

INSTITUTO CARO Y CUERVO

**SEMINARIO ANDRÉS BELLO
MAESTRÍA EN LINGÜÍSTICA**

**ANÁLISIS CUANTITATIVO DE LA MDO EN EL ESPAÑOL: TENSION
ARMÓNICA ENTRE ICONICIDAD Y ECONOMÍA**

STEPHANIE ALEJANDRA PUENTES VALBUENA

Trabajo de grado para optar por el título de Magister en Lingüística

Directores

SERGIO GONZALO JIMÉNEZ VARGAS

DANIEL EDUARDO CHÁVES PEÑA

BOGOTÁ

2021

Nota de aceptación

Firma presidente del jurado

Firma jurado

Firma jurado

Bogotá, 19 de mayo de 2021

CARTA DE AUTORIZACIÓN DE LOS AUTORES PARA LA CONSULTA Y PUBLICACIÓN ELECTRÓNICA DEL TEXTO COMPLETO

Bogotá, D.C., 19 de mayo de 2021

Señores
BIBLIOTECA JOSÉ MANUEL RIVAS SACCONI
Cuidad

Estimados Señores:

Yo, STEPHANIE ALEJANDRA PUENTES VALBUENA, identificada con C.C. No. 1053327986, autora del trabajo de grado titulado ANÁLISIS CUANTITATIVO DE LA MDO EN EL ESPAÑOL: TENSIÓN ARMÓNICA ENTRE ICONICIDAD Y ECONOMÍA presentado en el año de 2021 como requisito para optar el título de MAGISTER EN LINGÜÍSTICA; autorizo a la Biblioteca José Manuel Rivas Sacconi del Instituto Caro y Cuervo para que con fines académicos:

- Ponga el contenido de este trabajo a disposición de los usuarios en la biblioteca digital Palabra, así como en redes de información del país y del exterior, con las cuales tenga convenio el Seminario Andrés Bello y el Instituto Caro Y Cuervo.
- Permita la consulta a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea formato impreso, CD-ROM o digital desde Internet.
- Muestre al mundo la producción intelectual de los egresados de las Maestrías del Instituto Caro y Cuervo.
- Todos los usos, que tengan finalidad académica; de manera especial la divulgación a través de redes de información académica.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, ***“Los derechos morales sobre el trabajo son propiedad de los autores”***, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables. Atendiendo lo anterior, siempre que se consulte la obra, mediante cita bibliográfica se debe dar crédito al trabajo y a su (s) autor (es).



C.C.: 1053327986 de Chiquinquirá

DESCRIPCIÓN TRABAJO DE GRADO

AUTOR O AUTORES

Apellidos	Nombres
Puentes Valbuena	Stephanie Alejandra

DIRECTOR (ES)

Apellidos	Nombres
Jiménez Vargas	Sergio Gonzalo
Chávez Peña	Daniel Eduardo

TRABAJO PARA OPTAR POR EL TÍTULO DE: Magister En Lingüística

TÍTULO DEL TRABAJO: Análisis cuantitativo de la mdo en el español: tensión armónica entre iconicidad y economía.

NOMBRE DEL PROGRAMA ACADÉMICO: Maestría en Lingüística

CIUDAD: BOGOTA AÑO DE PRESENTACIÓN DEL TRABAJO: 2021

NÚMERO DE PÁGINAS: 57

TIPO DE ILUSTRACIONES: Ilustraciones ___ Mapas ___ Retratos ___ Tablas, gráficos y diagramas x Planos ___ Láminas ___ Fotografías ___

MATERIAL ANEXO (Vídeo, audio, multimedia):

Duración del audiovisual: _____ Minutos.

Número de casetes de vídeo: _____ Formato: $\frac{3}{4}$ ___ Mini DV ___ DV Cam ___ DVC Pro ___ Vídeo 8 ___

Hi 8 ___ Otro. Cual? _____

Sistema: Americano NTSC _____ Europeo PAL _____ SECAM _____

Número de casetes de audio: _____

Número de archivos dentro del CD (En caso de incluirse un CD-ROM diferente al trabajo de grado: _____)

PREMIO O DISTINCIÓN: Tesis Laureada

DESCRIPTORES O PALABRAS CLAVES:

ESPAÑOL

Marcación Diferencial de Objeto

español

Twitter

economía

iconicidad

tensión armónica

INGLES

Differential Object Marking

Spanish

Twitter

economy

iconicity

harmonic tension

RESUMEN DEL CONTENIDO Español (máximo 250 palabras):

La marcación diferencial de objeto es un fenómeno gramatical que, en español, refiere a la anteposición del morfema ‘a’ a ciertos objetos directos. La teoría actual plantea que este fenómeno está dominado por la tensión entre: la iconicidad (marcación) y la economía (omisión). Para probar esta hipótesis, se utilizó un corpus de 218 millones de tuits de distintas variedades del español. Se encontraron 4.967 casos de marcación, opcionalidad y omisión de la MDO del español de Colombia que se caracterizaron manualmente con 12 rasgos semánticos binarios y que fueron utilizados para estudiar las demás variedades. El análisis estadístico arrojó que el fenómeno está gobernado por dos fuerzas altamente correlacionadas ($r=0,92$) que se pueden asociar a la iconicidad (13,47%) y la economía (86,53%). Estas fuerzas no son independientes; están controladas por un factor común llamado aquí *tensión armónica*. Se reporta el grado de importancia de los 12 rasgos para proponer una nueva jerarquía que confirma empíricamente otras antes propuestas. Se concluye que la tensión armónica es un modelo plausible con soporte empírico para modelar el porcentaje de la MDO del español.

RESUMEN DEL CONTENIDO Inglés (máximo 250 palabras):

Differential Object Marking (DOM) in Spanish is a linguistic phenomenon in which the preposition ‘a’ is added before the direct object. Current theory state that this phenomenon is ruled by the tension between iconicity (markedness) and economy (omission), In order to test this hypothesis, a 218-million tweets corpora representing all varieties of Spanish was used. 4,967 DOM cases showing marking, omission and optional marking were found in the Colombian Spanish corpus, which were tagged manually with 12 binary features and used to study other varieties. Through the statistical analysis, it was discovered that this phenomenon is ruled by two highly correlated ($r=0,92$) tensions that can be associated to iconicity (13,47%) and economy (86,53%). This finding suggest that these tensions are not independent but rather controlled by a common factor named as harmonic tension. This study provides the degree of importance of the features analyzed, which is used to propose a new feature hierarchy that confirms empirically other small ones proposed before. It can be concluded that this harmonic tension is an empirical plausible model to demonstrate the percentage of DOM in Spanish.

Análisis cuantitativo de la MDO en el español: tensión armónica entre iconicidad y economía

Stephanie Puentes Valbuena

Maestría en Lingüística

Resumen

La marcación diferencial de objeto (MDO) es un fenómeno que se da en varias lenguas del mundo y que, en español, consiste en adicionar la preposición *a* antes del objeto directo en oraciones transitivas. La MDO ha sido ampliamente estudiada por medio de distintos modelos teóricos como la teoría de la marcación (Aissen, 2003; Bybee, 2011), en la cual se plantea que estos fenómenos están dominados por la tensión entre dos principios: la iconicidad, que promueve la marcación debido a ciertos rasgos de los verbos y objetos, y la economía, que promueve la omisión del marcado como una tendencia natural hacia la compacidad del lenguaje. Para poner a prueba esta hipótesis en términos cuantitativos, se utilizó un corpus de más de 200 millones de tuits representativos de todas las variedades del español del mundo. Como factor de comparación entre variedades dialectales, se recogieron 4.967 pares de {verbo conjugado, objeto} del español de Colombia, en el cual se encontraron casos de marcación, opcionalidad y omisión de la MDO, y que posteriormente fueron utilizados para estudiar las demás variedades. Este conjunto fue caracterizado manualmente con 12 rasgos binarios del verbo y del objeto que la literatura ha reportado como asociados al fenómeno. Los rasgos del verbo que se analizaron son aquellos que se asocian a la hipótesis de la transitividad de Hopper y Thompson (1980) y, los rasgos del objeto, los que comúnmente se han vinculado con la MDO en la literatura. En el análisis estadístico se encontró que el fenómeno está gobernado por dos factores que se pueden asociar a la iconicidad (13,47 %) y la economía (86,53 %). Igualmente, se observó que estos factores están altamente correlacionados ($r=0,92$) y que ocurren en la misma proporción para todas las variedades del español. Este hallazgo sugiere que la iconicidad y la economía no son efectos independientes, sino que están controlados por un factor común, el cual ha sido denominado en esta investigación como *tensión armónica*. Asimismo, se reporta el grado de importancia de los 12 rasgos estudiados a partir del cual se propone una nueva jerarquía de rasgos. Esta nueva jerarquía confirma empíricamente otras jerarquías más pequeñas reportadas en la literatura. Se concluye que la tensión armónica es un modelo plausible y con soporte empírico para modelar el porcentaje de la MDO del español.

Palabras clave: Marcación Diferencial de Objeto, español, corpus Twitter, economía, iconicidad, verbo, objeto directo, tensión armónica.

Quantitative analysis of DOM in Spanish: Iconicity and economy forces driven by a harmonic tension

Stephanie Puentes Valbuena

MA in Linguistics

Abstract

Differential Object Marking (DOM) is a linguistic phenomenon from some languages around the world, in which the preposition *a* is added before the direct object in a transitive construction, in the case of Spanish. DOM has been widely studied through different theoretical frameworks as in the markedness theory (Aissen, 2003; Bybee, 2011), in which it is stated that this phenomenon is ruled by the tension between two principles: iconicity, which fosters markedness based on the features of verbs and direct objects (DO), and economy, in which the marking is avoided favoring the compacity of language. In order to quantitatively test this hypothesis, a more than 200-million tweet corpora representing all varieties of Spanish around the world was used. As a comparison factor among dialectal varieties, 4,967 pairs {verb, direct object} were found in the Colombian Spanish corpus showing marking, omission and optional marking cases of DOM, and then they were used to study other varieties. This set of data was tagged manually with 12 binary features, 6 from verbs and 6 from direct objects, which have been related to this phenomenon. The verb features analyzed are based on Hopper & Thompson's (1980) transitivity hypothesis, and the DO features were based on those that the literature has proposed. Through the statistical analysis, it was discovered that this phenomenon is ruled by two tensions that can be associated to iconicity (13,47 %) and economy (86,53 %). It has also proven that these two forces are highly correlated ($r=0,92$) and that they occur in the same proportion throughout all Spanish varieties. This finding suggest that these tensions are not independent but rather controlled by a common factor named as *harmonic tension*, in this research. This study also provides the degree of importance of the features analyzed, which is used to propose a new feature hierarchy. This hierarchy confirms empirically other small ones proposed in the literature. It can be concluded that this *harmonic tension* is an empirical plausible model to demonstrate the percentage of DOM in Spanish.

Keywords: Differential Object Marking, Spanish, Twitter corpus, economy, iconicity, verb, direct object, harmonic tension.

TABLA DE CONTENIDO

1. Introducción	3
2. Referentes teóricos	4
2.1. La Marcación Diferencial de Objeto en el español	4
2.2. Antecedentes investigativos	12
3. Planteamiento del problema y objetivos de investigación	15
4. Metodología	18
4.1. Recolección de datos	18
4.2. Recolección de conteos en las variedades del español por país	21
4.3. Sistematización y filtrado de los datos: proceso de anotación	22
4.4. Normalización de los datos	24
4.5. Métodos estadísticos utilizados	26
5. Resultados y discusión	33
5.1. Distribución de probabilidad del porcentaje de marcado	33
5.2. Jerarquías de rasgos asociados a la MDO	38
5.3. Análisis de la MDO en los pares encontrados	40
6. Conclusiones	44
7. Referencias bibliográficas	47
Apéndice	51

1. Introducción

En español, la Marcación Diferencial de Objeto (MDO) se manifiesta a través de la anteposición de la preposición *a* a ciertos objetos directos (OD). Este tipo de marcación puede, o no, aparecer de manera obligatoria u opcional, dependiendo de las características del OD, el verbo de la oración e, incluso, la variedad de la lengua en la que ocurra este fenómeno.

La MDO se puede encontrar en diferentes lenguas del mundo como el rumano, el turco o el árabe, e incluso en algunas lenguas indígenas colombianas como el inga (Levinsohn y Tandioy 2000) o el ticuna (Montes, 2004), y ha sido ampliamente estudiada desde perspectivas generativistas, cognitivistas y variacionistas (Bossong, 1991; Torrego, 1998; Aissen, 2003; Laca, 1995; Rodríguez-Mondoñedo, 2007; López, 2012; Della, 2016; entre otros). Sin embargo, aún no se ha logrado dar una explicación que logre cubrir todos los casos de esta marcación ni que abarque todas las variedades del español. Adicionalmente, la mayoría de las investigaciones sobre este tema se realizan haciendo uso de pocos ejemplos, usualmente contruidos por el investigador, y solo algunos (ver sección 2.2) han optado por recurrir al uso de corpus del español que permitan analizar el fenómeno desde el uso real de los hablantes y a gran escala.

Asimismo, es posible encontrar un vacío respecto a estudios realizados sobre este fenómeno en el español de Colombia y, en general, son pocos los que analizan al mismo tiempo los efectos de la incidencia de las propiedades semánticas de los verbos y de los OD que suelen motivar la MDO. Por tal motivo, en el presente estudio se desarrolló una propuesta a través de la cual se busca comprender el grado de importancia, tanto de las propiedades de los verbos como de las del OD de la oración, y su incidencia en la MDO del español.

En esta investigación se utilizó un corpus de tuits de gran extensión que abarca la mayoría de las variedades del español del mundo. Gracias a su tamaño, homogeneidad y al uso del análisis estadístico, fue posible cuantificar e identificar relaciones entre las fuerzas que gobiernan la MDO, las cuales, en este estudio, se asociaron a los principios de iconicidad y economía. Además, usando el corpus del español de Colombia como base de comparación, se anotaron manualmente 12 rasgos de los verbos y objetos directos encontrados, de 4.967 casos frecuentes de la MDO en español; asimismo, estos rasgos se analizaron en 17

variedades de español de países hispanohablantes. El grado de importancia que se obtuvo sobre estos rasgos permitió establecer una jerarquía empírica para dar cuenta de las características de la MDO de cada una de las variantes del español que se analizaron.

En la primera sección de este artículo se revisan algunas consideraciones teóricas necesarias para comprender la naturaleza de este estudio. Posteriormente se presentan los antecedentes, el problema y los objetivos que llevaron a realizar esta investigación. Luego se explica la metodología implementada para alcanzar los objetivos, y los resultados obtenidos a partir de esta. Finalmente, se comentarán las conclusiones y los aportes que este estudio puede hacer a la investigación sobre este fenómeno.

2. Referentes teóricos

2.1. La Marcación Diferencial de Objeto en el español

Como se mencionó anteriormente, la MDO se presenta en aquellos casos en los que el morfema *a* aparece antes de algunos OD, dependiendo de sus características semánticas y pragmáticas. En general, se ha acordado que la MDO se utiliza cuando el referente del OD es animado, definido y específico (Camacho, 2018), como se puede observar en (1):

- (1) a. Luis ama *a* Luisa.
- b. Luis ama la comida de mar.

