*
- (12) *Consideraciones del abordaje Keyhole endoscópico para tumores de la base craneal*
- (13) *Contribución al conocimiento de la interacción Rhizobium-arroz (Oryza sativa L.). Oportunidades para la biofertilización del cultivo*

Kotwica (2018) describe el uso de estas etiquetas o palabras descriptivas que especifican el carácter del texto como parte de una “interesante tendencia” en la nomenclatura de los artículos científicos publicados en las revistas institucionales españolas en el período comprendido entre 1799 y 1920. En su opinión, permiten ver una progresiva especialización en cuanto a los contenidos de las publicaciones científicas, aludiendo a la actividad científica realizada en cada momento histórico (*observaciones, descripciones, estudios, etc.*) (Kotwica, 2018: 2013). Por consiguiente, a este aspecto deberá atender el estudio diacrónico.

Finalmente, encontramos un 21.5 % de uso de acortamientos gráficos (siglas y abreviaturas), dato que no pudimos comparar globalmente con otros estudios, pues no hallamos documentación de la frecuencia de uso de siglas, sino sólo del empleo de abreviaturas. Así, atendiendo sólo a las abreviaturas, el 6.5 % sí es comparable con la baja frecuencia reportada por Morales *et al.* en dos estudios: 10.8 %, en casos clínicos (Morales, Perdomo, Cassany, *et al.*, 2020a) y 3 % en revisiones sistemáticas (Morales, Perdomo, Cassany *et al.*, 2020b). Todas las abreviaturas correspondieron a nombres científicos (taxones botánicos o zoológicos), excepto una, la abreviatura de la palabra *doctor* (Dr.) (14-16):

- (14) Descubrimiento de ferroquelatasa 1 en *Nicotiana tabacum* L.: papel en la respuesta al estrés abiótico y el desarrollo de las plantas
- (15) Primera evidencia molecular de especies de micoplasmas hemotrópicos (*Mycoplasma* spp.) en animales de importancia económica y social en Cuba
- (16) Dr. Carlos Manuel Ramírez Corría, heraldo de la escuela neuroquirúrgica cubana

En nuestro caso, por el arco temporal estudiado (2022 a 2024), hallamos lógica la alta cifra que registraron las siglas COVID-19 y SARS-Cov-2 en la muestra, las cuales alcanzan 20 ocurrencias del total de 31 siglas. Las restantes siglas corresponden fundamentalmente a denominaciones de instituciones, de bases de datos, de medicamentos y medios de investigación (17-19):

- (17) Creación, resultados y desaparición del ICINAZ (1973-2011)


- (18) Producción científica cubana. Estudio bibliométrico comparativo en SciELO Citation Index y en el Directorio de Revistas en Acceso Abierto
- (19) UMELISA EGF: un medio para la investigación y el diagnóstico acompañante de terapia personalizada

## CONCLUSIONES








Nuestros resultados corroboran la tendencia que otros autores han descrito para los títulos científicos que se elaboran en la actualidad en lengua española, pero esta vez para las publicaciones cubanas. Los títulos científicos que se producen en la actualidad en Cuba se asemejan en extensión y estructura formal a los que se elaboran en otros países de habla hispana, incluso sin establecer diferencias entre géneros y campos disciplinares. Así mismo, tampoco hubo diferencias en cuanto al uso de los signos de puntuación, el empleo de artículos (determinado/indeterminado) y otros premodificadores del núcleo nominal o en cuanto al uso de acortamientos gráficos (siglas y abreviaturas).

Esta información nos permitirá comparar y describir, en diacronía, las prácticas de titulación en las publicaciones científicas cubanas como parte de una descripción integral –lingüística y retórica– de la evolución del estilo científico en Cuba. A fin de cuentas, el estilo científico no es estático ni atemporal, y pensar que es algo fijo, que trasciende toda situación social, es “*rhetorically naive and historically wrong*” (Bazerman, 1988: 315). En virtud de ello, muchas cuestiones relativas al estilo científico sólo pueden ser explicadas desde una perspectiva histórica, tomando en consideración el contexto cultural y social en el cual los textos son producidos.


## BIBLIOGRAFÍA

- Bazerman, Charles (1988), *Shaping written knowledge, The genre and activity of the experimental article in science*, Wisconsin, University of Wisconsin Press.
- Day, Robert A. y Barbara Gastel (2008), *Cómo escribir y publicar trabajos científicos*, Washington, OPS (Organización Panamericana de la Salud), consultado el 26 de agosto de 2024. 

- De la Fuente González, Miguel Ángel (1997), “Las funciones de los títulos en la descodificación lectora”, *Tabanque: Revista pedagógica*, núm 12-13, 185-202, consultado el 26 de agosto de 2024. [🔗](#)
- Genette, Gerard (2001), *Umbrales*, México, Siglo XXI Editores, consultado el 26 de agosto de 2024. [🔗](#)
- González Navarro, Lisset y Darwin Antonio Arduengo García (2022), “Anales: historia y patrimonio”, *Anales de la Academia de Ciencias de Cuba*, vol. 12, núm. 3, consultado el 26 de agosto de 2024. [🔗](#)
- Gross, Alan G., Joseph E. Harmon y Michael Reidy (2002), *Communicating Science: The Scientific Article from the 17th Century to the Present*, New York, Oxford University Press. [🔗](#)
- Hernández Guerrero, José Antonio y María del Carmen García Tejera (1994), *Historia breve de la Retórica*, Madrid, Síntesis.
- Hyland, Ken y Hang Joanna Zou (2022), “Titles in research articles”, *Journal of English for Academic Purposes*, vol 56, marzo 2022, 101094, consultado el 26 de agosto de 2024. [🔗](#)
- Jasrial, Dedi, Syahrul Ramadhan, Mukhaiyar Mukhaiyar y Joseph Benjamin Archibald Afful (2024), “Linguistic Characteristics of Research Article Titles in National and International ELT Journals”, *Journal of Languages and Language Teaching*, vol. 12, núm. 1, pp. 88-103, consultado el 26 de agosto de 2024. [🔗](#)
- Kotwica, Dorota (2018), “De observaciones a estudios. Títulos de los artículos científicos españoles entre 1799 y 1920”, *Normas: revista de estudios lingüísticos hispánicos*, vol. 8, pp. 201-14, consultado el 3 de septiembre de 2024. [🔗](#)
- Kress, Gunther (2003), *Literacy in the new media age* (routledge).
- McClellan III, James E. (1979), “The scientific press in transition: Rozier’s journal and the scientific societies in the 1770s”, *Annals of Science*, vol. 36, pp. 425-449, consultado el 17 de octubre de 2024. [🔗](#)
- Mestre, A., y J. J. Muñoz (1864), “Prospecto”, en *Anales de la Real Academia de Ciencias Médicas, Físicas y Naturales de La Habana*, I.
- Morales, Oscar Alberto, Bexi Perdomo, Daniel Cassany, Gina Burdiles, Jhon Álvarez Ahlgren y Yaymar Dávila (2022), “Estructura lingüística de títulos de casos clínicos odontológicos en español”, vol. 55, núm. 109, pp. 631-60, consultado el 3 de septiembre de 2024. [🔗](#)
- Morales, Oscar Alberto, Bexi Perdomo, Daniel Cassany, Joel Acevedo y Jhon Álvarez Ahlgren (2020a), “Estructura léxico-gramatical de títulos de artículos de investigación de odontología en español”, *RLA. Revista de lingüística teórica y*

- aplicada*, vol. 58, pp. 69-92, consultado el 17 de octubre de 2024. 
- Morales, Óscar Alberto, Bexi Perdomo, Daniel Cassany y Élix Izarra (2020b), “Estructuras morfosintácticas y funciones retóricas de títulos de revisiones sistemáticas publicadas en español en revistas odontológicas: un análisis de género discursivo”, *Panace@*, vol. 21, núm. 51, pp. 62-74, consultado el 14 de octubre de 2024. 
- Moya Méndez, Misael (2013), *Praxis editorial y lengua española. Introducción a una teoría de la edición ordinaria*, La Habana, Editorial Letras Cubanas.
- Posteguillo-Gómez, Santiago y Jordi Piqué-Angordans (2007), “El lenguaje de las ciencias médicas: comunicación escrita”, en Enrique Alcaraz-Varó, José Mateo-Martínez y Francisco Yus-Ramos (eds.), *Las lenguas profesionales y académicas*, Barcelona, Ariel Lenguas Modernas.
- Prieto Acosta, Diana Elena (2015), “La inscripción explícita del autor en el discurso científico: análisis diacrónico y perspectivas”, *Universidad de La Habana*, núm. 279, pp. 6-2, consultado el 14 de octubre de 2024. 
- Prieto Castro, María Rosa (2018), *Producción del discurso de divulgación científica en lengua española durante el primer tercio del siglo XX: contexto e ideología*, España, Editorial Universidad de Alcalá.
- Pruna Goodgall, Pedro M. (2011), *Ciencia y científicos en Cuba colonial. La Real Academia de Ciencias de La Habana 1861-1898*, La Habana, Academia de Ciencias de Cuba.
- Real Academia de la Lengua Española y Asociación de Academias de la Lengua Española (2009), *Nueva gramática de la lengua española*, Madrid, Espasa Libros.
- Salager-Meyer, Françoise y María Ángeles Alcaraz Ariza (2013), “Titles are serious stuff: a historical study of academic titles”, *Jahr–European Journal of Bioethics*, vol. 4, núm. 7, pp. 257-71, consultado el 14 de octubre de 2024. 
- Salager-Meyer, Françoise, María Angeles Alcaraz Ariza, Marianela Luzardo Briceño (2013), “Titling and authorship practices in medical case reports: A diachronic study (1840-2009)”, *Communication & Medicine*, vol. 10, núm. 1, pp. 63-80, consultado el 14 de octubre de 2024. 
- Segura, Joaquín (2001), “Los anglicismos en el lenguaje médico”, *Panace@*, vol. 2, núm. 3, pp. 52-57, consultado el 14 de octubre de 2024. 
- Soler, Viviana (2018), “Estudio exploratorio de títulos de tesis doctorales redactados en lengua española”, *Lebende Sprachen*, vol. 63, núm. 2, pp. 374-392, consultado el 14 de octubre de 2024. 



- Soler, Viviana (2009), “Títulos científicos en lengua española: Estudio exploratorio”, *Lebende Sprachen*, vol. 54, núm. 2, pp. 50-58, consultado el 14 de octubre de 2024. 
- Swales, John (2008), *Genre Analysis: English in academic and research settings*, Cambridge, Cambridge University Press.
- Swales, John M. y Christine B. Feak (2004), *Research genres: Explorations and applications*, Cambridge, Cambridge University Press.
- Trebucq, María Dolores, María Silvina Strieder y María Dolores Gonzalez Ruzo (2012), “Estudio diacrónico de la estructura retórica de la sección Resultados y Discusión de artículos de investigación en inglés”, ponencia presentada en VIII Congreso de la Asociación de Lingüística Sistemática Funcional, Montevideo.
- Valero González, Mercedes y Alejandro Jiménez Pérez (2022), “Anales de Academia: mirada en la tradición”, *Anales de la Academia de Ciencias de Cuba*, vol. 12, núm. 3.
- Vidal Ramírez, Gerardo (2022), *Usos y abusos de la palabra “retórica”*, Santa Fe, Ediciones UNL.
- Yang, Wenhsien (2019), “A diachronic keyword analysis in research article titles and cited article titles in applied linguistics from 1990 to 2016”, *English Text Construction*, vol. 12, núm. 1, pp. 84-102.

**DIANA ELENA PRIETO ACOSTA:** Doctora en Ciencias Lingüísticas; investigadora agregada del Instituto de Literatura y Lingüística José Portuondo Valdor, y profesora asistente de la Facultad de Artes y Letras de la Universidad de La Habana. Es miembro de varias sociedades científicas, entre las cuales se encuentran la Asociación de Lingüistas de Cuba, la Asociación Iberoamericana de Retórica, la International Pragmatic Association, la Asociación Latinoamericana de Estudios de la Escritura en Educación Superior y Contextos Profesionales (ALES) y de la *Association for Writing across the Curriculum (AWAC)*. Entre sus líneas de investigación se encuentran la pragmática lingüística, la fraseología y la retórica. Es editora de libros y revistas científicas.

**EVELYN RUBIO FIGUERAS:** Licenciada en Ciencias de la Información por la Facultad de Comunicación de la Universidad de La Habana. Fungió como Especialista en Vigilancia Tecnológica en el Centro de Investigación y Desarrollo Técnico, y ahora trabaja como Especialista en Políticas en la Dirección de Gestión Documental y Archivos del Ministerio de Ciencia Tecnología y

Medio Ambiente (DGDyA-CITMA). Es miembro del Programa Sectorial de Gestión Documental y Archivos, dirigido por el Archivo Nacional de la República de Cuba y la DGDyA- CITMA. Su trabajo se enfoca principalmente en el establecimiento de metodologías para normar la gestión de documentos digitales en el país. Es coautora del libro *Directrices y requisitos para la gestión de documentos archivísticos en el escenario de la transformación digital en Cuba*, actualmente en prensa. Colabora en la redacción del *Manual de procedimientos para la implementación de la base normativa del Sistema Nacional de Gestión Documental y Archivos de la República de Cuba*.

D. R. © Diana Elena Prieto Acosta, Ciudad de México, enero-junio, 2024.

D. R. © Evelyn Rubio Figueras, Ciudad de México, enero-junio, 2024.



## **Lexicon, Gastronomy and Culture: Loja (Ecuador) and Cerro Azul, Veracruz (Mexico)**

**SONIA E. MORETT ÁLVAREZ** 

El Colegio de México-Doctorado en Lingüística  
[sonia.morett@gmail.com](mailto:sonia.morett@gmail.com)

**ISIS NATHALY ZAMBRANO OJEDA** 

El Colegio de México - Doctorado en Lingüística  
[izambrano@colmex.mx](mailto:izambrano@colmex.mx)

**Abstract:** *This article exhibits a lexicological approach to different dishes belonging to the heritage culture of Loja and Cerro Azul. The main objective is to verify the links between language and culture, taking food as an empirical testimony. Through the information provided by six interviewees, we selected four referential fields common to the regional gastronomy of two geographically distant areas, but with shared elements. The peculiarities that characterize culinary preparations and associated traditions are considered. Our perspective is ethnolinguistic and historical, and seeks to motivate reflection on the importance of culture-centered approaches in lexical studies to the study of lexis.*

**KEYWORDS:** CULTURAL MEANING; ETHNOLINGUISTIC SIGN; GASTRONOMIC LEXICON; LEXICON OF AMERICAN SPANISH; TRADITIONAL DISHES; CULTURAL PARALLELS

**RECEPTION:** 05/08/2024

**ACCEPTANCE:** 20/09/2024

## Léxico, gastronomía y cultura: Loja (Ecuador) y Cerro Azul, Veracruz (México)

**SONIA E. MORETT ÁLVAREZ** 

El Colegio de México - Doctorado en Lingüística  
[sonia.morett@gmail.com](mailto:sonia.morett@gmail.com)

**ISIS NATHALY ZAMBRANO OJEDA** 

El Colegio de México - Doctorado en Lingüística  
[izambrano@colmex.mx](mailto:izambrano@colmex.mx)

**Resumen:** Este artículo presenta un acercamiento lexicológico a diferentes platillos pertenecientes a la cultura patrimonial de Loja y Cerro Azul. El objetivo principal es verificar los nexos entre lengua y cultura tomando la alimentación como testigo empírico. A través de la información proporcionada por seis entrevistados, seleccionamos cuatro campos referenciales comunes a la gastronomía regional de dos zonas geográficamente distantes, pero con elementos compartidos. Se consideran las peculiaridades que caracterizan preparaciones culinarias y tradiciones asociadas. Nuestra perspectiva es etnolingüística e histórica, y busca motivar la reflexión sobre la importancia de los enfoques centrados en la cultura para el estudio del léxico.

**PALABRAS CLAVE:** SIGNIFICADO CULTURAL; SIGNO ETNOLINGÜÍSTICO; LÉXICO GASTRONÓMICO; LÉXICO AMERICANO; PLATILLOS TRADICIONALES; PARALELISMOS CULTURALES

**RECEPCIÓN:** 05/08/2024

**ACEPTACIÓN:** 20/09/2024

## INTRODUCCIÓN<sup>1</sup>

La gastronomía forma parte del patrimonio y la identidad cultural de los pueblos, lo cual se refleja a través de las particularidades de cultivo, los productos, los platillos típicos, así como en las formas de servir y consumir (Fusté-Forné, 2016). Por tratarse de un mecanismo de transmisión cultural que conecta la memoria, el linaje y la herencia colectiva (Kershenovich Schuster, 2015), la gastronomía nos permite rescatar conocimientos, prácticas culturales (rituales, festejos, etc.) y destrezas culinarias que se sintetizan en un amplio inventario de vocablos que varían en cada país, región y localidad.

Por otra parte, nombrar implica reconocer lo que es relevante para una comunidad; igualmente, supone una clasificación simbólica socialmente compartida del entorno. Así, la lengua se va modelando por el conocimiento y las necesidades de la cultura de la que es parte. A través de las expresiones lingüísticas podemos conocer la forma en que sus usuarios comprenden, conocen y transforman su realidad. En particular, el léxico hace referencia a objetos, saberes e impresiones que varían de una cultura a otra, ya que “refleja el ambiente cósmico dentro del cual un pueblo con su cultura vive. Lo que no aparece en este contexto no tiene necesidad de una palabra” (Lenkendorf, 2010: 11).

Es por todo esto que nos ha parecido relevante estudiar las manifestaciones del léxico de la cultura alimentaria, desde lo regional, de dos países latinoamericanos que presentan una enorme diversidad cultural y, por ende, gastronómica: Ecuador y México. Si bien la gastronomía mexicana es reconocida mundialmente, los platillos que le dan fama internacional corresponden a una visión reducida y homogeneizada de su cocina, mientras que la ecuatoriana muy poco se conoce fuera del país. De ahí la necesidad de documentar y dar a conocer expresiones regionales del patrimonio cultural intangible de estas dos naciones. Específicamente, el estudio se centra, por una parte, en la ciudad de

<sup>1</sup> Este trabajo le debe su existencia a los seis colaboradores de Loja y Cerro Azul que accedieron a compartirnos su conocimiento, por lo que les agradecemos enormemente, así como a nuestros profesores Niktelol Palacios y Pedro Martín Butragueño, quienes nos aportaron ideas y revisaron los primeros borradores.

Loja, ubicada al sur de la región andina del Ecuador y, por el lado mexicano, en la población de Cerro Azul, Veracruz, en la Huasteca.

A partir de alimentos emblemáticos compartidos por ambas regiones: maíz, ají o chile, ganado porcino y vacuno, plátano y guineo, nos trazamos el objetivo de identificar los vocablos que, en torno a ellos, conforman campos referenciales por zona de estudio, y de penetrar en los significados culturales de los elementos con mayor relevancia social para cada comunidad (véase Moreno, 2021; López y Alcalde, 2023).

Se aborda la relación entre léxico y cultura desde una perspectiva etnolingüística e histórica, en el entendido de que “la cultura es histórica y ninguna sociedad puede comprenderse sin entender a su historicidad, a sus transformaciones” (Grimaldo, 2006: 42). En este sentido, esta investigación se vio motivada por el interés de describir y explicar tanto lo particular y diferencial como los puntos de convergencia entre los datos de las dos regiones, presunción inicial (derivada de la observación de ciertos paralelismos: recursos vegetales endémicos comunes, y la introducción y explotación extensiva de otros durante la colonia) que se fue corroborando durante el desarrollo de la misma.

## LOJA: HISTORIA Y GASTRONOMÍA

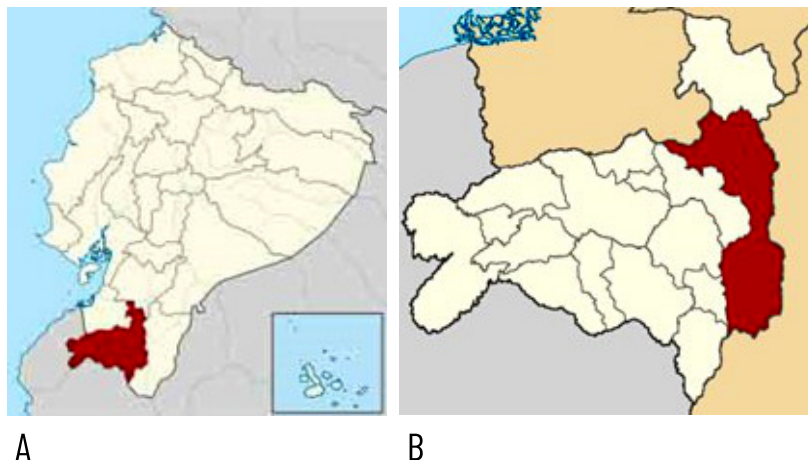
La ciudad de Loja se ubica al sur de los andes ecuatorianos (figuras 1 y 2); es cabecera del cantón Loja y capital de la provincia que recibe el mismo nombre. Según el Censo 2022, la ciudad de Loja contaba con 203 496 habitantes en la zona urbana (Instituto Nacional de Estadística y Censos, 2022).

La topografía irregular de la provincia se caracteriza por los climas frío-temperados (12 °C - 17 °C) como en la ciudad de Loja y climas cálidos (20 °C - 24 °C) como en los valles de Catamayo, Malacatos y Vilcabamba<sup>2</sup> que rodean a la ciudad. Ubicado al norte de Loja se encuentra el cantón Saraguro,

<sup>2</sup> El valle de Vilcabamba es conocido nacional y mundialmente como el ‘Valle de la longevidad’ debido a que sus habitantes generalmente sobrepasan los 100 años de edad. Este es uno de los valles que cuenta con una gran presencia extranjera; personas de varias nacionalidades se han radicado en la zona.

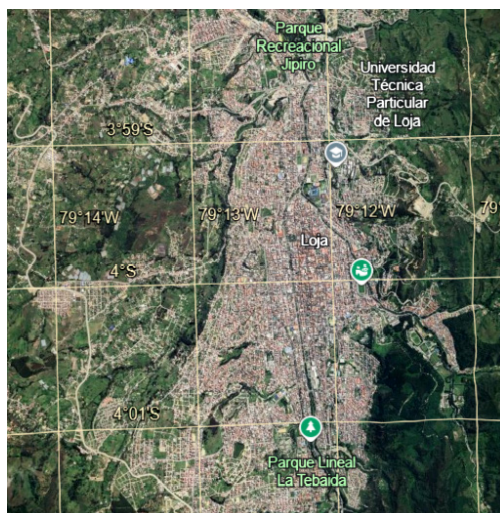
territorio de uno de los pueblos indígenas de la nacionalidad Kichwa. De ahí que en la capital lojana exista una presencia constante de los saraguros.

FIG. 1. A. MAPA DEL ECUADOR, PROVINCIA DE LOJA. B. MAPA DEL CANTÓN LOJA



Fuente: Sites Google, 2018.

FIG. 2. MAPA DE LA CIUDAD DE LOJA



Fuente: Google Earth, 2024.



La ciudad fue fundada en la época colonial por Alonso de Mercadillo, el 8 de diciembre de 1548, presumiblemente sobre una población preexistente. Actualmente, es conocida como “la capital musical y cultural del Ecuador” por sus diversos aportes en la música y las artes además de ser el lugar de nacimiento de muchos personajes ilustres del país y, también, por sus manifestaciones culturales destacadas en la arquitectura, la religión y la gastronomía.

Las principales actividades económicas de la zona son el comercio, la agricultura, la ganadería y la educación. La actividad pecuaria se caracteriza por la crianza de aves, *cuyes*,<sup>3</sup> ganado vacuno, porcino, ovino y mular; por otra parte, la producción agrícola incluye cultivos permanentes y transitorios como la caña de azúcar, maíz, arroz, banano, café y yuca, comercializados en las ferias libres y en los principales mercados del cantón (Lima, 2017: 3).

La gastronomía en Loja, como en otras provincias, fue influida por los españoles; sin embargo, tradiciones culinarias de los pueblos ancestrales que se asentaron en la provincia –los paltas y los saraguros– se mantienen hasta la actualidad. Dentro de los platillos más representativos de la ciudad destacan los que abordaremos en este estudio: la cecina de cerdo, el repe, los tamales lojanos y el ají de pepa, aunque la riqueza culinaria emblemática de la zona se extiende a muchos otros como *horchata*,<sup>4</sup> *cuy*,<sup>5</sup> *gallina cuyada*,<sup>6</sup> *bocadillos*,<sup>7</sup> *miel con quesillo*,<sup>8</sup> entre otros platillos que permiten observar que la población lojana ha conservado las tradiciones de generación en generación.

## CERRO AZUL: HISTORIA Y GASTRONOMÍA

Cerro Azul es una población con poco más de 100 años de historia, enclavada en la región huasteca, al norte del estado de Veracruz (figura 3).

<sup>3</sup> Conocido también como *cobayo* o *conejillo de indias*.

<sup>4</sup> Infusión preparada con alrededor de 22 hierbas y plantas medicinales.

<sup>5</sup> Atractivo gastronómico de la provincia que se prepara con diversas especias a la brasa.

<sup>6</sup> Gallina preparada de la misma manera que el cuy.

<sup>7</sup> Pequeños dulces en forma de cubo hechos con panela (producto elaborado a base de caña de azúcar) y maní.

<sup>8</sup> Postre hecho con miel de panela que se sirve acompañada con quesillo tierno. El quesillo es un tipo de queso fresco sin sal también típico de la provincia de Loja.

La Huasteca<sup>9</sup> comprende zonas de seis estados del país<sup>10</sup> y sus habitantes históricos pertenecen a los pueblos tének (o huasteco), nahua, tepehua, otomí, pame y totonaco (Instituto Nacional de Pueblos Indígenas, 2020), quienes desarrollaron sus culturas dentro de un geosistema que comprende sierra, llanos, costa, selva tropical, y una enorme diversidad de vegetación y fauna. El entorno natural de la Huasteca no recibió gran impacto sino hasta finales del siglo XIX, cuando las planicies del norte de Veracruz se transformaron en una gran cuenca ganadera; al iniciar el siglo XX, cambió nuevamente su fisonomía, al convertirse en la *Faja de Oro* de los recursos petrolíferos del país.

FIG. 3. UBICACIÓN DE CERRO AZUL EN LA HUASTECA BAJA (EN MORADO) DEL ESTADO DE VERACRUZ



Fuente: Tomado de Sistema de Información Estadística y Geográfica del Estado de Veracruz de Ignacio de la Llave [SIEGVER](2021). [[http://ceieg.veracruz.gob.mx/wp-content/uploads/sites/21/2019/06/Cerro-Azul\\_2019.pdf](http://ceieg.veracruz.gob.mx/wp-content/uploads/sites/21/2019/06/Cerro-Azul_2019.pdf)].

<sup>9</sup> Existen diferentes hipótesis sobre el origen de la palabra. Recogemos la información aportada por Montemayor (2007), para quien vendría del nombre náhuatl de la región, *Cuextecapan*; y Piña Chán (1989), según el cual podría derivarse de un caudillo: *Cuextécatl*.

<sup>10</sup> Tamaulipas, San Luis Potosí, Hidalgo, Querétaro, Puebla y Veracruz.

En 1916, brotó un prolífico pozo en lo que había sido la Hacienda de Cerro Azul; éste se convirtió en “el más productivo del mundo durante varios decenios del siglo xx” (Escobar y Carregha, 2002). La *fiebre del oro negro* atrajo –además de las compañías trasnacionales– a trabajadores de muchas partes del país,<sup>11</sup> que escapaban de los efectos económicos y sociales de la revolución. También llegaron chinos, árabes, europeos, centroamericanos, entre otros.

A decir de una cerroazulense (Rodríguez, 2012): “en la memoria de mi pueblo está gran parte de la historia petrolera del país”. Así, recuerdan las huelgas de 1934 y 1937, que rendirían frutos en el decreto de la expropiación petrolera. Con la fundación de Petróleos Mexicanos (PEMEX) llegaron los servicios básicos y aún más, se canalizaron recursos para la educación, se construyó el cine y se generó una identidad sociocultural, misma que ha sobrevivido al petróleo, cuya desmedida explotación hizo que en pocas décadas se agotara.

Actualmente, el municipio tiene una población de 25 011 habitantes (el 90 % se ubica en la cabecera) (Sistema de Información Estadística y Geográfica del Estado de Veracruz de Ignacio de la Llave [SIEGVER], 2021), de los cuales –de acuerdo con datos oficiales– sólo el 1.4 % de las personas mayores de 3 años se declara hablante de lengua indígena (SIEGVER, 2021).

El hecho de situarse en una zona culturalmente tan rica ha hecho que la comida huasteca sea la principal influencia culinaria en Cerro Azul (representada por *bocoles*,<sup>12</sup> tamales y enchiladas, entre otros); su cercanía con la costa, ríos y, en particular, con la laguna de Tamiahua, le aportan uno de sus ingredientes principales, los alimentos acuáticos (pescados, jaibas, camarones, ostiones, acamayás), y la ganadería, carne y lácteos.

Las diversas culturas de quienes fundaron el pueblo dejaron huellas en la configuración de la gastronomía local: los *tacos rojos* potosinos, el pozole del Bajío, la capirotada, y las torrijas de Semana Santa; los *bisquets* y el *pay de piña* creado por los chinos. En este escrito se presenta una muestra pequeñísima de toda esa riqueza.

<sup>11</sup> Las familias de nuestros entrevistados provenían de Huejutla, Hidalgo (en el corazón de la Huasteca) y Michoacán (C3); de Jalisco y Toluca, Edo. Mex. (C2); y de dos familias, una blanca y otra mulata (C1), arraigadas en la congregación de Juan Felipe (actual municipio de Cerro Azul), formada por colonos españoles. Los antepasados de la autora mexicana de este trabajo llegaron de los estados de San Luis Potosí, Tamaulipas y Colima.

<sup>12</sup> Del tének /bo. 'k'o:l/, tortilla gruesa de masa de maíz.

## MÉTODO DE TRABAJO Y CORPUS

La investigación que se reporta en este artículo es de carácter cualitativa y descriptiva. Nuestro enfoque ha sido histórico, comparativo y etnográfico,<sup>13</sup> con el empleo de las nuevas tecnologías de la información (TICs). Se cumplieron las siguientes etapas de trabajo: (1) aplicación de entrevistas semidirigidas a profundidad;<sup>14</sup> (2) transcripción de las entrevistas; (3) sistematización y análisis del material léxico obtenido (para lo cual se conformó una base de datos común, cuyo criterio de ordenamiento fue su frecuencia de aparición en las entrevistas); (4) identificación de conceptos nucleares y localización, dentro del corpus, de las unidades léxicas pertenecientes a su dominio semántico; dicho en otras palabras: conformación de un *corpus modular*;<sup>15</sup> (5) selección de los campos referenciales comunes a ambas zonas y sus elementos; (6) identificación de los significados culturales.

Retomamos de Niktelol Palacios el recurso de recuperar el léxico a partir “del discurso, en conversaciones semidirigidas en donde el vocabulario aparece en su contexto de uso” (Palacios, 2020: 81-82). Así, el corpus de la investigación se conformó por las transcripciones de las entrevistas: tres por población. En cuanto al sexo, cabe destacar que participaron tres mujeres y tres hombres. Se consideró como criterio de selección de los participantes que hubieran vivido gran parte de su vida en la zona respectiva y que fueran mayores de 50 años (lo que nos ha aportado información sobre cambios culturales y lingüísticos).

Los participantes de Loja fueron una mujer de 68 años (L1), profesora de Arte y Música en la Universidad Técnica Particular de Loja y con gran interés en la gastronomía local; un hombre de 52 años (L2), docente de gastronomía

<sup>13</sup> El trabajo realizado encuentra coincidencias con los métodos de la antropología social o cultural. Nuestra labor como investigadoras (siempre lo es cuando se trata de describir léxico) ha sido la de mediadoras culturales.

<sup>14</sup> Los participantes dieron sus testimonios a distancia por la situación de pandemia (las entrevistas se llevaron a cabo en marzo de 2021).

<sup>15</sup> “Un corpus modular es un fichero organizado de las palabras que se relacionan con la palabra nuclear. La selección de tales palabras se basa en criterios semánticos en un sentido amplio; más en un conocimiento del mundo de índole socio-histórica que en criterios estrictamente lexicográficos.” (Díaz de Atauri, 2008: 45). Un procedimiento semejante fue aplicado por Pérez Aguilar (2002) en su investigación sobre el “Léxico de los chicleros en Chetumal”.

en un instituto tecnológico de la ciudad y, por último, un hombre de 50 años (L3) con estudios de Administración y seguidor de la comida tradicional. En Cerro Azul, participaron una mujer de 91 años (C1) dedicada al hogar, quien se desempeñó como maestra de cocina durante ocho años en un Centro de Integración Familiar de Petróleos Mexicanos (siendo el único trabajo remunerado de su vida, a una edad mayor de 70), así como dos trabajadores jubilados de PEMEX (Petróleos Mexicanos): una mujer de 75 años (C2) y un hombre de 67 años (C3), ambos aficionados a la cocina regional.

Los ejes de las entrevistas fueron los platillos tradicionales de sal, los postres y las bebidas. Se consultó sus nombres, ingredientes y formas de preparación, así como quién los prepara típicamente; cuándo y para qué ocasiones se consumen (época del año, evento, hora del día).

Tanto la selección de las unidades léxicas como el modelo elegido para su análisis retoman la noción de *signo etnolingüístico* desarrollada por Quesada (2011). En cuanto al método expositivo, los datos léxicos se agrupan en campos referenciales que se presentan en un orden histórico, conforme los alimentos se fueron incorporando a la dieta y a la cultura de los pueblos americanos: dos conceptos corresponden a la época prehispánica (el maíz y el ají o chile) y dos al periodo colonial (el ganado porcino y vacuno, y el plátano).

