

Estado de la publicación: El preprint ha sido publicado como artículo en una revista  
DOI del artículo publicado: <https://doi.org/10.1590/1678-460x202238252279>

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

Stephanie A. Puentes, Daniel Eduardo Chaves Peña, Sergio Jimenez

<https://doi.org/10.1590/1678-460x202152279>

Enviado en: 2021-08-08

Postado en: 2021-08-09 (versión 1)

(AAAA-MM-DD)



<https://doi.org/10.1590/1678-460x202152279>

## **Análisis empírico de la MDO en el español: tensión armónica entre iconicidad y economía**

### **Empirical analysis of DOM in Spanish: Iconicity and economy forces driven by a harmonic tension**

Stephanie A. Puentes<sup>1</sup>

Daniel Eduardo Chaves Peña<sup>2</sup>

Sergio Jimenez<sup>3</sup>

#### **Resumen**

La marcación diferencial del objeto es un fenómeno gramatical de varias lenguas como el español, en la cual el morfema ‘a’ aparece ante 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, el cual se nombra en este estudio como 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.

Palabras clave: marcación diferencial de objeto, corpus de Twitter en español, economía, iconicidad, tensión armónica

<sup>1</sup> Instituto Caro y Cuervo. Bogotá – Colombia.

<https://orcid.org/0000-0003-4155-9751>. [stepha25@gmail.com](mailto:stepha25@gmail.com)

<sup>2</sup> Instituto Caro y Cuervo. Bogotá – Colombia.

<https://orcid.org/0000-0002-0912-5542>. [daniel.chaves@caroycuervo.gov.co](mailto:daniel.chaves@caroycuervo.gov.co)

<sup>3</sup> Instituto Caro y Cuervo. Bogotá – Colombia.

<http://orcid.org/0000-0003-3709-7881>. [sergio.jimenez@caroycuervo.gov.co](mailto:sergio.jimenez@caroycuervo.gov.co)

## Abstract

Differential Object Marking (DOM) is a linguistic phenomenon of many languages like Spanish, in which the preposition ‘a’ appears before direct objects. Current theory states 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 statistical analysis, it was discovered that this phenomenon is ruled by two highly correlated ( $r=0,92$ ) tensions that can be associated with iconicity (13,47%) and economy (86,53%). This finding suggests that these tensions are not independent but rather controlled by a common factor named, in this study, 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.

Keywords: differential object marking, Spanish twitter corpus, economy, iconicity, harmonic tension

## 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, del verbo de la oración e, incluso, de la variedad de la lengua en la que ocurra este fenómeno.

La MDO se puede encontrar en lenguas como el vasco (Odria, 2014) o el rumano (Tigău, 2010), e incluso en lenguas indígenas colombianas como el inga (Levinsohn y Tandioy 2000), el ticuna (Montes, 2004) o el miraña (Seifart, 2005), y ha sido ampliamente estudiada desde perspectivas generativistas y cognitivistas (Bossong, 1991; Torrego, 1998; Aissen, 2003; Laca, 1995; Rodríguez-Mondoñedo, 2007; López, 2012; Costanza, 2016; entre otros), así como variacionistas (Lizárraga y Mora-Bustos, 2010; Balach, 2011). 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 se realizan con pocos ejemplos, usualmente contruidos por el investigador, y solo algunas han optado por utilizar corpus que permitan analizar el fenómeno desde el uso real de los hablantes y a gran escala. Asimismo, es posible evidenciar un vacío respecto a estudios realizados sobre este fenómeno en el español de Colombia (además del propuesto por Montes, 2006) 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 con el fin de generar jerarquías conjuntas. Así, creemos que es necesario ampliar el conocimiento que se tiene de la MDO analizando una fuente grande de observaciones empíricas haciendo uso de técnicas estadísticas.

El presente estudio pretende analizar las construcciones sintácticas de un corpus obtenido de *Twitter*, una red social que permite dar cuenta del uso real y actual del español por parte de una gran cantidad de usuarios. Allí se busca encontrar de manera empírica qué factores semánticos establecen una variabilidad significativa sobre la presencia y la ausencia de la preposición *a* en las construcciones sintácticas encontradas, así como qué jerarquía se puede establecer entre estos rasgos, teniendo en cuenta su grado de incidencia en la MDO.