En la Nueva gramática de la lengua española (RAE y ASALE, 2010, pp. 4122-4151), se afirma que el OD encabezado por la preposición *a* puede asociarse con rasgos de animacidad, personificación, definitud o especificidad, así como con la clase semántica a la que pertenezca el verbo en cuestión. Así, esta gramática explica que la MDO aparece en diferentes contextos, algunos de los cuales se presentan en la Tabla 1.

Tabla 1. Ejemplos de la MDO.

Contexto	Ejemplo
Nombres propios	Camilo odia a Luis.
Pronombres personales tónicos	La atacaron <i>a ella</i> .
Interrogativos	¿A <i>cuál</i> persona estaba insultando?

Relativos	Juan es la persona <i>a la que</i> buscamos.
Algunos indefinidos que denotan persona como <i>nadie</i> .	No temas <i>a nadie</i> .
Definidos	Saludamos <i>a tu madre</i> hoy.
Con algunos nombres propios con uso metonímico	Ayer leímos <i>a Gabo</i> en clase.
Cuando se denominan tipos de individuos	Organizamos <i>a los</i> pacientes en habitaciones separadas.
Nombres propios de animales o animales con cercanía afectiva.	Ese perro atacó <i>a mi gato</i> .
Nombres de empresas	María critica <i>al Espectador</i> por sus columnas de opinión.
Topónimos	Ella ama <i>a Bogotá</i> .
Fernández Ramírez (1986, p. 151) añade el pronombre tónico reflexivo <i>sí</i> .	Se lo dijo <i>a sí</i> mismo.

1.1.1. Aspectos semánticos de los OD.

La MDO ha sido comúnmente asociada con propiedades internas de los OD como la animacidad, la definitud y la especificidad, entre otros. Estos rasgos pueden definir, en la mayoría de los casos, si un OD va marcado con la preposición *a*. Autores como Aissen (2003, p. 437) en (2) y Laca (2006, pp. 436-437) en (6) han propuesto unas jerarquías con base en la propuesta de Silverstein (1976), en las que se organizan los rasgos del OD de acuerdo con el grado de animacidad, definitud y especificidad que este objeto carga en su información interna. De acuerdo con Camacho (2018, p. 156), la animacidad expresa si el referente de una expresión nominal está vivo o no y la definitud expresa si un referente de una frase definida es único. En ese sentido, los OD que poseen aquellos rasgos son más susceptibles de recibir la MDO. Veremos más en detalle estos rasgos a continuación.

2.1.1.1. *Animacidad.*

Aissen (2003) establece la siguiente escala (2) para definir qué tipo de OD tiene mayor probabilidad de ser marcado con la preposición *a*, siendo el OD más propenso aquel que se ubica más a la izquierda de la escala. Así, por ejemplo, es más probable que se marque un OD animado que uno inanimado.

(2) a. Escala de animacidad: Humano > Animado > Inanimado

en donde > indica que el rasgo a la izquierda tiene mayor incidencia en la MDO.

Por su parte, Rodríguez-Mondoñedo (2007) afirma que expresiones como *nadie* o *alguien*, siempre van marcadas debido a su rasgo de animacidad, a diferencia de otras expresiones como *nada* o *algo*. Este autor plantea algunos ejemplos que pueden contradecir la relevancia de la animacidad en la MDO, como los que se plantean en (3).

- (3) a. Los ácidos atacan *a* los metales. [Molho 1958, p. 215 citado en Rodríguez-Mondoñedo, 2007]
- b. Un adjetivo califica *a* un sustantivo.

Entre otros ejemplos que contradicen la idea de que los OD con este rasgo van siempre marcados, encontramos los propuestos por Laca (2006, pp. 452) en (4), en los que se observan objetos inanimados marcados. En estos casos, se le atribuye un rasgo de personificación al OD.

- (4) a. Ama *a* un imposible. (DLNE, 1790-1800, 834 citado en Laca, 2006)
- b. Llevando en los rostros y en los ojos *a* la honestidad y en los pies *a* la ligereza (Quijote, 834 citado en Laca, 2006).

2.1.1.2. *Definitud y especificidad.*

La definitud es un rasgo que se puede determinar por factores sintácticos, como la presencia y ausencia de artículos definidos e indefinidos, y discursivos. Aissen (2003, p. 444) propone la siguiente escala (5) para clasificar aquellos OD que son más susceptibles de recibir la MDO.

- (5) Escala de definitud: Pronombre > Nombre > Definido > Específico indefinido > No específico

Por su parte, Laca (2006, pp. 436-437) propone una escala (6) de siete niveles que involucra estos rasgos semánticos.

- (6) Animacidad x Definitud: (I) Pronombre Humano > (II) Nombre Propio Humano || Pronombre Animado > (III) Definido Humano || Nombre Propio Animado || Pronombre Inanimado (IV) Humano Indefinido específico || Animado Definido || Nombre propio Animado (V) Humano no-específico || Animado Indefinido

específico || Inanimado Definido (VI) Animado no-específico || Inanimado Indefinido específico (VII) Inanimado no-específico

en donde || significa que cada nivel puede incluir más de un rasgo con la misma importancia y su aparición no representa un orden de jerarquía dentro del nivel.

Esta autora afirma que esta jerarquía (6) no pretende ofrecer explicaciones sobre el fenómeno, sino que permite formular predicciones sobre la obligatoriedad, opcionalidad e imposibilidad de la MDO.

2.1.1.3. Iconicidad y economía.

Existen dos principios que permiten describir varios fenómenos del lenguaje: la iconicidad y la economía. En lo que respecta a su incidencia en la MDO, se puede afirmar que, de acuerdo con el principio de iconicidad, un OD estará marcado en una construcción si posee los rasgos prototípicos de un sujeto, es decir, si es altamente individualizado (singular, definido), animado e incluso humano. Moravcsik (2011) afirma que, para todas las lenguas, dada una oposición entre dos nominales, un nominal que se encuentre más a la izquierda de la escala (por ejemplo, en (5)) muestra una mayor tendencia a aparecer marcado en relación con todos los otros nominales. Por otro lado, de acuerdo con el principio de economía, para todas las lenguas, ninguna frase debería presentar marcación.

En el marco de la lingüística cognitiva, el concepto de economía abarca distintos aspectos dependiendo del nivel de análisis. Las formas más económicas en términos de procesamiento cognitivo se corresponden con una distinción en el nivel gramatical: las formas menos complejas morfológica y sintácticamente suelen ser las más económicas dado que requieren menos carga de procesamiento. La idea de que existe un principio de economía (Næss, 2007) que rige el funcionamiento del lenguaje está relacionada con la economía en un plano gramatical puesto que hay formas que pueden caracterizarse mediante la presencia o ausencia de un rasgo.

De igual manera, la elección entre la forma prototípica y la forma económica se puede explicar con base en la carga de procesamiento: las formas menos complejas son las más frecuentes porque el sistema cognitivo opta por las estructuras lingüísticas relacionadas con las situaciones comunicativas más comunes o recurrentes; pero cuando el contexto y las

intenciones comunicativas lo requieren, el usuario de la lengua puede utilizar la forma marcada, es decir, la forma menos económica, para destacar algún contenido o dirigir la atención del oyente.

En el fenómeno de la MDO, la forma no marcada es la que se corresponde con los rasgos prototípicos pues el paciente del evento suele ser, por ejemplo algo no animado o no específico, mientras que el sujeto es el elemento destacado. Sin embargo, cuando la entidad denotada por el objeto directo posee los rasgos comunes del sujeto, el sistema gramatical elige la forma marcada para evitar ambigüedades con respecto a la función gramatical de los elementos lingüísticos. Por estos motivos, se espera que la frecuencia de las formas no marcadas sea mayor que las formas marcadas. Esto es lo que muestran los datos: formas económicas menos complejas gramaticalmente y más frecuentes, es decir, la omisión de la marcación de los OD.

Algunos autores han analizado el fenómeno de la MDO y su relación con estos dos principios. Entre ellos encontramos a Aissen (2003), quien, dentro del marco de la Teoría de la Optimalidad, afirma que la fuerza que ejerce el principio de iconicidad actúa cuando el objeto es prototípicamente prominente y causa que este se marque, mientras que, en el principio de economía, se evita realizar esta marcación en el OD por motivos de simplificación y ahorro de esfuerzo en el procesamiento de la información. De acuerdo con esta autora, estos dos principios entran en contraposición al momento de marcar un OD, y esta tensión sobre cuál de los dos principios rige cada caso se resuelve a través de jerarquías entre los rasgos del OD como los presentados en (5). El presente estudio pretende dar cuenta de la tensión que se da entre estas dos fuerzas en las tendencias de marcación del objeto en la variedad del español de Colombia, en contraste con las variedades de otros países hispanohablantes.

2.1.1.4. Otros rasgos del objeto.

En esta investigación también se tuvo en cuenta si el OD se presentaba como un *sustantivo escueto*. Un sustantivo escueto es aquel sintagma nominal que aparece sin un determinante (Garachana, 2008), el cual puede o no adicionarle el rasgo de definitud al mismo; por ejemplo, *ellos quieren casa*. Adicionalmente, se consideraron los rasgos de número (*plural* y *singular*), y si el OD se constituía como un *nombre propio*.

2.1.2. Clases de verbos que inciden en la MDO.

Respecto a las clases de verbos, es posible encontrar investigaciones y descripciones que buscan encontrar los efectos de la interacción entre las propiedades semánticas de los verbos con las propiedades semánticas que suelen motivar la MDO. La Nueva gramática de la lengua española describe cómo algunos verbos requieren, aceptan de manera opcional y rechazan la MDO. Entre estos, es posible encontrar la de verbos **intensionales** como *preferir* o *elegir*, los cuales requieren la MDO; verbos de **afectación** como *asustar* o *aburrir*, los cuales aceptan la MDO de manera opcional; o verbos de **causación** como *causar* o *producir*, los cuales suelen **rechazar** la MDO.

De igual manera, Fernández Ramírez (1986, p. 175-176) presenta verbos que se suelen construir con un complemento de nombre de persona, como los verbos *amar* o *besar*, es decir, con un OD que suele tener propiedades prototípicas de un sujeto; razón por la cual tienden a **requerir** la MDO. El autor también afirma que aquellos verbos que predicen deseo, necesidad, búsqueda o apetencia de algo suelen **rechazar** el uso de la preposición. Entre estos encontramos verbos como *buscar*, *conseguir*, *merecer*, *necesitar* o *tener*. Sin embargo, como se verá más adelante, varios de estos verbos suelen marcar sus OD en distintas variedades del español.

2.1.3 La hipótesis de la transitividad.

Con el fin de realizar una clasificación de verbos que permitiera hacer una búsqueda efectiva de casos de MDO en el corpus usado para este estudio, se optó por tomar como base de análisis la hipótesis de la transitividad de Hopper y Thompson (1980), quienes afirman que una acción se puede considerar más o menos transitiva a medida que cumpla con una serie de parámetros que determinan este aspecto. Estos parámetros se presentan en la Tabla 2.

Tabla 2. Parámetros de transitividad presentados por Hopper y Thompson (1980, p. 252).

Parámetro	Alta transitividad	Baja transitividad
A. Participantes	dos o más participantes	un participante
B. Kinesis	acción	no acción
C. Aspecto	télico	atélico

D. Puntualidad	puntual	no puntual
E. Volicionalidad	volicional	no volicional
F. Afirmación	afirmativa	negativa
G. Modo	real	irreal
H. Agentividad	agente alto en potencia	agente bajo en potencia
I. Afectación del Objeto	objeto totalmente afectado	no afectado
J. Individualización del Objeto	individualizado	no individualizado

La relación entre la transitividad y la MDO ha sido desarrollada por varios autores. Por un lado, Næss (2007) describe la transitividad como una categoría prototípica que se manifiesta a través de la combinación de los rasgos semánticos de una entidad. A través de la *hipótesis de máxima distinción de argumentos*, la autora indica que, en una construcción transitiva, es posible caracterizar el Agente y el Paciente de acuerdo con la presencia o ausencia de rasgos. De esta manera, según Næss (2007), un agente participante tendría los rasgos de *volicionalidad e instigación*, pero no de *afectación*, mientras que un paciente se caracterizaría por tener ese rasgo de *afectación*. Esta autora afirma que la presencia o ausencia de estos tres rasgos binarios en el agente y el paciente de una oración determinan el grado de transitividad que tiene una construcción. Dada esta hipótesis, la MDO podría aparecer como un mecanismo que permita diferenciar entre el sujeto y el objeto de una oración, en caso de que el objeto posea rasgos transitivos similares a los que tendría un sujeto; la marcación funcionaría como un mecanismo de desambiguación que permitiría distinguir entre estas dos funciones gramaticales en una misma oración.

Adicionalmente, Della (2016) indica que la transitividad se relaciona con la MDO dado que una construcción transitiva en la que el OD sea, por ejemplo, más afectado y delimitado, tiende a aparecer marcado (7), a diferencia de (8). Así, una construcción que cumpla con una mayor cantidad de rasgos de transitividad alta es más susceptible de aparecer con MDO.

(7) Juan rasca a Mairena.

(8) Juan rasca (*a) círculos en la arena (p. 22).

Por su parte, Torrego (1998) afirma que “uno de los factores que gobierna la presencia de la marcación con *a* precediendo objetos acusativos en la gramática del español es la clase aspectual a la cual el verbo pertenece” (p.15). La clase aspectual de un verbo se puede definir teniendo en cuenta si la acción es delimitada, es decir, si tiene telicidad; en ese sentido, la autora afirma que la telicidad intrínseca de verbos como los de logro (Bosque y Gutiérrez-Rexach, 2016) hace que este tipo de verbos deba ir acompañado de la marcación antes del OD: “la MDO puede causar que un predicado difiera en su clase aspectual, haciendo que un predicado se comporte como télico cuando el objeto está marcado (9), y como atélico cuando no lo está (10)” (p. 17).

(9) Laura escondió a un prisionero durante años. (Laura escondió cierto individuo que es un prisionero).

(10) Laura escondió un prisionero durante años. (Laura escondió un prisionero que no es definido).

Adicionalmente, Torrego (1998) afirma que el rasgo de la *afectación*, el cual ha sido vinculado con otros fenómenos gramaticales, se relaciona directamente con la delimitación del evento e influye en su medición y que, por tanto, requiere la marcación del OD. La *afectación* refiere al grado del cambio que se produce en el objeto a partir de la acción descrita por el verbo. Por ejemplo, *empujar* tiene un grado mayor de *afectación* que *percibir* o *juzgar*.

En el presente estudio se tuvieron en cuenta 6 rasgos asociados a la transitividad: el *aspecto*, que se refiere a la telicidad de la acción, es decir, si es un evento delimitado o completo; la *puntualidad*, es decir si la acción es instantánea o si su duración es prolongada (p. ej. *ganar* es un verbo con puntualidad mientras que *pensar* no); la *volicionalidad*, la cual se da en aquellas acciones en las que el agente de una acción tiene la intención de realizarla (p. ej. *estudiar* es un verbo que indica volicionalidad mientras que *olvidar* no); la *afirmación*, es decir, si la acción es afirmativa o negativa; el *modo*, el cual se refiere a la distinción entre *realis* e *irrealis*; y la *afectación*, es decir, si el paciente de la construcción verbal experimenta un cambio como efecto de la acción (p. ej. *empujar* es un verbo que afecta al paciente, mientras que *percibir* no). Por su parte, la *individualización* concierne al aspecto referencial del objeto y se define en términos del número y la definitud de este. No se consideró el rasgo

participantes, dado que todas las oraciones seleccionadas contienen un verbo transitivo, lo cual hace que estas siempre tengan dos participantes: el sujeto y el objeto.