## INTRODUCCIÓN AL SIGNIFICADO CULTURAL Y AL SIGNO ETNOLINGÜÍSTICO

Los significados culturales abarcan las funciones sociales y los valores simbólicos y afectivos que los miembros de una comunidad de habla indizan a los objetos y a los vocablos con que los nombran, algo que va más allá de la correspondencia unívoca entre referente y designación, la cual es sólo un modo de significar (propio de los lenguajes de especialidad), por más que se encuentre en el imaginario colectivo de muchos hablantes como el significado prototípico o único de las voces, al que se aspira acceder en una obra lexicográfica.<sup>16</sup> En consecuencia, debido a su característica central de nombrar objetos, acciones y relaciones [...] las palabras se convierten fácilmente en *símbolos sociales*; es

<sup>16</sup> Véase la explicación que Luis Fernando Lara elabora al respecto (2016: 65 y ss.).

decir, *trascienden su naturaleza de signos lingüísticos*” (Lara, 2006: 213), en el sentido saussureano.

Así pues, la idea anterior subyace al concepto de *signo etnolingüístico*, desarrollado por Quesada (2011), para quien se trata de un tipo de signo perteneciente a la clase de los *indicios*, entendidos como “una dualidad compuesta por un cuerpo, sea verbal (un sonido, una palabra, un sintagma, una locución, etc.) o no verbal (el fuego, las cigarras cantando, etc.), más un significado [cultural], o sea, su valor, o valores, dentro de la cultura en cuestión” (p. 128).

El correlato de esto en lexicografía es la posibilidad de dar cuenta de las palabras en tanto signos lingüísticos mediante definiciones de lengua (en cualesquiera de sus tipos), y la posibilidad de elaborar un acercamiento en cuanto a símbolos sociales a través de definiciones culturales, las cuales buscan contener la memoria social de la comunidad en cuestión y ser reflejo de la cultura que pretenden explicar (Rey, 2014).

Como elementos para la exposición de los significados culturales consideramos la información que aportaron los entrevistados en relación con la denominación (considerando el peso cultural de las etimologías populares); las particularidades que caracterizan la preparación y consumo de cada alimento; los hábitos sociales que vinculan tal objeto cultural con la forma de vida de la comunidad, así como las tradiciones asociadas y el significado simbólico colectivo que recibe en cada caso.

## MAÍZ

Existen dos hipótesis sobre el origen del cultivo del maíz: “Una sugiere que la domesticación pudo haber ocurrido en una región específica; otra propone que este proceso pudo haber ocurrido en diferentes momentos y lugares” (Kato *et al.*, 2009: 15). Dentro de la hipótesis multicéntrica, se manejan como las dos posibles zonas de procedencia Mesoamérica y los Andes. Independientemente de cuándo y cuál hubiera sido la fuente de irradiación del maíz, está ampliamente documentada la importancia alimentaria y cultural que el maíz tuvo y tiene para las regiones de América que nos ocupan en esta investigación.

En los Andes, desde la época prehispánica, el maíz ha mantenido un rol importante en la vida indígena tanto por su valor nutricional como simbólico. Cuando los peninsulares llegaron al actual Ecuador, el maíz se cultivaba en

todo el territorio, dada su amplia adaptación a los pisos ecológicos; para el siglo xx, la región andina ocupaba 60 700 hectáreas sembradas con esta gramínea, cuya totalidad se dedicaba al consumo humano (Echeverría y Muñoz, 1988). Es así que el maíz, como alimento, ha sido utilizado hasta la actualidad en sus diferentes formas. En particular, la harina de maíz que se obtiene moliendo el grano se la utiliza en la elaboración de tortillas, pan, tamales, arepas, *sango*<sup>17</sup> y *coladas*<sup>18</sup> (Bravo, 2005).

Por su parte, “a principios del siglo xvi el repertorio agrícola de la Huasteca se componía de maíz, base de la economía, de la sociedad, la religión y la cultura” (Ruvalcaba, 1998: 45). Sembraban, además, chile, frijol, camote, calabaza, yuca; frutales como la papaya, chicozapote, anona y piña. No obstante, este autor señala que la domesticación del maíz en la región debió ser muy posterior a la de la calabaza, el frijol y el chile (1998: 43). Actualmente, el maíz sigue siendo la base de la alimentación de las comunidades indígenas.

En cuanto a su significado simbólico en la región, los indígenas huastecos han elaborado una compleja concepción del maíz (Argüelles, 2010) y atribuyen su origen a un espíritu o deidad antropomorfa, que en su advocación de niño es conocido por los tének como Dhipák y por los nahuas como Chicomexóchitl (Sandstrom, 1998; Ochoa, 2000; Hernández Ferrer, 2004).

### Loja: *tamal lojano*

De los vocablos en estudio, *tamal* ocupó la tercera posición en número de ocurrencias en el corpus ecuatoriano, después de *guineo* y *repe*, lo cual es un indicador de su importancia en la gastronomía tradicional de la ciudad. Con respecto a las particularidades que lo diferencian del elaborado en otras zonas, resaltan el relleno y la hoja en la que se envuelve. Primero, los entrevistados coinciden en que el relleno tradicional es el que se prepara con la carne de la

<sup>17</sup> Sango: Harina de maíz seco, tostado y molido que se mezcla con agua, se fríe en un sartén y se añade queso y sal para obtener una crema.

<sup>18</sup> Colada: “Alimento que se prepara disolviendo harina de maíz en agua o leche caliente con azúcar o sal y cociendo la mezcla resultante hasta que adquiera una consistencia espesa” (*Diccionario de Americanismos*, en línea).

cabeza de cerdo. En este sentido, uno de ellos agregó la razón por la cual se utiliza esta parte del cerdo:

El [relleno] más típico, el original, llamémoslo así, es el que se hace con la cabeza de cerdo. Entonces, ya luego de que las familias hacían sus *pelas de chancho*,<sup>19</sup> se comían su cuerito, la *fritada*<sup>20</sup> y todo, pues, quedaba la cabeza y alguien dijo ¿qué vamos a hacer con la cabeza? Y, entonces, de ahí es donde se saca la carne para el tamal. Entonces, ¿por qué la cabeza? Primero, era un género que sobraba y teníamos que darle un uso y, segundo, pues porque tiene grandes cantidades de grasa. Entonces, la parte donde está la trompa del animal podemos encontrar secciones de carne que están adheridas con grasa. (L2)

Además, un toque particular para la masa se obtiene al:

Cocinar aparte la cabeza del cerdo, pero sin ponerle la trompa sino solo las partes que se las ve gordas con la piel y con ese jugo se hace la masa, con ese caldo. (L1)

Actualmente, también son comunes los tamales con relleno de pollo y cabe destacar que, además de la carne, ya sea de cerdo o pollo, el relleno incluye un refrito<sup>21</sup> con *arveja* y zanahoria picada.<sup>22</sup>

La siguiente característica de los tamales lojanos, apuntaron los hablantes, es el envoltorio. En otras zonas del país se utiliza, generalmente, hoja de plátano o de bijao, pero a los tamales lojanos es tradicional envolverlos en hojas de *achira*.<sup>23</sup>

<sup>19</sup> Pela de chancho: Reunión familiar festiva en la que una vez matado el cerdo se lo pela al aire libre para sacar poco a poco la carne y preparar sus diferentes productos (*cascaritas*, *fritada*, chicharrón, chanfaina, morcilla) para consumirlos durante todo el día. Se acompaña estos productos con yuca, mote, camote y cerveza. Se realiza en Carnaval, fiestas del pueblo o como celebración por cumpleaños o graduaciones.

<sup>20</sup> Platillo tradicional compuesto por costillas y otras partes del cerdo cortadas en pequeños trozos que se fríen en su propia manteca.

<sup>21</sup> Aceite frito con ajo, perejil, cebolla y condimentos que se utiliza como base de salsas, sopas, etc.

<sup>22</sup> En Cerro Azul se elabora un relleno semejante: “Los de pollo los hacemos con el pollo deshebrado. Se le pone zanahoria, *chicharo*, pasitas, chiles en vinagre picaditos y suficiente cebolla, y tomate picadito. A estos les dicen *tamales en blanco estilo San Luis*, pero se acostumbra mucho por aquí también” (C1).

<sup>23</sup> Planta originaria de América, *Canna indica*; en Cerro Azul se conoce como *papatla*.



Estas hojas le agregan sabor, aroma y es un plus extra. (L3)

Finalmente, la preparación concluye cuando una vez armado el tamal se lo cocina en las:

Famosas ollas tamaleras que son a vapor. (L3)

En este caso, se rescata el léxico derivado de *tamal* > *tamalera*. Éstas son ollas grandes de aluminio que incluyen una rejilla que la divide en dos, en las que generalmente se cocinan aproximadamente de 30 a 40 tamales al vapor.

En cuanto al consumo de los tamales, anteriormente se preparaban únicamente en las casas, especialmente en las fiestas de Navidad y fin de año, pero en la actualidad es frecuente consumirlos en cualquier época del año, generalmente en el desayuno o en la tarde, y siempre acompañados con *aji de pepa* y un café (figura 4).

FIG. 4. TAMAL LOJANO<sup>24</sup>



<sup>24</sup> Todas las fotografías correspondientes a Loja son de la autoría de Isis N. Zambrano Ojeda.

## Cerro Azul: *tamales*

El campo referencial del maíz fue –por mucho– el que ocupó el primer lugar en índice de frecuencia en el corpus. Consideramos dentro de él las palabras *masa*, *maíz*, y *elote*;<sup>25</sup> los platillos y productos que tienen como ingrediente insustituible el maíz (*zacahuil*,<sup>26</sup> *tacos*, *bocoles*, *molotes*,<sup>27</sup> *estrujadas*,<sup>28</sup> *tortillas*, *atole*, *pemoles*,<sup>29</sup> entre otros), así como la técnica del *nixtamal* y los utensilios *metate*, *molino*, *comal* y *batea*. Dentro de este ámbito, fundamental para la cocina y la vida, el vocablo con más ocurrencias fue *tamal*. Este platillo tiene en la Huasteca una función ceremonial vinculada con el homenaje a los difuntos; se ofrendan tamales como principal deleite en Todos Santos (o Día de Muertos, en el centro del país). En Cerro Azul –al igual que en toda la región– dicha celebración es la más importante del año.

El Día de Muertos es hermoso en nuestro pueblo porque para empezar hacen su altar; el que no llegue a hacer altar tiene de todos modos sus veladoras, las fotos de sus *muertitos* y su comida predilecta, que al que le gustó las *carnitas*,<sup>30</sup> el pan, la fruta; nada falta allá, y hacen tamales de todo, de los que tenga, hasta la gente más pobre hace sus tamales de sal.<sup>31</sup> (C2)

Me platica una comadre de un *ranchito*, que ahorran todo el año y van dejando lo mejorcito para festejar el Día de Muertos. Por ejemplo, las calabazas, el *frijol de chivo*<sup>32</sup> que le llaman aquí, para hacer los tamales (figura 5). Ese día

<sup>25</sup> Mazorca de maíz tierno.

<sup>26</sup> Platillo a base de maíz; su aspecto exterior se asemeja a un tamal de gran tamaño, que es como suele describirse en diferentes fuentes.

<sup>27</sup> Molote: Bulto de masa de maíz con algún relleno y frito en manteca.

<sup>28</sup> Estrujada: Tortilla gruesa de maíz pellizcada con los dedos y frita en manteca; se sirve bañada en salsa de chile y cubierta de otros ingredientes.

<sup>29</sup> Especie de galletas dulces elaboradas a base de harina de maíz y manteca de cerdo.

<sup>30</sup> “Carne de cerdo [en trozos] frita en su propia grasa: ‘Nos comimos unos ricos tacos de *carnitas* con guacamole’” (*Diccionario del Español de México*, en línea).

<sup>31</sup> Probablemente la colaboradora se refiera a los *piques* o *piques blancos* (para distinguirlos de los de frijol) (véase nota 36).

<sup>32</sup> Del tének CH'IB (Meléndez y Morett, 2024). Variedad de frijol de bejuco, cuya semilla es pequeña y redondeada; su color entre morado y negro.

las personas que pueden ¿verdad? engordan sus puercos, los matan esos días para hacer los tamales. (C1)

FIG. 5. TAMAL DE FRIJOL DE CHIVO Y CAFÉ<sup>33</sup>



Por otra parte, quienes fueron consultados para esta investigación coincidieron en señalar que, si bien se pueden elaborar en cualquier época del año y sin un evento especial de por medio, otra tradición es que los tamales se consuman en velorios y novenarios.<sup>34</sup> Y es que, para los habitantes de la Huasteca, los tamales están en el centro de los rituales de vida y muerte, y Cerro Azul no es la excepción. La importancia central de este alimento dentro del campo referencial del maíz puede advertirse en el hecho de que, de acuerdo con algunas versiones del mito de Chicomexochitl (Gómez, 2014), el maíz habría surgido de los restos de veinte tamales.<sup>35</sup>

<sup>33</sup> Todas las fotografías correspondientes a Cerro Azul son de la autoría de Sonia E. Morett Álvarez.

<sup>34</sup> En esto se observan coincidencias con el culto a Chicomexóchtli. “A cambio de los dones del alimento y de la música que los dioses han dispensado a los hombres, estos adquieren la obligación de venerarlos y propiciarles la comida y bebida necesaria” (Camacho, 2010: 72). De acuerdo con este autor, el ritual de alimentar a los dioses se extiende a otras prácticas, mismas que ha documentado en diferentes poblaciones de la Huasteca: “cuando una persona muere, durante el novenario se realizan diversas actividades entre las cuales se encuentra el alimentar a la sombra del fallecido”, lo cual puede suceder –nos dice– a través de un sustituto.

<sup>35</sup> Cuya materia prima habría sido amaranto, calabaza y pipián (véase nota 37).

En Cerro Azul se hacen tamales de muy diversos rellenos, incluyendo palmito o fauna nativa como el jabalí. (C3)

Lo tradicional son los de calabaza y de frijol de chivo, de puerco y de pollo. (C1)

También son tradicionales los *piques*,<sup>36</sup> que no llevan guisado. Para Día de Muertos se acostumbra de calabaza y *pipián*<sup>37</sup> (figura 6), con la cosecha que inicia en septiembre; hay que recordar –siguiendo a Ruvalcava (1998)– que la familia de las cucurbitáceas es el alimento con mayor antigüedad del que se tiene registro en la Huasteca.<sup>38</sup>

FIG. 6. PIPIÓN Y CALABAZA DE CASTILLA



A decir de los entrevistados, dependiendo del bolsillo, la cucurbitácea elegida se acompañará con frijoles o camarón.

<sup>36</sup> Pique: Del náhuatl *EPIKI*, aféresis de *tlaxpiktili* o *tlapiktili* ‘envuelto’ (Argüelles, 2010; Güemes, 2013). Tamal sin relleno. Se hace para aprovechar la masa que sobró después de haber envuelto los tamales; para su consumo, se rebanan y fríen. Una variante son los piques de frijol, que se preparan revolviendo frijoles con la masa.

<sup>37</sup> *Cucurbita argyrosperma*.

<sup>38</sup> De acuerdo con vestigios hallados en Tamaulipas, en la parte correspondiente a la Huasteca (Ruvalcava, 1998: 41).

Ahorita en Semana Santa y el Día de Muertos se hacen unos que tampoco se hacen en ninguna parte de la república y del mundo, que son de calabaza con camarón. (C2)

Éstos llevan chile guajillo, ajo, cilantro picadito, mucho cilantro y cebollita también, y se muele con el chile y ajo, la cebolla si la quieres picada o molida. Y la calabaza [...] se pone a hervir, ya se saca de ahí y se saca todo el chile con el cilantro picado, la cebolla picada, se le pone ajo al chile y el camarón [seco] se acitrona con la cebolla y el cilantro. Y ya cuando el camarón soltó el sabor se agrega la calabaza ya cocida y se deja que hierva bien para que suelte el sabor el camarón y el cilantro. (C1)

Una vez que se tienen listos los rellenos (generalmente en la víspera), se prepara la masa, que es *nixtamalizada* de maíz blanco; se amasa con manteca (tradicionalmente de res y de cerdo) y el caldo en donde se cocieron las carnes: de puerco, de pollo. Para envolverlos se emplean hojas *soasadas*, es decir, pasadas por fuego de leña, las cuales pueden ser de plátano o:

de otra hoja que venden, y que traen de por allá de Tepetzintla, que se llama *papatla*<sup>39</sup>, muy parecida a la hoja de plátano, nomás que es una hoja muy largueta, (C1)

más chica que la de plátano, pero con olor muy peculiar; también se usa para el pescado empapelado o envolver queso. (C3)

Los tamales se hierven al vapor; en la medida en que se van envolviendo se van acomodando en la *vaporera*<sup>40</sup> o lata de manteca;<sup>41</sup> lo importante es que al fondo de la olla haya siempre agua:

a mí me enseñó mi abuelita a ponerle una monedita de cobre en el agua para (...) mientras estaba sonando es que tenía todavía agua, entonces, dejaba de

<sup>39</sup> En Sudamérica recibe el nombre de *achira*.

<sup>40</sup> En Loja: *tamalera*.

<sup>41</sup> Antes no había vaporeras sofisticadas. (C2)

sonar y había que echarle un poquito más de agua, y hasta la fecha lo sigo haciendo. (C2)

Una vez que está llena la olla:

se ponen sus hojitas, después poníamos una toallita, un trapito, la tapadera y el *molcajete* arriba para que no se saliera el vapor. (C2)

Si durante el proceso de cocción una mujer embarazada llega a la casa, ésta debe abrir la olla y establecer contacto con los tamales (que puede ser introduciendo una cuchara entre ellos) porque de lo contrario, quedarán *pintos*.<sup>42</sup> Si los tamales llegaran a salir pintos, la manera de remediarlo es poniendo sobre la última cama de hojas unos cuchillos en cruz y llevando así la vaporera nuevamente al fuego.

Hay otra especie de tamales, los *chámitle*s.<sup>43</sup> (C1)

Estos tamales son de elote *amartajado* o molido y azúcar o piloncillo. Actualmente, hay quienes les agregan mantequilla, pasitas, y leche condensada a la hora de servir. Se acostumbra cotidianamente para merendar o cenar, por lo que muchas veces se compran hechos. Este tipo de tamal se envuelve en hoja de maíz.<sup>44</sup>

## AJÍ O CHILE

*Ají* o *chile* es el nombre que reciben en nuestros países diferentes especies del género botánico *Capsicum*. En la América precolombina su consumo estaba muy extendido; México se reconoce como el territorio donde se originó su domesticación. Entonces, se empleaba como condimento, por su sabor picante

<sup>42</sup> Estado de los tamales en el que algunas de sus partes están bien cocidas y otras crudas.

<sup>43</sup> Chámitle: Del náhuatl XAMITL 'adobe'. Vocabulario de Molina en *Gran Diccionario Náhuatl* (<https://gdn.iib.unam.mx/>).

<sup>44</sup> Sandstrom (1998) relata que, para los nahuas de la Huasteca veracruzana, el *xamitl* es la representación de un niño.

y astringente –y su aporte de vitaminas era significativo para la nutrición– pero también como medicamento, castigo, moneda y materia de tributo (Montes, 2010).

El vocablo *ají*, de origen taíno, fue una de las primeras palabras indígenas que los conquistadores incorporaron a su habla. El 15 de enero de 1493, Cristóbal Colón (en la edición de Anzoátegui, 2002: 132) escribía: “También hay mucho ají, que es su pimienta”. Al igual que otros vocablos que aprendieron en Las Antillas, en la medida en que se extendieron por territorio continental, fueron imponiéndolos sobre los locales. Así, en El Caribe y toda Sudamérica se emplea la palabra *ají*, mientras que en México la palabra que se impuso fue la derivada del náhuatl, *chilli*. Centroamérica alterna entre una y otra.

El chile o ají –habiendo surgido en América– es hoy un cultivo mundial y continúa siendo un elemento importante en la dieta de la población de nuestros respectivos países; es también aquí donde se localiza la mayor diversidad de sus especies.

## Loja: *ají de pepa*

En la cocina ecuatoriana siempre está presente el ají como una salsa que acompaña diferentes platillos, puesto que no interviene directamente en la composición de platos como ocurre en México (Pazos, 2017). A lo largo de todo el territorio, existe una gran variedad de salsas de ají. De acuerdo con cada zona, se combina el ají<sup>45</sup> con productos propios de cada región como los *chochos*,<sup>46</sup> el *tomate de árbol*,<sup>47</sup> el maní, la pepa de *sambo*<sup>48</sup> (*Cucurbita ficifolia*),

<sup>45</sup> Los tipos de ají varían desde las costas del Pacífico, hasta los Andes y la Amazonia. Todos son nombrados popularmente de acuerdo con su apariencia o por la reacción que provocan al comerlos, algunos de sus nombres son *ratón*, *rocoto*, *plátano amarillo*, *cerezo* (Lema, 2018).

<sup>46</sup> Chocho: “Planta anual de hasta 1 m de altura, con hojas digitadas, de forma oval a lanceolada, flores azules de cinco pétalos y frutos en vaina con semillas comestibles; antes de consumirse, estas semillas se ponen a remojar para extraerles el amargor y la acidez”. (*Diccionario de Americanismos*, en línea).

<sup>47</sup> “Fruto del tomate de árbol, de color rojizo, piel lisa, turgente y brillante, pulpa anaranjada y gelatinosa con numerosas semillas de color granate intenso; su sabor es agridulce y puede comerse crudo, en jugos o cocinado.” (*Diccionario de Americanismos*, en línea).

<sup>48</sup> Del kichwa SAMBU ‘calabaza’ (Stark y Muysken, 1977); “variedad de calabaza” (*Diccionario de Americanismos*, en línea).



entre muchos otros; esta mezcla resulta en una salsa que usualmente se la conoce como *ají*.

Particularmente, en la ciudad de Loja destaca una salsa de ají cuyo ingrediente principal es la pepa de sambo, de ahí que se la conozca tradicionalmente como *ají de pepa* o *ají de pepa de sambo* (figura 7). El compuesto sintagmático *ají de pepa* ocupó el cuarto lugar en los campos referenciales del corpus y, según apuntan los entrevistados, es característico usar este ají como acompañante de otros platillos típicos de la ciudad, como el cuy, la gallina cuyada, el repe o los tamales.

FIG. 7. PEPA DE SAMBO Y AJÍ DE PEPA



Cabe señalar que, para esta salsa, el ají, como ingrediente, no recibe tal importancia como la pepa de sambo, debido a que se puede incluir cualquier tipo de ají:

del rojo, del verde, del que se crea conveniente; (L1)

se le pone ají a tu elección el que pique más, el que pique menos. (L3)

De igual manera, la cantidad de ají puede variar e incluso se puede prescindir de él:



ahí sí va al gusto de las personas, puede hacer una parte sin ají y otra parte con ají. (L1)

Los entrevistados coinciden, además, en que la pepa de sambo tiene un alto costo económico. A pesar de que algunos restaurantes ofrecen una especie de imitación de esta salsa debido a que la pepa de sambo:

no es económica para el negocio. (L3)

Pero es indispensable conseguir las semillas para preparar el ají tradicional.

### **Cerro Azul: *enchiladas***

Después del campo referencial del maíz, el del chile fue el más extenso en el corpus correspondiente a México (es decir, ocupó la segunda posición) y –sin duda– fue el que obtuvo, en Cerro Azul, la mayor diversidad de expresiones complejas. Registramos 15 denominaciones para variedades específicas de chiles (algunas de ellas sinónimas) empleadas en la elaboración de salsas y como ingredientes de numerosos guisos y platillos; un producto usado como acompañamiento (los *chiles curtidos*), además de adjetivos que hicieron referencia a este vegetal por su color, tamaño, forma o la intensidad de su picante. Como platillos que tienen al chile como elemento central, registramos dos: los *chiles rellenos* y las *enchiladas*, una de las exquisiteces de la gastronomía de la región huasteca.

Las enchiladas tienen una versión sin chile, que son las *entomatadas*; respecto a las picantes, éstas pueden ser de chile verde, de *chile seco* (o *chiltepín*) (figura 8), o bien, a base de *pipián*, mismas que son motivo de orgullo para las personas del lugar.

A diferencia de lo que sucede en otras regiones, en la Huasteca –como hemos adelantado– se conoce como *pipián* tanto al fruto como a la semilla de una variedad de cucurbitácea,<sup>49</sup> aunque llega a aplicarse indistintamente

<sup>49</sup> Véase nota 37.

la denominación para hacer referencia a las semillas de diferentes variedades de cucurbitáceas:

Y la calabaza [...] se pone a hervir, así a vapor, ya después se saca de la cáscara la pulpa y del centro las semillas se le sacan, mucha gente las pone a secar y de ahí sale el *pipián*. (C1)

Y para hacer las *empipianadas*:

La semilla de pipián se tuesta, ya ves que brinca, y le pones una tapa que no se te queme, no más que veas que brinca y brinca le pones fuego muy bajito, ya cuando está tostada la pones en la licuadora y la mueles, te queda polvito, ya después le agregas agua, o si tú quieres agregarle chile, pues también [...] Y para que no se te corte, le agregas una poquita de leche. (C1)

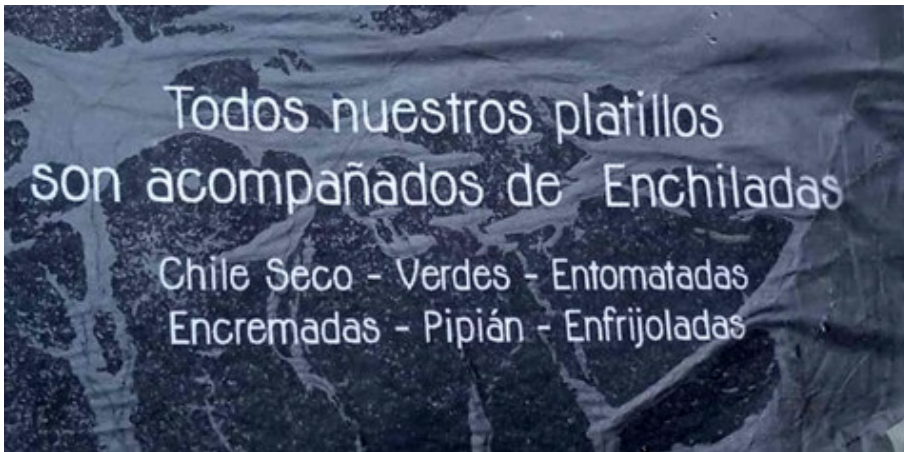
FIG. 8. ENCHILADAS DE CHILE SECO



Una vez hecha la salsa, se bañan en ella tortillas de maíz hechas a mano, e inmediatamente después se colocan en un plato; opcionalmente, pueden cubrirse con queso espolvoreado. En Cerro Azul, el acompañamiento característico de las *empipianadas*, y de cualquier tipo de enchilada, es un filete de carne, pollo o pescado, unos camarones, o bien, huevos en diferentes presentaciones, siempre con unos *frijolitos* (figura 9).

Es de llamar la atención que la cucurbitácea y el chile, los insumos característicos de este platillo, corresponden a cultivos cuya antigüedad es previa al del maíz en la región.

FIG. 9. LEYENDA EN EL MENÚ DE UN RESTAURANTE



## CERDO Y GANADO VACUNO

Tanto las investigaciones arqueológicas como los documentos de archivos manifiestan la existencia de camélidos mucho antes de la llegada de los europeos. La crianza de ganado autóctono en la región andina disminuyó considerablemente con la llegada de los españoles, al ser desplazado por la producción de ovejas, cabras, vacas y cerdos (Casas *et al.*, 2016). Por otra parte, como antecedente de la ganadería en la Huasteca, se ha destacado que “el jabalí fue tenido en los hogares, posiblemente castrado, y engordado cual cerdo” (Melgarejo, 1980: 170).

El ganado fue una necesidad para los conquistadores desde el primer momento, por sus usos para los desplazamientos, la guerra, la alimentación, las

labores del campo y de carga. Por eso Colón solicitó a los Reyes Católicos que lo proveyeran de “carneros vivos y aun antes corderos y cordericas, más fembras que machos, y algunos becerros y becerras pequeños son menester, que cada vez vengan en cualquier carabela que acá se enviare, y algunas asnas y asnos y yeguas para trabajo y simiente, que acá ninguna de estas animalias hay de que hombre se pueda ayudar ni valer” (Colón, 2002: 159).

En cuanto a México, en los primeros tiempos de la colonia, todo lo que entraba a la Nueva España era a través del puerto de Veracruz. Así sucedió con el ganado, que pronto se reprodujo en la zona. En el norte del estado hay registro de la existencia de tierras destinadas a la ganadería desde fines del siglo XVI (Melgarejo, 1980); no obstante, su práctica no se intensificó sino hasta entrado el siglo XIX, debido a la demanda interna. Las grandes planicies con profusa vegetación, la baja densidad de población y la falta de vías de comunicación para transportar productos perecederos hicieron de la ganadería vacuna la actividad económica más rentable de la región (Ariel, 1994). Actualmente, Veracruz es el estado con mayor producción de ganado bovino del país (Encuesta Nacional Agropecuaria, 2014) y la Huasteca aporta el 40 % de ella. Por su parte, el ganado porcino se cría mayormente para el consumo local.

Con respecto al cerdo, su predominio en la alimentación durante la colonia se debió a que fue la especie que primero y más se desarrolló, los ejemplares eran abundantes y, por tanto, muy económicos; el consumo de su carne se había extendido entre la población española asentada en América (Díaz, 2019). En la segunda mitad del siglo XVI, su cría fue generalizándose con rapidez a medida que la producción y consumo era frecuente entre la mayor parte de la población originaria, pues se obtenían múltiples productos elaborados a partir del cerdo, que se sometían a distintos procesos para obtener tocino, manteca, carne ahumada, entre otros, lo que hizo del cerdo un alimento de ingestión diaria (del Río Moreno, 1996).

### **Loja: *cecina***

Otro de los platillos tradicionales de la ciudad de Loja es la famosa *cecina* (L3) que consiste en la carne de cerdo asada que ha sido previamente secada al sol siguiendo, así, la tradición de más de cien años, ya que antiguamente la forma de conservar la carne era sazónarla y secarla al sol (Gallardo, 2012). A este respecto uno de los entrevistados señaló:

en sus inicios recordemos que no había métodos de conservación como el uso de refrigeradores. Entonces, lo que hacían era cubrir con sal, con condimentos para conservar la vida útil de este tipo y ahí es donde empieza a nacer esta adicción, llamémosle así, por secarle a la carne, por llevarla a orear, comúnmente le dicen orear, que básicamente es condimentarla y la dejan secarse [...] Entonces, eso es súper típico de la localidad. (L2)

Cabe suponer que los pueblos prehispánicos podrían haber consumido un antecedente de este platillo (a base de camélidos), tradición que se ha conservado principalmente en Loja, lo que hace a este plato uno de los más emblemáticos de la zona. Como señala Pazos (2017) la cecina es propia de esta provincia.

Una de las características principales para su preparación es que:

la carne de cerdo debe ser cortada lo más fina posible para obtener un faldón grande. (L1)

Los ingredientes para su preparación consisten únicamente en ajo, sal y *naranja agria*, que constituyen el condimento para la carne. Generalmente, se puede sustituir la naranja agria con limón, pero según identificó una de las entrevistadas:

La naranja agria es endémica de los valles de Vilcabamba, Malacatos y Catamayo<sup>50</sup> porque en otras provincias no hay la naranja agria. Por eso le digo que es endémica, es propia de acá. (L1)

Con respecto a la preparación, algunos libros de gastronomía nacional señalan que la cecina puede ser de res y que se la puede ingerir frita (Cuvi, 1994; Pazos, 2017), incluso hay restaurantes que la preparan de esta manera; sin embargo, los entrevistados enfatizan que:

hay que hacerla a la brasa; (L1)

<sup>50</sup> El clima apropiado y las propiedades que poseen los suelos en estos valles propician el cultivo de esta fruta.

la *cecina original*, la auténtica, debe ser asada pues mientras la carne va sudando los jugos caen en la brasa y esos vapores es lo que da ese sabor especial. (L2)

Sus guarniciones principales son la yuca,

que por excelencia debe acompañar a la cecina, (L2)

y el *encurtido*<sup>51</sup>, que hacen de este un platillo clásico (figura 10). Es común comerla en el almuerzo.

FIG. 10. CECINA LOJANA



Por último, es pertinente señalar que actualmente existen varios restaurantes típicos que ofrecen este platillo, por ende, no existe una época especial para su consumo y, también, se puede preparar en casa ya que según menciona un entrevistado:

hay hogares donde tienen unos cordeles y ahí la dejan secar por un tiempo para luego poderla asar. (L2)

<sup>51</sup> Encurtido de cebollas (o curtido): Cebolla roja, tomate y cilantro picados y condimentados con sal y limón.

## Cerro Azul: *carne seca, cecina y carne enchilada*

El campo referencial de la carne fue muy relevante en Cerro Azul: ocupó el tercer lugar tanto de los insumos considerados para esta investigación como de todos los ámbitos del corpus vinculados con la alimentación. Se ubicó por encima de los del frijol, los mariscos o la caña de azúcar. Incluimos en éste las menciones a las carnes rojas, ya sean de res o de cerdo y sus derivados, o bien, de las especies silvestres de jabalí y venado; así como algunos platillos como la *barbacoa* (a diferencia de lo que ocurre en otras regiones, en Tamaulipas y en el norte de Veracruz se hace con res).<sup>52</sup>

La carne más consumida en Cerro Azul es ésta; nos referiremos aquí a su preparación con mayor tradición: la *carne seca*, la cual es parte de la identidad gastronómica de la región, tal como se expresa en los famosos versos con que inicia el huapango *Las tres Huastecas*: “Para hablar de la Huasteca hay que haber nacido allá / saborear la *carne seca* con traguitos de mezcal”. La técnica de salar carnes para conservarlas y transportarlas es antiquísima y pudo haberse desarrollado en diferentes lugares. La primera evidencia que se tiene de su aplicación en México es que en “[e]l año de 1527, Hernán Cortés fue a España [...]. Compró dos navíos recién llegados a la Veracruz, y entre los avituallamientos figuró la cecina, seguramente ya hecha en la costa veracruzana” (Melgarejo, 1980: 44).