2.2. Antecedentes investigativos

Es de notar que las definiciones presentadas anteriormente no aplican de manera sistemática en todos los casos en los que un OD aparece marcado. A continuación, se presentan algunos trabajos de investigaciones sobre la MDO en distintas variedades del español y en el español de Colombia, incluyendo estudios de lenguas indígenas.

2.2.1 MDO en el español.

Respecto a investigaciones sobre la MDO en el español, podemos encontrar un amplio estado del arte realizado por Fábregas (2013), en el que se presentan las propiedades y condiciones asociadas a la MDO. El autor busca delimitar este fenómeno, teniendo en cuenta propiedades morfológicas, semánticas y sintácticas, así como ofrecer una descripción de las teorías que lo han estudiado. En esta segunda parte del texto, el autor presenta aspectos gramaticales relacionados con la MDO y algunos aspectos teóricos propuestos por los autores Torrego (1998), Rodríguez-Mondoñedo (2007) y López (2012). De igual manera, Fábregas (2013) explica que el contraste entre los OD marcados y no marcados debe implicar diferencias sintácticas o semánticas, y que ninguna de las condiciones semánticas descritas hasta ahora en la literatura es suficiente para explicar por sí sola la MDO.

Además de esta importante compilación, es posible encontrar trabajos realizados desde un enfoque variacionista. Sobre la variedad del español de México se encuentra el trabajo de Lizárraga y Mora-Bustos (2010) en el que se presenta un estudio de variación gramatical que tienen como objetivo establecer, en un corpus de habla espontánea, los rasgos semánticos, sintácticos o pragmáticos que muestran una tendencia de un OD a aparecer con *a*. Este trabajo se apoya en la hipótesis de la transitividad propuesta por Hopper y Thompson (1980) y en otros trabajos realizados con base en esta, para seleccionar algunos de los rasgos a analizar como la individuación y la afectación del objeto. Estos autores resaltan que no se había hecho antes una aproximación a este tema desde un enfoque variacionista y concluyen que “el objeto directo lleva la marca diferenciada cuando ocurren de manera simultánea y sistemática

los rasgos humano y definido; en segunda instancia, cuando la FN objeto es humana o definida y necesariamente individuada, es decir, singular” (Lizárraga y Mora-Bustos 2010).

Otro estudio de tipo variacionista es el realizado por Balasch (2011), en el que utiliza un corpus de 50.000 palabras del español de Mérida y de Madrid, y se realiza una comparación de los contextos que condicionan la aparición de la MDO en estas dos variedades del español. El objetivo de su investigación era determinar qué factores (animacidad, especificidad, topicalidad y número gramatical) tienen influencia sobre la aparición o ausencia de la MDO. Esta investigadora encontró una gran diferencia entre las tasas de aparición de la MDO en objetos animados e inanimados. Además, de las variables elegidas, se encontró que la definitud y la correferencia tienen una influencia significativa en la presencia y ausencia de la MDO, mientras que la especificidad y el número gramatical no tienen mayor influencia. También se concluye en esta investigación que, para determinar cómo la marcación se da con OD inanimados, es necesario utilizar un corpus significativamente más grande que el utilizado para este estudio.

Desde el enfoque generativista, encontramos los estudios de Aissen (2013), Torrego (1998) y Rodríguez-Mondoñedo (2007) mencionados anteriormente, los cuales trabajan aspectos más estructurales, sin dejar de lado el análisis de los rasgos semánticos que tienen incidencia en el fenómeno de la MDO. Asimismo, encontramos los estudios realizados por Montrul y Bowles (2009) y por Ormazábal y Romero (2013). El primero es un estudio que identifica las características de la MDO en estadounidenses hablantes de español de ascendencia hispana y descendientes de personas mexicanas, en comparación con hablantes nativos, mediante el uso de dos tareas de juicios de aceptabilidad en oraciones con OD animados marcados. Con base en los resultados de los dos experimentos, se concluyó que las oraciones agramaticales con OD animados que no estaban marcadas fueron juzgadas como agramaticales por los hablantes nativos en una mayor proporción que por los hablantes de español heredado. En la tarea de producción, los hablantes de español heredado realizaron un 29.1 % de omisiones de la MDO en comparación con los hablantes nativos, quienes no cometieron errores. Estos hallazgos permiten concluir que los hablantes de español heredado tienen un conocimiento inestable de la MDO incluso en casos en los que los hablantes nativos muestran seguridad absoluta respecto a la gramaticalidad de las oraciones con y sin MDO.

En el estudio de Ormazábal y Romero (2013), los autores buscan demostrar que la MDO está determinada por condiciones estructurales relacionadas con los fenómenos de Caso y concordancia (Case and agreement), y que la distinción entre el caso acusativo y el dativo no corresponde a diferentes sistemas de caso, sino que corresponde a una cuestión de concordancia entre los argumentos del verbo y el verbo. Los autores presentan varios ejemplos mediante los cuales sustentan la afirmación de que “la marcación con *a* es una consecuencia de una relación sintáctica que se activa por las características de ciertos objetos pero no de otros; los objetos que sí poseen características que activan la marcación no son solo los OD con los rasgos semánticos de especificidad y animacidad, sino también los OI en construcciones clíticas dativas e infinitivos que ascienden a objeto y predicados secundarios” (Ormazábal y Romero, 2013, p. 230).

2.2.1.1. MDO en el español de Colombia.

En relación con la variedad del español colombiano solo se ha encontrado el artículo de José Joaquín Montes llamado *Sobre el objeto directo preposicional*. En este, Montes (2006) describe la función de marcar el complemento directo, alude su origen al latín, describe algunos rasgos semánticos que pueden influenciar en su marcación, y presenta varios casos de esta construcción, con y sin preposición, tomados de fuentes escritas como libros, periódicos o revistas, clasificando estos ejemplos en cinco grupos semánticos. En las conclusiones del artículo se destaca que los únicos casos en los que la MDO es necesaria es frente a los pronombres personales. Adicionalmente, Montes (2006) indica que:

Lo que determina ante todo el uso de la preposición es la necesidad de evitar la ambigüedad que puede surgir cuando el complemento puede ser agente. Eludida esta ambigüedad por uno u otro medio, la libertad de uso u omisión de *a* es casi total (p.74).

Para este autor el factor más importante para realizar este tipo de marcación es aclarar los casos de ambigüedad para distinguir entre el sujeto y el objeto de la oración, en los casos que sea necesario; de otra manera, la marcación es opcional.

3. Planteamiento del problema y objetivos de investigación

Como se observó en la sección anterior, la MDO es un fenómeno analizado en diversas lenguas del mundo. Sin embargo, es posible encontrar un vacío respecto a estudios realizados sobre este fenómeno en el español de Colombia y, en general, son pocos los estudios que analizan los efectos de las propiedades semánticas de los verbos y de las propiedades semánticas de los OD a la vez.

De igual manera, solo algunos utilizan una muestra lo suficientemente amplia para poder formular generalizaciones sobre la caracterización de este fenómeno en distintas variedades del español. Por ejemplo, en Balasch (2011) se admite que es necesario trabajar con corpus más extensos que permitan definir de mejor manera algunos aspectos de la MDO. De igual manera, en Lizárraga y Mora-Bustos (2010), la muestra que se utilizó es de tan solo 14 entrevistas de un corpus sociolingüístico del habla del español de México. Por este motivo se considera pertinente realizar un análisis que permita lograr un acercamiento a la caracterización de este fenómeno en la variedad del español de Colombia y otros países hispanohablantes, recurriendo a la lingüística de corpus, la cual permite hacer un manejo mayor de datos y generar conclusiones más generalizadas.

Por otro lado, es importante comprobar el grado de incidencia que las propiedades semánticas tienen sobre el fenómeno de la MDO. Estas propiedades han sido estudiadas en otras variedades del español, pero como se ha dicho previamente, no se ha analizado cómo las propiedades del OD junto con las de los verbos afectan la MDO, y cuáles son aquellas que predominan. Respecto a los verbos que pueden tener relación en la MDO, varios autores (ver sección 2.1.2) han identificado algunos de estos que tienen relación con la MDO, ya sea porque la requieran, la admitan o la rechazan, o porque aspectos como la telicidad pueden generar la ausencia o presencia de la marcación. Sin embargo, estas propiedades semánticas, tanto de los verbos como de los OD, han sido estudiadas de manera independiente; aún no se ha buscado encontrar una jerarquía entre las propiedades de los OD marcados, junto con las de los verbos, que suelen desencadenar este tipo de marcación, o simplemente establecer si la marcación obedece a una motivación independiente de los factores mencionados.

El presente estudio surge de la necesidad de encontrar una explicación que permita abarcar de manera sistemática el mayor número de casos de la MDO posible, y como respuesta a la ausencia de descripciones de este fenómeno en la variedad del español colombiano. Por tal motivo, este estudio pretende analizar con qué propiedades de los OD y de los verbos se correlaciona la presencia de la preposición *a* en las construcciones sintácticas de un corpus obtenido de *Twitter*, una red social que permite dar cuenta del uso real y actual del español de Colombia y de otros países en los que se habla esta lengua; en este caso, el uso del español escrito en redes sociales, el cual no representa necesariamente el español hablado, pero sí un uso más espontáneo que aquel que se utiliza en textos escritos de carácter más formal. Varios estudios han comenzado a utilizar los datos obtenidos de las redes sociales como fuente de información en la investigación lingüística; tal es el caso de Rodríguez (2018) quien utiliza el mismo corpus seleccionado para este trabajo, o Pérez (2015) quien plantea en su estudio que la forma escrita utilizada en redes sociales tiene un estilo coloquial y ha comenzado a adquirir rasgos sintácticos del discurso oral, dada la inmediatez que requiere la comunicación a través de estas. Adicionalmente, el corpus recoge este uso de la lengua por parte de una gran cantidad de usuarios, lo cual se presenta como una ventaja, así como lo plantea Zappavigna (2012), quien afirma que “la noción de que la web pueda ser usada como un corpus es de gran atractivo dada la gran cantidad de datos disponibles y la facilidad de recopilación, en comparación con, por ejemplo, discurso hablado grabado”.

Teniendo en cuenta los criterios anteriormente establecidos, se busca encontrar cuáles de esos factores establecen una variabilidad significativa sobre la presencia y la ausencia de la preposición *a* en las construcciones sintácticas del corpus analizado, así como qué jerarquía se puede establecer entre estos rasgos, teniendo en cuenta el grado de incidencia que tienen sobre la MDO. Además, con un corpus suficientemente grande que abarque una cantidad considerable de variantes del español, se podría realizar un análisis estadístico que permita corroborar o refutar empíricamente aspectos teóricos relacionados con la MDO.

Adicionalmente, y con el fin de establecer esa jerarquía, se toman como base los postulados mencionados en la sección 2.1.3, en la cual se plantea que la marcación es más probable cuando un OD tiene rasgos prototípicos de un sujeto, es decir, cuando cumple con una mayor cantidad de rasgos asociados a la transitividad. De esta manera, se pretende comprobar si las

hipótesis de Hopper y Thompson (1980) y de Næss (2007) se pueden aplicar a los casos de MDO en el corpus del español de los países hispanohablantes analizados.

Finalmente, en esta investigación se propone dar cuenta del modo en que los principios de iconicidad y economía inciden en las tendencias de MDO en las variedades del español de Hispanoamérica. De esta manera, se busca comparar la incidencia que tienen estos dos principios en la distribución de la marcación y la omisión de marcado en los pares {verbo, objeto} que se encuentren en el corpus, como se explicará a continuación en la metodología de este estudio.

4. Metodología

4.1 Recolección de datos

4.1.1 El corpus.

Jiménez, Dueñas, Gelbukh, Rodríguez-Díaz y Mancera (2018) recolectaron un corpus de manera semiautomática en la página web de Twitter, compuesto por 217'928.418 tuits de más de 100.000 habitantes de 333 ciudades, en 21 países en los que se habla español. Las cantidades de datos obtenidos se presentan en la Tabla 3. Dicho corpus se utilizó en esta investigación con el objetivo de descubrir la manera en la que una gran cantidad de hablantes del español expresan un verbo transitivo con un objeto directo, es decir, en qué casos se antepone la preposición *a* ante un OD y en qué casos no. En primer lugar, se realizó la búsqueda de estas apariciones en el corpus de Colombia y, posteriormente, en los demás países hispanohablantes, como se explica a continuación.

Tabla 3. Estadísticas del corpus recolectado de Twitter

País	ISO	Ciudades	Palabras	Tuits	Vocabulario	Usuarios
Argentina	ARG	26	254'982.258	26'933.107	5'264.160	859.197
Bolivia	BOL	8	3'136.167	289.683	206.944	24.508
Chile	CHL	24	155'791.513	15'291.490	3'679.096	599.059
Colombia	COL	31	209'085.865	19'875.419	4'575.636	871.247
Costa Rica	CRI	5	43'905.034	4'272.517	674.130	97.211
Cuba	CUB	1	122.595	13.246	14.044	5.354
Ecuador	ECU	10	49'016.999	4'483.875	1'257.676	197.544
El Salvador	SLV	3	19'898.193	1'835.850	453.030	65.543
Guatemala	GTM	7	31'753.056	3'131.936	827.927	147.460
Honduras	HND	7	18'282.159	1'710.399	579.025	65.786
México	MEX	74	453'724.537	43'544.549	10'187.200	1'983.207
Nicaragua	NIC	4	10'982.904	1'222.135	321.567	25.862
Panamá	PAN	5	33'237.123	3'078.389	855.235	114.062
Paraguay	PRY	6	39'753.880	3'968.928	765.886	113.243
Perú	PER	14	35'355.182	3'329.937	973.957	181.880
Puerto Rico	PRI	3	35'230.113	3'863.552	666.343	94.589
Rep. Dominicana	DOM	5	86'657.210	8'608.484	1'603.572	245.348
España	ESP	36	499'630.471	45'276.446	10'771.631	1'646.083
EE. UU.	USA	35	59'974.018	6'172.521	2'759.849	956.255

Uruguay	URY	7	37'121.241	4'252.022	896.557	102.350
Venezuela	VEN	22	194'073.318	16'773.933	4'343.584	764.215
TOTAL	21	333	2.271'713.836	217'928.418	51'677.049	9'160.003

Nota. Tomado y traducido de Jiménez S., Dueñas G., Gelbukh A., Rodríguez-Díaz C. y Mancera S., 2018, *Advances in Artificial Intelligence. Vol. 11238*, p. 5.

4.1.2 Selección de las combinaciones {verbo, objeto} en el español de Colombia.

En esta sección se describe el método para encontrar una lista de pares de verbos y objetos relevantes para el estudio de la MDO en Colombia. Para esto, se realizó una exploración inicial del corpus de Colombia con el fin de identificar los verbos relevantes con el siguiente procedimiento:

- I. Método para encontrar los casos de **marcado** del OD en el corpus de tuits en Colombia.
 - a. Se segmentaron todos los tuits con el método *TweetTokenizer* de la librería NLTK¹ (Bird, Klein y Loper, 2009) para Python, con el fin de identificar unidades lingüísticas, en este caso, palabras, signos de puntuación, emoticones, etc.
 - b. Se identificaron todas las apariciones de las palabras *a* y *al*.
 - c. Para cada aparición, se recolectaron contextos de cinco unidades: dos palabras a la izquierda y dos a la derecha de las palabras *a* y *al*, adicionales a estos (por ejemplo, [...] *que destruyen a una persona*).
 - d. Se creó una tabla con los contextos que ocurrieron más de 20 veces con el fin de limitar el tamaño de la tabla de frecuencias y para eliminar los casos con poca evidencia, o que pudieran ser errores, en el corpus.
 - e. De manera manual se identificaron los verbos que aparecieran una o dos posiciones antes de *a* o *al*.
 - f. Se eliminaron los contextos que no tuvieran un verbo en estas posiciones, así como aquellos en los que el verbo antes de *a* o *al* fuera intransitivo. También se eliminaron los contextos que formaran conjuntos de verbo + *a/al* + verbo (por ejemplo, *ir a comer*) y los contextos en los que *a* o *al* fueran locativas o direccionales (Fábregas, 2013), por ejemplo, *tenerte a mi lado*. En general, se mantuvieron todos los

¹ NLTK es un conjunto de herramientas que permite procesar y analizar el lenguaje natural mediante Python.

contextos en los que se encontrara un OD después del verbo, sin importar si este estuviera marcado o no.

De esta búsqueda resultaron 129 verbos en distintas conjugaciones² relevantes para el estudio. Con estos verbos identificados, se buscaron todas las **apariciones** y **omisiones** del marcado del objeto en el corpus de Colombia, de la siguiente manera:

- g. De manera manual, se creó un listado de todas las conjugaciones de los 129 verbos encontrados.
- h. Con un programa de Python (Python Software Foundation, 2018) se buscaron contextos de seis unidades que consistieran en cada una de las conjugaciones de los verbos más dos palabras a la izquierda y tres a la derecha (por ejemplo, *les encanta abrazar un hombre*). Esto último se hizo porque el primer elemento a la derecha podía ser la *a* o el *al*, y era necesario tener al menos dos unidades más que permitieran determinar si el contexto era un caso de MDO o no.
- i. Se buscaron las apariciones de todos los contextos y para cada contexto se recolectó la frecuencia de omisiones del marcado, la frecuencia de marcado con *a* y la de marcado con *al* (por ejemplo, *les encanta abrazar un hombre* aparece 23 veces marcado y 12 veces sin marcar). En este proceso resultaron 54.029 contextos.
- j. Se eliminaron los casos en los que el verbo era un adjetivo (verbos en participio pasado (por ejemplo, *vivo enamorado de ti*), y los casos de subordinadas sustantivas (por ejemplo, *mejor espera a que llegue*). Nuevamente se buscaron y eliminaron los contextos que formaran conjuntos de verbo + *a/al* + verbo y los contextos en los que *a* o *al* fueran locativas o direccionales. Después de este paso, se obtuvo un total de 11.063 contextos.
- k. Para los contextos que se utilizaron en el análisis estadístico, no se incluyeron los casos de doblado pronominal (por ejemplo, *amarse a sí mismo*) pues estos siempre van acompañados de la marcación, ni los casos de verbo + *lo* + *que* (por ejemplo, *apreciar lo que te dan*) pues estos nunca aparecen marcados.

² Los verbos en infinitivo encontrados en el corpus se pueden ver en el Apéndice de este artículo.

- l. De manera manual se unieron todos los casos repetidos (que, por ejemplo, variaran en el primer o quinto token) y se sumaron las omisiones y apariciones de estos casos con el fin de dejar únicamente pares de {verbo conjugado, objeto}.
- m. Finalmente, se obtuvieron 4.967 pares los cuales se utilizaron para el análisis estadístico.

4.2 Recolección de conteos en las variedades del español por país

Los 4.967 pares {verbo conjugado, objeto} se utilizaron para buscar e identificar marcaciones y omisiones en el corpus completo, es decir, en todos los países que se presentan en la Tabla 3. El objetivo era encontrar un conjunto de combinaciones relevantes para los casos de marcado y omisión del objeto en el español en todas las variedades que este corpus contiene.

Luego de esta búsqueda se obtuvo una tabla con 4.967 filas para cada país del corpus, junto con el conteo de apariciones de marcaciones y omisiones. Debido al tamaño y al poco número de apariciones, se descartaron los países de Bolivia y Cuba. Igualmente, los conteos de los países de Nicaragua, Honduras y El Salvador se fusionaron en uno solo (identificado con el código NSH), dado que son países vecinos y tienen los corpus más pequeños³. Esta agrupación se hizo con el fin de uniformizar los tamaños de los corpus por país y reducir la brecha de tamaño entre los corpus más grandes y los más pequeños.

Después de obtener las combinaciones {verbo conjugado, objeto} en el corpus del español de Colombia, se buscaron estos mismos pares en los corpus de los demás países mencionados en la Tabla 3, con el fin de identificar en qué medida estas 4.967 combinaciones encontradas eran utilizadas en los demás países. En la Tabla 4 se presentan las cantidades de pares que aparecen, y los que no, en el corpus de cada país.

Tabla 4. Apariciones de combinaciones {verbo conjugado, objeto} en las variedades del español por país

País	ISO	Apariciones de pares en corpus	Pares que no aparecen
Argentina	ARG	4.065	902
Chile	CHL	4.115	852
Colombia	COL	4.967	0
Costa Rica	CRI	3.763	1.204
Ecuador	ECU	4.552	415

³ Al conjunto de datos de estos tres países se le denominó NSH en el código ISO de las tablas que se presentan en esta y las siguientes secciones.

Guatemala	GTM	4.360	607
México	MEX	4.387	580
Nicaragua, El Salvador, Honduras	NSH	4.492	475
Panamá	PAN	4.374	593
Paraguay	PRY	4.287	680
Perú	PER	4.280	687
Puerto Rico	PRI	4.145	822
Rep. Dominicana	DOM	4.380	587
España	ESP	4.091	876
EE. UU.	USA	4.520	447
Uruguay	URY	3.772	1.195
Venezuela	VEN	4.648	319

4.3 Sistematización y filtrado de los datos: proceso de anotación

Para la clasificación de 4.967 las secuencias {verbo conjugado, objeto} recopiladas por el script en Python, se tuvieron en cuenta los rasgos mencionados en la sección 1.1.3. y se establecieron los siguientes rasgos binarios:

- Rasgos del OD⁴: \pm humano, \pm animado, \pm definido, \pm nombre propio, \pm singular, y \pm sustantivo escueto.
- Rasgos del verbo: \pm modo, \pm afirmativo, \pm telicidad, \pm volicionalidad, \pm afectación, \pm puntualidad.

Para definir si cada una de las combinaciones {verbo conjugado, objeto} obtenidas cumplía o no con estos rasgos, se realizó un proceso de anotación manual en el que se asignaba un valor a cada combinación, para cada uno de estos 12 rasgos. En la Tabla 5 se presentan algunas consideraciones importantes para la anotación de estas propiedades.

Tabla 5. Anotación de rasgos en las combinaciones {verbo conjugado, objeto} obtenidas en el corpus

Rasgo ⁵	Método de anotación de combinaciones que sí cumplen con el rasgo	Ejemplo del corpus con rasgo binario +
1. Humano (H _o)	OD de persona	mis hijos, alguien
2. Animado (AN _o)	OD de entidades vivas o con movimiento	animales, la selección

⁴ Para este estudio no se tuvo en cuenta el rasgo de especificidad debido a que, por la manera en la que se recogieron los datos, no era posible acceder al tuit completo y, por esta falta de información, no se podía determinar en todos los casos si el OD se podía considerar \pm específico.

⁵ El subíndice en paréntesis en la nomenclatura de cada rasgo indica si el rasgo pertenece al objeto (o) o al verbo (v).

3. Definido (D _o)	OD con sustantivos acompañados de artículos definidos o pronombres posesivos	los recuerdos
4. Nombre propio (NP _o)	OD con sustantivos propios de personas, animales, películas, entre otros	Caperucita, Cartagena
5. Singular (S _o)	OD con sustantivos singulares	tu casa, Dora
6. Sustantivo escueto (SE _o)	OD con sustantivos sin determinantes que no sean nombres propios	rosas, vida
7. Modo (<i>Realis</i> e <i>Irrrealis</i>) (M _v)	Se marcan con + los verbos en modo <i>Realis</i> . El modo <i>irrealis</i> incluye formas verbales como las del subjuntivo, el imperativo, los condicionales, el tiempo futuro, entre otras	admiro, besar
8. Afirmativo (AFI _v)	Combinaciones {verbo conjugado, objeto} que no estuvieran acompañados de formas como <i>no</i> o <i>nunca</i>	cambiar, quiero
9. Telicidad (T _v)	Acciones con una delimitación o un final. Acciones no télicas son, por ejemplo, aquellas en las que el verbo se encuentra en infinitivo, o en tiempo presente, entre otros. Acciones en las que el OD determine su delimitación	quise, se dañó, ama a tu familia
10. Volicionalidad (V _v)	Acciones que se realicen con intención o que impliquen una decisión por parte del agente	destruir, elegir
11. Afectación (AFE _v)	Acciones que causan que el OD experimente un cambio de estado	herimos, lastimes
12. Puntualidad (P _v)	Acciones que ocurren de manera instantánea	gana, despierta

Para cada una de las combinaciones se asignó un valor + o – para cada rasgo, los cuales se representaron mediante los números 1 y 0, como se ejemplifica en la Tabla 6. De esta manera fue posible recopilar la información sobre las características de los verbos y OD obtenidos en la búsqueda del corpus, así como la cantidad de veces que el OD apareció marcado y no marcado.

Tabla 6. Ejemplos de combinaciones {verbo conjugado, objeto} anotados y cantidad de apariciones en los datos de Colombia.

{verbo conjugado, objeto}	H _o	AN _o	D _o	NP _o	S _o	SE _o	M _v	AFI _v	T _v	V _v	AFE _v	P _v	# marc	# no marc
{atraen, muchos clientes}	1	1	0	0	0	0	1	1	0	0	1	0	23	51
{esperar, el sábado}	0	0	1	0	1	0	1	1	0	1	0	0	56	17
{escoger, otra persona}	1	1	0	0	1	0	1	1	0	1	0	0	17	0
{vencerá, la inteligencia}	0	0	1	0	1	0	0	1	1	0	0	1	22	55
{quiero, una persona}	1	1	0	0	1	0	0	0	0	1	0	0	166	292
{viendo, mujeres}	1	1	0	0	0	1	1	1	0	1	0	0	0	125
{bendiga, Cartagena}	0	0	1	1	1	0	0	1	1	1	0	0	105	71

4.4 Normalización de los datos

4.4.1 Factor de escala asociado al tamaño del corpus.

Otro factor de escala que fue necesario controlar es el de las diferencias en el tamaño de los corpus de cada uno de los países, con el fin de evitar que sus dimensiones afectaran las estadísticas. Este factor de escala se evidenció en experimentos preliminares en los que las estimaciones de los parámetros de la distribución de probabilidad utilizada (ver sección 4.5.1.) resultaban altamente correlacionadas con el tamaño del corpus de cada país. Dado que la variable a predecir (porcentaje de marcado) es un cociente de dos números, estos serían de gran tamaño para los países de corpus grandes y, en contraste, pequeños para los corpus de menor tamaño. Así, el porcentaje de marcado en corpus grandes tiene una mayor resolución y una mayor posibilidad de aparición de valores diferentes en comparación con los valores obtenidos en un corpus pequeño. Dado que los valores del porcentaje de marcado se clasifican en tres grupos, el 100 %, el 0 % y los valores intermedios, es este tercer grupo el que se hace artificialmente más grande en corpus grandes. Al observar que estas diferencias en tamaño de corpus son considerables (ej. España con ~500 millones de palabras vs. Guatemala ~32 millones) se llevó a cabo el siguiente procedimiento mediante un script de Python para eliminar el factor de escala asociado al tamaño del corpus:

1. Se estableció el número de apariciones totales (número total de marcados y omisiones) para cada país. Este número se denotó como T_i , donde i denota cada uno de los 17 países estudiados.
2. Se determinó el número mínimo entre todos los T_i . Se denotó este número como T' .
3. El número de marcados normalizado para cada par {verbo conjugado, objeto} en cada país i se obtuvo usando la fórmula

$$\text{round}\left(\frac{M_{\{v,o\}} \cdot T'}{T_i}\right)$$

en donde $M_{\{v,o\}}$ es el número de marcaciones y $\text{round}()$ es la función de redondeo a números enteros.

4. El mismo procedimiento se realizó para el número de omisiones con la fórmula

$$\text{round}\left(\frac{O_{\{v,o\}} \cdot T'}{T_i}\right)$$

en donde $O_{\{v,o\}}$ es el número de omisiones.

En el siguiente ejemplo se ilustra el método de normalización descrito. El país en cuyo corpus se encontró el menor número de ocurrencias fue Guatemala, con ($T'=18.372$) casos de marcación diferencial de objeto. En comparación, Colombia y México tuvieron en total $T_{\text{col}}=103.570$ y $T_{\text{mex}}=155.051$ casos de marcado. El objetivo de la normalización es reducir de manera proporcional los valores de Colombia y México, de manera que el número total de marcaciones en estos países sea comparable al de Guatemala. Ahora, consideremos el par {elegir, alguien} en el que el objeto directo fue marcado 10 veces en Guatemala, 58 en Colombia, y 73 en México. La operación para normalizar las marcaciones de Colombia es $\frac{58 \times 18.372}{103.570} = 10,25$ y $\frac{73 \times 18.372}{155.051} = 8,65$ para México. Luego de aplicar la función de redondeo, los valores quedan así: Guatemala 10 (sin cambio), Colombia 10, y México 9. Así, se aprecia cómo la normalización remueve el efecto del tamaño del corpus y revela que “elegir a

alguien” ocurre proporcional y aproximadamente igual en Guatemala y Colombia, y ligeramente menos veces en México.

4.4.2 Variable a analizar: el porcentaje de marcado.

Puesto que el objeto de este estudio es establecer en qué grado los rasgos del objeto y del verbo determinan la presencia o la ausencia de la MDO, se optó por definir la variable numérica en términos del porcentaje de marcado que tuviera cada par {verbo conjugado, objeto}. Para esto, se realizó el cálculo del promedio de marcación de una expresión, esto es, el número de casos en que la expresión aparece con la a dividido por el número total de los casos de esa expresión. Esto permite eliminar el factor de escala de la frecuencia de apariciones de los pares y, así, evitar que el modelo explicativo y predictivo de cuenta de la frecuencia en lugar de explicar el fenómeno que queremos estudiar. Por ejemplo, si un par {verbo conjugado, objeto} tiene 25 omisiones y 5 marcaciones, este tendría el mismo porcentaje de marcado que otro par con 1.250 omisiones y 250 marcaciones, eliminando así este factor de escala de la frecuencia. Esto se realizó debido a que aquello que se quiere explicar a partir de los rasgos es la razón por la que ocurren las marcaciones u omisiones, mas no un factor asociado a la frecuencia de los casos de MDO.

4.5 Métodos estadísticos utilizados

En esta sección se presenta una breve definición de las herramientas estadísticas utilizadas para el análisis estadístico, adicionales a conceptos básicos como la media, la varianza, la correlación de Pearson y la significancia estadística. Dichos conceptos pueden ser consultados en Keeping (2010).