El procedimiento que se aplica a la carne para secarla consiste en untarla con el jugo de una variedad de naranja conocida como *de cucho* y exponerla al sol (figura 11).

Para su consumo, recibe diferentes preparaciones:

ya cocida como que la deshebras y le pones el huevo y luego una salsa de tomate nada más con chile y cebolla. Es como la *machaca* del norte. (C1)

Otro uso regional es en *huatape*:<sup>53</sup>

<sup>52</sup> “Veracruz, en teoría, por cálido, no podía ser tierra de barbacoa, pero la persistencia del rasgo cultural indígena se ha impuesto, y si no hay barbacoa de borrego, hacen de vaca y en forma generalizada”. (Melgarejo, 1980: 175).

<sup>53</sup> Huatape (o guatape): Del tének /wa.'t'ap/ `atole'. Caldo autóctono de la Huasteca que combina vegetales y, comúnmente, alguna proteína de origen animal, que puede ser camarón, acamaya, o bien, carne seca o huevo; se espesa con masa de maíz.



Hay muchas maneras de hacer el huatape, puede ser con carne seca, allá en Juan Felipe mis tías lo hacían con carne seca porque ahí no había carnicerías, así nomás mataban cada ocho días allá un animalito, lo hacían con carne seca, con costillita o la pura carne seca, y ya cuando guisaban el caldito, el *recaudo*, ya con la masa le iban moviendo moviendo, moviendo, que hierva bien para que quede la masa cocidita y no haga daño con rama de epazote y ya se le agregan los *jacubes*.<sup>54</sup> (C1)

FIG. 11. CARNE EXPUESTA AL SOL Y A LA VENTA EN CERRO AZUL



En cuanto a la denominación que recibe este producto, observamos que la entrevistada de mayor edad –y a la que consultamos primero– había dicho siempre *carne seca*, mientras que el informante de menor edad empleaba también *cecina*,<sup>55</sup> por lo que preguntamos:

<sup>54</sup> Del tének /ja.'kub/. Variedad de cactus comestible (*Acanthocereus tetragonus* / *Acanthocereus pentagonus*). Se vende y consume mucho en Cerro Azul. Junto con el chayote, la calabacita criolla, los nopales y el palmito es de los vegetales con más usos en guisos o como acompañamiento.

<sup>55</sup> De acuerdo con la etimología de Corominas y Pascual (1954), *cecina* derivaría de un latín vulgar *SICCĪNA* “carne seca”.



E1: ¿Cuándo usted dice *cecina* y *carne seca* se refiere a lo mismo o hay diferencias?  
 I (C3): Son diferentes: el contraste es que la *cecina* es más fresca y sólo está oreada, lo que significa que se pone al sol unas horas, es decir, contiene más humedad y la *carne seca* es una cecina que se deja colgada en lazos de yute en el interior de la cocina por días o semanas (mi padre, que en paz descanse, era fanático de la carne seca de venado). [Eso] permite que se seque mucho; como está impregnada con sal, limón y *naranja agria de cucho*, no sufre descomposición. Tanto la cecina como la carne seca se consume frita y acompañada por enchiladas o entomatadas y frijoles refrito con queso encima. El plus de la carne seca es que, si se *despica*<sup>56</sup> muy finamente, en automático obtenemos la *machaca*, que se puede preparar con huevos.<sup>57</sup>

Por su parte, la carne enchilada (que recibe un tratamiento similar al de la carne seca se condimenta y se deja orear) tiene como materia prima la carne de cerdo (figura 12).

FIG. 12. PUESTO DE CHORIZO Y CARNE ENCHILADA EN LA PLAZA



<sup>56</sup> Despigar: Aplicado a pescado, jaiba y carnes, desmenuzar finamente.

<sup>57</sup> A partir de esta respuesta y otros datos que hemos venido registrando, consideramos que se ha producido un caso de especialización semántica, algo que puede ocurrir cuando dos lexemas sinónimos –procedentes de dos variedades de lengua diferentes– entran en competencia en una misma región; *machaca*, como explicó la entrevistada, es un vocablo característico de los estados del norte.

## PLÁTANOS Y GUINEOS

Aunque hay controversia acerca de la existencia de algunas variedades silvestres de plátano en la época prehispánica<sup>58</sup> en general se acepta que se trata de un grupo de hierbas introducido a este continente. Restos arqueológicos sugieren que la tala de árboles habría favorecido su domesticación en Papúa Nueva Guinea hace aproximadamente 30 000 años<sup>59</sup> (Casas, 2016); después, el plátano se extendió por Asia y de ahí lo habrían trasladado a África Ecuatorial, a Guinea, y –en el siglo xv– a las Islas Canarias, desde donde se introdujo al continente americano hacia 1516 (Secretaría de Economía, 2012). Esta ruta explica el hecho de que diferentes variedades del fruto se conozcan en la región platanera del continente como *guineo*:<sup>60</sup> “Los negros y luego los blancos usaron en América de los nombres africanos para distinguir ciertas cosas de sus similares, según sus procedencias. Así se dice: gallina de Guinea, yerba de Guinea, plátano guineo” (Ortiz, 1978: 370).

La producción bananera requiere de un clima tropical, por lo cual se adaptó muy bien a las tierras templadas y húmedas de nuestro continente situadas entre tales latitudes. Asimismo, el plátano se asimiló paulatinamente a la identidad cultural y gastronómica de los países por donde se extendieron sus plantaciones (México, Centroamérica, el Caribe, Colombia, Ecuador, Venezuela, Brasil); a tal punto se asocia con el continente americano que, dentro de la obra escrita por Antonio de León Pinelo,<sup>61</sup> con el propósito de probar que el Edén estuvo en América del Sur, éste sostiene que el fruto prohibido era, en realidad, el plátano (citado por Ortiz, 1978: 105).

<sup>58</sup> En su *Ensayo político de la Nueva España*, Humboldt sostiene que los plátanos se conocían en América desde antes de la llegada de los europeos; posición que es defendida hoy día por algunos investigadores (Burgos, 2018).

<sup>59</sup> Diferentes fuentes coinciden en fechar el inicio de la agricultura hace unos 10 000 años.

<sup>60</sup> De acuerdo con el *Diccionario de Americanismos* (en línea), la denominación de *guineo* se emplea actualmente desde el sur de México (Tabasco y Chiapas) hasta Ecuador, Perú y Bolivia, pasando por Centroamérica y el Caribe hispanohablante.

<sup>61</sup> *El Paraíso en el Nuevo Mundo. Comentario Apologético, Historia Natural y Peregrina de las Indias Occidentales, Islas y Tierra-Firme del Mar Océano*.

Por otro lado, la explotación de este cultivo con fines de exportación promovió el monocultivo en detrimento de otras especies vegetales y fue causa de pobreza y gran derramamiento de sangre en la región durante el siglo xx (Galeano, 1970), situación que ha permeado en la visión hacia algunos de nuestros países reflejada en la denominación genérica de *repúblicas bananeras*.

En la década de 1950, Ecuador llegó a ser el primer exportador de plátano según la Organización de la Naciones Unidas para la Alimentación y la Agricultura (FAO), posición que ocupaba también en 2010 (Secretaría de Economía, 2012); igualmente, se trataba de su principal fuente de divisas, al menos, hasta 1970 (Galeano, 1970: 64). Actualmente, el plátano constituye el cuarto cultivo básico del mundo y representa el principal alimento para más de 400 millones de personas (Burgos, 2018). Ecuador y México (después de Brasil) encabezan la lista de países latinoamericanos productores de plátano (FAO, 2020).

En términos de consumo, es tan relevante el fruto como la hoja; de hecho, la palabra *plátano* proviene de la raíz griega *πλατος*, que refiere a una característica de ésta: el ser plana. En relación con ella, Fernando Ortiz (1978) señala que los cigarrillos de los indios consistían en “tripas” de tabaco envueltos en una capa de maíz o de plátano (p. 194). Actualmente, es empleada como envoltorio de alimentos frescos para su transporte y en la cocción de otros, o a manera de plato; tales usos los comparte, entre otras, con la *hoja de maíz*, y la de *achira* o *papatla* (*Canna indica*), nativa de América y especialmente significativa en las regiones en estudio.

## Loja: *repe*

Uno de los platillos que se vincula directamente con la gastronomía lojana es el *repe*. Esto lo demuestran también las entrevistas, pues el vocablo *repe* fue el más frecuente, al ocupar el primer lugar en el corpus. Los hablantes se referían a este platillo como:

el primero que sale a relucir en la comida lojana, (L1)

el platillo estrella, (L3)

debido a que es conocido a nivel nacional como el más emblemático de la región.

En cuanto al nombre del platillo, uno de los entrevistados mencionó su origen:

El *repe* hace referencia a cómo le denominaban, a cómo le llamaban las abuelitas al guineo. Le llamaban *repe*.<sup>62</sup> Ellas tenían esa manera de expresarse, decían voy a comprar *repe* y ahí es donde se va evolucionando y se queda con el nombre del *repe*. (L2)

En palabras de otra entrevistada, “al guineo también se lo llama *repe*”. En este sentido, se puede destacar que el ingrediente principal para este platillo es el *guineo*, vocablo que obtuvo el segundo lugar en el corpus. Cabe mencionar que, específicamente, en el dialecto lojano existe una diferencia crucial entre *guineo*, *plátano* y *banano*. El *guineo*, en este caso, el *guineo verde*, que se usa para las sopas, también tiene sus variedades como *guineo común* o *guineo maqueño*, pero el que se usa para este platillo es el denominado *guineo común*, mientras que el *guineo maduro* corresponde a la fruta comestible denominada *plátano*, *banano* o *banana* en otros lugares. Por el contrario, el *plátano*, aquel que no es posible comerlo crudo, es ingrediente principal de muchos otros platillos típicos del país.

Una peculiaridad de esta sopa es que adquiere una tonalidad blanca, debido a que todos sus ingredientes tienen el mismo color: guineo, quesillo y leche; de ahí que también se conozca tradicionalmente a este platillo como *repe blanco* (figura 13). Típicamente se consume en el almuerzo y siempre acompañado con aguacate.

La importancia cultural del *repe* se manifiesta cuando se dice que:

todas las familias lojanas deben saber preparar *repe*, (L2)

lo que nos habla de aquel conocimiento cultural gastronómico que tienen las familias al ser oriundas de la ciudad y que también se expresa como una huella identitaria cuando sus habitantes generalmente dicen:

yo soy más lojano/a que el *repe*. (L1)

<sup>62</sup> De etimología desconocida.

FIG. 13. REPE LOJANO



### Cerro Azul: plátano de Castilla

Veracruz es el tercer estado productor de plátano en México y el segundo con mayor extensión cultivada (Secretaría de Economía, 2012); su fruto es muy importante en la dieta diaria de los cerroazulenses y de la región.<sup>63</sup> En las entrevistas realizadas fueron reiteradas sus menciones como uno de los elementos del desayuno, la merienda o la cena. En relación con los otros campos referenciales considerados para este estudio, su frecuencia fue menor –con una diferencia notable– a la del maíz, el chile y la carne; es decir, ocupó el cuarto lugar.

Tradicionalmente, las variedades de plátano que se consumían crudos en Cerro Azul eran el *roatán* y el *manzano*, y para la cocina, una identificada por

<sup>63</sup> En el *Recetario totonaco de la costa de Veracruz*, su autora observa: “El plátano merece un apartado especial dentro de la alimentación actual en las regiones indígenas; aunque es una fruta introducida recientemente –el siglo pasado [léase el XIX]– su aprecio y gusto entre la gente lo colocan como una importante fuente alimenticia, nutricional y gastronómica” (Aguilera, 2000: 187).

los entrevistados como regional, el *plátano de Castilla*.<sup>64</sup> Estos plátanos se preparan, fritos (en tiras) o cocidos (enteros) y se sirven con queso espolvoreado (figura 14); o bien, se emplean como uno de los ingredientes del caldo de res. Es el postre característico de una *chicharronada*,<sup>65</sup> donde se fríen en manteca de puerco. Este plátano es muy apreciado por su sabor dulce:

es más chico y más delgado que el macho, que es más comercial, pero el Castilla tiene más sabor y dulzura. (C3)

La preferencia por el plátano de Castilla frente al macho (el más frecuente para cocinar en el país y que está desplazando al regional por ser más resistente y menos oneroso, a pesar de que no se produce localmente) adquiere connotaciones afectivas más allá del gusto:

Yo desayuno con unos *plátanos fritos*, porque allá tenemos un plátano parecido –que por cierto está en extinción y que ¡no sabes cómo me duele!–, parecido al plátano macho, pero es más chiquito y más delgado, y se llama *plátano de Castilla*. Ese plátano es pura azúcar. (C2)

<sup>64</sup> En una monografía del municipio de Álamo Temapache (vecino a Cerro Azul) se dice: “Cuando llegaron las compañías extranjeras al despuntar el siglo xx [...] se conjugaron las actividades y los campos se vieron convertidos en verdes vergeles, por un lado, las fincas plataneras de distintas especies, el plátano de castilla, manzano, roatán, sientto en boca, macho, guinea y morado” (Solano, 2006: 22). También se documentan registros de plátano con este nombre en otras zonas de Veracruz y en recetarios y textos gastronómicos correspondientes al estado de Oaxaca. No hemos logrado identificar estudios ni botánicos ni léxicos sobre esta variedad. El origen de su denominación resulta un misterio, pues es sabido que el complemento adnominal “de Castilla”, incorporado al nombre de diferentes elementos de la flora y fauna, fue una práctica común entre los indígenas para distinguir una nueva realidad de otra entidad semejante que le otorgaba nombre por extensión semántica (Parodi, 2007). Es de suponerse que esta variedad de plátano se contrastaría con una preexistente.

<sup>65</sup> Chicharronada: Reunión festiva en torno al sacrificio y la comida de un puerco (antiguamente criado en casa) que dura todo un día. Tradicionalmente, el animal se cocina en una paila al aire libre en fogón de leña en su propia manteca. En la medida en que van saliendo los diferentes productos (cueritos, chicharrón, carnitas, etc.) se van consumiendo con el acompañamiento de cerveza. “Se hacen por los cumpleaños, quinceañeras y hasta en los rezos de los difuntos”. (C3)



FIG. 14. PLÁTANOS DE CASTILLA



Como producto a base de plátano se identificó en el corpus el *machuco*; las dos entrevistadas coinciden en señalar que se trata de un elemento gastronómico que corresponde al pasado de Cerro Azul;<sup>66</sup> se evocó de la siguiente manera:

El *machuco* se hace con plátano verde. Yo recuerdo que de niña lo comía mucho. Allá mis tías, mi abuelita ponían a hervir el plátano verde y luego en el

<sup>66</sup> Algo que, al parecer, no sucede en otras comunidades: Tanto en el *Recetario nahua del norte de Veracruz* (Ramírez, 2001) como en el *Recetario totonaco de la costa de Veracruz* (Aguilera, 2000), así como en el denominado *Culinaria afrodescendiente de Tamiabua ...* (Careaga, 2018) se incluye la receta del machuco.

molino o en el metate lo molían ya cocido, y luego calentaban manteca, que estuviera bien caliente y le echaban el plátano ahí; le ponían salecita. Y luego lo servían con frijolitos y huevos revueltos; en vez de tortillas,<sup>67</sup> te comías el *machuco*. [...] No sé si en los *ranchos*<sup>68</sup> todavía lo harán, pero por aquí no he visto que se acostumbre. (C1)

Por otra parte, y como se mencionó, la hoja de plátano tiene muchos usos:

el que no tenía portaviandas, en una ollita de peltre, envolvía los taquitos en *hoja de plátano* o de papel [...] le ponían su arrozito, su *molito*, su carnita, y hasta sus frijolitos, pero las señoras todas, la mayoría que yo recuerde, hacía *lonche*.<sup>69</sup> (C2)

Otra entrevistada, a propósito de la elaboración del *zacahuil*, nos menciona un uso más y también habla de un cambio cultural en relación con él:

luego extienden la masa así en unas *hojas de plátano*; ahora le ponen un pedazo de plástico. (C1)

## DISCUSIÓN

Entre los elementos culturales más representativos de los pueblos están la lengua y la alimentación; ambos son particularmente importantes en la definición de sus identidades. Esta investigación nos permitió identificar vocablos significativos y sus interrelaciones en torno a cuatro campos referenciales comunes de la gastronomía tradicional de dos regiones de América Latina: Loja, Ecuador, y Cerro Azul, Veracruz (México). Pudimos rescatar conceptos

<sup>67</sup> Sobre este dato podría hacerse la lectura de que, en tiempos de escasez del maíz, el plátano funcionaba como sustituto del cereal, pero también cabría considerar que la familia materna de la entrevistada (a la que se refiere en este relato) era afrodescendiente.

<sup>68</sup> Comunidades rurales.

<sup>69</sup> Lonche (del inglés LUNCH): Alimento que consumen los trabajadores durante su jornada. Antiguamente, para que se ingiriera caliente, alguien lo llevaba al centro de trabajo, generalmente los hijos.



culturalmente complejos que refieren a la importancia social, cotidiana y simbólica de los alimentos para las comunidades estudiadas. Asimismo, obtuvimos comparaciones entre campos referenciales al interior de una misma región o al contrastar las dos. El enfoque histórico nos facilita explicar por qué para Loja el campo referencial de guineos y plátanos fue el más relevante, mientras que en Cerro Azul quedó en cuarta posición.

Una vez realizada esta investigación, se refuerza la aseveración de que la alimentación tiene un componente nutricional y otro marcadamente cultural; así, pudimos constatar cómo las mismas materias primas, una vez que pasan por el tamiz de la idiosincrasia de cada pueblo, se transforman en productos materiales, culturales y ceremoniales muy diversos.

En este sentido, se identificaron varias similitudes gastronómicas entre ambas zonas que, a pesar de estar ubicadas geográficamente en dos países distantes, coinciden en elementos naturales: el maíz, el chile (o *ají*) y la hoja de achira (o *papatla*); en platillos, preparaciones y técnicas ancestrales, como en el caso de los tamales o entre el *ají* de pepa y el pipián (ambas salsas a base de cucurbitáceas y chile), u otras más recientes y producto de una herencia común –pero igualmente arraigadas– como la técnica de secar la carne con naranja agria, como elemento auxiliar a la sal, o bien, en las celebraciones en torno al sacrificio de un cerdo: la *pela de chancho* o *chicharronada*.<sup>70</sup> En estos casos, es posible hablar de paralelismos culturales, al tiempo que se exhiben “datos relevantes para la descripción lexicológica de variedades dialectales y de culturas locales” (Palacios, 2020: 81).

También resulta interesante notar que, aunque ambos pueblos resolvieron, en los casos estudiados, sus necesidades alimentarias creativamente, los recursos y resultados fueron semejantes, mas no así las denominaciones. Las coincidencias de nombres (*tamal* o *maíz*), como el hecho de hablar una misma lengua dominante (sin obviar que ambas poblaciones se ubican en zonas de contacto lingüístico), sólo pueden explicarse desde una perspectiva histórica que nos remite a un pasado colonial común.

En cuanto a las diferencias, es destacable que mientras en Cerro Azul los plátanos manzano y de Castilla son elementos de una categoría, para los

<sup>70</sup> Esta similitud llama particularmente la atención; cabría suponerse un origen común, línea de investigación que podría explorarse.


habitantes lojanos, sus equivalentes guineo y plátano constituyen dos entidades del mundo bien diferenciadas. Asimismo, diferimos en los significados simbólicos; específicamente en torno al tamal.












Finalmente, el inventario de vocablos presentados en este trabajo es una muestra reducida de aquel “enorme conjunto de palabras que se forma como efecto de la necesidad de comunicación de los integrantes de una comunidad lingüística y cuyo valor, en consecuencia, se determina socialmente. En él se reúnen las memorias de experiencias prácticas, abstractas, históricas, políticas, técnicas o científicas compartidas” (Lara, 1993, citado en Palacios, 2014), como se ha podido constatar en las páginas anteriores.

## COMENTARIO FINAL









Consideramos que la exposición que se ha elaborado en este artículo sobre los significados culturales de nuestro interés es susceptible de presentarse en forma de definiciones, en el marco de un posible *diccionario cultural* por comunidad, cuya macroestructura –de acuerdo con el planteamiento de Alain Rey (2014: 200)– debería seguir un ordenamiento de las definiciones no semasiológico. Nuestra propuesta, en este sentido, es congruente con la secuencia expositiva seguida en el presente escrito: por agrupamientos léxicos en torno a ámbitos referenciales. En cuanto a la microestructura, para la arquitectura del artículo lexicográfico anclado en un diccionario cultural, se proponen los siguientes campos: hábitos de consumo, creencias, celebraciones asociadas y usos rituales relacionados con el objeto cultural, además de la información gramatical y enciclopédica pertinente para cada caso, incluyendo imágenes.








## BIBLIOGRAFÍA



- Aguilera, Rocío (2000), *Recetario totonaco de la costa de Veracruz*, México, CNCA (Consejo Nacional para la Culturas y las Artes), Cocina indígena y popular, núm. 44, consultado el 18 de septiembre de 2024. 
- Argüelles Santiago, Jazmín Nallely (2010), “El maíz en la identidad cultural de la Huasteca veracruzana”, *XVIII Jornadas Lascasianas Internacionales. Padre/*

- madre: nuestro maíz*, Acervo de la Biblioteca Jurídica Virtual, pp. 105-117, consultado el 10 de octubre de 2024. 
- Ariel de Vidas, Anath (1994), “La bella durmiente: el norte de Veracruz”, *Nuevo Mundo Mundos Nuevos*, consultado el 18 de septiembre de 2024. 
- Bravo, Ana, (2005), “El maíz en el Ecuador”, *Revista Semillas*, periodo 22/23, pp. 33-36, consultado el 3 de octubre de 2024. 
- Burgos-Hernández, Mireya (2018), “El plátano: la historia detrás de la fruta y otros secretos”, Centro de Investigación Científica de Yucatán, consultado el 18 de septiembre de 2024. 
- Camacho Díaz, Gonzalo (2010), “Dones devueltos. Música y comida ritual en la Huasteca”, *Itinerarios: Revista de estudios lingüísticos, literarios, históricos y antropológicos*, vol. 12, pp. 65-79, consultado el 18 de septiembre de 2024. 
- Careaga Gutiérrez, Dora Elena (2018), *Culinaria afrodescendiente de Tamiagua. Un discurso para iluminar a los afrodescendientes mexicanos*, Ciudad de México, Secretaría de Cultura, Dirección General de Culturas Populares, Indígenas y Urbanas, Arte y Culturas Populares.
- Casas, Alejandro *et al.* (2016), “Origen de la domesticación y la agricultura: cómo y por qué”, en Alejandro Casas, Juan Torres-Guevara y Fabiola Parra (coords.), *Domesticación en el continente americano*, México/Lima, Universidad Nacional Autónoma de México/Universidad Nacional Agraria La Molina (UNALM) del Perú, volumen 1. Manejo de biodiversidad y evolución dirigida por las culturas del Nuevo Mundo, pp.191-223, consultado el 18 de septiembre de 2024. 
- Colón, Cristóbal (2002). *Los cuatro viajes del almirante y su testamento*, 10ª ed., edición y prólogo de Ignacio B. Anzoátegui, Madrid, Espasa Calpe, consultado el 18 de septiembre de 2024. 
- Compendio Náhuatl (2012), “Compendio Enciclopédico Náhuatl”, Universidad Nacional Autónoma de México, consultado el 18 de septiembre de 2024. 
- Corominas, Joan y José Antonio Pascual (1954), *Diccionario crítico etimológico de la lengua castellana*, Madrid, Gredos, consultado el 18 de septiembre de 2024. 
- Cuvi, Pablo (1994), *Ecuador paso a paso: guía turística*, Colombia, Grupo Editorial Norma.
- del Río Moreno, Justo (1996), “El cerdo. Historia de un elemento esencial de la cultura castellana en la conquista y colonización de América (siglo xvi)”, *Anuario de Estudios Americanos*, 53(1), pp. 13-35, consultado el 26 de octubre de 2024. 
- Diccionario de Americanismos* (DA) (en línea), consultado el 4 de octubre de 2024. 




- Diccionario del Español de México* (DEM) (en línea), “Carnitas”, consultado el 4 de octubre de 2024. 
- Díaz Araujo, Yaxmareli (2019), “El cerdo en la conquista y la colonización: factor de poder y dominio occidental”, *Revista arbitrada del Centro de Investigación y Estudios Gerenciales*, núm. 39, consultado el 8 de mayo de 2021. 
- Díaz de Atauri, Juan (2008), “La incorporación al léxico de las voces bicicleta y ciclismo y el grupo de palabras que traen consigo (ensayo de un corpus modular)”, *Cuadernos del Instituto Historia de la Lengua*, núm. 1, pp. 45-61, consultado el 18 de septiembre de 2024. 
- Echeverría Almeida, José y Cristina Muñoz (1988), *Maíz: regalo de los dioses, Ecuador*, Instituto Otavaleño de Antropología (IOA), Curiñán, consultado el 19 de septiembre de 2024. 
- Instituto Nacional de Estadística y Geografía [INEGI] (2014), *Encuesta Nacional Agropecuaria 2014*, consultado el 9 de mayo de 2021. 
- Escobar Ohmstede, Antonio y Luz Carregha Lamadrid (2002), *El siglo XIX en las Huastecas*, México, Centro de Investigaciones y Estudios Superiores en Antropología Social / El Colegio de San Luis, Colección Huasteca.
- Fusté-Forné, Francesc (2016), “Los paisajes de la cultura: la gastronomía y el patrimonio culinario”, *Dixit*, núm. 24, pp. 4-16, consultado el 19 de septiembre de 2024. 
- Galeano, Eduardo (1970), *Las venas abiertas de América Latina*, Siglo XXI Editores, consultado el 19 de septiembre de 2024.
- Gallardo de la Puente, Carlos (2012), *Ecuador culinario*, Ecuador, UDLA (Universidad de las Américas). 
- Gómez Martínez, Arturo (2014), “Los tamales: ofrenda y simbolismo entre los nahuas de la Huasteca veracruzana, México”, *Anthropology of food*, vol. 9, consultado el 19 de septiembre de 2024. 
- Gran Diccionario Náhuatl* (GDN) (en línea), “Xamitl”, consultado el 1 de abril de 2021. 
- Grimaldo Muchotrigo, Miriam (2006), “Identidad y Política Cultural en el Perú”, *Liberabit. Revista de Psicología*, vol. 12, pp. 41-48, consultado el 19 de septiembre de 2024. 
- Güemes, Román (2013), *Hoy le canto a mi sustento. Romance para el sabor de la mesa huasteca*, Xalapa, Universidad Veracruzana, consultado el 8 de abril de 2022. 
- Hernández Ferrer, Marcela (2004), “Idhidh kwitol: niño maíz. Los niños en los

- rituales agrícolas de los teenek de la Huasteca potosina”, en Johana Broda y Catherine Good Eshelman (coords.), *Historia y vida ceremonial en las comunidades mesoamericanas: los ritos agrícolas*, México, Instituto Nacional de Antropología e Historia / Universidad Nacional Autónoma de México, pp. 215-233.
- Instituto Nacional de Estadística y Censo (INEC) (2022), *Base de Datos-Censo de Población y Vivienda 2022*, consultado el 19 de septiembre de 2024. 
- Instituto Nacional de Pueblos Indígenas (INPI) (2020), “Bienvenidos al Atlas de los Pueblos Indígenas de México”, *Atlas de los pueblos indígenas de México*, consultado el 19 de septiembre de 2024. 
- Kato Yamakake, Takeo Angel *et al.* (2009), *Origen y diversificación del maíz. Una revisión analítica*, México, Universidad Nacional Autónoma de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, consultado el 19 de septiembre de 2024. 
- Kershenovich Schuster, Paulette (2015), “Jewish Food in Mexico: Reflections of a Community’s History, Culture, and Values”, *Revista de Estudos da Religião*, 15(1), pp. 58-79, consultado el 4 de octubre de 2024. 
- Lara Ramos, Luis Fernando (2006), *Curso de lexicología*, México, El Colegio de México.
- Lara Ramos, Luis Fernando (2016), *Teoría semántica y método lexicográfico*, México, El Colegio de México.
- Lema, Ninfa (2018), El ají, variedades, técnicas y usos aplicados a la cocina moderna ecuatoriana, tesis de grado en Gastronomía, Quito: Universidad de las Américas, consultado el 26 de octubre de 2024. 
- Lima, Luis (2017), Estudio gastronómico del cantón Loja, recopilación de recetas tradicionales y ruta gastronómica, tesis de grado, Loja, Universidad Internacional del Ecuador, consultado el 10 de octubre de 2024. 
- López López, Juan de Dios e Ignacio Alcalde Sánchez (2023), “Palabras que se lleva el río. Historia oral y memoria léxica de las actividades productivas en el Guadalquivir”, *El Futuro del Pasado*, vol. 14, pp. 549-584, consultado el 3 de octubre de 2024. 
- Mares Figueras, Úrsula (2020), La muerte de Chicomechitl. Experiencias de apoyo mutuo y modernización en el ejido de Tepetzintla, Ver., 1974-1995, tesis de maestría en Historia moderna y contemporánea, México, Instituto Mora, consultado el 3 de octubre de 2024. 
- Meléndez, Lucero y Sonia Morett (2024), “Préstamos léxicos tének en el español

- huasteco”, ponencia presentada en *XX Congreso Internacional de la Asociación de Lingüística y Filología de América Latina. Proyecto Lenguas en Contacto*, Concepción (Chile), Universidad de Concepción, 24 de enero.
- Melgarejo Vivanco, José Luis (1980), *Historia de la ganadería en Veracruz*, Xalapa, Ediciones del Gobierno de Veracruz, consultado el 3 de octubre de 2024. 
- Montemayor, Carlos (coord.) *et al.* (2007), *Diccionario del náhuatl en el español de México*, México, Gobierno del Distrito Federal/UNAM.
- Montes Hernández, Salvador *et al.* (2010), *Recopilación y análisis de la información existente de las especies del género Capsicum que crecen y se cultivan en México*, Bioteca (CONABIO), consultado el 8 de octubre de 2024. 
- Moreno Moreno, María Águeda (2021), “Introducción. Un mundo en retazos léxicos”, en M. Águeda Moreno (ed.), *Un mundo en retazos léxicos. Ambientes lingüísticos en la literatura oral de Jaén*, Madrid / Frankfurt, Iberoamericana / Vervuert, pp. 7-20, consultado el 8 de octubre de 2024. 
- Murra, John V. (1975), *Formaciones económicas y políticas del mundo andino*, Lima, Instituto de Estudios Peruanos.
- Ochoa, Ángela (2000), “Las aventuras de Dhipaak o dos facetas del sacrificio en la mitología de los teenek (huastecos)”, *Dimensión Antropológica*, vol. 20, pp. 101-123, consultado el 8 de octubre de 2024. 
- Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) (2020), “Perspectivas a mediano plazo. Perspectivas para la producción y el comercio mundial de bananos y frutas tropicales 2019-2018”, *FAO*, pp. 1-16, consultado el 19 de septiembre de 2024. 
- Ortiz, Fernando (1978), *Contrapunteo cubano del tabaco y el azúcar*, La Habana, Fundación Biblioteca Ayacucho, consultado el 8 de octubre de 2024. 
- Ovando, Félix (2018), *Cerro Azul, un pueblecito de la Huasteca veracruzana*, Instituto Tecnológico de Cerro Azul.
- Ovando, Félix (s.f.), *Historia sucinta de Juan Felipe*, Veracruz, [s.l.], [s.e.].
- Palacios, Niktelol (2014), Estudio lexicológico del vocabulario textil común en el español de México, tesis de doctorado en Lingüística, México, El Colegio de México, consultado el 8 de octubre de 2024. 
- Palacios, Niktelol (2020), “Los nombres de los panes en la ciudad de Puebla”, en Jorge Lázaro y Rafael Saldívar Arreola (coords.), *Estudios sobre léxico del español de México basados en corpus*, Ciudad de México, Fontamara, pp. 81-99.
- Paredes, Octavio *et al.* (2006), Los alimentos mágicos de las culturas indígenas

- mesoamericanas, *La ciencia para todos*, núm. 212, México, Fondo de Cultura Económica (FCE).
- Parodi, Claudia (2007), “La semántica cultural y la indianización de América: un análisis del contacto lingüístico”, en B. Mariscal y A. González (eds.), *Actas del XV Congreso de la Asociación Internacional de Hispanistas*, México, Fondo de Cultura Económica (FCE), pp. 211-223, consultado el 8 de octubre de 2024. 
- Pazos Barrera, Julio (2017), *Elogio de las cocinas tradicionales del Ecuador*, Quito, Pontificia Universidad Católica del Ecuador.
- Pérez Aguilar, Raúl Arístedes (2002), “Léxico de los chichleros en Chetumal”, *Nueva Revista de Filología Hispánica*, vol. L, núm. 1, pp. 15-41.
- Piña Chán, Román (1989), “El desarrollo de la tradición huasteca”, en Lorenzo Ochoa (ed.), *Huastecos y totonacos. Una antología histórico-cultural*, México, Consejo Nacional para la Cultura y las Artes.
- Quesada, Miguel Ángel (2011), “El mundo de las creencias en América Central: aproximación etnolingüística”, *Temas Iberoamericanos*, pp. 127-154.
- Ramírez, Marina (2001), *Recetario nahua del norte de Veracruz*, México, Consejo Nacional para la Cultura y las Artes / Dirección General de Culturas Populares, Cocina indígena y popular.
- Rey, Alain (2014), “El diccionario cultural”, traducción de Luz Fernández Gordillo, *Andamios. Revista de investigación social*, vol.11, núm. 26, pp. 143-219. (Trabajo original publicado en 1978).
- Rodríguez Sepúlveda, Teresa (2012), *Cerro Azul en imágenes*, [s.l.], Programa de desarrollo cultural de la Huasteca.
- Ruvalcaba Mercado, Jesús (1998), “Notas sobre las plantas y los animales domésticos de la Huasteca”, en Jesús Ruvalcaba (coord.), *Nuevos aportes al conocimiento de la Huasteca*, México, CIESAS / CIHSLP / IPN / UACH / INI.
- Sandstrom, Alan (1998), “El nene lloroso y el espíritu nahua del maíz: el cuerpo humano como símbolo clave en la Huasteca veracruzana”, en Jesús Ruvalcaba (coord.), *Nuevos aportes al conocimiento de la Huasteca*, México, CIESAS / CIHSLP / IPN / UACH / INI.
- Secretaría de Economía (2012), “Monografía del sector plátano en México: situación actual y oportunidades de mercado”, *Gobierno de México*, consultado el 10 de octubre de 2024. 
- Sistema de Información Estadística y Geográfica del Estado de Veracruz de Ignacio



- de la Llave (SIEGVER) (2021) “Cerro Azul”, *Cuadernillos municipales 2021*, consultado el 19 de enero de 2022. 
- Solano, Felipe (2006), “Monografía histórica de la ciudad de Álamo, Veracruz”, consultado 10 de mayo de 2021. 
- Stark, Louisa y Pieter Muysken (1977), *Diccionario español-quichua quichua-español*, Guayaquil: Archivo Histórico del Guayas.
- Vázquez Gómez, Wendy Anel (2008), Elaboración de corpus y prototipo de diccionario cultural culinario de Guerrero, México, tesis de maestría en Lingüística Aplicada, México, Universidad Nacional Autónoma de México, consultado el 10 de mayo de 2021. 