4.5.1 Función de densidad de probabilidad (Beta).

La distribución Beta (Johnson, Kotz & Balakrishnan, 1995) es una distribución continua cuyo dominio de la variable aleatoria es el intervalo $[0,1]$, a diferencia de otras distribuciones como la distribución normal cuyo dominio no está delimitado. Esto hace que esta distribución sea adecuada para modelar variables que representan porcentajes o proporciones, es decir, es

adecuada para la variable “porcentaje de marcado”. La función de densidad de probabilidad de la distribución Beta está definida por la expresión

$$f(x; \alpha, \beta) = K \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1}$$

en donde K es una constante de normalización que sirve para asegurar que el área bajo la función sume 1, según la siguiente expresión:

$$K = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)}$$

en donde $\Gamma()$ es la función gamma.

Cuando $\alpha < 1$ y $\beta < 1$, la distribución Beta tiene forma de “U”, lo que significa que los valores de la variable más probables están en los extremos cercanos a 0 y a 1. En el caso de este estudio, se considera esta condición porque, tal como se presentará más adelante, la variable “porcentaje de marcado” se comporta de esa manera; es decir, los valores más probables son de claro marcaje (valor cercano a 1) o clara omisión (valor cercano a 0), y los valores intermedios (p. ej. valor cercano a 0.5) son poco probables.

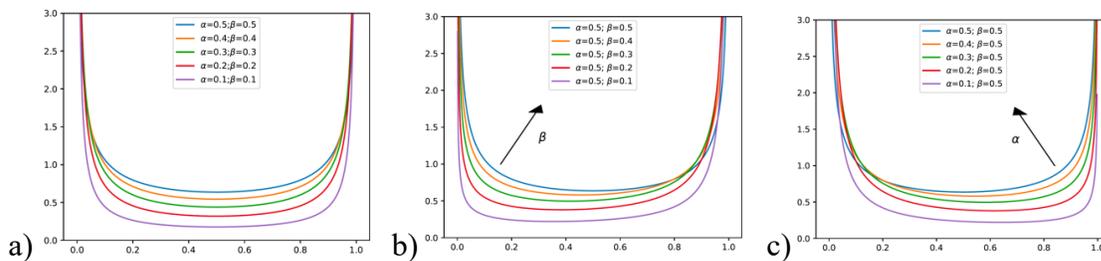


Figura 1. Parámetros de la distribución Beta en forma de U a) caso simétrico, b) caso asimétrico a la izquierda controlado por β , y c) caso asimétrico a la derecha controlado por α

Como se aprecia en la Figura 1, los parámetros α y β establecen la forma de la distribución Beta. En el caso en el que α es igual a β (Figura 1a) la forma de la “U” es simétrica. Además, si los valores de α y β aumentan, la “U” sube, haciendo que aumente la probabilidad de los valores intermedios de la variable y disminuyendo la probabilidad de los extremos 0 y 1. En contraste, cuando α y β disminuyen, la “U” baja, haciendo que los valores intermedios sean menos probables y los extremos 0 y 1 sean más probables.

Cuando α y β son diferentes, la forma de la “U” es asimétrica. La Figura 1b muestra que el codo izquierdo de la “U” es controlado por β . A medida que β aumenta, el “codo” se hace más redondeado y se desplaza hacia arriba haciendo que la probabilidad de los valores cercanos a 0 aumente. De manera similar, α controla el codo derecho y la probabilidad de los valores cercanos a 1 (Figura 1c).

Tal como lo veremos más adelante, la variable “porcentaje de marcado” obedece a un patrón asimétrico porque, en todas las poblaciones estudiadas, el número de pares {verbo conjugado, objeto} cuyo porcentaje de marcado es cercano a 0 (omisión) es mucho mayor que el número de pares con porcentaje cercano a 1 (marcación). Entonces, el fenómeno estudiado se podría modelar con la distribución Beta con $\alpha < 1$, $\beta < 1$ y $\alpha < \beta$.

La estimación de los parámetros de la distribución Beta se hace por el método de los momentos (Owen, 2008. Ecuaciones 2.19 y 2.20) con las expresiones

$$\hat{\alpha} = \bar{x} \left(\frac{\bar{x}(1 - \bar{x})}{\bar{v}} - 1 \right) \text{ y } \hat{\beta} = (1 - \bar{x}) \left(\frac{\bar{x}(1 - \bar{x})}{\bar{v}} - 1 \right)$$

donde \bar{x} es el promedio y \bar{v} la varianza de la población.

En la Figura 2a se muestra un histograma de frecuencias de la variable *porcentaje de marcado* correspondiente a los datos de Colombia. El histograma muestra una forma de “U” que parece apropiada para ser modelada con la distribución Beta. En la Figura 2b, se muestra la forma de la distribución Beta con los valores estimados de α y β para los datos de Colombia. En todas las figuras se usó una escala logarítmica en la ordenada (eje vertical) para mejorar la visualización.

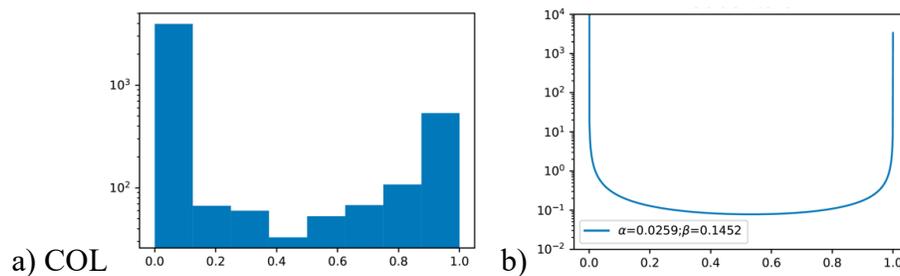


Figura 2. a) Histograma de frecuencias del porcentaje de marcado en Colombia, b) función de densidad de probabilidad Beta estimada para Colombia.

En la Figura 3 se presentan ejemplos de histogramas de otros países que se habla español (3a y 3c), así como la forma de la distribución Beta (3b y 3d) con los valores estimados de α y β . Esto con el fin de mostrar que para los demás países también se obtuvieron formas de “U” en los histogramas y en la distribución de probabilidad Beta.

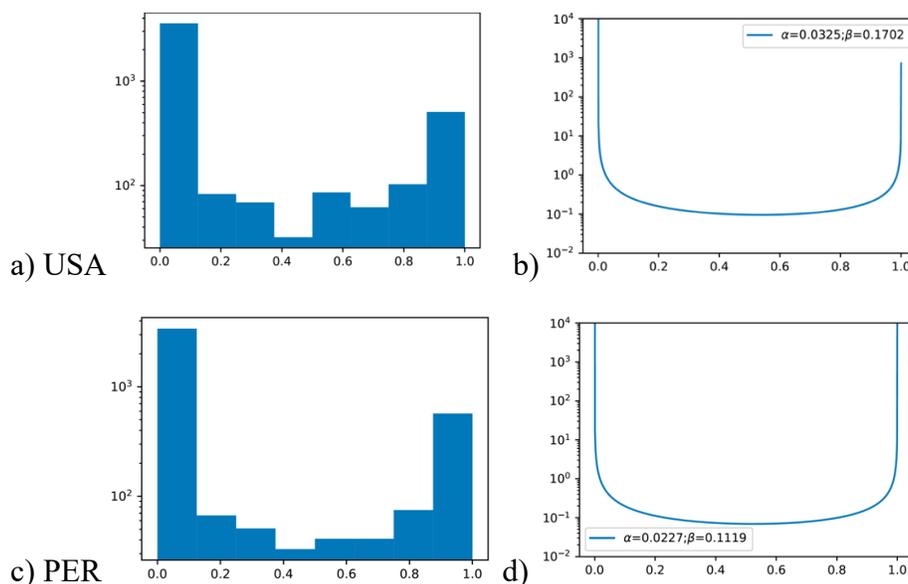


Figura 3. a) Histograma de frecuencias del porcentaje de mercado en Estados Unidos, b) Histograma de frecuencias del porcentaje de mercado en Estados Unidos, c) función de densidad de probabilidad Beta estimada para Perú, d) función de densidad de probabilidad Beta estimada para Perú.

Es importante notar que muy pocos fenómenos de las ciencias naturales, sociales y humanas han sido reportados en la literatura cuya distribución esté acorde con la distribución Beta en forma de “U”. Algunos ejemplos de estos casos se dan en meteorología (Falls, 1974), economía (Panas, 2005) y ecología (Damgaard y Irvine, 2019).

4.5.2 Prueba de bondad de ajuste de Kolmogorov-Smirnov.

La prueba de bondad de ajuste de Kolmogorov-Smirnov (Mayorga, 2004) es un test estadístico utilizado para determinar si una muestra de datos se ajusta o no a una determinada distribución de probabilidad continua, tales como la distribución normal, Poisson, Beta, entre otras. Este método consiste en comparar la distribución de probabilidad acumulada teórica a probar con la distribución acumulada de los datos. El valor de la estadística de prueba es la máxima diferencia vertical entre las dos funciones, la cual se indica con la flecha en la Figura 4.

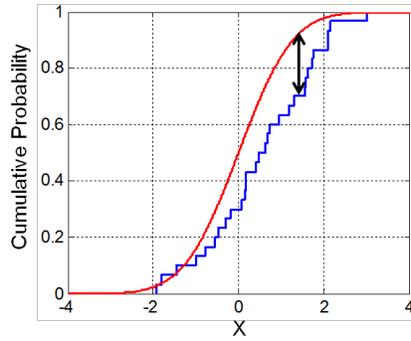


Figura 4. Ilustración de la estadística Kolmogorov-Smirnov, (Recuperada de https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Smirnov_test)

Si el valor de la estadística de la prueba, representada por la máxima diferencia vertical entre las dos funciones, es pequeña, entonces las dos funciones son similares y se puede concluir que la línea roja en la Figura 4 es un buen modelo teórico para los datos representados por la línea azul. Dado que la distribución de probabilidad Beta no tiene disponible un test de ajuste específico, la prueba de Kolmogorov-Smirnov es una alternativa conveniente dado que, en principio, se puede aplicar a cualquier distribución de probabilidad. La Figura 5 muestra el equivalente de la Figura 4 para los datos del porcentaje de marcado correspondientes a Colombia, Costa Rica y Argentina, con el fin de mostrar que, aunque los datos de los tres países varían, cualitativamente se puede apreciar que la distribución Beta provee un buen modelamiento de los datos.

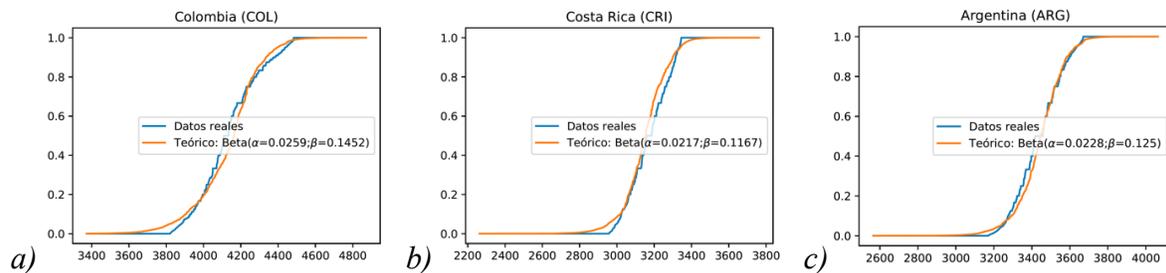


Figura 5. Comparación de la distribución acumulada de probabilidad de los datos del “porcentaje de marcado” de a) Colombia, b) Costa Rica y c) Argentina versus la distribución Beta con sus parámetros estimados.

Para que unos datos superen la prueba de ajuste de Kolmogorov-Smirnov se requiere el valor de la estadística de prueba (*Test KS*) sea pequeño y que el valor *p* de la significancia

estadística sea superior a 0.05. En la Sección 4 reportamos los resultados de *Test KS* y el valor p obtenidos para las pruebas de ajuste para los 17 países estudiados.

4.5.3 Regresión lineal multivariada.

La regresión lineal multivariada es un modelo que permite construir un predictor de una variable dependiente Y (p. ej. porcentaje de mercado) a partir de un conjunto de p variables o rasgos independientes x_1, x_2, \dots, x_p , en este caso, los rasgos presentados en la Tabla 4. Este modelo se define por la siguiente ecuación⁶:

$$Y_i = \delta_0 + \delta_1 X_{i1} + \delta_2 X_{i2} + \dots + \delta_p X_{ip} + \epsilon_i$$

en donde $i=1,2,\dots,n$, n es el número de observaciones, p es el número de rasgos, δ_0 es el sesgo (o intercepto), los restantes δ son los coeficientes de cada rasgo, y ϵ_i es el error en la predicción de la i -ésima observación. El objetivo es encontrar los valores para los parámetros δ que minimizan los errores ϵ_i obtenidos para cada observación. Una de las limitaciones de este modelo es que se asume que la relación entre las variables independientes y la variable dependiente es lineal. Una ventaja de este modelo consiste en que es computacionalmente muy sencillo y se pueden obtener los valores óptimos de los parámetros de manera eficiente para conjuntos de datos con muchas observaciones y rasgos. Otra ventaja se puede observar cuando el número de observaciones es mucho mayor que el número de rasgos, es decir, el modelo es robusto al sobreajuste, lo cual hace que este sea una buena alternativa para predecir valores de la variable dependiente para nuevas observaciones no tenidas en cuenta en el proceso de obtención de los parámetros.

Para determinar la calidad del modelo se puede utilizar una medida de rendimiento como la correlación de Pearson para comparar los valores reales de la variable independiente y las predicciones del modelo.

⁶ En lugar de usar el tradicional símbolo β para hacer referencia a los coeficientes de la regresión lineal, se utilizó el símbolo δ para evitar confusiones con el parámetro β de la distribución de probabilidad Beta.

4.5.4 Análisis de la importancia de rasgos mediante un estudio de ablación.

Un estudio de ablación (Cohen y Howe, 1988) es un método para determinar la importancia de cada uno de los rasgos en un modelo predictivo. Se realizó un estudio de este tipo en esta investigación con el fin de mostrar la importancia de cada uno de los rasgos de verbo y objeto (ver sección 4.3) involucrados en la MDO. Para esto, primero se determina la calidad del modelo predictivo cuando este se construye con todos los rasgos. Luego, para determinar la magnitud de la contribución de un rasgo, se remueve dicho rasgo del modelo, se recalculan los parámetros del modelo y se determina nuevamente su calidad. La diferencia entre la calidad del modelo con todos los rasgos y la del modelo con el rasgo removido corresponde a la contribución del rasgo.

El estudio de ablación es una alternativa al método tradicional para determinar la importancia de los rasgos en un modelo lineal utilizando los coeficientes $\delta_0, \delta_1, \dots, \delta_p$, de la regresión lineal. A diferencia de un método como el de los coeficientes, el estudio de ablación es robusto ante posibles colinealidades en los rasgos y su interpretación es mucho más clara. Además, cuando la medida de calidad está estandarizada en un rango fijo (p. ej. la correlación de Pearson entre 0 y 1) las contribuciones de los rasgos se pueden escalar fácilmente para representarlas como porcentajes. En el siguiente apartado veremos los resultados de este tipo de estudio con los rasgos propuestos en esta investigación.