**SONIA ELISA MORETT ÁLVAREZ.** Licenciada en Lengua y Literaturas Hispánicas por la Universidad Nacional Autónoma de México (UNAM), maestra en Lexicografía Hispánica por la Asociación de Academias de la Lengua Española (ASALE), maestra en Modelado y Simulación de Sistemas por la Universidad de Los Andes (Venezuela), doctoranda en Lingüística por El Colegio de México. Se ha desempeñado como docente en diferentes instituciones de educación superior (Universidad de Los Andes, Venezuela, Instituto Superior de Intérpretes y Traductores, México y Universidad Autónoma Metropolitana, México), impartiendo asignaturas como Lectura y escritura, Lingüística de corpus, Terminología, Sociolingüística y Lingüística del discurso. Ha participado en proyectos como investigadora o lexicógrafa. Sus áreas de interés son la lingüística aplicada, así como la variación, el cambio y el contacto lingüístico, particularmente en el nivel léxico. Es autora del capítulo dedicado a México dentro del libro *Actitudes lingüísticas de los hispanohablantes hacia el idioma español y sus variantes*, y del artículo “La valoración lingüística en los años escolares. Estudio de caso”, publicado en *Lingüística Mexicana. Nueva Época*.

**ISIS NATHALY ZAMBRANO OJEDA.** Licenciada en Lingüística Aplicada con Mención en Traducción en la Pontificia Universidad Católica del Ecuador. Doctoranda en Lingüística por El Colegio de México. Miembro del Programa de Investigación Oralidad Modernidad desde 2015, del que fue asistente de investigación en un subproyecto sobre léxico etnográfico de plantas medicinales y, también, capacitadora de software lingüístico ELAN, Lexique-Pro y Praat a miembros de comunidades indígenas y a estudiantes de Licenciatura en Lingüística. Trabajó como miembro del equipo de investigación FUNPROEIB



Andes en la búsqueda, revisión y sistematización de activistas digitales de lenguas indígenas de Ecuador en redes sociales. Sus áreas de interés son lenguas indígenas, documentación y revitalización lingüística, y léxico. Es autora del artículo “¡Achachay, qué frío, ve! Subjetivización de ‘ve’ como un marcador discursivo”, publicado en *Cuadernos de Lingüística de El Colegio de México* y coautora del capítulo el “El kichwa en productos alimenticios: ¿sumak kawsay o mercantilización?”, en *Paisajes lingüísticos de las dos orillas. Miradas trasatlánticas*.

D.R. © Sonia Elisa Morett Álvarez, Ciudad de México, enero-junio, 2024.

D.R. © Isis Nathaly Zambrano Ojeda, Ciudad de México, enero-junio, 2024.

## *The personal pronouns yo and tú in the oral interaction of Chinese Spanish-speaking learners*

LIU, MUZI 

Universidad de La Habana - FENHI  
[paula\\_liu@outlook.com](mailto:paula_liu@outlook.com)

LIDIA ESTER CUBA VEGA 

Universidad de La Habana - FENHI  
[lidiacuba@fenhi.uh.cu](mailto:lidiacuba@fenhi.uh.cu)

MARITZA GUERRERO HIDALGO 

Universidad de La Habana - FENHI  
[mguerreohidalgo57@gmail.com](mailto:mguerreohidalgo57@gmail.com)

**Abstract:** *The objective of this article is to characterize the use of the personal pronouns yo and tú in the oral interaction of a sample of Sino-speaking learners, in which difficulties have been observed in its use. The method to access the required information was class observation, from which samples of use (sometimes anomalous) of the pronouns under study were obtained. We consider that these results will contribute to the understanding of the phenomenon, which –in turn– will allow for theoretical and practical reflections that contribute to the improvement of the teaching-learning of Spanish in Sino-speakers.*

**KEYWORDS:** GRAMMAR; COMMUNICATIVE EXCHANGE; MANDARIN; LINGUISTIC DISTANCE; PRONOMINAL SUBJECT

**RECEPTION:** 13/08/2024

**ACCEPTANCE:** 02/10/2024

## Los pronombres personales *yo* y *tú* en la interacción oral de aprendientes sinohablantes de español

LIU, MUZI 

Universidad de La Habana - FENHI

[paula\\_liu@outlook.com](mailto:paula_liu@outlook.com)

LIDIA ESTER CUBA VEGA 

Universidad de La Habana-FENHI

[lidiacuba@fenhi.uh.cu](mailto:lidiacuba@fenhi.uh.cu)

MARITZA GUERRERO HIDALGO 

Universidad de La Habana-FENHI

[mguerreohidalgo57@gmail.com](mailto:mguerreohidalgo57@gmail.com)

**Resumen:** El presente artículo tiene como objetivo caracterizar el empleo de los pronombres personales *yo* y *tú* en la interacción oral de una muestra de aprendientes sinohablantes, ya que, en esta población, se aprecian dificultades para su utilización. El método para acceder a la información requerida fue la observación de clases, a partir de la cual se obtuvieron muestras de uso (a veces anómalo) de los pronombres objeto de estudio. Consideramos que estos resultados contribuirán a la comprensión del fenómeno, lo que –a su vez– permitirá realizar reflexiones teóricas y prácticas que aporten al perfeccionamiento de la enseñanza-aprendizaje del español en sinohablantes.

**PALABRAS CLAVE:** GRAMÁTICA; INTERCAMBIO COMUNICATIVO; MANDARÍN; DISTANCIA LINGÜÍSTICA; SUJETO PRONOMINAL

**RECEPCIÓN:** 13/08/2024

**ACEPTACIÓN:** 02/10/2024

## INTRODUCCIÓN<sup>1</sup>

La enseñanza de la gramática es una de las prioridades en la adquisición del español como segunda lengua (L2), ya que –por un lado– la profundización en el conocimiento detallado de las estructuras de la lengua sirve como fundamento para acceder a otros conceptos lingüísticos, y –por otro– constituye un recurso necesario para el desarrollo de las habilidades orales y escritas que permiten arribar a la competencia comunicativa. En lo que respecta al proceso comunicativo, éste es altamente complejo, pues supone una interacción permanente (en la que se intercambian mensajes) entre los interlocutores, lo que, al mismo tiempo, les impone asumir una conducta.

Una de las formas en la que se concreta la comunicación es la oralidad, caracterizada –desde la perspectiva educativa– como el proceso didáctico en el que se establece una toma de contacto entre varios participantes basada en interacciones verbales, que dependen de los interlocutores y del contexto (Mamani, 2020).

La oralidad, asimismo, se concibe como una habilidad enseñable y aprendible. En tal sentido, Hernández (2023: 53) asume que ésta podría promover “intercambios orales vinculados con la producción de sentido, la confrontación de puntos de vista, y el reconocimiento del otro (...), lo que posibilitaría posicionar la acción de los participantes en los eventos comunicativos.”

Por su parte, acertadamente, Cevallos (2022: 332) la considera “un eje transversal en todos los procesos pues les permite a los estudiantes intercambiar ideas, criterios, puntos de vista y con ello, las posibilidades para enriquecer y perfeccionar su léxico para que se conviertan en comunicadores eficientes”.

Por otra parte, y de acuerdo con lo expresado por Xiao (2023), la interacción oral se reconoce como un soporte desde el cual se expresa, amplía, confronta, reelabora y desarrolla el conocimiento, por lo que se convierte en una habilidad esencial para poner en práctica la comunicación en el contexto académico.

<sup>1</sup> Este artículo constituye un resultado parcial del proyecto de investigación “La variedad cubana del español: estudios léxico-gramaticales y pragmáticos. Su aplicación al español como lengua extranjera (LE) y segunda (L2)”, dirigido por la Dra. Lidia Ester Cuba Vega, proyecto asociado al Programa Nacional Las Ciencias Sociales y las Humanidades. Desafíos ante la estrategia de desarrollo de la sociedad cubana.

Para esta autora (Xiao, 2023: 1), adicionalmente, “las intervenciones orales de los aprendientes reflejan cómo es el proceso de comprensión e interpretación, qué competencias poseen de las propias de su formación disciplinar y cómo ponen en práctica aspectos como la claridad de la expresión, el uso adecuado de la lengua, el respeto al interlocutor, la cordialidad con los participantes, etc.”.

En relación con este tema, Rodríguez *et al.* (2023: 3) asumen que “la expresión e interpretación de marcas pragmáticas, la negociación de significados y los turnos de palabras constituyen elementos de orden lingüístico que intervienen en la interacción”, y que resultan “de vital importancia”, puesto que tales aspectos contribuyen a la concreción y al ejercicio adecuado de la competencia oral, especialmente, dentro del proceso de enseñanza-aprendizaje.

En el marco de este último, la interacción oral se define:

como una acción didáctica fundamental en la generación de aprendizaje, así como para el desarrollo competencial de los alumnos, acción en la que dos alumnos o más llevan a cabo una interacción con un objetivo comunicativo, que va más allá de la práctica formal de determinados elementos del sistema (Franco y Batlle, 2023: 249).

Como se ha podido observar, en el contexto de la enseñanza del español como lengua extranjera (ELE) o español como segunda lengua (EL2), la interacción oral es fundamental.

En el Plan Curricular del Instituto Cervantes (PCIC)<sup>2</sup> (2011), en relación con la habilidad comunicativa que aquí nos ocupa, se manifiesta que interactuar en una lengua extranjera supone poder comunicar a un interlocutor concreto y en un momento determinado, aquello que se piensa o se necesita, de la forma más efectiva posible conforme a las expectativas del interlocutor y a lo que la situación comunicativa demanda.

<sup>2</sup> El Plan Curricular del Instituto Cervantes para la enseñanza del ELE se publicó en 1994 y, desde entonces, ha constituido un documento básico de trabajo para todos los profesores inmersos en este ámbito. Éste desarrolla y fija los niveles de referencia para el español, según el Marco Común Europeo de Referencia para las Lenguas (MCER) (Xiao, 2023: 50).

Igualmente, Xiao (2023) refiere, siguiendo a Portolés (2004), que la interacción oral tiene como singularidad que, por lo general, los participantes se encuentran en una copresencia personal, al compartir un entorno cognitivo más grande. En este tipo de intercambio, la prosodia codifica gran cantidad de información, lo que condiciona su concreción con mayor claridad en una situación contextual dada, en la que se instituyen las restricciones o beneficios que se ponen en juego en el evento comunicativo (Charaudeau, 2004: 28; como se citó en Portolés, 2004: 110-111).

Por su parte, expresan al respecto Franco y Battle (2023: 29):

La importancia de la práctica de la interacción oral en el aula de ELE implica la necesidad de tenerla en consideración en la formación docente. El hecho de conocer cómo se caracteriza la práctica de la interacción oral permite al docente tener un mayor conocimiento de lo que ocurre cuando propone una actividad de interacción y, en consecuencia, le permite al profesor estrechar la distancia entre lo que planifica y lo que ocurre en el aula.

Las ideas plasmadas en el pasaje anterior hacen evidente la necesidad de propiciar interacciones en el aula, dado su impacto positivo en el desarrollo de la habilidad para expresarse oralmente en cualquier ámbito por parte de los aprendientes.

Pero dicha interacción oral no sólo funciona como vehículo de transmisión de información entre los interlocutores (al hacer del conocimiento del receptor algún aspecto de la realidad o de la forma de pensar del emisor, etc.), sino que también permite, entre otros muchos aspectos, comprender el mundo de relaciones de quienes intervienen en la comunicación. En este sentido, es de llamar la atención la manera en que los aprendientes sinohablantes de español (cuya lengua materna es muy distante lingüísticamente de la hispana) hacen uso de los pronombres personales *yo* y *tú* en sus interacciones orales.

En este punto, cabe recordar que la presencia/ausencia del sujeto explícito mediante pronombres personales es un aspecto del español que ha sido ampliamente estudiado desde distintas perspectivas; entre ellas, la variación motivada por diferencias dialectales. En este tipo de estudios, se destaca que las hablas del Caribe tienden a hacer mayor uso de los sujetos pronominales.

Se coincide con Liu (2016) en que esta problemática ha sido analizada a través de los enfoques de la gramática tradicional, de la sintaxis generativa,

de la pragmática y de la sociolingüística. A su vez, los estudios han tenido en cuenta tanto variedades dialectales del español peninsular como de países americanos, tales como Venezuela (Bentivoglio, 1987), Puerto Rico (Cameron, 1992, 1993 y 1996), España (Prieto, 2011) y Estados Unidos (Silva-Corvalán, 1982, 1994). En Cuba, se cuenta con investigaciones sobre esta dicotomía como la de Pérez (2006), la de Sánchez (2015) y, más recientemente, la de Ramírez *et al.* (2020).

En lo que toca de manera particular a la presencia/ausencia de los pronombres personales de sujeto *yo* y *tú* en el aula de ELE, cabe decir que no son muchos los estudios existentes al respecto. Entre los más destacados, está el trabajo de López (2015), en el que la autora aborda las dificultades que supone para los aprendientes de habla inglesa la omisión del pronombre personal sujeto en español, ya que su presencia es de uso obligado en su lengua materna; asimismo, se refiere a la falta de tratamiento adecuado de este aspecto en las clases y manuales de ELE.

Por su parte, los sinohablantes tampoco están ajenos a presentar dificultades en la adquisición de este aspecto del español, como expresa Liu (2016: 5):

Es notoria la dificultad que enfrentan los estudiantes sinohablantes al determinar si emplean o no sujeto expreso en sus actos de habla, toda vez que en el chino mandarín lo más frecuente es el uso obligatorio del sujeto expreso dada la ausencia de desinencias verbales que puedan desambiguar a qué persona se hace referencia, aunque hay situaciones, esencialmente en el habla coloquial, en el que (*sic*), si el contexto lo aclara, es permisible la elisión.

En este sentido, Roselló (2023: 342) señala que “se observa que los aprendientes no hacen un uso agramatical del pronombre sujeto, aunque sí inapropiado, ya que lo utilizan en contextos donde un nativo, normalmente, no haría uso de él”. Si bien estos usos anómalos de los pronombres, en general, no afectan la comprensión de lo expresado por ellos, sí inciden negativamente en su competencia lingüística, lo que resulta inadecuado para el futuro profesional de quien empleará el español como lengua de trabajo.

A pesar de lo anterior, no hay trabajos específicos que den cuenta de la presencia/ausencia de los sujetos pronominales *yo* y *tú*<sup>3</sup> en la interacción

<sup>3</sup> En el presente estudio, se tomó la decisión de analizar exclusivamente los pronombres *yo* y *tú* siguiendo

oral de los aprendientes sinohablantes, salvo el de Liu (2019), principal antecedente de este estudio. Aunque se revisaron publicaciones dedicadas al estudio de la adquisición del español por aprendientes sinohablantes, como *SinoELE* y *DidactiRed* del Centro Virtual Cervantes (CVC), no se localizaron investigaciones con este objeto de estudio. En *SinoELE* se encontró un trabajo que tiene en cuenta los sujetos tácitos en español y chino (Li, 2020), pero se analizan muestras de hispanohablantes que aprenden chino, y no a la inversa.

Dada la escasez de información estructurada al respecto, cabe preguntarse: ¿cómo puede caracterizarse el uso de los pronombres personales *yo* y *tú* en aprendientes sinohablantes?; ¿predomina la presencia o la ausencia de éstos en función de sujeto?; ¿en qué contextos prevalece la presencia o la ausencia del sujeto pronominal?

Considerando todo lo anterior, el objetivo de este artículo es caracterizar el empleo de los pronombres personales *yo* y *tú* en la interacción oral en una muestra de aprendientes sinohablantes.

## LOS PRONOMBRES PERSONALES YO Y TÚ DEL ESPAÑOL

En la *Nueva Gramática de la Lengua Española* (NGLE), la Real Academia Española y la Asociación de Academias de las Lengua Española (2010) definen los pronombres personales de la siguiente forma:

(...) se denominan así porque presentan rasgos gramaticales de personas (...). Se caracterizan por designar a los participantes en el discurso. Esta propiedad reduce considerablemente su contenido léxico y los convierte además en categorías deícticas. (...) Son, además, elementos definidos, propiedad que comparten con los artículos determinados y con los nombres propios (2010: 1161).

---

el criterio expresado por Ramírez *et al.* (2020) relativo a que de los textos examinados y cotejados sobre este tema se colige que el número singular suscita la explicitación de los pronombres personales de sujeto, y a que son los de primera y segunda persona singular los de mayor frecuencia. Además, mantener estos mismos pronombres conlleva la posibilidad de establecer comparaciones con el mencionado estudio y otros que también abordan la oposición entre sujeto expreso y nulo en español.



Adicionalmente, la *NGLE* sostiene que los pronombres personales mantienen una estrecha relación con los nombres propios, particularmente los de primera y segunda persona, ya que se emplean para designar de modo unívoco a su referente, por lo que ni unos ni otros pueden ser reemplazados por otras expresiones. Asimismo, muestran similitud en su resistencia a recibir modificadores restrictivos.

Autores como Castañeda *et al.* (2008) y Prieto (2011) consideran que, en sentido general, las formas de primera y segunda persona del paradigma pronominal en español se corresponden con los pronombres intrínsecamente personales, debido a que sus referentes suelen formar parte de la interacción verbal por estar presentes –de algún modo– en el contexto discursivo: *yo*, *tú*, *usted* y sus plurales se construyen mutuamente en la interacción; así, la segunda persona es el resultado de su enunciación por parte de la primera. Al mismo tiempo, estos pronombres resultan reversibles en el marco del intercambio comunicativo: *yo* se transforma en *tú* y viceversa, por lo que –de alguna manera– se oponen.

Por añadidura, Castañeda *et al.* (2008) y Prieto (2011) asumen que la riqueza expresiva de estos pronombres está dada por su capacidad para cumplir las funciones de sujeto y complemento; en este último caso, adoptan otras formas: *me*, *mí* y *conmigo* (para la primera persona), y *te*, *ti* y *contigo* (para la segunda).

Como explica Liu (2016), la función sintáctica que realizan los pronombres objeto de esta investigación es la de sujeto, el cual –de acuerdo con la *NGLE*– “designa una función sintáctica y se aplica también –por extensión– al elemento que la desempeña. (...) [E]l pronombre *yo* es el sujeto de *estaba en Yo estaba allí*, y la ciudad lo es de *tiene en La ciudad tiene sus casas y sus calles y sus habitantes*” (2010: 2557).

Para la *NGLE*, adicionalmente, el término *sujeto* alude a una noción semántica, cuando se opone a *predicado*. Así, en el ejemplo anterior *la ciudad* actúa como sujeto del predicado *tiene sus casas y sus calles y sus habitantes*. La *RAE* menciona que esta interpretación puede identificarse como *sujeto de predicación*, y que resulta más abarcadora que la sintáctica, ya que no requiere obligatoriamente de la presencia de un verbo, como en *Perro ladrador, perro mordedor*.

En resumen, es posible definir el *sujeto* como la función sintáctica u oracional cuyo núcleo concuerda con el verbo en número y persona; desde

un punto de vista semántico, designa una realidad (persona, cosa, animal o entidad abstracta) de la que se dice algo a través del predicado (Liu, 2016).

## Variación dialectal y presencia/ausencia de sujetos pronominales

En relación con el uso de los pronombres personales en función de sujeto, vale precisar lo concerniente a las particularidades dialectales que éste presenta en el español del Caribe. Al referirse a esta situación, Orozco y Guy (2008: 70) apuntan que los resultados de estudios realizados “revelan tasas más altas de sujetos explícitos en el castellano americano que en el peninsular y en el Caribe que en dialectos continentales como los de México y Ecuador. También existen diferencias dialectales respecto a los factores que afectan la expresión pronominal”.

Coincidentemente, Sánchez (2012) aduce que un rasgo lingüístico característico de los dialectos hispanocaribeños es una mayor frecuencia en la expresión de los sujetos pronominales de primera y segunda persona frente a lo que se registra en otras zonas de habla hispana, en las que la elisión de los sujetos supone la norma.

Al respecto, Méndez y Moreno (2014) exponen que, en el español caribeño, el rasgo dialectal consistente en la elisión y la aspiración de /s/ en la segunda persona del singular conlleva marcar explícitamente el sujeto, con la finalidad de compensar y evitar una posible ambigüedad; es decir, el sujeto pronominal se emplea como elemento desambiguador.

Por su parte, Vandermeulen (2011: 19), en su estudio sobre el pronombre sujeto en el español del Caribe, específicamente en el de Puerto Rico y Venezuela, expresa: “Son varios los estudios que consideran la presencia de sujetos explícitos, y más específicamente, la de sujetos antepuestos, como una característica prototípica del habla caribeño (*sic*)”.

Para el caso específico del español de Cuba, un estudio realizado por Sánchez (2012), sobre un corpus oral inédito de catorce entrevistas aplicadas a adultos, concluye que los factores que más inciden en el uso de los sujetos pronominales en esta variante dialectal “son los de índole pragmática, particularmente el cambio de referencia con respecto al sujeto de la cláusula anterior, así como el énfasis y la marcación de contraste” (p. 149). En segundo orden de importancia, la autora señala que, entre los rasgos formales, destaca el pronombre de primera persona singular, cuya presencia, en muchos casos, se asocia con una función semántica desambiguadora, dado el sincretismo

que se presenta en español entre la primera persona y la tercera persona del singular. En tercer lugar, se añade que los verbos copulativos y los verbos de cambio de estado son las clases verbales que mayormente se asocian con la presencia de los sujetos pronominales.

Posteriormente, Ramírez *et al.* (2020), con una muestra de habla y una perspectiva muy diferentes a las del estudio anterior, también hicieron suyo este objeto de estudio. En su caso, analizaron cartas familiares manuscritas de un autor cubano, las cuales corresponden a los primeros años del siglo xx (1900-1904). En las conclusiones de esta investigación, se destacan como aspectos característicos de la muestra una mayor presencia de sujetos explícitos cuando sus referentes son de primera persona del singular, y que el cambio de referencia resulta significativo en la explicitación de las formas pronominales, ya que éstas desempeñan la función de presentar a la nueva persona gramatical en la que se focalizará el enunciado.

Como ha podido observarse, estos estudios revelan que en el español de Cuba prevalece la presencia de los sujetos pronominales frente a la ausencia de los mismos, y que los factores de carácter pragmático tienen una alta incidencia en su expresión en relación con otros factores gramaticales y semánticos.

## VARIABLES ANALIZADAS EN ESTUDIOS PRECEDENTES SOBRE PRESENCIA/AUSENCIA DE LOS PRONOMBRES *YO/TÚ*

En este apartado se reportan los hallazgos de una revisión bibliográfica que permitió identificar las variables que se han tomado en cuenta para el análisis de la presencia/ausencia del sujeto pronominal *yo/tú* en estudios precedentes, como los de Bentivoglio (1987), Contreras (2004), Quesada y Blackwell (2009), Bentivoglio *et al.* (2011), Prieto (2011), Liu, W. (2016) y Liu, M. (2019), así como en el de Ramírez *et al.* (2020). Se aprecia coincidencia en algunas de las variables dependientes o lingüísticas; otras, son particulares de cada trabajo.

Así, Contreras (2004) consideró pertinentes las variables *tipo de discurso, tema, intención comunicativa, género y edad*; por su parte, Bentivoglio *et al.* (2011) determinaron como elementos de análisis para PRESEEA, entre otras, *frases idiomáticas, usos enfáticos con mismola, y sujetos focales*; en tanto que, para Quesada y Blackwell (2009), resultaron definitorias las variables centradas en reglas pragmáticas: *prominencia del referente, foco de atención, foco de contraste, peso pragmático y suspensión del juicio epistémico*.

En cuanto a Prieto (2011), seleccionó *tipo de cláusula, tiempo verbal, modo verbal, aspecto, sintaxis verbal, semántica verbal* y la *información nueva frente a la conocida*; mientras que Liu, W. (2016) presentó como factores de análisis *sujeto expreso frente a sujeto nulo, persona y número del sujeto expreso, posición del sujeto expreso, tipo de sujeto nulo, especificidad del sujeto* (específico, no específico o genérico), clase verbal y correferencialidad.

Por su parte, los aspectos observados por Liu, M. (2019) fueron *presencial/ ausencia del sujeto pronominal; formas verbales correctas e incorrectas* que acompañan al pronombre; *tipo de construcción oracional* en que aparece el sujeto pronominal explícito o implícito; *posición del sujeto expreso* en la oración; *empleo con verbos reflexivos, empleo de tú frente a usted*. Finalmente, Ramírez *et al.* (2020), retomaron, con algunos ajustes, las variables establecidas por Prieto (2011), a las que añadieron el énfasis.

Como puede observarse, en las investigaciones mencionadas confluyen variables tales como *referente, tipo de discurso / construcción / cláusula, información nueva / foco de contraste* y énfasis. En la investigación que se reporta en este artículo se trabajó con aquellas variables que se consideraron más productivas, dadas la naturaleza del tipo de estudio, por una parte, y las dificultades observadas en la muestra de aprendientes estudiada, por otra. Se partió de los aspectos tenidos en cuenta por Prieto (2011), Liu (2016) y Ramírez *et al.* (2020).

Concretamente, las variables lingüísticas objeto de análisis fueron: a) tiempo, b) modo, c) sintaxis verbal, d) semántica verbal, e) información nueva o conocida, f) cambio de referente del enunciado, g) énfasis y h) especificidad del sujeto (específico, no específico o genérico).

Para el análisis de la semántica verbal, aunque se ha considerado la clasificación en *estativos* y *dinámicos* de Prieto (2011), se ha sumado, siguiendo a Ramírez *et al.* (2020), la tipología de Bentivoglio (1987) de *verbos de actividad cognoscitiva y mental, verbos de percepción y sensación, verbos de dicción o pronunciación y verbos de volición*.

## METODOLOGÍA

En este estudio, se siguió una metodología mixta, que combinó las perspectivas cualitativa y cuantitativa con una finalidad descriptiva; se emplearon

como métodos la observación de clase y la encuesta a profesores, y como instrumento, el cuestionario.

La población estudiantil con la que se trabajó la conforman estudiantes de la carrera de Lengua Española para No Hispanohablantes de la Universidad de La Habana de segundo y tercer grado, población que asciende a 33 matriculados. Todos ellos tienen como lengua materna el chino mandarín y han estudiado español durante tres y cuatro años respectivamente,<sup>4</sup> en un contexto de inmersión lingüístico-cultural, en el ámbito comunicativo cubano. De acuerdo con el plan de estudios (currículo) de la carrera, deben poseer un nivel de competencia en la lengua correspondiente a B1, en el caso de los de segundo, y a B2, en el de los de tercero.

Como en la mayor parte de los trabajos consultados en los que analiza la producción oral de aprendientes de español como L2 (Blanco, 2012; Dong, 2018; Pérez, 2023; García, 2024), se constituyó un corpus con muestras de oralidad en contextos reales de interacción, siguiendo las pautas aportadas por García (2024: 119): “grabaciones de interacciones espontáneas o semi-espontáneas (...), conversaciones entre participantes con la misma categoría funcional (generalmente aprendientes de la L2) obtenidas de manera explícita por el investigador mediante una grabación no secreta, con cierta planificación y en contextos controlados”.

Así pues, se seleccionó una muestra constituida por 22 informantes: 11 de segundo año, de un total de 17 matriculados, lo que representa un 64.7 %, y 11 de tercero; en este caso, de una población total de 16, es decir, un 68.7 % de la matrícula. No se tuvo en cuenta la variable independiente *sexo*, pues la población de estudiantes de los dos años académicos es mayoritariamente masculina, por lo que se desestimó este factor como significativo para establecer un contraste. La variable edad tampoco fue determinante, en tanto ésta es concomitante con el año académico, que era lo que más interesaba a los efectos de esta investigación.

A partir de la variable independiente *año académico*, se le asignó un código de identificación a cada informante, donde LE equivale al nombre de la

<sup>4</sup> Cada cohorte de estudiantes debe cursar primero la preparatoria, con un año académico de duración, cuyo objetivo principal es desarrollar la competencia en lengua española, de manera que les sirva como lengua vehicular para la aprehensión de los conocimientos en los cuatro años de la carrera.

licenciatura (lengua española),<sup>5</sup> el segundo bloque corresponde al año académico y el tercero, a un número consecutivo asignado arbitrariamente a cada participante. Así, el código LE\_A2\_01 nos refiere al colaborador de segundo año de la carrera identificado con el número 01; si se tratara del código LE\_A3\_01, éste nos conduciría a un estudiante del tercer año de la carrera.

En lo que respecta al trabajo con los profesores, se realizó una encuesta a partir de un cuestionario (véase el anexo 1), mismo que se aplicó a los docentes de la licenciatura referida con el propósito de determinar su grado de consciencia sobre la problemática estudiada, es decir, qué tanto perciben las dificultades que afrontan sus alumnos en cuanto al uso del sujeto pronominal en la primera y segunda persona del singular, pues una mayor consciencia de la situación facilitaría una intervención didáctica intencionada que combinara aspectos pragmáticos con gramaticales.

Para este ejercicio fue encuestado el 100 % de los profesores de las disciplinas propias del ejercicio de la profesión que trabajaron con segundo y tercer años de la carrera durante el segundo semestre del curso (10 docentes en total).

En cuanto a la observación de clase, ésta se llevó a cabo con el propósito de registrar la presencia/ausencia de los pronombres *yo* y *tú* en la interacción oral aprendiente-aprendiente, aprendiente-profesor, y caracterizar dichos usos alternantes en la población objeto de estudio.

En total, se realizó el seguimiento de 11 clases de 90 minutos cada una, lo que equivale a 16 horas y 50 minutos de clases observadas. Se seleccionó la modalidad de clases prácticas, fundamentalmente de las disciplinas Lengua y Comunicación y Práctica Preprofesional, que –por sus propias características– requieren de la interacción de los aprendientes entre ellos y con sus profesores. Las clases observadas fueron videograbadas con la autorización del docente. Ello permitió recurrir a la videograbación siempre que hubo necesidad de esclarecer algún aspecto.

La información de nuestro interés se registró sistemáticamente en una planilla por clase. Por ser relevante para la interacción oral, se anotó cada participación de los estudiantes identificándola con el nombre o código de su

<sup>5</sup> Esta notación es importante, porque la investigación tiene un espectro más amplio, ya que se tiene en cuenta también a los maestrantes de la facultad. LE distingue así a los estudiantes de licenciatura de los de la maestría (M).

emisor, de manera que se pudieran determinar las variables extralingüísticas que habrían incidido en cada caso.

Siguiendo a Sánchez (2015), se calificaron como casos aquellas ocurrencias de verbos en forma personal, cuyos sujetos tienen referentes humanos y muestran una probabilidad de alternancia entre la no expresión y la expresión con un pronombre personal del caso nominativo. De manera que las variantes o formas alternantes de la variable estudiada son: un pronombre personal con función de sujeto y un sujeto tácito o nulo.

Posteriormente, se seleccionaron los contextos discursivos en los que se encontraba el sujeto pronominal *yo/tú* explícito o nulo, y se realizó su transcripción ortográfica. Por último, se analizó cada una de las apariciones.

## RESULTADOS

### Cuestionario a profesores

Con la pregunta 1 del cuestionario (véase anexo 1), se consultó a los docentes sobre los principales problemas sintácticos presentados por sus estudiantes. Las respuestas obtenidas aluden a dificultades de concordancia de género y de número, así como al uso de artículos, de los verbos irregulares, de los tiempos y modos verbales y de marcadores discursivos, además de particularidades en la construcción de oraciones subordinadas. Sólo un profesor mencionó el empleo del sujeto pronominal (ausencia/presencia), lo que equivale a 10 % de los encuestados.

En relación con la pregunta 2 (en anexo 1), dirigida a conocer cómo se trata el sujeto pronominal en el aula, se evidencia que los docentes se limitan a los aspectos gramaticales y semánticos. Por ejemplo, mencionaron que, en lo que atañe a la segunda persona, se instruye en el empleo de la expresión del sujeto para desambiguar cuando se produce elisión de /s/ en la forma verbal; también se refirieron a la posibilidad de omisión del sujeto. Ninguno aludió a aspectos pragmáticos.

Con la tercera pregunta, se pudo constatar que los encuestados dominan la información relativa a los tipos de sujeto en español y, en particular, la que corresponde al sujeto nulo; sin embargo, no aprecian esto como una dificultad que deba ser atendida.

Con la cuarta y última pregunta, se demostró que, salvo una profesora de traducción, los docentes que están en contacto con estudiantes sinohablantes no conocen las particularidades del sujeto expreso/nulo en chino mandarín. El hecho de no contar con una preparación previa que permita detectar las posibles interferencias del chino mandarín en la adquisición de las particularidades del sujeto pronominal en español, influye en la baja consciencia, por parte de los profesores, sobre las dificultades que afrontan los aprendientes en este aspecto.

En este punto, se concluye que sería importante que el docente supiera que el chino es una lengua aislante en la que casi no se encuentran procedimientos flexivos, lo que la distingue morfosintácticamente del español. “En una oración simple, el verbo del chino no puede mostrar los rasgos de persona ni de número, y pierde la capacidad de indicar la referencia de su sujeto” (Li, 2020: 133-134). Y cuando no se cuenta con información del contexto concreto, el sujeto tácito de una oración simple no es identificable, salvo en el caso de la oración impersonal.

Contar con este conocimiento previo posibilitaría dirigir, intencionadamente, la enseñanza de todo lo concerniente a la presencia/ausencia del sujeto pronominal de primera y segunda persona del singular en español, tanto en las clases de lengua y comunicación como en las de gramática española.

## **Análisis de las observaciones de clase: resultados globales**

Tras haber realizado los pasos que se desarrollan en el apartado metodológico, el corpus quedó conformado por 374 verbos que presentan la posibilidad de admitir, como sujeto, un pronombre de primera o segunda persona del singular (expreso), o bien, de que éste permanezca ausente (nulo).

Los datos muestran (tabla 1) que es más frecuente la omisión del sujeto pronominal (52.1 %) que su expresión (47.8 %). Sin embargo, se registran diferencias de distribución entre el sujeto de primera persona y el de segunda. Para la segunda persona, su expresión pronominal es más recurrente que la nula (+ 6.3 %); mientras que con la primera persona sucede lo contrario: hay un mayor porcentaje de sujetos tácitos (pronombre personal sujeto (PPS) nulo) que expresos (con igual diferencia de + 6.3 %). Además, es más frecuente el sujeto expreso *tú* (58.6 %) que el *yo* (41.3 %). En general,



hay un predominio de los PPSs de segunda persona (55.3 %) sobre los de primera (44.6 %).

TABLA 1. PPSs EXPRESOS Y NULOS EN LA MUESTRA GENERAL

Tipo de sujeto	# de ocurrencias	%	PPS	# de ocurrencias	%
Expresos	179	47.8	Yo	74	41.3
			Tú	105	58.6
Nulos	195	52.1	Yo	93	47.6
			Tú	102	52.3
Total (ambos tipos)	374	100	Yo	167	44.6
			Tú	207	55.3

A continuación, se reportan los resultados desgregados por año escolar.

### Análisis de las observaciones de clase: segundo año

El análisis por año académico evidencia que, en el caso de segundo (tabla 2), al igual que ocurre con la muestra global, predominan los PPSs ausentes (53.6 %) sobre los explícitos (46.3 %).

Asimismo, el PPS *tú*, tiene una mayor frecuencia de aparición total de expresos y nulos que el pronombre *yo*.

TABLA 2. PPSs EXPRESOS Y NULOS EN LA MUESTRA DE SEGUNDO AÑO

Tipo de sujeto	# de ocurrencias	%	PPS	# de ocurrencias	%
Expresos	95	46.3	Yo	40	42.3
			Tú	55	57.8
Nulos	110	53.6	Yo	50	45.4
			Tú	60	54.5
Total (ambos tipos)	205	100	Yo	90	43.9
			Tú	115	56

### Análisis de las observaciones de clase: tercer año

En lo que respecta a los datos de tercer año, es similar el número de ocurrencias de los PPSs expresos que el de los nulos, con una ligera superioridad de estos últimos en un + 0.5 % (tabla 3).

También aquí es mayor la cantidad de sujetos que corresponden al PPS *tú* (54.4 %) que a *yo* (45.5 %).

TABLA 3. PPSs EXPRESOS Y NULOS EN LA MUESTRA DE TERCER AÑO

Tipo de sujeto	# de ocurrencias	%	PPS	# de ocurrencias	%
Expresos	84	49.70	Yo	34	40.4
			Tú	50	59.5
Nulos	85	50.20	Yo	43	50.5
			Tú	42	49.4
Total (ambos tipos)	169	100	Yo	77	45.5
			Tú	92	54.4

### Análisis comparativo de las dos submuestras

En cuanto a la variable año académico, se observó (tabla 4) que los estudiantes del nivel inferior producen PPSs en mayor cantidad (54.8 %) que los de nivel superior, quienes lo hacen en un 45.1 %. Igualmente, en segundo año, resultó más frecuente la elisión del sujeto pronominal, en un 53.5 %, frente a la de tercer año, que es de un 50.2 %. En ambos niveles, prevalece el sujeto nulo en relación con el sujeto expreso; pero, igualmente, en el año inferior la elisión en la segunda persona supera en 4.4 % a la elisión en el tercer año.

TABLA 4. ANÁLISIS COMPARATIVO DE LAS DOS SUBMUESTRAS

Año	Expresos		Nulos		Total					
	yo	tú	yo	tú						
2º año	40	19.50 %	55	26.80 %	50	24.30 %	60	29.20 %	<b>205</b>	<b>54.80 %</b>

TABLA 4. (CONT.)

Año	Expresos		Nulos		Total					
	yo	tú	yo	tú						
3er. año	34	20.10 %	50	29.50 %	43	25.40 %	42	24.80 %	<b>169</b>	<b>45.10 %</b>
<b>Total</b>	<b>74</b>	<b>19.70 %</b>	<b>105</b>	<b>28.00 %</b>	<b>93</b>	<b>24.80 %</b>	<b>102</b>	<b>27.20 %</b>	<b>374</b>	<b>100 %</b>

## Análisis de las variables restantes

De las otras variables consideradas para esta investigación, las que arrojaron resultados más significativos fueron *tiempo verbal*, *sintaxis verbal*, *semántica verbal* y *cambio de referencia*. En cuanto a la primera, resultó más frecuente el empleo de los PPSs en el tiempo presente, seguido por su uso en pretérito y, finalmente, en futuro.

A continuación, se muestran los resultados por variable, para cuyo análisis se han seguido los procedimientos empleados por Ramírez *et al.* (2020).

### Tiempo verbal y modo verbal

Para el análisis, se agruparon estas dos variables por su interconexión y, sobre todo, porque el número de aparición de formas verbales en modo subjuntivo fue muy bajo y no se reportó ningún empleo en modo imperativo.

En lo que respecta al modo subjuntivo, sólo se constataron ocho formas verbales correspondientes a la primera o segunda persona del singular; de ellas, tres con sujetos expresos (1), lo que equivale al 37.5 % del total, y cinco con sujetos nulos (2), es decir, un 62.5 %. La mayor cantidad de ocurrencias se aprecia en la segunda persona (6 casos, equivalentes al 75 % del total de formas en modo subjuntivo).

(1) (...) tú debieras estar de acuerdo, porque estabas allí cuando sucedió y lo viste todo. (LE\_A2\_03).

(2) (...) si estuviera en el aula, hago lo mismo...(LE\_A3\_09).

Contrariamente a lo anotado por Méndez y Moreno (2014) sobre la función del sujeto expreso en el español caribeño como recurso desambiguador en los

casos de elisión del fonema /-s/ correspondiente a la segunda persona del singular, en este corpus no se documentó ninguna elisión de este tipo, algo que coincide con los resultados de Ramírez *et al.* (2020). En este punto, es importante destacar que, aunque la mayor parte de los colaboradores sólo han estado en contacto con hablantes cubanos, no es común en ellos la elisión de /-s/ final de sílaba o palabra, propia de la variante dialectal en cuestión. Todas las formas verbales restantes corresponden al modo indicativo (366) y se distribuyen en los modelos temporales verbales como se muestra en la tabla 5.

TABLA 5. PRESENCIA/AUSENCIA DE LOS PPSs SEGÚN TIEMPO VERBAL

Tiempo verbal	PPSs presentes		PPSs ausentes		Total
	#	%	#	%	#
Presente	118	45.9	139	54.08	<b>257</b>
Pretérito	48	52.1	44	47.8	<b>92</b>
Futuro	13	52	12	48	<b>25</b>

Las frecuencias para cada tiempo verbal se reflejan en la tabla 6.

TABLA 6. PRESENCIA/AUSENCIA DE LOS PPSs SEGÚN MODO VERBAL

Tiempos verbales	Presentes		Ausentes		Total
	#	%	#	%	#
Presente	118	45.9	139	54.08	<b>257</b>
Pretérito	27	45.7	32	54.2	<b>59</b>
Antepresente*	5	41.6	7	58.3	<b>12</b>
Copretérito	14	77.7	4	22.2	<b>18</b>
Pospretérito	2	66.6	1	33.3	<b>3</b>
Futuro	13	52	12	48	<b>25</b>

\* Nota: A diferencia de Ramírez, Valdés y Domínguez (2020), aquí se considera el tiempo antepresente un tiempo del modelo temporal pretérito, tanto por su uso (para hacer referencia a hechos pasados que tienen relación con la zona temporal en la que se encuentra el hablante) como por la nomenclatura que recibe de la RAE (pretérito perfecto).

Como ya se mostró, el tiempo más empleado por los estudiantes es el presente (3), seguido por los tiempos de pretérito (4) y, por último, se localizó el futuro; no obstante, la diferencia entre estos dos respecto del primero es notable, ya que al tiempo presente le corresponde el 68.7 % de todos los tiempos empleados en las 374 formas verbales de primera y segunda persona del singular que se analizaron.

(3) De sujeto expreso: Por favor, profesora, yo no estoy de acuerdo con lo que mi compañero dice. (LE\_A3\_11).

(4) De sujeto nulo: No entiendo, profesora, lo que mi compañero quiere decir. (LE\_A3\_04) // Llegaste a Cuba hace dos años y no has vuelto todavía (LE\_A2\_05).

Asimismo, se observa que la presencia de los tiempos verbales copretérito, antepresente y pospretérito es baja, pues entre los tres no alcanzan el 9 % respecto del total: 4.8 %, 3.2 % y 0.8 % respectivamente. Llama la atención que, en los casos de futuro, copretérito y pospretérito, la presencia del sujeto pronominal *yo* o *tú* es superior a su omisión.

### Semántica verbal

En cuanto a la semántica verbal, la mayoría de los casos expresos de *tú* y *yo* en el discurso de los estudiantes corresponden a los verbos estativos (5) considerados por Prieto (2011); es decir, a aquellos en que persiste un único estado por cuanto no se produce un cambio de uno a otro (*ser, tener, amar, pensar*), con el 67.3 % de apariciones frente al 32.6 % de las correspondientes a los verbos dinámicos (6) del tipo *bailar, caminar o decir*, en coincidencia con los resultados obtenidos por esa autora y por Ramírez *et al.* (2020).

(5) Yo creo en el beneficio del estudio. (LE\_A3\_7) // Tú no puedes tener la misma opinión mía... (LE\_A3\_10).

(6) ¿Te vas? (LE\_A2\_08) // Sí, paseaba por el Malecón porque me recordaba el paseo de mi provincia. (LE\_A3\_02).

En relación con los restantes tipos de verbos que se tuvieron en cuenta, coincidentemente con el trabajo de Ramírez, Valdés y Domínguez (2020),

predominan, de modo significativo, los sujetos expresos con verbos de actividad cognoscitiva y mental (7), los cuales representan el 78.3 % frente al 21.6 % de los restantes, que se distribuyen fundamentalmente entre los de dicción o pronunciación oral (8), 14.4 %, de volición (9), 5.08 %, y de percepción y sensación (10), 2.13 %.

(7) Yo creo en el beneficio del estudio. (LE\_A3\_7) // Por favor, profesora, yo no estoy de acuerdo con lo que mi compañero dice. (LE\_A3\_11).

(8) Tú dices eso porque no es contigo. (LE\_A2\_01) // Tú hablas de lo que no te gusta... (LE\_A2\_05).

(9) Yo quiero salir bien en todas las asignaturas. (LE\_A2\_11).

(10) ¿Tú escuchas esa música? (LE\_A3\_04).

### Sintaxis verbal

De acuerdo con la variable sintaxis verbal, se observó el predominio de los PPSs con verbos transitivos (11), lo que corrobora lo apuntado por Prieto (2011) y Ramírez *et al.* (2020). En este corpus, los sujetos expresos de primera y segunda persona del singular con verbos transitivos constituyen el 79.9 %.

(11) Yo pienso que es mejor estudiar solo, no en grupo, porque cuando hay muchas personas se entretienen. (LE\_A3\_08)

Contrariamente a lo que sustentan las fuentes estudiadas, incluyendo a Ramírez *et al.* (2020), predominan los PPSs expresos en las construcciones predicativas (12), con un 80.4 % de aparición, y no en las copulativas.

(12) Yo tengo una opinión distinta, pero tú tienes esa opinión. (LE\_A2\_10)

### Cambio de referente

La bibliografía consultada da cuenta de que los factores pragmáticos influyen en la presencia/ausencia del sujeto pronominal *yo* y *tú*, pues cuando dos oraciones seguidas comparten referente, éste suele omitirse en la segunda (13); sin embargo, cuando los referentes son distintos, el sujeto se explicita en ambas (14).

(13) Yo soy china y Ø vivo en un apartamento con mi amiga... (LE\_A2\_6).

(14) Yo vine por primera vez a Cuba hace tres años. Tú tienes sólo dos años aquí. (LE\_A3\_02).

Sin embargo, en la muestra analizada, este aspecto, aunque favorece al sujeto expreso cuando se producen cambios de referencia, no evidencia una diferencia amplia (sólo 1.1 % más que cuando no hay cambio de referente).

### Énfasis

En el corpus, los PPSs expresos ascienden a 179 casos (47.8 % del total), y el 41.3 % de ellos corresponde a la primera persona (véase la tabla 1). La explicación de esto parece ser el énfasis que intencionadamente marca el hablante, algo que se ha descrito en la bibliografía estudiada: el énfasis es “una variable que justifica la aparición del pronombre” (Ramírez *et al.*, 2020: 277). Entre los contextos en que predomina la presencia del sujeto pronominal con valor enfático están PPSs seguido de la conjunción *pero* (como en el ejemplo 12) y *en oraciones subordinadas* (15 y 16).

(15) Tú dijiste que ibas porque tú estabas apurado. (LE\_A2\_09)

(16) (...) Tú lo preguntaste porque tú no sabías si era así. (LE\_A3\_05)

Llama la atención la recurrencia del PPSs expreso en las oraciones subordinadas adverbiales causales introducidas por la conjunción *porque* (16 y 17).

### Especificidad del sujeto

Existen teorías que explican el empleo del pronombre personal *tú* en los sujetos expresos cuando éste adquiere un valor no específico o genérico, como presenta Liu (2016: 63) en los ejemplos (17)<sup>6</sup> y (18):<sup>7</sup>

<sup>6</sup> Ejemplo 2 del trabajo de Lui (2016).

<sup>7</sup> Ejemplo 3 en Lui (2016).

(17) (...) en alguna etapa de tu vida Ø te das cuenta de que en realidad no era lo que tú esperabas de él... (LHAB\_H13\_073)

(18) Un amigo es el que en todo momento que tú lo necesites, lo mismo en lo bueno que en lo malo... (LHAB\_M31\_031)

En la citada investigación de Liu (2016) hay una alta frecuencia de sujetos expresos *tú* con valor no específico o genérico (63.2 % de los casos). Sin embargo, en el corpus de la actual investigación no se documentó ninguno.

Otro elemento significativo que pudo apreciarse, a pesar de que no se tuvo en cuenta como variable lingüística de análisis, fue una mayor ocurrencia de sujetos nulos en las oraciones interrogativas.

## DISCUSIÓN Y CONCLUSIONES

El análisis de las encuestas a los docentes evidenció que la problemática descrita en este trabajo y que afecta a los estudiantes chinos no es percibida por sus profesores; igualmente, se pudo observar que no existe, en las clases, un tratamiento intencionado hacia la presencia/ausencia de *yo* y *tú* de manera general y, en particular, en lo que se refiere a la interacción oral desde un abordaje pragmático.

Por otra parte, y en lo que respecta a la observación de clase, con la presente investigación se pudo corroborar la productividad de algunas de las variables de análisis en correspondencia con trabajos de los autores referenciados, aunque no de todas. Por ejemplo, a diferencia de investigaciones como las de Ramírez *et al.* (2020), en la actual muestra la presencia de los pronombres *yo* y *tú* se ve favorecida en la segunda persona. Hay coincidencia en relación con el tiempo presente, los verbos transitivos y el cambio de referencia, así como en la mayor frecuencia de los pronombres implícitos o sujeto nulo.

Una observación interesante fue que los estudiantes del nivel académico inferior (y supuestamente con menor dominio de la lengua) hicieron un uso más cercano a las peculiaridades de los sujetos expresos/nulos en el español de Cuba.

Hay una alta correspondencia entre los resultados aquí obtenidos en lo que respecta al empleo de los PPS por parte de los aprendientes sinohablantes y







los documentados para los hispanohablantes caribeños, lo que corrobora la importancia del proceso de inmersión lingüístico-cultural para la enseñanza-aprendizaje de segundas lenguas.









Sin embargo, se aprecian ciertas anomalías en el empleo de estos PPSs en relación con lo descrito en las fuentes consultadas, como es el caso de la mayor presencia de éstos con la segunda persona del singular, en tanto lo más recurrente en estudios con hablantes nativos ha sido que prevalezcan en la primera persona.







Por otra parte, y contrariamente a lo previsto, el sujeto no específico o genérico no se registró en la muestra de sinohablantes, a pesar del predominio de los PPSs para *tú*. Tampoco han sido pertinentes para los sujetos nulos las variables *construcciones con formas no personales del verbo* y las *construcciones imperativas*.









Cabe seguir profundizando en la investigación para precisar si las causas de los usos anómalos antes mencionados obedecen a factores vinculados con interferencias lingüísticas –dada la distancia morfológica y sintáctica existente entre el chino mandarín y el español–, o bien, si son atribuibles a vacíos didácticos en el proceso de enseñanza -aprendizaje.

## BIBLIOGRAFÍA

- Bentivoglio, Paola (1987), *Los sujetos pronominales de primera persona en el habla de Caracas*, Caracas, Universidad Central de Venezuela, consultado el 24 de agosto de 2024. 
- Bentivoglio, Paola, Luis A. Ortiz y Carmen Silva-Corvalán (2011), “La variable expresión del sujeto pronominal. Guía de codificación” [documento metodológico del Proyecto Presseas], consultado el 18 de junio de 2019. 
- Blanco Canales, Ana (2012), “Corpus oral para el estudio de la adquisición y aprendizaje del componente fónico del español como lengua extranjera”, *Revista de Lingüística Teórica y Aplicada*, vol. 50, núm. 2, pp. 13-37, consultado el 14 de octubre de 2024. 
- Cameron, Richard (1992), *Pronominal and Null Subject Variation in Spanish: Constraints, Dialects, and Functional Compensation*, disertación en Lingüística, Universidad de Pennsylvania, consultado el 24 de agosto de 2024. 
- Cameron, Richard (1993), “Ambiguous agreement, functional compensation, and

- nonspecific tú in the Spanish of San Juan, Puerto Rico, and Madrid, Spain”, *Language Variation and Change*, vol. 5, núm. 3, pp. 305-334, consultado el 24 de agosto de 2024. 
- Cameron, Richard (1996), “A community-based test of a linguistic hypothesis”, *Language in Society*, vol. 25, núm. 1, pp. 61-111, consultado el 24 de agosto de 2024. 
- Castañeda Castro, Alejandro, Jenaro Ortega Olivares, Lourdes Miquel López, Rosario Alonso Raya, José Plácido Ruiz Campillo y Pablo Martínez Gila (2008), “Pronombres personales en la gramática básica del estudiante de español. Guía para el profesor”, *MarcoELE. Revista de Didáctica*, núm. 5, s/p.
- Centro Virtual Cervantes (1994), *Plan curricular del Instituto Cervantes*, Madrid, Instituto Cervantes.
- Cevallos Tuárez, Jennifer Estefanía (2022), “Estrategia didáctica para desarrollar la comunicación oral entre el docente y los estudiantes de sexto año de Educación General Básica”, *Didáctica y Educación*, vol. 13, núm. 4, pp. 341-361, consultado el 24 de agosto de 2024. 
- Contreras, Alicia (2004), Presencia y ausencia de los pronombres ‘yo’ y ‘tú’ en el habla culta de Santiago de Chile, tesis para optar al grado de magíster en Lingüística con mención en Letras Hispánicas, Universidad de Chile, consultado el 24 de agosto de 2024. 
- Dong, Yang (2018), “El uso de corpus orales de español espontáneo para la enseñanza de dicha lengua”, *Monográficos SinoELE*, núm. 17, pp. 226-237, consultado el 14 de octubre de 2024. 
- Franco Cordon, Ana Indira y Jaume Batlle Rodríguez (2023), “Conceptuación de la interacción oral por parte de los profesores de español como lengua extranjera en formación inicial”, *Revista Lengua y habla*, vol. 27, pp. 248-265, consultado el 24 de agosto de 2024. 
- García Ruiz-Castillo, Carlos (2024), “Corpus orales de aprendientes de ELE para el estudio de la conversación en la L2 disponibles en Internet”, *広島大学学術情報リポジトリ Hiroshima University Institutional Repository*, núm. 27, pp. 117-132, consultado el 14 de octubre de 2024. 
- Hernández Rincón, Marleny (2023), Didáctica del género discursivo en la educación inicial. Comprensiones en comunidades mixtas de profesores desde la enseñanza remota de emergencia, tesis de doctorado en Educación y sociedad, Bogotá, Facultad de Ciencias de la Educación. Universidad de La Salle, consultado el 24 de agosto de 2024. 

- Li, Ge (2020), “Sujeto tácito del chino y del español desde la perspectiva contextual”, *Monográficos SinoELE*, núm. 20, pp. 133-142, consultado el 14 de octubre de 2024. 
- Liu, Muzi (2019), Presencia y ausencia de los pronombres personales yo y tú en la comunicación oral y escrita de estudiantes chinos de la FENHI, tesis de maestría en Estudios Lingüísticos del Español como Lengua Extranjera, Facultad de Español para No Hispanohablantes. Universidad de La Habana.
- Liu, Wei (2016), Usos del sujeto pronominal expreso/nulo en una muestra del corpus PRESEEA, La Habana, tesis de maestría en Estudios Lingüísticos del Español como Lengua Extranjera, Facultad de Español para No Hispanohablantes, Universidad de La Habana.
- López Rueda, Susana (2015), “Adquisición de los sujetos pronominales en español por aprendientes anglófonos”, *MarcoELE. Revista didáctica de ELE*, núm. 20, pp. 1-18, consultado el 25 de agosto de 2024. 
- Mamani Quispe, Marisol (2020), “La comunicación oral en el idioma inglés: la construcción de una revisión teórica”, *Paidagogo. Revista de Investigación en Ciencias de la Educación*, vol. 2, núm. 1, pp. 26-40, consultado el 11 de noviembre de 2024. 
- Méndez Rivera, Nelson José y Saúl Leonardo Moreno Mancipe (2014), “La expresión de los pronombres personales sujetos en español: evidencia desde el lenguaje nativo y no nativo y sus implicaciones para la clase de Español Lengua Extranjera”, *IX Conferencia Internacional sobre Lenguas Extranjeras, Comunicación y Cultura (WEFLA)*, pp. 1-11, consultado el 25 de agosto de 2024. 
- Orozco, Rafael y Gregory R. Guy (2008), “El uso variable de los pronombres sujetos: ¿qué pasa en la costa Caribe colombiana?”, en Maurice Westmoreland y Juan Antonio Thomas (eds.), *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*, Somerville (Massachusetts), pp. 70-80, consultado el 14 de octubre de 2024. 
- Pérez Rodríguez, Marisela del Carmen (2006), Comportamiento de los pronombres personales en muestras de habla culta de ciudad de La Habana, tesis de doctorado, La Habana, Universidad de La Habana.
- Pérez, María del R. (2023), “Estudios de pragmática desde perspectivas contemporáneas”, *Signos Lingüísticos*, vol. XIX, núm. 37, enero-junio, pp. 8-19, consultado el 14 de octubre de 2024. 
- Portolés Lázaro, José (2004), *Pragmática para hispanistas*, Madrid, Síntesis.
- Prieto Muñoz, Sonia Yalily (2011), Los pronombres personales sujetos en el español

- de América y España: estudio dialectal comparativo, tesis de doctorado en Estudios Hispánicos, Puerto Rico, Universidad de Puerto Rico.
- Quesada, Margaret y Sarah E. Blackwell (2009), “The L2 Acquisition of Null and Overt Spanish Subject Pronouns: A Pragmatic Approach”, en Joseph Collentine, Maryellen García, Barbara Lafford y Francisco Marcos (eds.), *Selected Proceedings of the 11<sup>th</sup> Hispanic Linguistics Symposium*, pp. 117-130, consultado el 15 de octubre de 2024. 
- Ramírez Fuentes, Yesenia, Damian Valdés Santiago y Marlen Aurora Domínguez Hernández (2020), “Análisis de los pronombres sujeto de primera y segunda persona del singular en cartas privadas”, *Lenguaje*, vol. 48, núm. 2, pp. 261-288, consultado el 25 de agosto de 2024. 
- Real Academia Española y Asociación de Academias de las Lengua Española (2010), *Nueva Gramática de la Lengua Española*, Espasa, consultado el 25 de agosto de 2024. 
- Rodríguez Rodríguez, Malena, Mirian López Serrano y Diana Rosa Cáceres León (2023), “TIC y desarrollo de la interacción oral en L2. Una experiencia en la FENHI”, *Órbita Científica. Revista Científico Pedagógica*, vol. 29, núm. 124, pp. 1-11, consultado el 15 de octubre de 2024. 
- Roselló, Jorge, (2023), “El corpus PRESEEA aplicado a la enseñanza de la lengua: la expresión del sujeto pronominal”, *Normas*, vol. 13, núm. 1, pp. 341-354, consultado el 15 de octubre de 2024. 
- Sánchez Arroba, María Elena (2012), “Presencia del sujeto pronominal en el español de Cuba”, en Patrizia Botta *et al.* (coords), *Rumbos del hispanismo en el umbral del Cincuentenario de la AIH*, vol. 8, [s. l], Bagatti Libri, pp. 140-149, consultado el 15 de octubre de 2024. 
- Sánchez Arroba, María Elena (2015), “La variable expresión de sujetos pronominales en el español de Cuba”, *Actas del XIV Simposio de Comunicación Social: retos y perspectivas Centro de Lingüística Aplicada*, pp. 233-238, consultado el 25 de agosto de 2024. 
- Silva-Corvalán, Carmen (1982), “Subject expression and placement in Mexican-American Spanish”, en John Amastae y Lucía Elías-Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, [Cambridge], Cambridge University Press pp. 93-120, consultado el 25 de agosto de 2024. 
- Silva-Corvalán, Carmen (1994), *Language Contact and Change: Spanish in Los Angeles*, Oxford, Oxford University Press.
- Vandermeulen, Nina (2011), *El pronombre personal sujeto en el español caribeño*.

Un estudio comparativo entre el puertorriqueño y el venezolano, tesis de maestría en Lingüística y literatura: lenguas iberorrománicas, Universidad de Gante, consultado el 15 de octubre de 2024. [🔗](#)

Xiao, Jing (2023), Perífrasis verbales del español empleadas en la interacción oral por estudiantes sinohablantes, tesis de maestría en Estudios Lingüísticos del Español como Lengua Extranjera, Facultad de Español para No Hispanohablantes. Universidad de La Habana.

## ANEXO 1

### Cuestionario a profesores

Estimado profesor:

En el marco de una investigación que estamos realizando sobre la competencia gramatical de los aprendientes sinohablantes de la carrera, nos resultará de muchísima utilidad su participación, respondiendo el cuestionario que presentamos a continuación. Muchas gracias por su colaboración.

Datos generales:

Graduado de:

Años de experiencia en ELE:

Categoría docente y científica:

Responda las siguientes preguntas:

1. Mencione los principales problemas sintácticos presentados por los aprendientes en la interacción oral en sus clases.
2. ¿Cómo trata el sujeto pronominal en sus clases? ¿Qué aspectos tiene en cuenta?
3. Mencione los distintos tipos de sujeto en español.
4. ¿Cuáles son las peculiaridades del sujeto expreso/nulo en el chino mandarín?

**LIU, MUZI:** Máster en Estudios Lingüísticos del Español como Lengua Extranjera por la Universidad de La Habana. Graduada de Lengua Española en la Universidad Internacional de Heilongjiang. Es doctoranda del Programa Doctoral de Ciencias Lingüísticas en la Universidad de La Habana. Es miembro del Proyecto

de Investigación “La variedad cubana del español: estudios léxico-gramaticales y pragmáticos. Su aplicación al español como lengua extranjera (LE) y segunda (L2)”, asociado al Programa Nacional Las Ciencias Sociales y las Humanidades. Desafíos ante la estrategia de desarrollo de la sociedad cubana, al cual se adscribe su tema doctoral. Investiga y publica en el ámbito de la lingüística aplicada a la enseñanza de la gramática del español para no hispanohablantes.

**LIDIA ESTER CUBA VEGA:** Máster en Lingüística Hispánica y Doctora en Ciencias Lingüísticas por la Universidad de La Habana. Ha sido docente en la Universidad de Ciencias Pedagógicas de La Habana y en la Facultad de Artes y Letras y en la Facultad de Español para No Hispanohablantes de la Universidad de La Habana, de la cual es decana desde hace 13 años. Dirige el Proyecto de Investigación “La variedad cubana del español: estudios léxico-gramaticales y pragmáticos. Su aplicación al español como lengua extranjera (LE) y segunda (L2)”, asociado al Programa Nacional Las Ciencias Sociales y las Humanidades. Desafíos ante la estrategia de desarrollo de la sociedad cubana. Sus investigaciones y publicaciones se inscriben fundamentalmente en esta temática. Dirige SIELE y SICELE en la Universidad de La Habana.

**MARITZA GUERRERO HIDALGO:** Doctora en Ciencias Pedagógicas por la Universidad de Ciencias Pedagógicas de La Habana, donde participa en la formación doctoral del Programa de Ciencias Lingüísticas. Fue docente de la Facultad de Lenguas Extranjeras y de la Facultad de Español para No Hispanohablantes de la Universidad de La Habana. En esta última se desempeñó como Jefa del Departamento de Estudios Lingüísticos y Culturales durante cinco años. Dirigió el Proyecto Institucional de Investigación “Los estudios gramaticales en la enseñanza del español como lengua extranjera y segunda lengua”.

D. R. © Liu, Muzi, Ciudad de México, enero-junio, 2024.

D. R. © Lidia Ester Cuba Vega, Ciudad de México, enero-junio, 2024.

D. R. © Maritza Guerrero Hidalgo, Ciudad de México, enero-junio, 2024.







NOTA

## The Bible as a Corpus for Language Technologies

GERARDO E. SIERRA MARTÍNEZ 

Instituto de Ingeniería - UNAM

[gsierram@iingen.unam.mx](mailto:gsierram@iingen.unam.mx)

GEMMA BEL ENGUIX 

Instituto de Ingeniería - UNAM

[gbele@iingen.unam.mx](mailto:gbele@iingen.unam.mx)

**Abstract:** *This work aims to create an aligned corpus of eleven Spanish translations of the Bible to advance computational linguistics in Spanish. The use of this corpus is essential for applications such as paraphrase detection, lexical grouping identification, and language model evaluation for search systems. In this way, the study covers various aspects of natural language processing, including similarity, lexical extraction, and bias analysis, with the goal of promoting the development of language technologies in Spanish.*

**KEYWORDS:** LINGUISTIC CORPUS; NATURAL LANGUAGE PROCESSING; PARAPHRASE DETECTION; SEMANTIC CLUSTERING; QUESTION-ANSWERING SYSTEMS

**RECEPTION:** 16/08/2024

**ACCEPTANCE:** 22/08/2024

# La Biblia como corpus para tecnologías del lenguaje

**GERARDO E. SIERRA MARTÍNEZ** 

Instituto de Ingeniería - UNAM

[gsierram@iingen.unam.mx](mailto:gsierram@iingen.unam.mx)

**GEMMA BEL ENGUIX** 

Instituto de Ingeniería - UNAM

[gbele@iingen.unam.mx](mailto:gbele@iingen.unam.mx)

**Resumen:** El presente trabajo tiene como objetivo presentar un corpus alineado de once traducciones de la Biblia al español, para avanzar en la lingüística computacional en español. El uso de este corpus resulta fundamental en aplicaciones como la detección de paráfrasis, la identificación de grupos léxicos y la evaluación de modelos de lenguaje para sistemas de búsqueda. De esta manera, este trabajo abarca estudios en procesamiento de lenguaje natural, como la similitud, extracción léxica, y análisis de sesgos, con el propósito de impulsar el desarrollo de tecnologías del lenguaje en español.

**PALABRAS CLAVE:** CORPUS LINGÜÍSTICO; PROCESAMIENTO DE LENGUAJE NATURAL; DETECCIÓN DE PARÁFRASIS; AGRUPACIÓN SEMÁNTICA; SISTEMAS PREGUNTA RESPUESTA

**RECEPCIÓN:** 16/08/2024

**ACEPTACIÓN:** 22/08/2024

## INTRODUCCIÓN

La Biblia es el texto más leído en la historia de Occidente y el más traducido del mundo, ya sea en versiones completas o parciales, como el Nuevo Testamento. Aunque es difícil tener un inventario acabado, Wikipedia afirma que hay traducciones completas en, al menos, 700 idiomas (Wikipedia, s.f.).