5. Resultados y discusión

A partir de la implementación de la metodología anteriormente expuesta, se obtuvieron los resultados que se describen a continuación. Estos resultados se enfocan, en primer lugar, en aquellos obtenidos al realizar el análisis estadístico propuesto usando la distribución Beta para modelar el fenómeno observado. Posteriormente, se presentan los resultados obtenidos en relación con las jerarquías de rasgos obtenidas y, finalmente, unas anotaciones sobre los resultados respecto a la marcación opcional de OD en los países observados.

5.1 Distribución de probabilidad del porcentaje de marcado

La Tabla 7 muestra los resultados de las estimaciones de los parámetros α y β para cada uno de los países analizados. En esta se puede observar que, para todos los países, α y β son menores de 1, lo cual indica que la distribución Beta tiene forma de “U” en todos los casos. Allí también se pueden observar los valores de la media (\bar{x}) y la varianza (\bar{v}) como referencia, dado que con estos valores se obtienen las estimaciones de α y β , los cuales son los parámetros adecuados para la interpretación de los datos en este tipo de distribución.

En la columna “Test KS” se reportan los valores de la estadística de la prueba de bondad de ajuste de Kolmogorov-Smirnov (ver sección 4.5.2), los cuales son todos menores de 0,0187; esto indica que el ajuste de los datos en todos los países fue bueno. En la última columna se muestran los valores de la significancia estadística del test (valor p) en todos los países, los cuales son considerablemente mayores del valor crítico de 0,05; esto indica que no se puede rechazar (es decir, se acepta) la hipótesis nula de que los datos se distribuyen de acuerdo con la distribución Beta⁷.

Tabla 7. Estadísticas de la función de densidad de probabilidad del porcentaje de marcado en las variedades del español.

ISO	$\hat{\alpha}$	$\hat{\beta}$	\bar{x}	\bar{v}	Test KS	p^*
ARG	0,0228	0,1250	0,1543	0,1137	0,0187	0,1155
CHL	0,0235	0,1255	0,1577	0,1156	0,0122	0,5729
COL	0,0259	0,1452	0,1511	0,1096	0,0122	0,4661
CRI	0,0217	0,1167	0,1569	0,1162	0,0091	0,9158

⁷ La implementación utilizada del Test KS se realizó mediante el uso del paquete *Scipy* para Python (<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kstest.html>)

ECU	0,0292	0,1450	0,1674	0,1187	0,0119	0,5384
GTM	0,0240	0,1155	0,1722	0,1251	0,0109	0,6798
MEX	0,0232	0,1132	0,1703	0,1243	0,0144	0,3233
NSH	0,0295	0,1494	0,1651	0,1169	0,0155	0,2299
PAN	0,0255	0,1273	0,1670	0,1207	0,0100	0,7697
PRY	0,0260	0,1282	0,1688	0,1216	0,0104	0,7403
PER	0,0227	0,1119	0,1685	0,1235	0,0093	0,8539
PRI	0,0281	0,1528	0,1554	0,1111	0,0171	0,1768
DOM	0,0293	0,1655	0,1504	0,1069	0,0127	0,4768
ESP	0,0239	0,1205	0,1655	0,1207	0,0141	0,3935
USA	0,0325	0,1702	0,1604	0,1119	0,0159	0,2027
URY	0,0261	0,1440	0,1535	0,1110	0,0124	0,6097
VEN	0,0275	0,1448	0,1595	0,1144	0,0152	0,2336

*El valor p del test de Kolmogorov-Smirnov indica que hubo un buen ajuste cuando $p > 0,05$

Con este resultado es posible abstraer todos los casos correspondientes a cada país y analizarlos solamente por los valores de las estimaciones de los parámetros α y β . Según lo comentado para la Figura 1 en la sección 4.5.1, $\hat{\alpha}$ y $\hat{\beta}$ se pueden interpretar como fuerzas que dan forma a la función de distribución de probabilidad. En la Tabla 7 también se puede observar que todos los valores de $\hat{\alpha}$ son menores que los de $\hat{\beta}$, indicando que el fenómeno de MDO tiene un comportamiento asimétrico, pues este está inclinado hacia la omisión del mercado.

De igual manera, como se indicó en la sección 2.1.2.3, la teoría actual sugiere que el fenómeno de la MDO está sujeto a las fuerzas opuestas de iconicidad y economía. Dado que α da forma al codo derecho de la función de distribución de probabilidad, cuyos valores están cercanos al valor máximo (100 % del porcentaje de mercado), es pertinente relacionar a α con la fuerza de iconicidad, es decir, con la fuerza que representa el mercado del objeto. De manera análoga, el codo izquierdo de la función de distribución de probabilidad es controlado por β , y es posible asociarlo a la fuerza de economía.

A partir de estas asociaciones, los resultados muestran que la fuerza de economía es predominante sobre la iconicidad en todos los países estudiados. Este resultado confirma de manera empírica lo que ya se sabía o se había predicho en la teoría de la marcación (Greenberg, 1963; Croft, 2003; Bybee, 2011) con respecto a la alta frecuencia de uso del elemento no marcado (omisión) en comparación con la frecuencia del elemento marcado.

Este hecho puede explicarse en términos de economía cognitiva pues, por lo general, el elemento marcado se corresponde con una mayor complejidad estructural y una información excepcional o prominente. Por su parte, el elemento no marcado suele ser una forma regular y ya establecida en la memoria a largo plazo.

En la Figura 6 se muestra un gráfico de dispersión entre los valores de $\hat{\alpha}$ y $\hat{\beta}$ para todos los países estudiados. Allí es posible observar que existe una clara relación entre la fuerza de iconicidad, representada por α y la de economía representada por β . La correlación de Pearson entre los valores de $\hat{\alpha}$ y $\hat{\beta}$ es de $r=0,923$, la cual está muy por encima del valor crítico de $r=0,606$ para una significancia estadística de $p<0,01$ y un tamaño de muestra de 17 países. Este resultado provee una evidencia considerable de que la iconicidad y la economía no son fuerzas independientes, sino que están ligadas la una a la otra. A partir de la revisión de todos los estudios que se han propuesto sobre la MDO, se pudo determinar que es la primera vez que se reporta una relación de este tipo entre las fuerzas que dominan este fenómeno de marcación, las cuales se asumían como independientes hasta ahora.

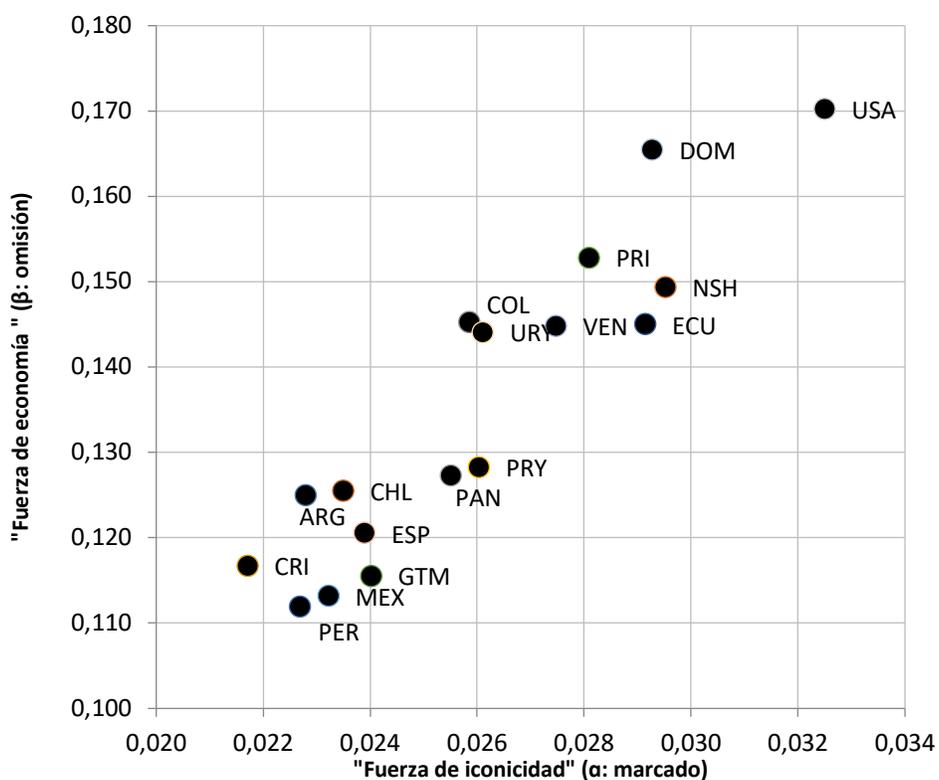


Figura 6. Variación de la “Fuerza de iconicidad” representada por la estadística a versus la “Fuerza de economía” representada por β , para las variedades del español por país.

Esta fuerza que controla la iconicidad y la economía es de gran intensidad en (USA) y es débil para para otros países como Costa Rica (CRI) y Perú (PER), dado que cuando esta fuerza es de gran intensidad, la función de distribución de probabilidad se hala hacia arriba haciendo que aumente la probabilidad de aparición de {verbo conjugado, objeto}, cuyo porcentaje de marcado es intermedio, y reduciendo la probabilidad de los extremos de marcación y omisión total. Esto muestra que, a medida que aumenta esta fuerza, los extremos de total marcación y de total omisión se reconcilian, por tanto, es posible nombrar esta fuerza **tensión armónica**. En el caso contrario, cuando la tensión armónica es débil, los valores intermedios del porcentaje de marcado se vuelven menos probables y son los extremos los que aumentan su probabilidad; esto se puede interpretar como una menor armonía o una conciliación entre los extremos de total marcación y omisión.

Es posible denominar la tensión armónica con la variable H y definirla como la suma de las tensiones de economía e iconicidad:

$$H = \text{economía} + \text{iconicidad}$$

Utilizando la regresión lineal y los $\hat{\alpha}$ y $\hat{\beta}$ de la Tabla 7, es posible obtener los siguientes modelos para la relación de dependencia entre las tensiones de economía e iconicidad con la tensión armónica:

$$\text{economía} = 0,8653H - 0,0042$$

$$\text{iconicidad} = 0,1347H + 0,0042$$

Estos modelos muestran que la tensión armónica H se divide proporcionalmente entre la tensión de economía en un 86,53 % y en un 13,47 % para la iconicidad. Adicionalmente, se encuentra un pequeño sesgo de 0,0042 que se remueve de la economía y favorece la iconicidad.

La Figura 7 ilustra la dinámica del modelo de **tensión armónica** haciendo una analogía mecánica de la acción de dicha tensión sobre la forma de la función de distribución de probabilidad del porcentaje de marcado. De izquierda a derecha se ilustran los casos de una

tensión armónica fuerte (7a), media (7b) y débil (7c). Un arreglo de cuerdas conecta la tensión armónica con los codos de la función, la cual se representa como una línea flexible.

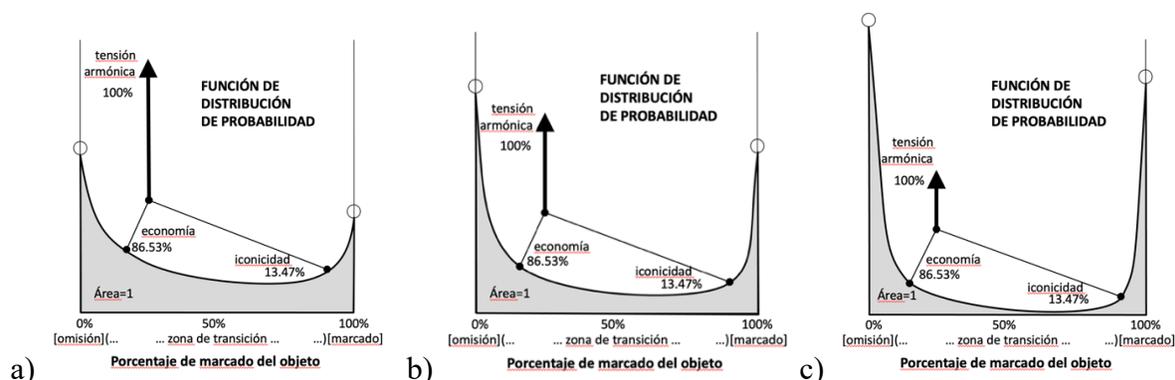


Figura 7. Ilustración mecánica de tres tipos de variación en la fuerza de tensión armónica sobre las fuerzas de iconicidad y economía.

Para ilustrar la asimetría del fenómeno, las cuerdas son de diferente longitud para dividir la tensión armónica de manera desbalanceada y proporcional entre la economía (86,53 %) y la iconicidad (13,47 %). Dado que se trata de una función de distribución de probabilidad, el área sombreada bajo la curva debe ser siempre igual a 1. Cuando la tensión armónica es fuerte (7a) la función es halada hacia arriba haciendo que los codos de la función sean redondeados, mientras que, simultáneamente, los extremos de porcentaje de mercado descienden (i.e. "omisión" al 0 % y "marcado" al 100 %).

Al disminuir la tensión (7b), la función desciende, sus codos se hacen más agudos y los extremos ascienden. Este comportamiento se da por la restricción de que el área bajo la curva se debe mantener constante, lo que se hace más evidente cuando la tensión es débil (7c). En el caso extremo, cuando la tensión armónica es nula, la función se reduce a una distribución de probabilidad binominal donde el caso de la omisión tiene una probabilidad p , el caso de "marcado" $1-p$ y no existen casos en la *zona de transición*, es decir, en donde se encuentran valores intermedios del porcentaje de marcado. De este modelo se puede inferir que la existencia de pares {verbo conjugado, objeto} cuyo porcentaje de marcado en una población sea diferente al 0 % o al 100 % es una consecuencia de la acción de la tensión armónica.

Esta analogía ayuda a interpretar los resultados de la Figura 6, en donde el fenómeno de la MDO en el español de los Estados Unidos (USA) y de República Dominicana (DOM) se asemeja a la Figura 7a reflejando una conciliación o **armonía** entre los extremos de omisión

y marcación. En estos dos casos la zona de transición es considerable en comparación con la de otros países como Costa Rica (CRI) y Perú (PER), en donde la MDO se puede considerar como polarizada o **débilmente armonizada**.

5.2 Jerarquías de rasgos asociados a la MDO

Para determinar la importancia que tienen los rasgos del verbo y del objeto analizados en esta investigación, se realizó una regresión lineal y un estudio de ablación para cada rasgo (ver secciones 4.5.3 y 4.5.4). Después de realizar las regresiones lineales por país, se obtuvo como resultado promedio 0,724 (con desviación estándar de 0,015) en la correlación de Pearson. En esta regresión se medía la efectividad del modelo primero con todos los rasgos y, posteriormente, se iba quitando cada rasgo, uno a uno, con el fin de revisar qué rasgo aportaba más a este modelo. Además, se observó que el modelo que mejor predice la marcación diferencial por país es aquel en el que intervienen todos los rasgos.

Por otro lado, el método de ablación permitió sopesar el nivel de importancia de los rasgos considerados de manera independiente en el porcentaje de marcado. Como se observa en la Tabla 8, se puede señalar que los rasgos del objeto son los que definen la marcación en mayor proporción pues, de los rasgos del verbo, únicamente el de volicionalidad resultó significativo.

Tabla 8. Importancia relativa de los rasgos del verbo y del objeto en el porcentaje de marcado.