La Biblia es una compilación de un número variable de libros, que se dividen en dos grupos principales: Antiguo Testamento y Nuevo Testamento. Hay una diferencia en el canon del Antiguo Testamento para las tradiciones protestante (39 libros), católica (46 libros) y ortodoxa (51 libros). El Nuevo Testamento incluye 27 libros para las tres principales tradiciones cristianas.

En general, las versiones de la Biblia en distintas lenguas tienen el objetivo de capturar el mensaje exacto que el texto quiere transmitir. Sin embargo, existen diferentes enfoques de traducción (Klein *et al.*, 2017). Por un lado, la perspectiva *formalmente equivalente* intenta ser fiel al texto, al asignar una traducción literal palabra por palabra del original, prefiriendo la precisión a la claridad o inteligibilidad para los lectores actuales. Por otro lado, las traducciones dinámicamente o *funcionalmente equivalentes* se preocupan más por la claridad de las oraciones, al reformular los textos e interpretar su significado en contexto. Entre estas dos posiciones radicales se encuentra un enfoque óptimamente equivalente, que busca alcanzar la mejor traducción posible para cada pasaje, favoreciendo uno u otro planteamiento principal.

Muchos conjuntos de datos incluyen la Biblia en sus repositorios. Para el caso de las tecnologías del lenguaje, por sus numerosas traducciones y versiones, ésta se utiliza como un recurso valioso para un vasto campo de aplicaciones en áreas como la traducción automática, el procesamiento de lenguaje natural (PLN) y los estudios lingüísticos comparativos.

Por ejemplo, las múltiples versiones de la Biblia en diferentes idiomas resultan útiles para entrenar modelos de traducción automática, ya que la comparación de textos paralelos facilita la realización de estudios comparativos tendientes a analizar cómo se interpretan y expresan los mismos conceptos en diferentes lenguas y culturas. Este repertorio de multiplicidad de formas para un mismo sentido favorece el estudio de fenómenos como el cambio lingüístico, la variabilidad léxica y las diferencias sintácticas. Asimismo, debido

a que su disposición en capítulos y versículos permite crear bases de datos estructuradas, puede utilizarse para desarrollar y evaluar sistemas de búsqueda y recuperación de información. Adicionalmente, las distintas versiones y traducciones de la Biblia ofrecen un campo fértil para la estilometría, el estudio de estilos literarios mediante métodos cuantitativos, y –en general– para tareas de PLN, como el análisis de sentimiento, el etiquetado de partes del discurso y el reconocimiento de entidades nombradas.

Lo anterior, justifica el desarrollo del corpus que se describe en el presente escrito, mismo que consta de once traducciones de la Biblia a la lengua española; éstas abarcan diferentes siglos y regiones geográficas. Las once versiones se han alineado en función de la distribución de los versículos y la equivalencia de sentido. Se presentan también tres aplicaciones de este recurso: a) *detección de paráfrasis*, b) *identificación de grupos léxicos*, y c) evaluación de los *modelos grandes de lenguaje* (LLM, por sus siglas en inglés), para búsqueda de información y sistemas preguntas-respuesta.

## CORPUS PARALELO DE BIBLIAS EN ESPAÑOL

Nuestro corpus contiene once versiones diferentes de la Biblia en español, cuyas ediciones fueron creadas en diferentes períodos, aunque la mayoría se publicaron o editaron durante el siglo xx en distintos países de habla hispana (Sierra *et al.*, 2024). A continuación, se enlistan y describen brevemente tales publicaciones.

**La Biblia, que es, los Sacros Libros del Viejo y Nuevo Testamento (OSO).** Esta primera Biblia fue traducida por Casiodoro de Reina (OSO) y se publicó por primera vez en Suiza en 1569. La traducción se basó principalmente en el texto hebreo masorético, pero también tomó en consideración otras versiones en latín, griego e, incluso, judeo-español.

**La Santa Biblia (REV).** Versión de Cipriano de Valera de 1602; consiste en una revisión y actualización de la traducción de Casiodoro de Reina.

**La Sagrada Biblia (PET).** En 1823, Torres Amat (1772-1847) lanzó su traducción de la Biblia conocida como PET. Esta traducción se basa en gran medida en el trabajo del jesuita Petisco, quien había realizado una versión hacia finales del siglo xviii. A diferencia de las ediciones anteriores, que fueron

traducidas directamente del hebreo, arameo y griego, la de Torres Amat se basa principalmente en la *Vulgata*.<sup>1</sup>

**La Sagrada Biblia (JNM).** Nuestro corpus tiene un vacío de cien años entre el de Torres Amat y el siguiente texto, que es una traducción latinoamericana hecha en Chile por Wilhelm Jünemann. Este autor tradujo tanto el Nuevo Testamento como el Antiguo; no obstante, el primero salió a la luz en 1928, mientras que el Antiguo Testamento se publicó más recientemente, en 1992. Este último volumen fue traducido del griego koiné y se completó antes de 1938, el año de la muerte de Jünemann. A pesar de que no se editó antes, el volumen debe datarse, a todos los efectos, en la década de 1920 o 1930.

**La Sagrada Biblia (NAC).** La Sagrada Biblia de Nacar y Colunga (NAC) se publicó en España en 1944 y ocupa una posición prominente entre las versiones católicas, tras haberse difundido a través de más de treinta ediciones. Esta traducción se deriva de los textos originales en hebreo, arameo y griego. En el prólogo, los autores expresan su intención de mantener la fidelidad al original, aunque señalan que no necesariamente siguen servilmente sus letras; pues si bien se guían por las palabras del texto, atienden principalmente al sentido de las frases. A pesar de que la versión original incluye 73 libros canónicos, nuestro corpus comprende exclusivamente 66 libros.

**La Biblia de Jerusalén (JER).** Como un proyecto de traducción colaborativa, llevada a cabo por la *École biblique et archéologique française* de Jérusalem, el objetivo principal de este esfuerzo de traducción es facilitar la exégesis bíblica. Con principios de crítica histórica, se incluyen en el texto comentarios de esta índole, así como lingüísticos, los cuales proporcionan valiosos análisis e información. Desde su publicación inicial en 1966, la Biblia de Jerusalén ha experimentado varias revisiones. Sin embargo, la versión incorporada a nuestro corpus es la edición original de 1966. La traducción se basa en los textos originales en hebreo, arameo y griego, y sigue las directrices empleadas en la traducción francesa.

**La Santa Biblia (EMN).** Evaristo Martín Nieto dirigió y lanzó en 1961 esta versión en español realizada por un equipo de traductores que buscó

<sup>1</sup> Traducción de la Biblia al latín, según la tradición realizada por San Jerónimo y Santa Paula en el siglo IV.

proporcionar una representación fiel de los textos originales, adhiriéndose estrechamente a la redacción y estructura de los idiomas fuente.

**La Nueva Biblia - Edición pastoral para Latinoamérica (LAT).** Versión orientada específicamente a los lectores latinoamericanos. Se imprimió por primera vez en 1972, después del Concilio Vaticano II, tiene una ideología progresista en el marco de la Iglesia Católica. Fue considerada una herramienta de la teología de la liberación. Es muy popular entre la gente de América Latina. Está traducida del hebreo, arameo y griego con un estilo de equivalencia funcional.

**Nueva Biblia Española (ESP).** Esta versión también es una traducción con un enfoque de equivalencia funcional a partir de los textos originales; realizada por Alonso Shöckel y Juan Mateos, vio la luz en 1975. Con el propósito de hacer el texto más accesible para las personas y ayudarlas a conectar con el mensaje de la Biblia, acorde con su propio contexto cultural y lingüístico, los traductores emplearon un estilo libre que utilizaba expresiones y lenguaje que contribuyera a cerrar la brecha cultural entre el mundo moderno y el contexto histórico en el que se escribieron los textos bíblicos.

**La Biblia (SER).** Bajo la coordinación de Serafín de Ausejo, en 1975 se publicó esta versión regida por los principios de equivalencia formal. Realizada por un equipo de traductores, tuvo como objetivo mantener una representación fiel de los textos originales, enfatizando la precisión en la traducción de los idiomas fuente. A la postre, es considerada la versión en lengua española más literal.

**La Biblia de las Américas (AME).**<sup>2</sup> El título *Biblia de las Américas* refleja adecuadamente el público y el propósito de esta edición, ya que está dirigida, principalmente, a individuos de habla hispana que residen en el continente americano. La traducción en sí es resultado de un esfuerzo colaborativo que involucró a individuos de diversas corrientes protestantes, quienes trabajaron a partir de los textos originales. Esta versión se publicó por primera vez en 1986, lo que la convierte en una adición relativamente reciente dentro del conjunto de traducciones bíblicas disponibles en español.

<sup>2</sup> En adelante, se identificará cada versión de la Biblia con el código alfabético de tres caracteres que se muestra, entre paréntesis y tras su nombre, en este listado: OSO, REV, PET, JNM, NAC, JER, EMN, LAT, ESP, SER Y AME (*N. del Ed.*).

El principal criterio para la selección de los libros fue que las ediciones estuvieran disponibles íntegramente en formato digital, con posibilidad de ser descargadas y convertidas en documentos de texto. Al conformar el corpus, se excluyeron subtítulos y notas, con la finalidad de conservar únicamente el contenido de los versículos.

Dado que el número de libros en las diversas ediciones varía de 64 a 73, se suscitaron discrepancias sobre la inclusión o exclusión de algunos. Otro aspecto problemático para establecer paralelismos lo constituyen las diferencias en el número de capítulos dentro de un mismo libro, así como en el número de versículos dentro de cada capítulo, diferencias que se deben a factores como enfoques de traducción diferentes, desafíos en la interpretación del texto, empleo de textos fuente diferentes o divergencias ideológicas entre los traductores. Por ejemplo, mientras los libros de Judas o Lamentaciones muestran coincidencias en la cantidad de versículos y contenido en todas las ediciones, el libro de Salmos presenta diferencias significativas, ya que en éste se observaron hasta diez variantes.

Las disparidades representaron un reto importante al intentar alinear de manera paralela el contenido de las ediciones que conforman el corpus, ya que –como se ha dicho– éste no siempre coincide y el número de versículos es variable. Para facilitar la comparación de los textos de las diferentes ediciones, se utilizó un código creado en el lenguaje de programación Python, que ha permitido etiquetar los versículos en cada libro y categorizar los capítulos en función de las diferencias en la numeración de los versos. Esto facilitó, finalmente, obtener una alineación automática, tras la cual se llevó a cabo una revisión manual.

Como resultado del proceso anterior, se pudo determinar cuál fue la versión con el mayor número de versículos para cada segmento de la obra, misma que sirvió como guía para el libro o capítulo en cuestión, de manera que ninguna edición en particular se empleó como referencia de alineación para toda la obra. Los capítulos o versículos omitidos en una versión particular fueron marcados y etiquetados como **texto omitido**.<sup>3</sup>

<sup>3</sup> Para facilitar la lectura de los diferentes elementos de esta nota, se establecieron los siguientes usos tipográficos: Negritas (para resaltar las etiquetas empleadas en el proceso de marcación del texto). Subrayados (para destacar un elemento dentro del ejemplo (versículo), marca que corresponde al analista y no a la fuente de origen (*N. del Ed.*))



También hubo casos en los que faltaba el último versículo de un capítulo en una versión, pero el contenido omitido, en realidad, existía al principio del capítulo siguiente. Ante tales situaciones, se reorganizó la información para presentarla en la posición que ocupaba en otras versiones.

En (1), ejemplo procedente de Lucas 17: 36, la Biblia OSO (a) contiene una información adicional que no está presente en otras versiones, como en ESP (b) o en LAT (c), lo que da como resultado un hueco en el contenido:

- (1) a. Dos mujeres estarán moliendo juntas; la una será tomada, y la otra será dejada. Dos estarán en el campo; el uno será tomado, y el otro será dejado (OSOLUC017036).<sup>4</sup>
- b. Estarán dos moliendo juntas, a una se la llevarán y a la otra la dejarán (ESPLUC017036).
- c. Dos mujeres estarán moliendo juntas, pero una será llevada y la otra dejada (LATLUC017036).

En total, nuestro corpus comprende 754 libros, los cuales abarcan una colección voluminosa de 7 970 269 palabras. Para facilitar el acceso al mismo y la recuperación de sus materiales con fines académicos, la colección está disponible en la plataforma en línea GitHub,<sup>5</sup> así como en nuestro gestor de corpus GECO.<sup>6</sup>

El corpus alineado de traducciones de la Biblia en español tiene múltiples aplicaciones en tareas de procesamiento de lenguaje natural (PLN). A continuación, se detallan algunas líneas de investigación que se han llevado a cabo usando este recurso como base textual. En concreto, se explica cómo se

<sup>4</sup> Los caracteres alfanuméricos que se muestran entre paréntesis tras los ejemplos corresponden a los datos de identificación del versículo y la fuente exacta de dónde fueron extraídos. El código se forma de la siguiente manera: las tres primeras letras son de la versión de la Biblia, las tres siguientes letras corresponden al libro (LUC para Lucas, por ejemplo), los siguientes tres números refieren al capítulo y los últimos tres al versículo.

<sup>5</sup> <https://github.com/GIL-UNAM/SpanishParaphraseCorpora/tree/main/Biblias>

<sup>6</sup> <http://www.geco.unam.mx/geco3/proyecto/CPBE>

implementan las aplicaciones en las áreas de detección de paráfrasis, agrupación semántica y evaluación de modelos grandes de lenguaje (LLM).

## APLICACIÓN PARA DETECCIÓN DE PARÁFRASIS

Se considera que dos textos, o dos oraciones, son paráfrasis cuando son semánticamente equivalentes, independientemente de la causa de dicha equivalencia (Das y Smith, 2009). La detección automática de paráfrasis es una tarea de PLN que sirve de base para múltiples aplicaciones, tales como detección de plagio, sistemas de pregunta-respuesta o traducción automática (Kong *et al.*, 2020).

Se pueden categorizar las paráfrasis teniendo en cuenta el tipo y la profundidad de variantes entre dos textos que comparten un significado (Mota-Montoya *et al.*, 2016), lo que da lugar a dos clases principales. En la denominada *paráfrasis baja* se dan preferentemente cambios de tipo léxico, como la sustitución de una palabra por un sinónimo o hiperónimo, entre otros. La *paráfrasis alta* emplea, además, otro tipo de recursos, como cambios en el orden, estrategias de reescritura sintáctica y modificaciones de tipo discursivo, como el reordenamiento lógico del texto.

Para modelar la detección de paráfrasis de manera automática se necesitan corpus etiquetados sobre los que poder entrenar algoritmos de aprendizaje automático. Entre los más usados por los investigadores se encuentran el Microsoft Research Paraphrase Corpus (MRPC) (Dolan y Brockett, 2005), el Quora Question Pairs (QQP) y el PAWS (Zhang *et al.*, 2019).

El MRPC consta de 5801 pares de oraciones escogidos de sitios web dedicados a las noticias. Los pares se muestran etiquetados con 0 (que se aplica a las no paráfrasis) y 1 (para las paráfrasis). Por otra parte, el QQP contiene casi 800 mil pares de oraciones etiquetados también de manera binaria; éstos fueron extraídos de la conocida red social Quora, la cual tiene por objetivo compartir conocimiento en formato pregunta/respuesta.<sup>7</sup> Por último, PAWS es una colección de 65 000 pares de oraciones etiquetadas de Wikipedia.<sup>8</sup>

<sup>7</sup> <https://quoradata.quora.com>

<sup>8</sup> <https://dumps.wikimedia.org>

Es importante hacer notar que todos los corpus de este tipo se encuentran en inglés. En realidad, con anterioridad al nuestro, no existía un recurso en español que pudiera servir de referencia para el entrenamiento de modelos de aprendizaje automático para la detección de paráfrasis.

Entre las posibles estrategias para elaborar corpus de paráfrasis, una muy utilizada es tomar diferentes traducciones de un texto fuente hacia una misma lengua. Ésta es una práctica aconsejable específicamente para libros que hayan sido bastante versionados, en diferentes épocas y geografías, a la lengua destino. A pesar de sus evidentes ventajas, esta forma de recopilación tiene el problema de que algunas versiones no ofrecen una traducción suficientemente adherida a la literalidad del texto, sino que reformulan muy libremente sus ideas generales.

Como puede advertirse, la Biblia es un recurso muy valioso para llevar a cabo todo tipo de estudios lingüísticos: comparativos, cuantitativos, diacrónicos, etc. Entre los corpus bíblicos existentes en diferentes lenguas, destacan el conformado por la colección anotada de los 66 libros del canon protestante de la Biblia en ocho lenguas, y por los del Nuevo Testamento en doce lenguas, elaborado por Resnik *et al.* (1999); el corpus de traducciones del Nuevo Testamento en lenguas Indoeuropeas, construido por el proyecto PROIEL (Pragmatic Resources in Old Indo-European Languages), con información morfológica, sintáctica y discursiva (Eckhoff *et al.*, 2018); el corpus multilingüe de Mayer and Cysouw (2014), compuesto por 994 traducciones alineadas, las cuales abarcan 76 familias lingüísticas; el corpus alineado de 4272 traducciones de la Biblia en 1611 lenguas (incluidas 27 versiones en inglés), y el corpus multilingüe de Christodouloupoulos and Steedman (2015), elaborado por McCarthy *et al.* (2020), el cual incluye traducciones totales o parciales de la Biblia en cien lenguas, por lo que se trata de una de las fuentes textuales más extensamente usadas en PLN.

No son tan frecuentes los corpus de traducciones de la Biblia a una sola lengua y, sin embargo, hay algunos casos reseñables. En español, se cuenta con el trabajo de Enrique-Arias y Pueyo Mena (2008), quienes prepararon un corpus con 14 traducciones alineadas de la Biblia en español medievales; mientras que en inglés, es posible acceder a las 27 versiones del corpus de McCarthy (2020).

Dicho lo anterior, el corpus alineado de traducciones de la Biblia en español resulta un recurso fundamental para el tratamiento de la paráfrasis. Diversos

aspectos confluyen para ello. En primer lugar, es sabido que los traductores de la Biblia adoptan los principios de la traducción formalmente equivalente, o bien funcionalmente equivalente. A pesar de encontrarse en extremos opuestos de los enfoques de traducción, en ambas perspectivas el traductor intenta respetar el texto original de la manera más estricta, ya que se considera un texto sagrado. Por lo tanto, la libertad del traductor no pone en peligro, en principio, la equivalencia del texto.

En segundo lugar, la división de los textos bíblicos en versículos es una característica que facilita enormemente el trabajo en un gran número de tareas de PLN. A pesar de la gran ventaja que supone esta numeración, la alineación paralela automática que podría derivarse de ella no es perfecta. Por distintas razones, ya sean de carácter pragmático, ideológico o filológico, los versículos no siempre se alinean de forma perfecta en todas las versiones. Pero, en general, éstos suponen una gran ayuda para el investigador. Como ya se ha explicado, en el corpus que presentamos, se ha llevado a cabo un trabajo pormenorizado de revisión de la alineación, lo que permite experimentar a partir de pares o grupos de versículos con identidad semántica.

Computacionalmente, el problema de la detección de paráfrasis no es otro sino el consistente en evaluar la similitud semántica entre dos entradas. Para atacar el problema desde el aprendizaje automático, se necesita un conjunto de datos etiquetado, el preprocesamiento y vectorización de los textos y, finalmente, unos algoritmos de clasificación, en este caso, la arquitectura de una red neuronal siamesa (Reimers y Gurevych, 2019). A continuación, se reporta el proceso de elaboración de un *dataset*<sup>9</sup> para la detección automática de paráfrasis a partir del corpus alineado de biblias y los parámetros generales de configuración de los experimentos realizados con este recurso.

Como se ha mencionado, para determinar si hay o no paráfrasis se necesita una doble entrada de dos textos que puedan ser comparados. No es necesario que el algoritmo defina si uno de ellos es la fuente y el otro la copia, sino, únicamente, si ambas entradas son equivalentes o no. Como el corpus con el que trabajamos está completamente alineado, la elaboración del *dataset* es sencilla. En este artículo, se toman como textos de inicio dos libros, 1 Corintios y Génesis, y dos traducciones, la Biblia Latinoamericana (LAT) y la Biblia de

<sup>9</sup> Conjunto de datos específicamente preparado para una tarea concreta de PLN.

las Américas (AME). El *dataset* consta de pares de versículos de 1 Corintios y pares de versículos de Génesis. Primero, se empareja cada uno de los versículos de un libro (por ejemplo, 1 Corintios) en una traducción (AME) con todos los versículos de la otra traducción (LAT).

En el conjunto de datos se marcan sólo como paráfrasis aquellos versículos con la misma etiqueta (excepto el código que identifica cada biblia). El resto, no son considerados paráfrasis. Así, en el ejemplo (2), se obtiene el par LAT1CO001001-AME1CO001001, que se etiqueta automáticamente con el valor positivo para paráfrasis (es decir, con un 1) debido a la coincidencia en el código del versículo:

- (2) a. De Pablo, apóstol Cristo Jesús por decisión Dios que lo ha llamado, y Sóstenes nuestro hermano (LAT1CO001001).
- b. Pablo, llamado a ser apóstol de Jesucristo por la voluntad de Dios, y Sóstenes, nuestro hermano (AME1CO001001).

También se forman pares con etiquetas no coincidentes, que se marcan como no-paráfrasis (con valor 0), como sucede en (3) con (a) y (b) (véase el último dígito de las etiquetas LAT1CO001001 y AME1CO001002).

- (3) a. De Pablo, apóstol Cristo Jesús por decisión Dios que lo ha llamado, y Sóstenes nuestro hermano (LAT1CO001001).
- b. A la iglesia de Dios que está en Corinto, a los que han sido santificados en Cristo Jesús, llamados a ser santos, con todos los que en cualquier parte invocan el nombre de nuestro Señor Jesucristo, Señor de ellos y nuestro (AME1CO001002).

Al comparar el contenido de (3a) y (3b), resulta evidente que dichos textos no son paráfrasis entre ellos. Para evitar la profusión de pares anotados como no paráfrasis que, en realidad, están muy alejados entre sí, se implementó la siguiente estrategia: Se decidió que el 25 % de los pares del *dataset* serían paráfrasis, y se considerarían todos los posibles pares. Así, a partir de los datos etiquetados con el valor 1, se calculó el número de pares etiquetados con 0 para cada libro. Para elegir cuáles de los pares 0 entrarían a formar parte del

conjunto de datos se realizó la prueba de la distancia coseno. Los pares con el coeficiente más alto se incluyeron en el *dataset*, cuya configuración se puede observar en la tabla 1.

TABLA 1. NÚMERO DE PARES DE CADA SUBCONJUNTO EN RELACIÓN CON EL CONJUNTO DE DATOS PARA EL EXPERIMENTO

<b>Libro</b>	<b>Entrenamiento (64 %)</b>	<b>Validación (16 %)</b>	<b>Prueba (20 %)</b>
<b>Génesis</b>	3921	981	1266
<b>1 Corintios</b>	1118	280	350

Una vez elaborado el conjunto de datos con el que se realizarían los experimentos, éste se dividió en tres partes: entrenamiento (64 %), validación (16 %) y prueba (20 %). Con los pares de entrenamiento, se alimenta el sistema de aprendizaje automático. Con los pares de validación, se comprueba si el algoritmo ha generalizado. Finalmente, el conjunto de prueba sirve para evaluar la eficiencia del proceso.

El siguiente paso consistió en la vectorización de las entradas textuales y en la extracción de características. Se usaron dos tipos de métricas para evaluar la distancia entre las dos oraciones de entrada, las *conjuntistas* (Dice, Jaccard, Jaro-Winkler, Levenshtein y coseno), y las *vectoriales* (diferencia de longitud, n-gramas, tf-idf y WMD). Los experimentos usaron diferentes combinaciones de estas métricas con el fin de buscar la mejor configuración.

Por lo que respecta a la implementación neuronal, se optó por una red siamesa (Reimers y Gurevych, 2019), con dos entradas y dos ramas iniciales de procesamiento independientes. La red siamesa se adecua muy bien a la tarea que consiste en procesar dos entradas de manera independiente para después compararlas. En este caso, la red consta de una capa de entrada, una capa de vectorización y una capa LSTM (Long Short Term Memory), esto es, un tipo de redes neuronales recurrentes. Una vez la entrada es procesada por la red, se comparan los vectores resultantes mediante la distancia euclidiana.

Para la representación matemática de los textos en la capa de vectorización se usó un modelo BERT (Bidirectional Encoder Representations from Transformers) (Devlin *et al.*, 2018).

Los resultados obtenidos están alrededor del 0.95 de F1-score, tanto si se usan modelos de aprendizaje automático clásicos como si se emplean modelos

basados en *transformers*. Como la medida del F1-score está entre 0 y 1, los resultados son altamente satisfactorios.

## APLICACIÓN PARA AGRUPACIÓN SEMÁNTICA

El objetivo principal del agrupamiento semántico es identificar pares de palabras que puedan ser utilizadas de manera intercambiable en un cierto contexto, de manera que las palabras de un grupo están fuertemente asociadas, ya que comparten características semánticas, mientras que palabras de otro grupo mostrarán características diferentes.

Contar con grupos de palabras semánticamente similares resulta de gran utilidad para las tecnologías del lenguaje debido a su relevancia en varias aplicaciones, como la detección de plagio, la respuesta a preguntas, el entendimiento textual, el resumen y la evaluación de los sistemas de traducción automática. Un método para encontrar grupos semánticos a partir del alineamiento de definiciones fue desarrollado por Sierra y McNaught (2000), específicamente diseñado para identificar pares de palabras donde uno de sus miembros puede ser reemplazado por el otro en el contexto de una definición sin alterar el concepto subyacente.

Este método, mediante la alineación de definiciones que expresan el mismo concepto, pero a través de diferentes formulaciones, logra determinar las correspondencias más probables entre las palabras en la oración de destino y las palabras en la oración de origen. De manera análoga, gracias a la disponibilidad de nuestro corpus paralelo de biblias en español, es posible alinear los versículos de las diferentes versiones para identificar equivalencias de palabras dentro de los versículos dados. Véase, por ejemplo, en (4), el versículo de Marcos 5: 34 para las biblias ESP (a), LAT (b), JUN (c) y NAC (d).

- (4) a. Él le dijo: Hija, tu fe te ha curado. Vete en paz y sigue sana de tu tormento (ESPMAR005034).
- b. Jesús le dijo: “Hija, tu fe te ha salvado; vete en paz y queda sana de tu enfermedad” (LATMAR005034).
- c. Y él díjola: “Hija, tu fe te ha salvado, ve a paz, sé sana de tu plaga” (JUNMAR005034).

- d. Y Él le dijo: Hija, tu fe te ha salvado; vete en paz y seas curada de tu mal (NACMAR005034).

Al observar las oraciones, se pueden identificar varios pares y agrupamientos de palabras, a saber: *él, Jesús; curado, salvado; sigue, queda; sana, curada, y tormento, enfermedad, plaga, mal*. Estos grupos exhiben correspondencia semántica, lo que indica que las palabras pueden sustituirse entre sí sin alterar significativamente el significado.

Una ventaja del algoritmo de alineación es que no depende de grandes cantidades de datos y puede generar agrupamientos incluso cuando la alineación de palabras es infrecuente, ya que compara las palabras en dos versículos de manera secuencial, estableciendo correspondencias entre ellas. Básicamente, identifica el costo mínimo asociado con cada operación requerida para transformar una frase en otra, incluidas las sustituciones de palabras, inserciones y eliminaciones, lo que se conoce como distancia de Levenshtein (Levenshtein, 1966). Conforme aumenta el número de palabras iguales circundantes a un par de palabras diferentes, mayor será la similitud semántica del par en cuestión, lo que coincide con la hipótesis distribucional de Harris (1954), que postula que las palabras con un uso contextual similar tienden a compartir significados similares.

Para el ejemplo de Marcos 5: 34, en el alineamiento (5) formado por las biblias ESP y LAT, podemos ver a continuación que los pares de palabras diferentes de (a) y (b), que denominamos *pares correspondientes*, tienen varias palabras alrededor que son iguales en (a) y (b).

- (5) a. Él le dijo Hija tu fe te ha curado Vete<sup>10</sup> en paz y sigue sana de tu tormento (ESPMAR005034).

- b. Jesús le dijo Hija tu fe te ha salvado vete en paz y queda sana de tu enfermedad (LATMAR005034).

<sup>10</sup> En estos ejemplos, se mantiene el uso de mayúsculas acorde con la fuente de origen, pero se han omitido los signos de puntuación.



Véase que el par correspondiente *curado-salvado* tiene 7 palabras iguales a la izquierda y 4 a la derecha, por lo que puede considerarse que su ventana de similitud es de 11 palabras. Por su parte, los pares *Él-Jesús* y *sigue-queda* tienen una ventana de 7 palabras, mientras que *tormento-enfermedad* tiene una ventana de similitud de 3.

A continuación, se muestra una selección de grupos semánticos derivados de haber realizado la tarea de agrupar o unir los pares coincidentes identificados en el libro de Marcos a lo largo de las biblias ESP, JNM, NAC y LAT. Estos agrupamientos demuestran las asociaciones entre palabras que exhiben equivalencia semántica dentro de los textos analizados. Y son los siguientes:

- Mente, inteligencia, entendimiento
- Departamento, pieza
- Diestro, derecho
- Decir, preguntar, responder, contestar, asegurar, contar
- Mañana, madrugada
- Impuro, malo, inmundo
- Enojar, indignar
- Ir, venir, volver, regresar, llegar, bajar, pasar, acercar
- Permitir, dejar
- Luego, después
- Lago, orilla, mar
- Germinar, brotar
- Misericordia, compasión
- Tierra, suelo
- Hablar, conversar
- Príncipe, jefe, oficial
- Predicar, anunciar
- Hombre, gente, nación, gentío, pueblo, aldea, multitud

Se observa que ciertos grupos semánticos coinciden con vocablos que pudieran encontrarse relacionados en cualquier diccionario de sinónimos. Por ejemplo, sustantivos como *mente*, *inteligencia* y *entendimiento*, así como verbos como *enojar* e *indignar*. En otros casos, los grupos semánticos consisten en *sinónimos indirectos*, como en el caso de *departamento* y *pieza*, pues si bien no son sinónimos, ambos comparten *habitación* como sinónimo.

Además, se presentan casos en los que los agrupamientos semánticos comprenden cohipónimos, como *príncipe*, *jefe* y *oficial* para sustantivos, o implicación léxica, como *decir*, *preguntar*, *responder*, *contestar*, *asegurar* y *contar* para verbos. Estas agrupaciones dan cuenta de relaciones de inclusión o implicación, donde una palabra abarca o implica el significado de otra dentro del mismo grupo.

Sonia Morett (2006) realizó un estudio sintáctico de un corpus de pares de definiciones en el área de física para extender el método de alineamiento con el fin de ir más allá de la alineación consecutiva de palabra por palabra y mejorar el rendimiento de obtención de grupos semánticos. A partir de su estudio de las relaciones y funciones de los constituyentes de la oración, propuso realizar un preprocesamiento del corpus para considerar, entre otros, perífrasis gramaticales (locuciones, términos compuestos y perífrasis verbales), las palabras que funcionan como nexos (conjunciones, preposiciones y verbos copulativos), las relaciones léxicas (sinonimia, hiponimia e hiperonimia) y la negación. Un ejemplo de locución y relación semántica se tiene en el alineamiento (6), de Marcos 1: 13, entre ESP (a) y LAT (b):

- (6) a. Estuvo en el desierto cuarenta días: Satanás lo ponía a prueba, estaba con las fieras y los ángeles le servían (ESPMAR001013).
- b. Estuvo cuarenta días en el desierto y fue tentado por Satanás. Vivía entre los animales salvajes y los ángeles le servían (LATMAR001013).

Por su parte, Natalia Guerrero (2022) observó otra serie de fenómenos lingüísticos que ocurren entre pares de versículos y que se detectan con el algoritmo de agrupación semántica, tales como la derivación y la composición, la sustitución léxica (hiperónimos e hipónimos, merónimos y holónimos, antónimos). Algunos de estos fenómenos son particularmente frecuentes en algunas versiones de la Biblia. Un ejemplo es la enclisis en el siguiente pasaje (7) de Marcos 1: 17, muy común en JNM.

- (7) a. Jesús les dijo: “Síguenme y yo los haré pescadores de hombres” (LATMAR001017).

- b. Y díjoles Jesús: “Venid en pos de mí, y haréos ser pescadores de hombres” (JNMMAR001017).

## APLICACIÓN PARA EVALUACIÓN DE LLM PARA PREGUNTA-RESPUESTA

Los modelos grandes de lenguaje (LLMs, por sus siglas en inglés) han transformado el acceso y procesamiento de la información, especialmente en sistemas de pregunta/respuesta. Estos modelos entienden y responden consultas en lenguaje natural, lo que facilita el diseño de interfaces intuitivas y accesibles para los usuarios. Su capacidad para analizar el contexto y producir respuestas relevantes y coherentes es crucial en todas las áreas de conocimiento.

Los LLM, entrenados con grandes conjuntos de datos, poseen la capacidad de responder consultas específicas extrayendo y sintetizando conocimientos de sus bases de datos internas. Sin embargo, surge un desafío significativo cuando la información requerida es restringida, es decir, contiene detalles específicos que podrían no estar en su base de datos de entrenamiento, o bien, diferir, en algunos aspectos, de la almacenada. Por ejemplo, en el caso de manuales técnicos, las regulaciones de construcción que presentan diferentes estipulaciones que varían de un país a otro; o cuando se trata de políticas y procedimientos de empresas, los cuales pueden diferir sustancialmente de las regulaciones generales.

En este sentido, el desafío radica en la capacidad de estos modelos para interpretar y aplicar su conocimiento a casos que, aunque superficialmente se parezcan a situaciones familiares, difieren en aspectos críticos que pueden afectar la precisión de las respuestas generadas. La capacidad de ajustarse a estas diferencias sutiles pero fundamentales es lo que pone a prueba la versatilidad y efectividad de los LLMs en aplicaciones prácticas.

Considerando lo anterior, en esta sección se reportan distintos experimentos realizados con el fin de evaluar la adecuación de los LLMs para inferir respuestas basadas exclusivamente en documentos restringidos, limitando el uso de información previamente almacenada en sus bases de conocimiento. En estas pruebas se ha intentado encapsular el funcionamiento de los modelos de lenguaje, aislándolos de su conocimiento anterior. Los LLMs usados en

nuestros experimentos fueron Llama2 (Chowdhery *et al.*, 2022), ChatGPT (Brown *et al.*, 2020) y PaLM (Touvron *et al.*, 2023).

Como estudio pivote, se seleccionó el dominio bíblico como tema central, y se determinó que las respuestas a las preguntas debían basarse no en el conocimiento general, sino estrictamente en capítulos específicos de la Biblia; en concreto, de la versión LAT de nuestro corpus. El enfoque del experimento consiste en probar la eficiencia de los modelos de lenguaje utilizados para esta tarea, no en la interpretación religiosa de los textos. Estas preguntas requieren que los LLMs consideren la Biblia como una fuente de información, no como directrices religiosas o como información que contradice, o no, el conocimiento preentrenado del modelo. Los experimentos tienen como objetivo evaluar si los LLMs pueden:

- a) limitarse a la información y a los *prompts*<sup>11</sup> proporcionados;
- b) ignorar el contexto y corregir la respuesta usando su conocimiento preentrenado, y
- c) verificar si la información proporcionada afecta la exactitud de la respuesta.

Para tal efecto, en primera instancia, se escribieron una serie de preguntas sobre diversos pasajes de la Biblia, tanto del Antiguo como del Nuevo Testamento. En total, se seleccionaron 36 preguntas para la versión final del experimento. La mayoría de ellas referidas a eventos, nombres, momentos y lugares específicos. Dividimos estas cuestiones en dos grandes grupos: A) aquellas preguntas que pueden ser contestadas correctamente con la información existente en la Biblia, y B) aquellas para las que el texto fuente no proporciona información suficiente que permita llegar a una solución válida.

El primer tipo de preguntas, que corresponde a la mayor parte de nuestro cuestionario, engloba 31 *inputs* (tabla 2). En general, la respuesta se puede localizar en la información contenida en un capítulo, como sucede con las consultas 1 y 3, que corresponden a Éxodo 14 y 1 Samuel 17. En muy pocos casos, la respuesta se llega a encontrar en un solo versículo. Un ejemplo de esto último corresponde a la pregunta 10, cuya solución se localiza en Daniel 1:

<sup>11</sup> Los *prompts* son informaciones o consultas simples proporcionadas a un LLMs para obtener una respuesta.

1. Con el fin de probar el razonamiento de información factual de un LLM, algunas preguntas requieren de interpretación, como la pregunta 16.

**TABLA 2. PREGUNTAS DEL GRUPO A, QUE PUEDEN SER RESPONDIDAS DIRECTAMENTE CON INFORMACIÓN DE LA BIBLIA**

<b>#</b>	<b>Pregunta</b>	<b>Respuesta</b>
1	¿Qué mar fue abierto por Moisés y para qué?	El Mar Rojo
2	¿Qué ídolo erróneamente veneran los israelitas y por qué?	Un becerro de oro
3	¿Cómo mató David a Goliat?	Con una piedra de su honda
4	¿Quién tuvo el sueño de las vacas gordas y las vacas flacas?	El Faraón de Egipto
5	¿A quién le dijo Rut las palabras: donde tú vayas, iré yo, y donde tú vivas, viviré yo; tu pueblo será mi pueblo y tu Dios será mi Dios?	A su suegra Noemí
6	¿Cómo se llamaba el jefe del ejército a quien derrotaron los israelitas bajo el mando de la jueza Débora?	Sísara
7	¿Quién mató a Holofernes?	Judit
8	¿Quién despojó de su cabellera a Sansón y por qué lo hizo?	Su esposa, Dalila, para que perdiera su fuerza
9	¿Qué comían los israelitas en el desierto?	Maná
10	¿Quién era Nabucodonosor?	El rey de Babilonia
11	¿Qué oficio tenía Melquisedec?	Sacerdote
12	¿Qué construyó Noé?	Un arca
13	¿De dónde era Ciro?	Ciro era de Persia
14	¿Quién tentó a Jesús en el desierto?	Satanás
15	¿En qué monte fue crucificado Jesús?	En Gólgota, también llamado Calvario
16	¿Qué pareja acompañó a Pablo en algunos de sus viajes?	Aquila y Priscila
17	¿Quién se quedó sin oreja la noche que murió Jesús?	Malco

TABLA 2. (CONT.)

#	Pregunta	Respuesta
18	¿Quién estaba siendo juzgado junto con Jesús por los romanos?	Barrabás
19	¿Qué le hizo Juan Bautista a Jesús?	Lo bautizó
20	¿Por cuántas monedas Judas traiciona a Jesús?	Treinta piezas de plata
21	¿Quién hizo que decapitaran a Juan Bautista?	La hija de Herodías
22	¿Qué amigo le escribe dos cartas a Timoteo?	Pablo de Tarso
23	¿Quién niega a Jesús? ¿Cuántas veces?	Pedro. Tres
24	¿A qué hora murió Jesús?	A las tres de la tarde
25	¿En qué ciudad hizo Pablo de Tarso su discurso "Al Dios no conocido"?	Atenas
26	¿En qué fueron grabados los diez mandamientos y cuáles son esos?	Fueron dados en dos tablas de piedra y son: 1. Amarás a Dios sobre todas las cosas. Sólo hay un Dios, creador y todopoderoso, al cual adorar. 2. No tomarás el nombre de Dios en vano. 3. Santificarás las fiestas. 4. Honrarás a tu padre y a tu madre. 5. No matarás. 6. No cometerás actos impuros. 7. No robarás. 8. No darás falso testimonio ni mentirás. 9. No consentirás pensamientos ni deseos impuros. 10. No codiciarás los bienes ajenos.
27	¿Cuántos hijos tuvo Jacob? ¿Cómo se llamaban?	Doce. Rubén, Simeón, Leví, Judá, Isacar, Zabulón, Dan, José, Benjamín, Neftalí, Gad, Aser
28	¿Cuál era el oficio de Mateo antes de unirse a los seguidores de Jesús? ¿Y de Pedro?	Cobrador de impuestos (publicano). Pescador
29	¿Cuántos candeleros hay en Apocalipsis y a qué se refiere?	Siete. A las siete iglesias
30	¿A quién se tragó el gran pez?	A Jonás
31	¿A quién le fue revelado el libro del Apocalipsis?	A Juan

El grupo B incluye solamente 5 preguntas (tabla 3). Para contestarlas se requiere más información que la proporcionada por el contexto, pues –en este caso– se precisa de interpretaciones y opiniones formuladas por académicos o autoridades eclesiásticas. Por tanto, se espera que los LLMs no puedan proporcionar respuestas, ya que éstas no se encuentran explícitamente en los libros de la Biblia.

En algunas ocasiones, el modelo ha sido capaz de contestar estas cuestiones, y esto prueba cómo hace uso de su conocimiento previo; se trata, por tanto, de una malfunción del sistema. En algunas ocasiones, en cambio, se podría defender que un sistema inferencial suficientemente potente sería capaz de deducir la respuesta. Por ejemplo, para la pregunta 33, el sistema podría llegar a la conclusión de que David es el autor de los Salmos, debido a las constantes referencias a este rey que se ponen en boca de Dios en los textos cuando el autor y el creador están en diálogo, como en los versículos de (8).

- (8) a. Por mi propia santidad lo juré una vez, ¿cómo podría yo mentir a David? (SAL089035)
- b. Señor, ¿dónde están tus favores pasados que juraste a David por tu fidelidad? (SAL089049)
- c. Señor, tenle en cuenta a David su fervoroso celo (SAL132001)

**TABLA 3. PREGUNTAS DEL GRUPO B, QUE REQUIEREN MÁS INFORMACIÓN QUE LA PROPORCIONADA POR EL TEXTO PARA SER RESUELTAS**

#	Pregunta	Respuesta
32	¿Qué libro de la Biblia cuenta el amor de los esposos?	El Cantar de los Cantares
33	¿Quién es considerado el autor de los Salmos?	El rey David
34	¿Qué son Isaías, Jeremías, Ezequiel y Daniel?	Profetas: los profetas mayores
35	¿Qué profeta escribió el libro de Lamentaciones?	Jeremías
36	¿Cuál era el más escéptico de los discípulos de Jesús?	Tomás

Como se mencionó anteriormente, el cuestionario está diseñado para probar las capacidades de cada modelo más allá de extraer respuestas explícitas del texto. El proceso para llegar a la extracción tiene dos pasos: el primero consiste en la *recuperación de contextos*, y el segundo, en la *extracción de la respuesta correcta de cada modelo*. A continuación, se describen ambos procesos.

Para llevar a cabo la recuperación de contextos, se seleccionó un enfoque de *generación aumentada por recuperación* (RAG) (Lewis *et al.*, 2021). La metodología RAG es ampliamente usada como método para mejorar las respuestas de los LLM a una pregunta dada. Su funcionamiento consiste en expandir el conocimiento de un modelo de lenguaje mediante el análisis de una pregunta determinada, alimentando así el LLM con información relacionada, la cual se extrae de los datos sobre los que actúa.

Para garantizar su consistencia y eliminar el ruido que puede derivarse del proceso de recuperación de cada modelo, la parte de recuperación se realizó mediante *embeddings*<sup>12</sup> estandarizados. Las preguntas y los capítulos se representaron utilizando un modelo de *embeddings* consistente para todos los LLMs, seguido de una recuperación por similitud. Los textos recuperados fueron utilizados posteriormente por cada LLM para responder las preguntas correspondientes. A continuación, se implementó el modelo para la extracción de las respuestas. Para medir la incidencia del RAG en el resultado, se hicieron dos experimentos diferentes. Primero, se formularon las preguntas sin proporcionar contexto adicional, confiando únicamente en los datos de entrenamiento de los LLMs. Esto sirvió como una línea base para comparar con el sistema RAG completo.

La segunda evaluación incluyó los pares de texto y pregunta recuperados, utilizando *prompts* diseñados para instruir al modelo a responder basándose únicamente en el contexto proporcionado. Cuando éste resultaba insuficiente, se instruyó al modelo a indicarlo con un comentario como “información no disponible”. La tabla 4 muestra los *prompts* utilizados para cada modelo LLM con RAG.

Cada par conformado por pregunta y contexto se ingresó en el modelo de manera iterativa, generando una sola respuesta por iteración. Las respuestas, junto con las preguntas y los contextos correspondientes, se recopilaron para

<sup>12</sup> Vectores numéricos que representan palabras y que capturan información léxica.



la evaluación. Los resultados se representan como tuplas de tres elementos: **pregunta**, **contexto** y **respuesta**.

Teniendo en cuenta los dos tipos de evaluaciones, con contexto y sin contexto, y los tres LLM utilizados (LLAMA, ChatGPT y PaLM), es posible analizar los resultados que se generan para cada uno de los grupos de preguntas: grupo A y grupo B. Los resultados obtenidos se muestran en la tabla 5, donde FI (**falsa inferencia**) significa que el LLM ha recurrido a su conocimiento externo, y 0 significa que no hubo respuesta.

TABLA 4. PROMPTS UTILIZADOS PARA CADA MODELO

PaLM	LLaMA	GPT
Responde la pregunta dado el siguiente contexto {{context:}}. \n Si la respuesta\ \ no está dentro del {{context:}} y no estás seguro, por favor responde "Información no disponible".	Dado el siguiente contexto: '{retrieved_info}', responde: '{question}'. \n \n Si la respuesta no está dentro de la información proporcionada y no estás seguro de la respuesta, por favor responde "Información no disponible en el contexto".	Si la respuesta no está dentro del {{context:}} y no estás seguro, por favor responde "Información no disponible".

TABLA 5. RESULTADOS PARA LAS 31 PREGUNTAS DEL GRUPO A

	LLAMA CONTEXTO	LLAMA NO CONTEXTO	CHATGPT CONTEXTO	CHATGPT NO CONTEXTO	PALM CONTEXTO	PALM NO CONTEXTO
<b>Correctas</b>	19	15	17	17	19	23
<b>Incorrectas</b>	8	16	3	6	6	2
<b>FI</b>	4	-	1	-	1	-
<b>0</b>	-	-	10	8	5	6

En el caso de las consultas con contexto, para que una respuesta se considere correcta, no sólo es necesario que sea acertada, sino que haya sido adecuadamente inferida a partir del RAG. En muchas ocasiones el RAG no arroja un contexto de donde se pueda extraer la solución. Si se falla en este primer paso, y el algoritmo funciona de forma adecuada, la solución coherente es la de que no existen suficientes datos (0). En algunas ocasiones, en cambio, la respuesta es la que se esperaba, aunque no se haya inferido del RAG. Esto

quiere decir que el LLM ha acudido a su conocimiento más allá del texto fuente. Para los propósitos de este experimento, se considera una solución incorrecta, que se marca como FI.

La tabla 6 muestra el ejemplo de la pregunta 18: “¿Quién estaba siendo juzgado junto con Jesús por los romanos?” En este caso, el RAG ofrece todo el capítulo 15 de Marcos, que incluye la información que se buscaba. Por tanto, el contexto es óptimo para que los tres modelos puedan encontrar una respuesta satisfactoria. En cambio, ni LLAMA ni ChatGPT lo hacen, ya que confunden el pasaje del juicio por parte de Pilatos con el de la crucifixión. En cambio, PaLM RAG encuentra la información solicitada sin problemas.

TABLA 6. RAG Y CONTESTACIONES DE LOS MODELOS PARA LA PREGUNTA 18: “¿QUIÉN ESTABA SIENDO JUZGADO JUNTO CON JESÚS POR LOS ROMANOS?”

RAG	LLAMA RAG	ChatGPT RAG	PaLM RAG
MARCOS 15	En cuanto a tu pregunta, según el texto que me has proporcionado, Jesús fue juzgado junto con otros dos ladrones, uno a su derecha y otro a su izquierda, por los romanos. El texto especifica que los tres fueron crucificados juntos en un lugar llamado Gólgota o Calvario	Información no disponible	Barrabás
<b>Correcto</b>	<b>Incorrecto</b>	<b>Incorrecto</b>	<b>Correcto</b>

La tabla 7 muestra los resultados con RAG para la pregunta 5: “¿A quién le dijo Rut las palabras: donde tú vayas, iré yo; y donde tú vivas, viviré yo; tu pueblo será mi pueblo y tu Dios será mi Dios?” En esta ocasión, el RAG obtenido corresponde a un texto completamente alejado de la pregunta. En cambio, LLAMA es capaz de encontrar la respuesta correcta: *Noemí*, a pesar de que la identifica como madre y no como suegra; aunque esto no puede desprenderse de la información del RAG. Por tanto, la etiqueta asignada es FI (**falsa inferencia**). Aquí se observa que el modelo no ha funcionado porque claramente ha salido de su dominio. En cuanto a ChatGPT y PaLM, dicen no tener información disponible, por lo que se les adjudica la etiqueta 0; en cambio, esto es coherente con lo que reporta el RAG.

TABLA 7. RAG Y CONTESTACIONES DE LOS MODELOS PARA LA PREGUNTA 5: "¿A QUIÉN LE DIJO RUT LAS PALABRAS: DONDE TÚ VAYAS, IRÉ YO; Y DONDE TÚ VIVAS, VIVIRÉ YO; TU PUEBLO SERÁ MI PUEBLO Y TU DIOS SERÁ MI DIOS?"

RAG	LLAMA RAG	ChatGPT RAG	PaLM RAG
Lev 18	La persona a quien Rut le dijo las palabras "donde tú vayas, iré yo; y donde tú vivas, viviré yo; tu pueblo será mi pueblo y tu Dios será mi Dios" fue Noemí, su madre	Información no disponible	No hay información disponible en el contexto
<b>Incorrecto</b>	<b>F1</b>	<b>0</b>	<b>0</b>

Por lo que respecta a las preguntas del grupo B, no se espera que el sistema sea capaz de inferir las contestaciones, ya que no se desprenden directamente del texto. La tabla 8 muestra los resultados que indican que hubo claramente un mejor rendimiento en aquellos experimentos donde la respuesta no dependía del RAG sobre aquellos para los que sí se requería, donde tanto LLAMA y PaLM, con dos respuestas correctas, como ChatGPT, que no consiguió responder ninguna, tienen un desempeño muy pobre.

TABLA 8. RESULTADOS PARA LAS 5 PREGUNTAS DEL GRUPO B

	LLAMA contexto	LLAMA no contexto	ChatGPT contexto	ChatGPT no contexto	PaLM contexto	PaLM no contexto
<b>Correctas</b>	2	4	-	4	2	5
<b>Incorrectas</b>	2	1	1	-	3	-
<b>F1</b>	1	-	-	-	-	-
<b>0</b>	-	-	4	1	-	-

Una observación clave que se desprende de este estudio es la demostración de la capacidad de estos modelos para responder preguntas basadas en un contexto dado. La variabilidad en las respuestas correctas entre los modelos *con* y *sin* contexto depende de factores como la correcta recuperación

de información, el tamaño del modelo y la disponibilidad de datos. Cabe destacar que no podemos garantizar la ausencia de la Biblia en los datos de preentrenamiento de ninguno de los tres modelos.

Esta metodología facilita la adaptación al dominio de manera sencilla y, con constantes actualizaciones de los modelos, se puede utilizar sin necesidad de un ajuste fino.

Es importante mencionar la relevancia de contar con modelos de código abierto, ya que –de otra forma– el costo computacional y económico de generar modelos de este tipo es accesible sólo para unas pocas de las empresas más grandes. Con la metodología RAG, podemos aprovechar estos sistemas para evitar la necesidad de entrenar un modelo de este tamaño.

## REFLEXIONES FINALES

La elaboración de corpus, conjuntos de datos y recursos de diferentes tipo, es un elemento fundamental y previo para la lingüística de corpus, el PLN y, en definitiva, para las tecnologías del lenguaje que, finalmente, nutren los sistemas de inteligencia artificial, que en los últimos años han venido para quedarse en la sociedad, las relaciones, las formas de comunicación y la vida cotidiana.







La mayor parte de los corpus de investigación se encuentran en inglés, y cada vez resulta más difícil la implementación de tecnologías del lenguaje en otras lenguas. El español, a pesar de ser una lengua extendida, apenas puede competir con los avances que se encuentran basados en las estructuras, el léxico y la cultura inglesa.

El corpus alineado de traducciones de la Biblia en español, elaborado por el Grupo de Ingeniería Lingüística (GIL) del Instituto de Ingeniería de la UNAM, es un recurso pensado para posibilitar el avance de la lingüística computacional en español. El corpus presenta un mismo texto en once variantes cercanas semánticamente equivalentes. La elaboración de este corpus ha sido un trabajo colaborativo de los investigadores del GIL con estudiantes, alumnos de servicio social y becarios de investigación.





A partir de este corpus, se pueden llevar a cabo estudios realizados por becarios de investigación, con fondos del proyecto CF-2023-G de CONAHCyT y en diferentes áreas de la lingüística y del PLN, como similitud,

extracción léxica, detección de paráfrasis, estudio de sesgo, cambio lingüístico y dialectología. En este trabajo, se han presentado, en líneas generales, tres aplicaciones: agrupación semántica, evaluación de sistemas de pregunta-respuesta y detección de paráfrasis. El corpus se ofrece pública y gratuitamente para que se puedan evaluar en él diferentes algoritmos para todo tipo de tareas de PLN y, de esta manera, impulsar el desarrollo de tecnologías del lenguaje en español.







## BIBLIOGRAFÍA

- Brown, Tom B. *et al.* (2020), “Language Models are Few-Shot Learners”, pp. 1-75, consultado el 24 de agosto de 2023. 
- Chowdhery, Aakanksha *et al.* (2022), “PaLM: Scaling Language Modeling with Pathways”, pp. 1-87, consultado el 24 de agosto de 2024. 
- Christodouloupoulos, Christos y Mark Steedman (2015), “A massively parallel corpus: the bible in 100 languages”, *Language Resources and Evaluation*, vol. 49, núm. 2, pp. 375–395, consultado el 24 de agosto de 2024. 
- Das, Dipanjan y Noah A. Smith (2009), “Paraphrase identification as probabilistic quasi-synchronous recognition”, *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pp. 468–476, consultado el 24 de agosto de 2024. 
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee y Kristina Toutanova (2019), “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171–4186, consultado el 24 de agosto de 2024. 
- Dolan, William B. y Chris Brockett (2005), “Automatically Constructing a Corpus of Sentential Paraphrases”, *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)*, pp. 9-16, consultado el 24 de agosto de 2024. 
- Eckhoff, Hanne M. *et al.* (2018), “The PROIEL treebank family: a standard for early attestations of Indo-European languages”, *Language Resources & Evaluation*, vol. 52, pp. 29-65.
- Guerrero Cerón, Natalia (2023), Análisis de paráfrasis en traducciones de la Biblia en español, tesis de licenciatura en Lengua y Literaturas Hispánicas, México,

- Facultad de Estudios Superiores Acatlán, Universidad Nacional Autónoma de México, consultado el 24 de agosto de 2024. 
- Harris, Zellig S. (1954), “Distributional structure”, *Word*, vol. 10, núm. 2-3, pp. 146-162, consultado el 24 de agosto de 2024. 
- Klein, William W., Craig B. Blomert y Robert L. Hubbard Jr. (2017), *Introduction to Biblical Interpretation* (3ª. ed.), California, Zondervan.
- Kong, Leilei, Zhongyuan Han, Yong Han y Haoliang Qi (2020), “A deep paraphrase identification model interacting semantics with syntax”, *Complexity*, vol. 2020, pp. 1-14, consultado el 24 de agosto de 2024. 
- Levenshtein, Vladimir I. (1966), “Binary codes capable of correcting deletions, insertions, and reversals”, *Soviet physics doklady*, vol. 10, núm. 8, pp. 707–710.
- Lewis, Patrick *et al.* (2021), “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”, consultado el 24 de agosto de 2024. 
- Mayer, Thomas y Michael Cysouw (2014), “Creating a massively parallel Bible corpus”, *International Conference on Language Resources and Evaluation (LREC'14)*, pp. 3158–3163, consultado el 24 de agosto de 2024. 
- McCarthy, Arya D. *et al.* (2020), “The Johns Hopkins University Bible corpus: 1600+ tongues for typological exploration”, *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 2884-2892, consultado el 24 de agosto de 2024. 
- Morett Álvarez, Sonia Elisa (2006), Inclusión de una perspectiva sintáctica a la obtención de pares semánticos, tesis de licenciatura en Lengua y Literaturas Hispánicas, México, Universidad Nacional Autónoma de México, consultada el 24 de agosto de 2024. 
- Mota Montoya, Margarita A., Iria Da Cunha y Fernanda López-Escobedo (2016), “Un corpus de paráfrasis en español: metodología, elaboración y análisis”, *RLA. Revista de lingüística teórica y aplicada*, vol. 54, núm. 2, pp. 85-112, consultado el 24 de agosto de 2024. 
- Reimers, Nils e Iryna Gurevych (2019), “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks”, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 3982–3992, consultado el 24 de agosto de 2024. 
- Resnik, Philip, Mari Broman Olsen y Mona Diab (1999), “The Bible as a parallel corpus: Annotating the ‘book of 2000 tongues’”, *Language Resources and Evaluation*, vol. 33, pp.129-153.

- Sierra, Gerardo y John McNaught (2000), “Extracting semantic clusters from MRDs for an onomasiological search dictionary”, *International Journal of Lexicography*, vol. 13, núm. 4, pp. 264-286, consultado el 24 de agosto de 2024. 
- Sierra, Gerardo *et al.* (2024), “An aligned corpus of Spanish bibles”, *Language Resources & Evaluation*, consultado el 24 de agosto de 2024. 
- Touvron, Hugo *et al.* (2023), “LLaMA: Open and Efficient Foundation Language Models”, pp. 1-27, consultado el 24 de agosto de 2024. 
- Zhang, Yuan, Jason Baldridge y Luheng He (2019), “PAWS: Paraphrase Adversaries from Word Scrambling”, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 1298–1308, consultado el 24 de agosto de 2024. 

## Corpus y repositorios en línea

- Enrique-Arias, Andrés y Francisco Javier Pueyo Mena (2008), “Biblia medieval”, *Corpus Biblia Medieval*, consultado el 24 de agosto de 2024. 
- Grupo de Ingeniería Lingüística – Universidad Nacional Autónoma de México (s.f.), *Spanish Paraphrase Corpora*, consultado el 23 de octubre de 2024. 
- Iyer, Shankar, Nikhil Dandekar y Kornél Csernai (s. f.), *First Quora Dataset Release: Question Pairs*, consultado el 23 de octubre de 2024. 
- Sierra G., Bel-Enguix G., Díaz-Velasco A., Guerrero-Cerón N., Bel N. (2024). Corpus paralelo de Biblias, consultado el 23 de octubre de 2024. 
- Wikipedia the Free Encyclopedia (s.f.), *Bible translations*, consultado el 10 de agosto de 2024. 
- Wikipedia the Free Encyclopedia (s.f.), *Wikimedia Downloads*, consultado el 30 julio 2024. 

## Biblias que conforman el corpus

- Casidoro de Reina, Jerónimo (trad.) (1569), *La Biblia, que es, los Sacros Libros del Viejo y Nuevo Testamento*, [s.e.], Basilea.
- de Ausejo, Serafín (trad.) (1975), *La Biblia*, Barcelona, Editorial Herder.
- de Valera, Casidoro (trad.) (1960), *La Santa Biblia*, Amsterdam, Editorial Sociedades Bíblicas Unidas. (Original publicado en 1602).
- Escuela Bíblica de Jerusalén (trads.) (1967), *La Biblia de Jerusalén*, París, Desclée de Brouwer. (Original publicado en 1956).

- Jünemann, Wilhem (1928), *La Sagrada Biblia*. Concepción (Chile), Editorial Diocesana de Concepción.
- Lockman Foundation (trad.) (1986), *La Biblia de las Américas*, La Habra (California), Lockman Foundation.
- Martín Nieto, Evaristo (trad) (1961), *La Santa Biblia*, Madrid: San Pablo.
- Nácar, Eloíno y Alberto Colunga (trads.) (1944). *Sagrada Biblia*. Madrid, Biblioteca de Autores Cristianos.
- Ricchiardi, Ramón y Hurault, Bernardo (trads.) (1972), *La Nueva Biblia. Edición pastoral para Latinoamérica*. San Pablo/Verbo Divino.
- Schökel, Luis Alonso, y Juan Mateos (trads.) (1975), *Nueva Biblia Española*, Madrid, Ediciones Cristiandad.
- Torres Amat, Félix (trad.) (1825), *La Sagrada Biblia*, Montaner y Simón Editores, Barcelona (Original publicado en 1823).

**GERARDO E. SIERRA-MARTÍNEZ:** Investigador titular del Instituto de Ingeniería en la UNAM, en donde es fundador y líder del Grupo de Ingeniería Lingüística. Realizó su doctorado en lingüística computacional, en UMIST, Inglaterra. Ha sido responsable de continuos proyectos patrocinados de tecnologías del lenguaje. Sus áreas de interés se enfocan a la lexicografía computacional, terminológica, extracción conceptual, corpus lingüísticos, minería de textos y lingüística forense. Es autor del libro *Introducción a los corpus lingüísticos*; coautor de los libros *Tratamiento de información textual y generación de taxonomías* y *Lingüística computacional en México: Investigación y desarrollo*, y ha publicado más de doscientos artículos en revistas, capítulos de libros y artículos arbitrados de congresos.

**GEMMA BEL-ENGUIX:** Doctora en Lingüística por la Universitat Rovira i Virgili (Tarragona, Catalunya). Ha trabajado como investigadora en las universidades de Georgetown y Milano-Bicocca y Aix-Marseille. Ha sido investigadora Ramón y Cajal en la URV (Rovira i Virgili). Es experta en modelos bioinspirados para el lenguaje natural y en aproximaciones formales a la lingüística. Desde 2016 trabaja en la UNAM, en el Grupo de Ingeniería Lingüística,



desarrollando distintos temas de investigación dentro del procesamiento de lenguaje natural: estudio de redes sociales, detección de paráfrasis, redes léxicas, análisis de sentimientos, entre otros. Es coautora de ocho libros y tiene numerosas publicaciones en el campo de la lingüística matemática y computacional.

**D. R. © Gerardo E. Sierra Martínez, Ciudad de México, enero-junio, 2024.**

**D. R. © Gemma Bel-Enguix, Ciudad de México, enero-junio, 2024.**



## RESEÑAS

**AUZA BENAVIDES, ALEJANDRA (2021), *SEMILLAS DEL LENGUAJE: DESARROLLO TÍPICO Y ATÍPICO EN PEQUEÑOS HABLANTES DEL ESPAÑOL*, MINDEN, (NEVADA), BARKER AND JULES, 149 PP.**

**E**l libro *Semillas del lenguaje. Desarrollo típico y atípico en pequeños hablantes del español* es una obra dirigida a terapeutas de lenguaje y familias, aunque también puede ser de gran utilidad para docentes de nivel preescolar e inicial y para todos aquellos que, de manera profesional o personal, deseen conocer el proceso del desarrollo de lenguaje en los primeros tres años de vida.

Alejandra Auza es terapeuta del lenguaje y doctora en lingüística. Esta combinación le permite, en principio, tener una mirada práctica y enfocada en su audiencia meta y, a la vez, el rigor para presentar, en forma sencilla y bien documentada, un proceso altamente complejo.

Este libro aborda un tema de salud pública poco estudiado en nuestro país, sobre el que se cuenta con muy pocos especialistas, y con una creciente prevalencia. El banco de datos *Lingmex* (Barriga Villanueva, 2021), que reúne trabajos relacionados con la investigación lingüística en México y que se actualiza permanentemente, arroja treinta y dos estudios sobre los trastornos de lenguaje y tres sobre desarrollo atípico/problemas de lenguaje (uno de ellos es el que aquí se reseña). La formación de terapeutas de lenguaje en México es muy escasa. En general, la carrera se imparte en instituciones privadas con muy baja matrícula y aún son pocas las institutos y universidades estatales que ofrecen formación en comunicación humana o terapia de lenguaje.

El más reciente censo muestra un incremento en el porcentaje de la población con dificultades para hablar o comunicarse; mientras que en 2010 representó el 8 % del total de personas con discapacidad o dificultades, éste es ahora del 15 % (INEGI, 2010; 2020). De ahí que la obra de Auza sea de gran relevancia, pues ofrece a los especialistas y no especialistas una perspectiva científica sobre el desarrollo del lenguaje en sus fases iniciales y sobre los indicadores y desafíos de su desarrollo atípico.

El libro incluye una introducción y ocho capítulos. En la “Introducción”, Auza plantea dos preguntas alrededor de las cuales gira el libro:

¿Cómo saber si el retraso expresivo del lenguaje de nuestros hijos debe preocuparnos?

¿Hasta dónde es necesario pensar en un programa de actividades de intervención para promover el desarrollo del lenguaje?

En el trabajo clínico de terapia de lenguaje, así como en la formación de docentes en servicio, estas preocupaciones se expresan de diferentes maneras:

¿Por qué no habla mi hijo/a, o mi alumno/a?

¿Por qué no se le entiende como a otros?

¿Tendrá algún problema?

¿Necesitará algún apoyo?

¿Se le quitará con el tiempo?

Auza sugiere que los profesionales de primer contacto, médicos, docentes y psicólogos, dada su formación en áreas distintas del lenguaje, suelen reforzar en el conocimiento colectivo la idea de que, si no hay problemas orgánicos o del comportamiento, el lenguaje se desarrollará sin problema. Aunque no están del todo equivocados, estos criterios no son suficientes, pues los trastornos específicos del lenguaje existen y se deben identificar a tiempo para ser atendidos. Este libro, por tanto, es un excelente recurso para que los profesionales de primer contacto profundicen en sus conocimientos sobre diferentes facetas de la adquisición lingüística.

En los primeros tres capítulos, Auza ofrece un recorrido por el desarrollo temprano del lenguaje: desde “Las semillas” (capítulo 1), plantadas y cuidadas a través de la interacción entre la madre o cuidador primario y el niño o la niña, pasando por “Los primeros frutos de la comunicación” (capítulo 2), que son los recursos no verbales (gestos) y las primeras palabras que permiten al niño o la niña participar más activamente en los eventos de comunicación, hasta la interrogante “¿Qué esperar de esta primera floración?” (capítulo 3), donde —además de presentar los principales hitos del lenguaje entre el nacimiento y los tres años—, la autora proporciona datos sobre la población mexicana y pautas para monitorear el desarrollo en este periodo.

En los siguientes capítulos, “Cuando el lenguaje no anda bien” (capítulo 4) y “Actuemos pronto” (capítulo 5), Auza presenta información científica sobre los indicadores del desarrollo atípico del lenguaje. Argumenta que identificar a una niña o niño como hablante tardío no es suficiente, pues, además, hay

que detectar la causa, ya que, en al menos una cuarta parte de las niñas y niños hablantes tardíos, la recuperación no será espontánea y sí tendrá efectos no deseados en otras áreas, como la escolaridad y la socialización. También aborda un tema de gran actualidad por sus efectos en el desarrollo infantil: el uso de pantallas. Seguidamente, a partir de hallazgos actualizados sobre desarrollo neurológico, plasticidad cerebral y los periodos sensibles y críticos del desarrollo infantil, fundamenta la importancia de la estimulación temprana.