ISO	Rasgos											
	H _o	AN _o	V _v	D _o	SE _o	AFI _v	NP _o	S _o	T _v	AFE _v	M _v	P _v
ARG	67,29%	9,99%	11,22%	7,21%	2,51%	0,89%	0,29%	0,19%	0,00%	0,23%	0,07%	0,10%
CHL	63,61%	13,83%	10,42%	7,04%	2,76%	1,27%	0,07%	0,15%	0,39%	0,01%	0,40%	0,05%
COL	67,91%	11,01%	9,67%	4,30%	3,48%	1,10%	1,59%	0,34%	0,53%	0,01%	0,02%	0,02%
CRI	59,34%	15,84%	12,18%	7,03%	3,02%	0,82%	1,00%	0,31%	0,02%	0,25%	0,17%	0,02%
ECU	66,09%	11,39%	10,75%	7,68%	2,06%	1,16%	0,19%	0,07%	0,14%	0,11%	0,24%	0,12%
GTM	67,68%	13,14%	8,74%	6,02%	2,50%	1,04%	0,38%	0,24%	0,07%	0,10%	0,05%	0,04%
MEX	65,78%	13,29%	9,21%	7,65%	2,36%	1,01%	0,06%	0,25%	0,13%	0,09%	0,15%	0,02%
NSH	69,68%	10,70%	9,10%	5,55%	3,01%	1,00%	0,61%	0,16%	0,05%	0,05%	0,02%	0,07%
PAN	69,55%	10,29%	10,42%	5,57%	2,71%	0,83%	0,14%	0,16%	0,00%	0,09%	0,07%	0,17%
PRY	65,74%	12,46%	10,76%	6,71%	2,69%	0,79%	0,39%	0,06%	0,00%	0,06%	0,14%	0,20%
PER	65,66%	12,59%	10,77%	6,46%	2,75%	0,77%	0,17%	0,41%	0,00%	0,21%	0,10%	0,11%
PRI	70,60%	7,43%	12,15%	4,80%	2,91%	0,62%	1,04%	0,07%	0,01%	0,05%	0,09%	0,23%
DOM	71,28%	6,03%	12,26%	3,71%	3,58%	0,91%	1,42%	0,29%	0,51%	0,01%	0,00%	0,00%

ESP	65,66%	9,64%	14,43%	5,99%	1,89%	0,72%	0,57%	0,09%	0,01%	0,67%	0,24%	0,08%
USA	61,86%	14,42%	11,92%	6,49%	2,82%	0,81%	0,74%	0,33%	0,37%	0,15%	0,07%	0,02%
URY	69,90%	8,14%	10,62%	5,76%	4,00%	0,88%	0,00%	0,16%	0,01%	0,20%	0,05%	0,28%
VEN	68,75%	10,87%	9,89%	6,03%	2,77%	1,02%	0,11%	0,04%	0,37%	0,04%	0,11%	0,02%
promedio	66,85%	11,24%	10,85%	6,12%	2,81%	0,92%	0,51%	0,20%	0,15%	0,14%	0,12%	0,09%

En general, la jerarquía de rasgos para las variedades del español que se obtuvo en esta investigación después de analizar los rasgos con el método de ablación es la que se observa en (11).

6 Humano > Animado > Volicionalidad > Definitud > Sustantivo Escuerto > Afirmación > Nombre propio > Singularidad > Telicidad > Afectación > Modo > Puntualidad

La predominancia de los rasgos *humano* y *animado* en la marcación se ajusta a lo que plantea la teoría. Cuando un OD posee los rasgos propios de un sujeto, entonces se recurre a la estrategia de marcación como un mecanismo de desambiguación. Algo semejante puede afirmarse con respecto a la importancia de la *volicionalidad*, ya que los verbos que poseen este rasgo reafirman la propiedad de agentividad característica del sujeto prototípico en una construcción transitiva. Con este tipo de verbos, si tenemos un OD con rasgos atípicos, es decir, con rasgos que no suelen ser prototípicos para un objeto (humano, animado, definido), la ambigüedad es mayor y el conflicto se resuelve con la marcación.

La definitud (ver sección 2.1.2.2) también es un rasgo significativo en los datos extraídos para la marcación. En la medida en que un objeto sea más definido, entonces será más individualizado y tendrá mayor probabilidad de ser marcado. Sin embargo, en los resultados se observa que el porcentaje de este rasgo no es tan alto como el de humano o animado, lo cual sugiere que este no siempre es tan relevante al momento de decidir si un OD se marca o no. Por otro lado, un OD que se considere *sustantivo escuerto* rasgo asociado a la definitud como se indicó en 2.1.1.4, no es tan relevante para determinar la marcación. Sin embargo, este rasgo tiene mayor relevancia que el de número, es decir, no hay mayor diferencia entre un OD singular o plural para determinar si este aparece marcado o no. Este rasgo *singular* va de la mano con el de telicidad, pues un OD singular suele provocar que la construcción gramatical en la que se encuentre sea télica. Se puede observar en los datos que ninguno de estos dos rasgos fue relevante en el modelo de análisis propuesto.

Como se mencionó anteriormente, los rasgos del verbo, de manera individual, tienen menor importancia que los del objeto para determinar la MDO. Después de la *volicionalidad*, el rasgo de *afirmación* fue el rasgo del verbo con un mayor grado de importancia. Sin embargo, al contrario de lo mencionado en la teoría, no se encontró evidencia de que la *telicidad* (ver sección 2.1.3.1) tuviera mayor incidencia en la MDO, de acuerdo con los resultados. Consecuentemente, el rasgo de *puntualidad*, asociado a la telicidad, tampoco mostró un mayor grado de importancia. Adicional a este, los rasgos de *modo* y *afectación* fueron los que mostraron menor incidencia en la MDO de los casos encontrados en el corpus.

A pesar de lo mencionado, es importante notar que los 6 rasgos del verbo en conjunto suman una importancia relativa de 12,27 %, valor superior al segundo rasgo en importancia, lo cual puede confirmar postulados como los de Della (2016) referentes a las hipótesis de la transitividad o de Næss (2007), de máxima distinción de argumentos (ver sección 2.1.3), los cuales indican que, a mayor cantidad de rasgos asociados a la transitividad, mayor es la posibilidad de que un OD aparezca marcado.

5.3 Análisis de la MDO en los pares encontrados

Con el fin de comprender de mejor manera las tendencias de la marcación de objeto en los 17 países analizados, en la Figura 8 se muestra la variación del porcentaje de marcado para los pares {verbo conjugado, objeto}. Allí se presentan estos 4.967 pares ubicados en un plano cartesiano, en donde la horizontal representa el promedio del porcentaje de marcado en los 17 países y, la vertical, la desviación estándar. Se observa que la mayoría de los puntos se encuentran en los extremos del porcentaje de marcado en donde la desviación estándar es mínima. Esto quiere decir que, en los casos en las que hay acuerdo total sobre la marcación o la omisión, la variación entre países es mínima. A pesar de que la visualización no permite apreciar que la mayoría de los casos se encuentran en el lado izquierdo (omisión del marcado), se visualiza que la mayor variación la tienen los pares {verbo conjugado, objeto} en los valores intermedios del porcentaje de marcado, es decir, en la zona de transición. Esta situación está relacionada con la *tensión armónica* identificada, dado que es esta la que genera esta zona de transición, y su ausencia o debilidad hace que dicha zona se desvanezca en el gráfico.

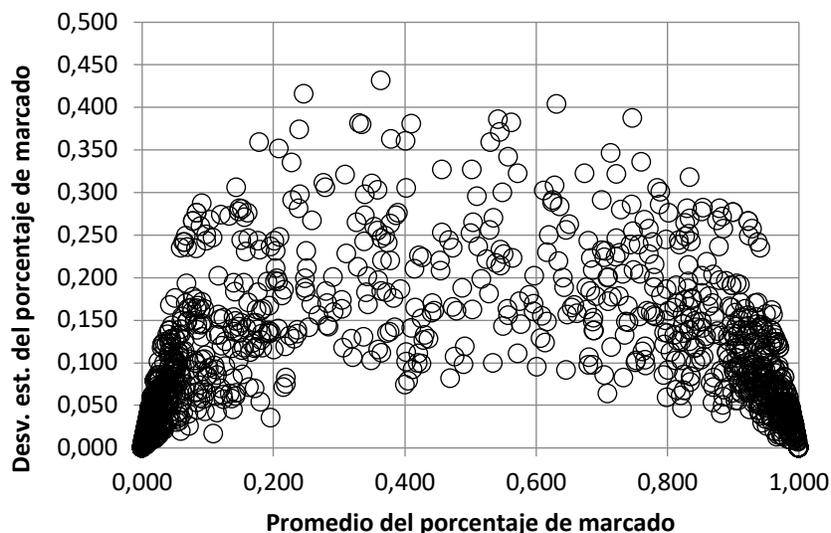


Figura 8. Comparación del promedio vs. desviación estándar de los pares {verbo conjugado, objeto} analizados.

En un intento por explorar aquellos pares de {verbo conjugado, objeto} que hacen parte de esta zona de transición de marcado, se reportan en la Tabla 9 los 39 casos que obtuvieron la mayor desviación estándar, es decir, casos ubicados en la zona de transición, en la parte superior. Para proveer una base de comparación, en la Tabla 10 se listan los 64 pares {verbo conjugado, objeto}, cuyo porcentaje de marcado fue del 100 % en los 17 países.

Tabla 9. Pares {verbo conjugado, objeto} con la mayor desviación estándar en su porcentaje de marcado en las variedades del español observadas.

Verbo conjugado	Objeto	\bar{x}	\sqrt{v}	Verbo conjugado	Objeto	\bar{x}	\sqrt{v}
tener	otra persona	0,3636	0,4312	perdí	un amigo	0,4568	0,3268
ve	una pareja	0,2462	0,4162	bendice	cada uno	0,6736	0,3226
recuerde	algunas personas	0,6310	0,4042	temas	nada	0,5719	0,3222
conozco	mi gente	0,7464	0,3874	soporto	esa gente	0,7222	0,3216
tengo	la mejor mujer	0,5417	0,3857	oye	la gente	0,3095	0,3209
ver	un pobre	0,5624	0,3820	perseguir	el amor	0,8333	0,3183
encontré	la que	0,4097	0,3811	odio	ese maldito	0,2758	0,3111
quiere	el hombre	0,3343	0,3804	tiene	su novio(a)	0,3498	0,3106
deja	tu corazón	0,2395	0,3738	enamora	varias mujeres	0,6278	0,3086
vi	los ojos	0,5446	0,3712	encuentran	el amor	0,2792	0,3063
odio	esos hombres	0,3788	0,3630	viendo	la vuelta	0,1436	0,3059
tengo	la mamá	0,4012	0,3607	hacen	la persona	0,4026	0,3054
ver	un mujeriego	0,5303	0,3593	besado	personas	0,7847	0,3052

odio	Romeo Santos	0,1786	0,3589	quiera	nadie	0,3582	0,3028
amo	mi pueblo	0,2083	0,3518	extraña	nadie	0,6119	0,3023
pierde	un ser	0,7135	0,3461	espera	la persona	0,7889	0,3010
esperar	el sábado	0,5573	0,3419	creó	Pocahontas	0,5487	0,3000
besa	nadie	0,7602	0,3360	mata	miles	0,2406	0,2978
olvidar	aquel hombre	0,2273	0,3354	usan	una mujer	0,3396	0,2977
amo	mi selección	0,5023	0,3272				

En los pares presentados en la Tabla 9, es posible encontrar OD que, a pesar de tener rasgos que siempre deben aparecer marcados según la teoría, acá aparecen con una variación muy alta. Observamos que OD como *algunas personas* en el tercer caso de la Tabla, que tienen los rasgos humano y animado, en promedio se marcan el 63 % de las veces cuando va acompañado del verbo conjugado *recuerde*. Sin embargo, este par tiene una desviación estándar muy alta, lo cual indica que existe un gran desacuerdo para este par en las 17 variedades del español estudiadas. Algo similar sucede con los pares {*besa, nadie*}, {*besado, personas*} {*espera, la persona*} o {*conozco, mi gente*}.

En contraste, la Tabla 10 muestra aquellos pares en los que la marcación fue total, es decir, los pares {verbo conjugado, objeto} que siempre aparecieron unidos mediante la preposición *a* en los datos de todos los países. Allí encontramos OD muy similares a los de la Tabla 9 como *la/una persona* o *su novio(a)*, los cuales comparten los mismos rasgos de los OD descritos anteriormente.

Tabla 10. Pares {verbo conjugado, objeto} con porcentaje de marcado 100 % y desviación estándar 0 en las variedades del español observadas.

Verbo conjugado	Objeto	Verbo conjugado	Objeto	Verbo conjugado	Objeto
conviertas	alguien	evalúa	los demás	trata	su mujer
enfrentarte	cosas	odias	los envidiosos	consentir	su novio(a)
ayudar	Dora	castigar	los hombres	trate	sus amigas
llamar	el árbitro	insultar	los idiotas	matar	todas las mujeres
juzgues	el caído	mirase	los ojos	maldices	tu amigo(a)
culpando	el mundo	llamarás	mi corazón	mirar	tu amigo(a)
perdonar	el novio	mirar	mi mamá	soportar	tu amigo(a)
ayudar	el otro	conquistar	mil personas	perdona	tu enemigo
descuides	el que	prostituir	mis amigos(as)	mirar	tu hermano
amarás	el señor	lastimes	nadie	tratarán	tu hija
adoro	este retrasado	recuerdan	nadie	altera	tu madre

temo	la oscuridad	escoger	otra persona	altera	tu mamá
golpeando	la otra persona	prefiera	otra persona	mirar	tu mejor amigo(a)
abrazas	la persona	humillar	otros	maldecir	tu subconsciente
cuidar	la persona	arruinar	quien	trates	tus amigas
temo	la soledad	dejes	quien	descifrar	una mujer
llama	las emisoras	herimos	quien	oirás	una mujer
juzgues	las personas	lastimar	quienes	subestimen	una mujer
odies	las personas	queriendo	quienes	destruyen	una persona
jugaste	los amores	llamando	su ex	perdonado	una persona
atraerá	los chinos	encuentra	su madre	subestimes	una persona
detener	los conformes				

Por otro lado, encontramos OD como *el sábado* o *mi pueblo*, los cuales generan duda a los hablantes y, en ocasiones, aparecen marcados a pesar de no tener los rasgos de objeto que suelen activar la marcación. De igual manera, los OD *la oscuridad* y *las emisoras*, los cuales tampoco poseen rasgos como *animacidad* o *humano*, siempre aparecen marcados. En estos cuatro casos, se podría afirmar que es el rasgo de definitud el que determina que estos OD vayan marcados.

Un rasgo que ha sido comúnmente asociado a la MDO es el de la definitud, el cual, como se mostró en la Tabla 8, resultó no ser tan significativo para el modelo implementado en este estudio; consecuentemente, en estas dos tablas es posible encontrar OD que cumplen con el rasgo de definitud. En la Tabla 9 se observan OD como *la gente*, *los ojos*, o *la persona*, que no siempre aparecen marcados y que generan duda; sin embargo, en la Tabla 10 es posible encontrar una mayor cantidad de OD definidos que van acompañados de pronombres posesivos y artículos definidos. En contraste, la Tabla 10 muestra una gran cantidad de OD indefinidos como *nadie*, *alguien*, *una persona* o *una mujer*, los cuales deberían rechazar la marcación por no poseer este rasgo, pero siempre aparecen marcados.

Respecto a los verbos, y teniendo en cuenta la clasificación presentada en 2.1.2, encontramos en las dos tablas verbos de afectación como *besar*, *temer* o *soportar*, y verbos intensionales como *esperar*, *recordar*, *olvidar*, *extrañar* o *querer*, los cuales sí son compatibles con la MDO. Sin embargo, también se observan verbos de causación como *matar*, *dejar*, *crear*, *enamorar*, *golpear*, *hacer* o *castigar*, de los cuales se dice que suelen rechazar la MDO, pero que en los datos aparecen como marcados en el 100 % de los casos. De igual manera, es

posible encontrar pares {verbo conjugado, objeto} con el rasgo de *volicionalidad*, esto indica que tanto en los pares que siempre aparecen marcados, como para aquellos en los que hay duda, este rasgo suele estar presente.

Adicionalmente, en la Tabla 9 podemos observar pares con un rasgo alto de telicidad como *tengo la mejor mujer, perseguir el amor o esperar el sábado* que generan duda en los hablantes. De igual manera, se encuentran verbos conjugados como *conviertas, descuides, juzgues* o *mirase*, de oraciones en modo subjuntivo, que suelen rechazar la MDO pero que aparecen siempre marcados. Estos resultados muestran que no hay un único rasgo e, incluso, una combinación fija de estos que permita explicar de manera general los casos en los que se da este tipo de marcación.

6. Conclusiones

A partir del análisis de un corpus extenso y de la implementación de las metodologías propuestas, fue posible proponer una hipótesis empírica de la distribución estadística de los fenómenos de marcación. Se concluyó que el porcentaje de la MDO del español se puede modelar como una variable aleatoria cuya distribución se ajusta adecuadamente a la función de distribución de probabilidad Beta en forma de “U”, y que los parámetros de esta distribución, α y β , modelan adecuadamente las nociones teóricas de iconicidad y economía. Esta metodología estadística se recomienda para analizar otros fenómenos de marcado de diferentes lenguas del mundo en los que se encuentren valores graduales de la presencia de esos fenómenos.

Asimismo, se observó que los resultados obtenidos en relación con las tendencias de marcado en el español de los 17 países observados corroboran las intuiciones de estudios previos en los que se resalta la predominancia de la economía sobre la iconicidad. Adicionalmente, se pudo concluir que los principios de iconicidad y economía en la MDO del español son dependientes de una única fuerza denominada en este trabajo **tensión armónica**, la cual se convierte en el mayor aporte de esta investigación. Además, gracias a la metodología estadística propuesta se pudo cuantificar por primera vez el desbalance entre las tensiones de

economía (86,53 %) e iconicidad (13,47 %) para el fenómeno de marcación de la MDO del español.

Respecto a los rasgos de los OD y de los verbos, se logró corroborar empíricamente que la jerarquía propuesta en este estudio coincide con las propuestas en la literatura; además, se adicionaron rasgos que antes no habían sido considerados en estas jerarquías. Es de resaltar que la *volicionalidad* es el rasgo asociado a los verbos que más aporta al modelo utilizado, lo cual lo ubica en un tercer lugar de importancia, luego de los rasgos *humano* y *animado*. La distribución de estos rasgos en la jerarquía es acorde con lo que plantean la hipótesis de transitividad (Hopper y Thompson, 1980) y la de máxima distinción de argumentos (Næss, 2007), según las cuales el objeto directo de una oración estará marcado cuando posea los rasgos prototípicos del sujeto. De este modo, se confirma que la MDO es un dispositivo gramatical de desambiguación en tanto que permite distinguir entre las funciones gramaticales de una misma construcción. Adicionalmente, se pudo concluir que todos los rasgos del verbo tienen una importancia relativa de 12,27 %, lo cual indica que estos rasgos, en conjunto, permiten predecir la MDO de una manera más efectiva que de forma individual.

Adicionalmente, esta investigación confirma que aún no es posible generar una hipótesis sobre los rasgos que siempre desencadenan la MDO o sobre los OD que nunca la acepten. Esto se concluye a partir de la obtención de pares {verbo conjugado, objeto} con rasgos prototípicos de un OD marcado que generan duda en los hablantes, así como OD con rasgos que, de acuerdo con la teoría, suelen rechazar la marcación pero que siempre aparecen marcados. La variación de este fenómeno hace que no sea posible encontrar una sola causa que aplique a todas las variedades del español o a todos los casos, como ya se ha mencionado en la literatura. Por el contrario, sí se puede concluir que hay una fuerte tendencia a omitir la marcación (86,53 % de omisión en los datos).

Este estudio se consolida como base y ejemplo para futuras investigaciones sobre la marcación diferencial de objeto. En este se ha presentado una alternativa de metodología que permite abarcar la mayor cantidad de datos que se ha utilizado para estudiar este fenómeno en todas las variedades del español, del que se tenga conocimiento. Al utilizar una sola fuente de datos para representar todas las variedades del español, este estudio es uniforme a nivel de estilo, tiempo y modalidad en la información recolectada. El corpus obtenido de Twitter

representa el uso informal de la lengua española, lo cual ha sido poco estudiado hasta el momento y es de relevancia por su actualidad. De igual manera, se ha logrado analizar un mayor número de rasgos al que comúnmente se utiliza para los estudios de este fenómeno, dado que se incluyeron rasgos tanto de los verbos como de los objetos. Esta metodología de recolección y análisis de los datos puede servir como modelo de investigación para el análisis de la MDO en otras lenguas.

Resulta deseable que en las investigaciones futuras se integre el análisis del rasgo de especificidad, el cual ha sido considerado por varios autores pero que, por la manera en la que se recogieron los datos, no fue posible incluir en este estudio. De igual manera, para analizar los casos que generan dudas en los hablantes, se podrían incluir factores de tipo pragmático que se pueden encontrar al analizar los tuits completos y no únicamente los elementos que aparecen antes y después de la preposición.

Finalmente, se consolida una base de pares de 4.967 verbos y objetos anotados que pueden ser fuente de estudio de este u otros fenómenos. Asimismo, los hallazgos obtenidos abren una nueva perspectiva de investigación interesante para proponer hipótesis que permitan explicar el origen o generar una explicación de la tensión armónica que regula la MDO en el español.

7. Referencias bibliográficas

- Aissen, J. (2003). Differential object marking: iconicity vs. economy. *Natural Language and Linguistic Theory* 21 (3), pp. 435-483. <http://dx.doi.org/10.1023/A:1024109008573>
- Balasch, S. (2011). Factors determining Spanish differential object marking within its domain of variation. En J. Michnowicz & R. Dodsworth (eds.), *Selected proceedings of the 5th Workshop on Spanish Sociolinguistics* (pp. 113-124). Somerville, Cascadilla Press.
- Bird, S. Klein, E. y Loper, E. (2009) *Natural Language Processing with Python*. O'Reilly Media.
- Bosque, I., y Gutiérrez-Rexach, J. (2016). *Fundamentos de sintaxis formal*. Madrid: Akal.
- Bossong, G. (1991). Differential object marking in Romance and beyond, in D. Wanner & D. Kibbee (eds.), *New analyses in Romance Linguistics: Selected papers from the XVIII Linguistics Symposium on Romance Linguistics*. Amsterdam, John Benjamins, pp. 143-170.
- Bybee, J. (2011). Markedness: Iconicity, Economy, and Frequency. En Jung Song, J. (Ed.) *The Oxford handbook of Linguistic Typology* (pp. 131-147). Nueva York: Oxford University Press.
- Camacho, J. (2018). *Introducción a la sintaxis del español*. Rutgers University, New Jersey: Cambridge University Press.
- Cohen, Paul R.; Howe, Adele E. (1988-12-15). "How Evaluation Guides AI Research: The Message Still Counts More than the Medium". *AI Magazine*. 9(4). pp. 35-43. doi:10.1609/aimag.v9i4.952.
- Croft, W. (2003). *Typology and Universals*. Cambridge: Cambridge University Press.
- Damgaard, C. F. y Irvine, K. M. (2019). Using the beta distribution to analyse plant cover data. *Journal of Ecology*, 107(6), 2747-2759.
- Della Costanza, Mario Andrés. (2016). *La marcación diferencial del objeto (DOM) en español - ¿Una construcción con varios significados?* Universidad of Zurich, Facultad de Artes.

- Fábregas, A. (2013). Differential Object Marking in Spanish: state of the art. *Borealis – An International Journal of Hispanic Linguistics*, 2(2), 1-80. <https://doi.org/10.7557/1.2.2.2603>
- Falls, L. (1974). The beta distribution: A statistical model for world cloud cover. *Journal Of Geophysical Research*, 79(9), 1261-1264. doi: 10.1029/jc079i009p01261
- Fernández Ramírez, S. (1986 [1951]). Gramática española 4. *El verbo y la oración*. Madrid, Arco Libros.
- Garachana, M. (2008). Gramática y pragmática en el empleo del artículo en español. *Revista de didáctica ELE*, 7, 1-19. <https://marcoele.com/descargas/7/garachana.pdf>
- Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. En Greenberg (ed.), *Universals of Language*, pp. 73–113.
- Hopper, P., y Thompson, P.J. (1980). Transitivity in Grammar and Discourse. *Language*, 56, 251 - 299.
- Jiménez S., Dueñas G., Gelbukh A., Rodríguez-Díaz C.A. y Mancera S. (2008). Automatic Detection of Regional Words for the Pan-Hispanic Spanish on Twitter. En Simari G., Fermé E., Gutiérrez Segura F., Rodríguez Melquiades J. (eds), *Advances in Artificial Intelligence - IBERAMIA 2018*. IBERAMIA 2018. Lecture Notes in Computer Science, 11238. Springer, Cham. doi: 10.1007/978-3-030-03928-8_33
- Johnson, N., Kotz, S., y Balakrishnan, N. (1995). Capítulo 25: Beta Distributions. *Continuous univariate distributions* Vol. New York: J. Wiley.
- Keeping, E. S. (2010). Introduction to Statistical Inference. Dover Publications.
- Lizárraga, G. y Mora-Bustos, A. (2010). Variación en la marcación diferenciada de objeto en español. *Forma y Función*, 23(1), pp. 9-38.
- Laca, B. (1995). Sobre el uso del acusativo preposicional en español, in C. Pensado (ed.), *El complemento directo preposicional*. Madrid, Visor, pp. 61-91.
- Laca, B. (2006). El objeto directo. La marcación preposicional, in C. Company (dir.), *Sintaxis histórica de la lengua española. Parte 1. La frase verbal*. México D. F., Universidad Autónoma de México / Fondo de Cultura Económica, pp. 423-479.

- Levinsohn, S. y Tandioy, F. (2000). Inga. En: González de Pérez, M. y Rodríguez de Montes, M. (Eds.), *Lenguas indígenas de Colombia: una visión descriptiva* (pp. 121-134). Bogotá: Instituto Caro y Cuervo.
- López, L. (2012). *Indefinite objects. Scrambling, choice functions and differential marking*. Cambridge (Mass.), MIT Press.
- Mayorga, J. (2004). *Inferencia Estadística* (p. 268). Bogotá, Colombia: Universidad Nacional de Colombia.
- Montes Giraldo, J. (2006). Sobre el objeto directo preposicional. *Boletín de Filología*, 41, 63-76.
- Montes, M. (2004) Lengua ticuna: resultados de fonología y morfosintaxis. *Forma y Función* 17. 147-178.
- Moravcsik, E. (2011). Explaining language universals. En Jung Song, J. (Ed.) *The Oxford handbook of Linguistic Typology* (p. 87). Nueva York: Oxford University Press.
- Montrul, S., y Bowles, M. (2009). Back to basics: Incomplete knowledge of Differential Object Marking in Spanish heritage speakers. *Bilingualism: Language and Cognition*, 12(3), 363-383. doi:10.1017/S1366728909990071
- Næss, A. (2007). *Prototypical transitivity*. Amsterdam: Benjamins.
- Odria, A. (2014). Differential Object Marking and the nature of dative Case in Basque varieties. *Linguistic Variation* 14(2), 289-317.
- Ormázabal, J., y Romero, J. (2013). Differential Object Marking, Case and Agreement. *Borealis – An International Journal of Hispanic Linguistics*, 2(2), 221-239. doi: 10.7557/1.2.2.2808
- Owen, C. E. B. (2008). *Parameter estimation for the beta distribution*. (Tesis de maestría). Brigham Young University, Provo, Utah.
- Panas, E. (2005). Generalized beta distributions for describing and analysing intraday stock market data: testing the U-shape pattern. *Applied Economics*, 37(2), 191-199. doi: 10.1080/00036840412331315006

- Python Software Foundation (2018). Python Language Reference, version 3.6. Disponible en <http://www.python.org>
- Real Academia Española [RAE] y Asociación de Academias de la Lengua Española [ASALE] (2009). *Nueva gramática de la lengua española*. Madrid, Espasa Libros.
- Rodríguez, C. (2018). *Dialectonos: búsqueda de fronteras dialectales estadísticamente significativas usando corpus léxicos geolocalizados*. Tesis de maestría, Bogotá, Instituto Caro y Cuervo.
- Rodríguez-Mondoñedo, M. (2007). *The syntax of objects: agree and Differential Object Marking*. Tesis doctoral, Storrs, University of Connecticut.
- Silverstein, M. (1976). 'Hierarchy of Features and Ergativity', in R. M.W. Dixon (ed.), *Grammatical Categories in Australian Languages*, Australian Institute of Aboriginal Studies, Canberra. Pp. 112–171.
- Seifart, F. (2005) *The structure and use of shape-based noun classes in Miraña (North West Amazon)* (Tesis de doctorado). Max Planck Institute for Psycholinguistics. Nijmegen, Países Bajos.
- Torrego, E. (1998). *The dependencies of objects*. Cambridge (Mass.), MIT Press.
- Tigău, M. (2010) Towards an Account of Differential Object Marking in Romanian. *Bucharest Working Papers in Linguistics 12*(1). Pp. 137- 158.
- Zappavigna, M. 2012. *Discourse of Twitter and Social Media: How to Use Language to Create Affiliation on the Web*, London: Continuum

Apéndice

Verbos (en infinitivo) relacionados con la MDO obtenidos en el corpus de Colombia

Abandonar, abrazar, aburrir, acelerar, aceptar, admirar, adorar, aguantar, ahogar, ahuyentar, alejar, alimentar, alterar, amar, aplastar, apoyar, apreciar, arruinar, asesinar, asfixiar, atacar, atraer, ayudar, bendecir, besar, borrar, buscar, cambiar, castigar, comer, comparar, comprender, condenar, confundir, conocer, conquistar, consentir, conservar, convertir, corregir, crear, creer, criticar, cuidar, culpar, dañar, definir, dejar, descifrar, describir, descuidar, despertar, destruir, detener, disfrutar, distinguir, dividir, elegir, enamorar, encontrar, enfrentar, engañar, engrandecer, enriquecer, entender, entrenar, envolver, escoger, esconder, escuchar, esperar, evaluar, extrañar, golpear, hacer, herir, humillar, hundir, ignorar, imitar, impresionar, insultar, interrumpir, invitar, jugar, juntar, juzgar, lastimar, llamar, maldecir, maltratar, matar, merecer, mirar, morder, necesitar, negar, odiar, oír, olvidar, pegar, perder, perdonar, perseguir, preferir, presentar, prostituir, proteger, quemar, querer, rechazar, recordar, respetar, sacar, saludar, salvar, seducir, seguir, soltar, soportar, sorprender, sostener, stalkear, subestimar, sujetar, superar, temer, tener, traicionar, tratar, usar, valorar, vencer, ver.