Los tres últimos capítulos del libro, “Lo que sí importa” (capítulo 6), “Estrategias de intervención temprana” (capítulo 7) y “*¡Manos a la obra!*” (capítulo 8), están dedicados a presentar diversas formas de intervención temprana en niñas y niños con desarrollo atípico del lenguaje. El enfoque de estos tres capítulos es altamente empoderante para madres, padres, cuidadores, docentes y terapeutas. La autora no sólo aporta ideas generales, sino que generosamente ofrece ejemplos muy puntuales de cómo intervenir. Así, inicia planteando los rasgos deseables de toda comunicación entre adulto y niño; inmediatamente, ofrece recomendaciones y técnicas para apoyar el lenguaje durante el juego y las situaciones de la vida cotidiana, ya sea en forma individual o grupal. Además, brinda varios ejemplos de programas científicamente probados a los que se puede recurrir. Frente a la gran cantidad de información disponible, que no toda útil, especialmente la procedente de internet, Auza ofrece algunos criterios para elegir un programa de intervención. Esta guía es otro recurso para el empoderamiento de las familias, ya que las previene de ser víctimas de la publicidad no honesta de programas que, si bien pueden ser muy populares, no son necesariamente útiles.




Hacia el final de la obra, la autora presenta una puesta en práctica de los principios que ha desarrollado a lo largo todo el libro. Ofrece elementos para tomar decisiones en la evaluación del desarrollo del lenguaje y criterios para valorar la efectividad de un programa de intervención breve para hablantes tardíos con el apoyo de sus madres, padres u otros cuidadores primarios.

El libro concluye con una enumeración de los mitos más comunes sobre los problemas del lenguaje, en los que se simplifica el proceso de adquisición del lenguaje y se le otorga un peso preponderante a la función motora del habla, cuando el lenguaje es, ante todo, un proceso cognitivo, o se responsabiliza a los niños y niñas de no hablar y comunicarse según lo esperado. Estos mitos y muchos otros son resultado de una comprensión superficial del lenguaje y su adquisición.

Es mucho lo que hay que agradecerle a Alejandra Auza por obsequiarnos este libro. El trabajo de revisión exhaustiva de la literatura y su conversión a un

estilo amable en un discurso organizado y coherente, con frecuentes referencias a situaciones con las que los padres, las madres, los cuidadores, las educadoras, los docentes y cualesquiera personas se pueden identificar, hace que este libro sea en sí mismo una *semilla* para que florezcan mejores formas de entender el proceso de desarrollo típico como referencia y el desarrollo atípico como un fenómeno que requiere del soporte de profesionales comprometidos y dispuestos a formar equipos de colaboración con padres y maestros, para que, entre todos, utilizando formas de interacción amables y situadas, facilitemos el tránsito de las niñas y los niños hacia el complejo mundo de la comunicación y el lenguaje maduros.

## BIBLIOGRAFÍA

- Barriga Villanueva, R. (Coord.) con la colaboración de Valeria Chávez Granados (2021), *Lingmex: Bibliografía Lingüística de México desde 1970*, 26ª. ed. [página web adscrita a El Colegio de México], consultado el 24 de febrero de 2022. 
- Instituto Nacional de Estadística y Geografía (INEGI) (2010), “Las personas con discapacidad en México: una visión al 2010”, [página web del INEGI], consultado el 24 de febrero de 2022. 
- Instituto Nacional de Estadística y Geografía (INEGI) (2020), “Discapacidad. Porcentaje de la población con discapacidad según dificultad en la actividad 2020”, [página web del INEGI], consultado el 16 de octubre de 2024. 

SILVIA ROMERO-CONTRERAS 

Universidad Autónoma de San Luis Potosí

[silvia.romero@uaslp.mx](mailto:silvia.romero@uaslp.mx)

**SILVIA ROMERO CONTRERAS:** Profesora Investigadora de la Universidad Autónoma de San Luis Potosí (Facultad de Psicología); doctora en Educación por la Universidad de Harvard. Terapeuta de lenguaje. Investigadora Nacional Nivel II (SNII-II). Sus principales líneas de investigación son la evaluación e intervención en psicopedagogía de la lengua, la educación inclusiva, la formación docente, y la educación comunitaria sustentable.

D. R. © Silvia Romero-Contreras, Ciudad de México, enero-junio, 2024.

**ARAIZA, KARINA Y STEPHEN A. MARLETT (2021), *CONCEPTOS DE LINGÜÍSTICA DESCRIPTIVA: UN ACERCAMIENTO A LAS LENGUAS DEL MUNDO*, MÉXICO, INSTITUTO DE LINGÜÍSTICA DE VERANO, 164 PP.**

**E**l comportamiento del lenguaje, su diversidad y universalidad son aspectos fascinantes para adentrarse al estudio de la lingüística. Araiza y Marlett crean una obra de carácter aplicado cuya principal aportación es brindar un material sumamente accesible tanto para hablantes nativos –como lo mencionan los propios autores– como para estudiantes de las ciencias del lenguaje y estudiosos de las lenguas en general.

El libro *Conceptos de lingüística descriptiva: Un acercamiento a las lenguas del mundo*, sin duda, será una obra de consulta obligatoria y necesaria para todo aquel que desee iniciarse en la lingüística descriptiva, o bien, para todo aquel que necesite familiarizarse con el estudio del lenguaje.

Esta obra aporta contribuciones en distintos ámbitos. En esta reseña abordaremos algunas de ellas; sin embargo, es importante señalar que el orden de aparición no refleja una jerarquía en su relevancia.

Uno de los aportes más notables, que incluso sus propios autores señalan, es que con este material se promueve el desarrollo lingüístico y la capacitación de los hablantes nativos de lenguas originarias, ya que es un libro que se desarrolló para la clase de Introducción a la lingüística dentro del Diplomado Internacional de Lingüística Aplicada (DILA).

Otra contribución es el formato y la accesibilidad de la obra, ya que, por su alcance, puede tener un impacto en la formación y preparación de estudiantes, pues los autores hacen una notable labor al explicitar la terminología y los conceptos lingüísticos que son relevantes para el estudio de las lenguas del mundo. Todo ello ilustrado con ejemplos de treinta y tres lenguas; la mayoría, lenguas nacionales habladas en México.

Por otra parte, es sumamente importante destacar que no existe una obra similar escrita en español, pues mucha de la literatura sobre el tema se encuentra en inglés, por lo que este texto cobra relevancia en el contexto latinoamericano y de habla hispana.



Igualmente, debemos señalar que, en esta publicación, se analizan y ejemplifican los conceptos con lenguas tanto indoeuropeas como indígenas (entre ellas, algunas desplazadas), lo cual implica que todas deben ser valoradas de la misma forma, por el simple hecho de ser lenguas.

Además del libro, los autores crearon paralelamente un sitio web (Araiza y Marlett, 2021), donde se albergan ejercicios y recursos de interés tales como videos y materiales interactivos, entre otros; asimismo, hay enlaces a otras páginas web. Se prevé que el sitio se siga actualizando con información complementaria para el libro. Todo ello resulta además de interesante y beneficioso, muy innovador.

En cuanto a su estructura, la obra consta de seis capítulos: “Conceptos preliminares”, “Semántica”, “Pragmática”, “Morfología”, “Sintaxis” y “Fonología”. Cada capítulo o unidad cuenta con un apartado de terminología, donde –a manera de glosario– se presentan los términos más importantes empleados en el capítulo; una sección de repaso, en la que se ponen a prueba los conceptos clave; un apartado de ejercicios, en el cual se proponen una serie de actividades para poner en práctica lo aprendido en la unidad; una sección denominada “Otros recursos”, en la que se enlistan una serie de videos o páginas de interés relacionadas con el tema; finalmente, se proporciona una bibliografía recomendada, misma que se presenta en orden cronológico como apoyo para la unidad.

En el capítulo 1, “Conceptos preliminares”, se discuten términos y conceptos fundamentales de los estudios del lenguaje, algunos básicos como las nociones de *lenguaje*, *lengua* y *dialecto*, y otros más especializados como *lingüística sincrónica*, *diacrónica* y *aplicada*, entre otros. Se presentan también, los niveles lingüísticos que serán abordados en los capítulos subsecuentes.

El capítulo 2, “Semántica”, presenta la bifurcación clásica entre semántica léxica y semántica de las oraciones (Lyons, 1995). También describe las relaciones semánticas que pueden existir entre distintas palabras, tales como sinonimia, antonimia, homonimia o polisemia, entre otras; todo ello ejemplificado con lenguas tan diversas como el seri, el me’phaa de Malinaltepec, el español –tanto mexicano como ibérico– y el inglés. Es importante mencionar que, en cada subtema, se desarrolla la utilidad y la complejidad de cada concepto con fines lexicográficos.

El capítulo 3, “Pragmática”, se inicia describiendo brevemente la historia del término y los principales exponentes de la disciplina: Austin (1962),

Searle (1969), Grice (1975) y Sperber y Wilson (1995). Asimismo, se centra en conceptos como referencia (definida e indefinida), deixis (personal, locativa, temporal, social y discursiva), actos de habla, implicatura conversacional, principio de cooperación y teoría de la relevancia. Si bien este capítulo tiene pocos ejemplos, busca ilustrar los puntos centrales recurriendo al español, al seri y al zapoteco de Comaltepec.

El capítulo 4, “Morfología”, aborda la unidad mínima de estudio en este nivel de análisis, que es el morfema, las clases de palabras, las clases mayores y menores, las palabras compuestas, la estructura de la palabra, los tipos de morfemas, los tipos de lenguas de acuerdo con su estructura morfológica, las categorías semánticas como género, número, caso, diminutivo, aumentativo, y persona, además de marcación en el sustantivo, marcación en el verbo, deixis, TAM, otros morfemas derivativos y flexivos como la pasiva, antipasiva, reflexivo, recíproco, causativo, aplicativo, desiderativo, negación, movimiento, nominalización, infinitivo, alomorfía, y reduplicación, entre otros. Ésta es la unidad que más ejercicios contiene incluyendo lenguas como español, seri, yagua, zapoteco del istmo, popoluca de la sierra, náhuatl de la sierra de Puebla, chinanteco de Lalana, quechua de Imbabura, madija, huichol, me’phaa de Tlacoapa, náhuatl del Norte de Puebla, y mixe del istmo.

El capítulo 5, “Sintaxis”, presenta la información en tres macrorrubros, que son *la estructura, el orden y la relación*. Dentro de la estructura, se definen las frases, complementos y adjuntos; algo que los autores relacionan con temas que se presentaron en el capítulo anterior (sobre morfología), como la voz pasiva, pero ahora desde la transitividad. En la sección de orden se enmarcan los distintos órdenes que se pueden presentar en las lenguas del mundo. En cuanto a la relación, se hacen explícitas las combinaciones entre los distintos tipos de frases; se incluyen términos como *concordancia, control, y correferencia*.

Además, se abordan las frases adposicionales, frases determinantes, frases nominales, la recursividad, omisión de pronombres, entre otros. Aquí se encuentran ejercicios de las lenguas rapanui, seri, español, mixteco del norte de Tlaxiaco, awa, mandarín, otomí de la sierra, hixkaryana, amuzgo de Xochistlahuaca, chinanteco de Comaltepec, yavapai, purépecha, mualang, haida, mixteco de Mixtepec, náhuatl de Morelos, Zapoteco de Aloápam, mixteco

de Coicoyán, mazateco de Eloxochitlán, mazateco de Mazatlán, me'phaa de Teocuitlapa, zapoteco de Loxicha, mixteco de Amoltepec, yaqui, maya yucateco, zapoteco de Quiquitani, Gavião, cora, kikapú, tarahumara, zoque de San Miguel Chimalapa, y griego koiné.

El capítulo 6 “Fonología”, como se menciona desde el prólogo, excluye temas de fonética, no por ser menos relevantes sino por cuestiones del mapa curricular. Para abordar este capítulo se recomienda tener nociones básicas de fonética. Se inicia con una definición del concepto central del capítulo, y retoma otros que también son clave para la disciplina, como fonema, rasgo distintivo y sílaba, entre otros. Asimismo, se centra en la presentación y descripción de los inventarios de sonidos de las lenguas, recurriendo a leguas como zapoteco de Comaltepec, español de México, quechua de cusco, zapoteco de Quiégolani, quichua de Salasaca, mixteco de Ayutla de los libres, coreano, y seri, entre varias otras.


Como profesores de lingüística, e interesados en la formación de personas atraídas por la diversidad del país, hemos recibido esta obra como una celebración para nuestras asignaturas y discusiones en clases, por lo que decidimos incluir este material en nuestros programas de las licenciaturas en lingüística y en antropología antes de hacer esta reseña.

Todo ello responde a la solicitud de los estudiantes de tener lecturas en español donde se especifiquen los conceptos clave de lingüística, así como a la de tener a la mano instrumentos prácticos para la autoevaluación. Estamos seguros que será también un excelente material para formar a hablantes en distintos ámbitos como la educación, así como a actores sociales de las lenguas de México y el mundo a través de talleres y diplomados.

Esta obra recopila los supuestos básicos de la lingüística y los presenta, como sus propios autores mencionan, de forma holística. Cuando revisamos este texto no dejamos de equipararlo a los consolidados libros introductorios de la lingüística, como *Understanding Morphology* de Haspelmath (2002) para morfología, *Introductory phonology* de Hayes (2009) para fonología o *Describing Morphosyntax* de Payne (1997) para la sintaxis; no obstante, estos textos, al ser libros especializados, cubren más contenido.

Reconocemos la gran labor de publicar un material como éste y recomendamos ampliamente su uso como una guía para el estudio del lenguaje, la lingüística y las lenguas con el que se podrán iniciar futuros lingüistas.

## BIBLIOGRAFÍA

- Araiza R., Karina & Marlett, Stephen A. (2021). *Conceptos de lingüística descriptiva*. 
- Austin, John L. (1962), *How to Do Things with Words*, Oxford, Clarendon Press, consultado el 24 de agosto de 2024. 
- Grice, H. Paul (1975), “Logic and Conversation”, *Syntax and Semantics*, vol. 8, pp. 41-58, consultado el 24 de agosto de 2024. 
- Haspelmath, Martin (2002), *Understanding Morphology*, Oxford, University Press, consultado el 24 de agosto de 2024. 
- Hayes, Bruce (2009), *Introductory Phonology*, Londres, Blackwell, consultado el 24 de agosto de 2024. 
- Lyons, J. (1995). *Linguistics Semantics: An Introduction*. Cambridge, Cambridge University Press.
- Payne, Thomas (1997), *Describing Morphosyntax: A Guide for Field Linguists*, Cambridge, Cambridge University Press.
- Searle, John (1969), *Speech Acts*, Cambridge, Cambridge University Press.
- Sperber, Dan y Deirdre Wilson (1995), *Relevance: Communication and Cognition*, Oxford, Blackwell.

GABRIELA GARCÍA SALIDO 

Centro de Estudios Antropológicos - FCPyS - UNAM  
[ggsalido@politicas.unam.mx](mailto:ggsalido@politicas.unam.mx)

MANUEL ANDRÉS PÉREZ-PÉREZ 

Posgrado de Lingüística Hispánica - UNAM  
[ma.perez@politicas.unam.mx](mailto:ma.perez@politicas.unam.mx)

**GABRIELA GARCÍA SALIDO:** Es profesora titular de tiempo completo en el Centro de Estudios Antropológicos de la Facultad de Ciencias Políticas y Sociales de la Universidad Nacional Autónoma de México (CEA-FCPYS-UNAM). Es doctora en Lingüística por la Universidad de Texas en Austin (2014). Sus intereses de investigación incluyen: documentación lingüística, morfosintaxis, tipología y lenguas yuto-aztecas. Ha presentado y publicado numerosos textos sobre las estructuras gramaticales de las lenguas tepehuanas. Ha creado una base de datos de más de 100 horas del tepehuano del sur con financiamiento de organizaciones nacionales e internacionales (*National Science Foundation*,

*Endangered Language Fund, Foundation for Endangered Languages, Academia Mexicana de Ciencia-CONACYT-Consejo Consultivo de Ciencias y UNAM).*

**MANUEL ANDRÉS PÉREZ-PÉREZ:** Es licenciado en lingüística por la Escuela Nacional de Antropología e Historia. Actualmente concluye la maestría en Lingüística Hispánica en la Universidad Nacional Autónoma de México. Ha coordinado tres proyectos de documentación, dos de ellos de carácter internacional bajo el financiamiento de la *Endangered Language Fund* y ha formado parte de más de diez proyectos nacionales en instituciones como la Universidad Nacional Autónoma de México (UNAM) y el Centro de Investigación y Estudios Superiores en Antropología Social (CIESAS). Asimismo, ha dedicado su labor académica a la documentación y registro de lenguas nacionales, así como a la investigación y a la docencia. Sus intereses de investigación se centran en la fonología de las lenguas otópames y en la adquisición de la lengua escrita.

D. R. © Gabriela García Salido, Ciudad de México, enero-junio, 2024.

D. R. © Manuel Andrés Pérez-Pérez, Ciudad de México, enero-junio, 2024.



## Normas editoriales

**A**l someter un texto a la revista, el autor se compromete a no enviarlo a ninguna otra publicación nacional o extranjera. NO se aceptan colaboraciones que estén en proceso de dictamen, hayan aparecido o estén por aparecer en otras publicaciones impresas o electrónicas.

*Signos Lingüísticos* está bajo una licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional. Está permitida la reproducción y difusión de los contenidos de la revista para fines educativos o de investigación, sin ánimo de lucro, siempre y cuando éstos no se mutilen, y se cite la procedencia (*Signos Lingüísticos*) y al autor.

Los derechos patrimoniales de los artículos publicados en *Signos Lingüísticos* son cedidos por el autor a la Universidad Autónoma Metropolitana una vez que los originales hayan sido aceptados para que se publiquen y distribuyan tanto en la versión impresa como digital de la revista. Sin embargo, tal y como lo establece la ley, el autor conserva sus derechos morales. El autor recibirá una forma de cesión de derechos patrimoniales que deberá firmar una vez que su original haya sido aceptado. En el caso de trabajos colectivos es necesario que todos los autores firmen el documento.

Los autores podrán usar el material de su artículo en otros trabajos o libros publicados por ellos mismos, con la condición de citar a *Signos Lingüísticos* como la fuente original del texto. Todo artículo firmado es responsabilidad de su autor; las opiniones expresadas en él no necesariamente representan la posición de *Signos Lingüísticos*.

**COLABORACIONES** Los **ARTÍCULOS** serán resultado de una investigación lingüística original e inédita, tendrán una extensión mínima de 20 cuartillas y máxima de 40 cuartillas (10 000 palabras).

Las **NOTAS** presentarán descripciones u opiniones críticas en relación con temas de lingüística en forma de reporte de investigación o artículo divulgativo (10 000 palabras).

Las **RESEÑAS** presentarán una valoración crítica de obras de publicación reciente (hasta 5 años previos a la fecha de envío de la reseña) en un máximo de 20 cuartillas (5000 palabras).

**No se aceptan colaboraciones que hayan aparecido o estén por aparecer en otras publicaciones**

En la primera página se indicará el título del manuscrito, el nombre del autor, número de ORCID, correo electrónico, grado académico, adscripción institucional y cargo. El original deberá ir acompañado de:

- a) El **TÍTULO DEL TRABAJO** (en español e inglés) deberá dar una idea clara del contenido del artículo y no excederá 110 caracteres. En el caso de las reseñas el título será la ficha bibliográfica completa del libro reseñado.
- b) **RESUMEN** (en español e inglés) en el que se destaquen: el objetivo, las aportaciones y los alcances del trabajo, con un máximo de diez renglones (100 palabras).
- c) **CINCO PALABRAS CLAVE** (en español e inglés), no repetidas en el título o en el resumen, que expresen el contenido específico del mismo, que no sean frases y separadas por punto y coma (;).

Las **COLABORACIONES** se enviarán en formato Word y PDF (fuente Arial, 12 puntos, interlineado a 1.5) a la redacción de *Signos Lingüísticos*: [slingui@gmail.com](mailto:slingui@gmail.com) o [sili@xanum.uam.mx](mailto:sili@xanum.uam.mx).

Las **RESEÑAS** deberán ir encabezadas con la ficha bibliográfica completa del libro reseñado, según el siguiente modelo:



Irma Munguía Zatarain (2009), *Líneas y perfiles de la investigación y la escritura*, México, Universidad Autónoma Metropolitana, Unidad Iztapalapa.

- Citas** Cuando una CITA sea mayor a cinco líneas, estará fuera de texto con su referencia al final entre paréntesis anotando (Apellido, año: páginas). Para las citas dentro de texto, se indicará entre paréntesis el apellido del autor, seguido de una coma para anotar el año y dos puntos antes del número de página(s), por ejemplo:
- a) Niedzielski y Preston (2000: 26) proponen una estructura triangular para entender las interrelaciones entre los aspectos mencionados.
  - b) Esta disciplina descansa sobre el hecho de que las comunidades lingüísticas poseen configuraciones mentales sobre otras regiones y variedades lingüísticas: [...] como señala Cassidy (1989, p. ix), las personas disponen de cierta especie de mapas mentales acerca de las diferentes partes del país, sustentados en estereotipos, fragmentos de información exacta, conciencia de las diferencias étnicas y la manera de hablar de la gente. (Erdösová, 2011: 70)

**Notas al pie** Las NOTAS AL PIE se indicarán con números arábigos, en superíndice, en orden consecutivo y al pie de página; cuando contengan referencias bibliográficas deberán indicar: nombre del autor, año y número de páginas, al igual que en las citas. Las citas dentro de la nota al pie, sin importar la extensión, no irán fuera de texto.

**Ilustraciones y Gráficos** Todas las ILUSTRACIONES y GRÁFICOS deben estar preparados para su reproducción en formato .jpg, .tiff o .png y numeradas consecutivamente, con una resolución de 300 dpi. Deben consignar con exactitud la fuente, y los permisos correspondientes. El autor es el responsable de tramitar los permisos para su reproducción.

**Tablas** Las tablas deben enviarse en un archivo Word para su edición, además, deben entregarse en un archivo PDF para usarse como guía en el proceso editorial. Cada tabla debe consignar con exactitud la fuente y/o permisos correspondientes.

**Glosas** Se requiere el uso del sistema de glosas Leipzig Glossing Rules, disponible en línea en la siguiente dirección:  
<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>  
Para los ejemplos que se encuentran en una lengua diferente al español, los autores deben proporcionar glosas interlineales, cuidando que cada palabra esté alineada, con la fragmentación necesaria en la segunda línea, su respectiva glosa en la tercera línea y con una traducción libre en la cuarta.

Por ejemplo, en una glosa de correspondencia morfema por morfema:

Hakha Lai:

1.		aniiláay
a-	nii	-láay
3SG	reír	FUT
		“Él/Ella reirá”

El Instituto Lingüístico de Verano provee algunas recomendaciones para glosar en el siguiente vínculo:  
<https://bir.ly/3WjTBAK>

**Transcripción fonética** Se solicita el uso de la fuente Doulos SIL del Alfabeto Fonético Internacional, disponible en línea para descarga gratuita en la siguiente dirección:  
<http://software.sil.org/doulos/download/>

**BIBLIOGRAFÍA** La BIBLIOGRAFÍA deberá incluirse al final de los artículos y se ordenará alfabéticamente. Cuando el autor tenga más de una obra, se repetirá el nombre completo y se ordenarán del año más reciente al más antiguo. Si se repite el año, el primero

que se consigne en el texto será “a” y los siguientes seguirán las letras del alfabeto.

Por ejemplo:

Lee Zoreda, Margaret y Javier Vivaldo Lima (coords.) (2014a), *Construyendo una disciplina: una mirada plural al estudio de las lenguas y las culturas extranjeras*, México, Universidad Autónoma Metropolitana, Unidad Iztapalapa, Biblioteca de Signos.

## Libros

Apellido(s) del autor, Nombre(s) del autor (año), *Título en cursivas*, Ciudad de impresión, Editorial, Nombre de la colección [en caso de que sea parte de una], Número de la colección [en caso de que lo tenga], consultado el (día, mes y año). DOI o URL (si fue consultado o descargado de la web).  
Por ejemplo:

Bello, Andrés (1984), *Gramática de la lengua castellana*, Madrid, Editorial EDAF, consultado el 22 de julio de 2024. Disponible en: <https://biblioteca.org.ar/libros/131258.pdf>

## Libros con más de un autor, editor, colaborador, director y otros roles

Apellido(s) del autor principal, Nombre(s) del autor principal, Nombre(s) de autor secundario, Apellidos de autor secundario (año), *Título en cursivas*, Ciudad de impresión, Editorial, Nombre de la colección [en caso de que sea parte de una], Número de la colección [en caso de que lo tenga], consultado el (día, mes y año). DOI o URL (si fue consultado o descargado de la web).

Por ejemplo:

Bosque, Ignacio y Violeta Demonte (dirs.), *Gramática descriptiva de la lengua española*, vol. I, Madrid, Espasa-Calpe, consultado el 10 de junio de 2024. Disponible en: <https://www.espanolavanzado.com/recursos/2771-gramatica-descriptiva-de-la-lengua-espanola>

**Libros con más de una editorial**

Apellido(s) del autor, Nombre(s) del autor (año), *Título en cursivas*, Ciudad de impresión, Editorial 1 [comenzando por las académicas]/ Editorial 2 [comercial o institucional secundaria].

Por ejemplo:

Moreno Fernández, Francisco (2009), *Sociolingüística cognitiva. Propositiones, escolios y debates*, Madrid/Fránkfort, Iberoamericana/Vervuert.

**Artículos**

Apellido(s) del autor, Nombre(s) del autor (año), “Título del artículo entre comillas”, *Nombre de la revista en cursivas*, vol., núm., periodo, pp. rango de páginas separado por guion, consultado el (día, mes y año). DOI o URL.

Por ejemplo:

Lizárraga Navarro, Glenda Zoé y Armando Mora-Bustos (2010), “Variación en la marcación diferenciada de objeto en español”, *Forma y Función*, vol. XXIII, núm. 1, pp. 9-38, consultado el 4 de agosto de 2024. DOI: <https://doi.org/10.2307/j.ctv6jmw1.23>

**Capítulos**

Apellido(s) del autor, Nombre(s) del autor (año), “Título del capítulo o colaboración”, en Nombre(s) Apellido(s) del ed./coord./dir./comp., *Título del libro en cursivas*, Ciudad de impresión, Editorial, Nombre de la colección [en caso de que sea parte de una], Número de la colección [en caso de que lo tenga], pp. Rango de páginas separado por guion, consultado el (día, mes y año). DOI o URL (si fue consultado o descargado de la web).

Por ejemplo:

Heusinger, Klaus von y A. Kaiser Georg (2003), “The Interaction of Animacy, Definiteness, and Specificity in Spanish”, en Klaus von Heusinger y Georg A. Kaiser (eds.), *Proceedings*

*of the Workshop Semantic and Syntactic Aspects of Specifiers in Romance Languages*, Constanza, Universität Konstanz, pp. 41-65, consultado el 10 de junio de 2024. Disponible en: [https://www.researchgate.net/publication/228736494\\_The\\_interaction\\_of\\_animacy\\_definiteness\\_and\\_specificity\\_in\\_Spanish](https://www.researchgate.net/publication/228736494_The_interaction_of_animacy_definiteness_and_specificity_in_Spanish)

**Capítulo de libro en series de más de un volumen** Apellido(s) del autor, Nombre(s) del autor (año), “Título del capítulo o colaboración”, en Nombre(s) Apellido(s) del ed./coord./dir./comp., Título de la serie, vol./tomo (número): *Nombre del volumen/tomo en cursivas*, Ciudad de impresión, Editorial, pp. rango de páginas separado por guion.  
Por ejemplo:

Martín Zorraquino, María Antonia y José Portolés (1999), “Los marcadores del discurso”, en Ignacio Bosque y Violeta Demonte (coords.), Gramática descriptiva de la lengua española, vol. 3: *Entre la oración y el discurso. Morfología*, Madrid, Espasa Calpe, pp. 4051-4214.

**Tesis y disertaciones** Apellido(s) del autor, Nombre(s) del autor (año), Título de la tesis o disertación, tesis de Grado en Área, Ciudad, Institución, consultado el (día, mes y año). DOI o URL (si fue consultado o descargado de la web).  
Por ejemplo:

Muñoz Cruz, Héctor (2008), Reflexividad sociolingüística de hablantes de lenguas indígenas: concepciones y cambio sociocultural, tesis de doctorado en Lingüística Hispánica, México, Centro de Estudios Lingüísticos y Literarios-El Colegio de México, consultado el 4 de agosto de 2024. Disponible en: <https://repositorio.colmex.mx/concern/theses/pk02c9945?locale=es>

**Conferencias publicadas** Apellido(s) del autor, Nombre(s) del autor (año), “Título de la conferencia”, en Nombre(s) Apellido(s) del ed./coord./dir./comp., *Título de la recopilación en cursivas*, Ciudad en la

que se realizó el evento, Fecha en la que se realizó el evento, Ciudad de impresión, Editorial, pp. rango de páginas separado por guion.

Por ejemplo:

Peregrina Llanés, Manuel (2006), “Caracterización de la voz media en la narrativa del náhuatl de la Huasteca Veracruzana”, en Zarina Estrada (ed.), *Memorias del Octavo Encuentro Internacional de Lingüística en el Noroeste*, tomo 1, Universidad de Sonora, del 15 al 17 de noviembre de 2004, Hermosillo, Editorial UniSon, pp. 199-220.

**Conferencias inéditas** Apellido(s) del autor, Nombre(s) del autor (año), “Título de la conferencia”, ponencia presentada en *Nombre del evento del que fue parte*, Ciudad en la que se realizó el evento, Sede en la que se realizó el evento, Fecha.

Por ejemplo:

Herrera Castro, Samuel (2011), “Construcciones pasivas en huave de San Mateo del Mar, Oaxaca”, ponencia presentada en el *XI Congreso Nacional de Lingüística*, Quintana Roo, Universidad de Quintana Roo, 8 de noviembre de 2011.

**Página de Internet** Apellido, Nombre del autor (año), “Título del artículo o texto”, *Página o blog en la que se encuentra el texto*, consultado el (día, mes y año). URL.

Por ejemplo:

Stănescu, Nichita (1982), *Limba română este patria mea, e-Bibliotheca Septentrionalis*, consultado el 28 de abril de 2021. Disponible en: <https://ebibliothecaseptentrionalis.wordpress.com/2018/03/31/limba-romana-este-patria-mea/>

**Arbitraje** La aceptación de los ARTÍCULOS dependerá de la evaluación confidencial de dos especialistas anónimos. La aceptación de las NOTAS y RESEÑAS dependerá de la evaluación por parte

del Consejo de Redacción. De acuerdo con estas, el Consejo podrá solicitar cambios o modificaciones al autor. Una vez aceptado, el texto no podrá modificarse.

Se notificará al autor la recepción de su artículo en menos de 30 días naturales después de recibir el original y se iniciará el proceso de evaluación una vez que el artículo se ajuste a las normas mencionadas. El resultado del arbitraje se comunicará al autor en un plazo no mayor a seis meses.

Al someterse a los presentes lineamientos editoriales, los autores conceden la licencia no exclusiva para el uso de los derechos patrimoniales de la obra, en todas las modalidades de explotación, en todos los soportes, ya sea texto, imagen o electrónico para la UAM y los terceros que esta casa de estudios decida.

Para cualquier duda sobre la presentación de originales puede escribir a: [slingui@gmail.com](mailto:slingui@gmail.com) o [sili@xanum.uam.mx](mailto:sili@xanum.uam.mx)





biblioteca de  
**signos**

EL LENTO CAMBIO. CONSENSOS, MEDIACIONES Y REGULACIONES PARA ARRAIGAR DISEÑOS  
MULTILINGÜES INTERCULTURALES

MUÑOZ CRUZ, HÉCTOR (COORD.), UAM-IZTAPALAPA/EDICIONES DEL LIRIO, NÚM. 102 (2023).

*Evocar para no olvidar a 100 años de la creación de la Secretaría de Educación  
Pública (1921-2021)*

ALBERTO ENRÍQUEZ PEREA (COORD.), UAM-IZTAPALAPA/EDICIONES DEL LIRIO, NÚM. 98 (2022).

*LA EDUCACIÓN EN LA LITERATURA ESPAÑOLA A LO LARGO DE LOS SIGLOS*

MUÑOZ COVARRUBIAS, PABLO (COORD.), UAM-IZTAPALAPA/EDICIONES DEL LIRIO, NÚM. 94 (2022).

*De ANIMALIBUS. LA PRESENCIA ZOOLOGICA EN LA LITERATURA (DE LA ANTIGÜEDAD A LA ÉPOCA  
CONTEMPORÁNEA)*

CRUZ MARTÍNEZ, XOCHIQUETZALLI; FERNÁNDEZ IZAGUIRRE, PENÉLOPE MARCELA (COORDS.), UAM-IZTAPALAPA/  
EDICIONES DEL LIRIO, NÚM. 106 (2023).

# Contenido

<b>Artículos</b>	8	Variación en anáfora de frase nominal: objetos nulos y pronombres de objeto directo LAURA MELANY FLORES MORALES
	38	Los títulos de las revistas científicas cubanas: estudio exploratorio DIANA ELENA PRIETO ACOSTA, EVELYN RUBIO FIGUERAS
	58	Léxico, gastronomía y cultura: Loja (Ecuador) y Cerro Azul, Veracruz (México) SONIA E. MORETT ÁLVAREZ, ISIS NATHALY ZAMBRANO OJEDA
	104	Los pronombres personales <i>yo</i> y <i>tú</i> en la interacción oral de aprendientes sinohablantes de español LIU, MUZI, LIDIA ESTER CUBA VEGA, MARITZA GUERRERO HIDALGO
<b>Nota</b>	136	La Biblia como corpus para tecnologías del lenguaje GERARDO E. SIERRA MARTÍNEZ, GEMMA BEL ENGUIXO
	170	<b>Reseñas</b>
	181	<b>Normas editoriales</b>