Adicionalmente, y con el fin de establecer esa jerarquía, se toman como base los postulados de Hopper y Thompson (1980) y Næss (2007), en los cuales 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.

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, y si existe alguna fuerza que permita, por primera vez, establecer una relación de dependencia entre estos dos principios, como se explicará a continuación en la metodología de este estudio.

## 2. Referentes teóricos

### *La Marcación Diferencial de Objeto en el español*

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.

### *Aspectos semánticos de los OD*

La MDO ha sido comúnmente asociada a 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) han determinado 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 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.

### *Animacidad y definitud*

Aissen (2003) establece las escalas (2) de animacidad 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.

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

Por su parte, 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 (3) para clasificar aquellos OD que son más susceptibles de recibir la MDO.

(3) Escala de definitud: Pronombre > Nombre > Definido > Específico indefinido > No específico en donde > indica que el rasgo a la izquierda tiene mayor incidencia en la MDO.

Sin embargo, algunos autores han encontrado casos que contradicen la idea de que los OD con este rasgo van siempre marcados, como los de Laca (2006, pp. 452) en (3), 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).

### *Iconicidad y economía*

Los principios de iconicidad y economía pueden verse como dos fuerzas que regulan el procesamiento de la información lingüística. Tanto en el plano sintáctico como en el morfológico, las formas más frecuentes en el uso se expresan de manera más simplificada que aquellas que destacan cierto tipo de información, ya sea por razones semánticas o pragmáticas del evento o la entidad denotadas. La iconicidad es el principio que obliga a utilizar una marca o una estructura distinta para poner de relieve la anomalía o novedad de lo expresado. En nuestro caso, la iconicidad se diferencia de la prototipicidad en que esta última define los rasgos prototípicos del sujeto en una oración transitiva, o sea una categoría a nivel general, mientras que la iconicidad es el reconocimiento o no por parte de un hablante o una población de la pertenencia de casos particulares a dicha categoría (Neset, 2000). En el análisis de las estructuras oracionales, un ejemplo claro de este principio es la contraposición entre la forma activa (no-marcada) y la forma pasiva de las oraciones. En términos de presentación de la información, la forma más económica es sin duda la forma declarativa, pero dados los propósitos pragmáticos del usuario del lenguaje, puede resultar más importante destacar a uno de los participantes del evento (el paciente) frente al agente y por tanto se utiliza la forma pasiva. La naturaleza de los rasgos y mecanismos que determinan la iconicidad de una estructura sintáctica o morfológica depende obviamente del fenómeno en cuestión. Desde la teoría de la optimalidad, Aissen (2003) afirma que, con respecto a la MDO, la fuerza que ejerce el principio de iconicidad actúa cuando el objeto es prominente y causa que este se marque, mientras que, en el principio de economía, se evita realizar esta marcación 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 ellos rige cada caso se resuelve a través de jerarquías entre los rasgos del OD como

los presentados en (2) y (3). Entre más animado, específico y definido sea un objeto, es más probable que resulte marcado.

#### *Otros rasgos del objeto*

En esta investigación también se tuvo en cuenta si el OD se presentaba como un *sustantivo escueto*, es decir, un sintagma nominal que aparece sin determinante (Garachana, 2008), y que 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*.

#### *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 que se utilizó en 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 1.

**Tabla 1** — Parámetros de transitividad presentados por Hopper & Thompson (1980, p. 252)

<b>Parámetro</b>	<b>Alta transitividad</b>	<b>Baja transitividad</b>
. Participantes	dos o más participantes	un participante
. Kinesis	acción	no acción
. Aspecto	télico	atélico
. Puntualidad	puntual	no puntual
. Volicionalidad	volicional	no volicional

Afirmación	afirmativa	negativa
Modo	real	irreal
Agentividad	agente alto en potencia	agente bajo en potencia
Afectación del Objeto	objeto totalmente afectado	no afectado
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 *volitionalidad* 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. Otra manera de designar este mecanismo de desambiguación es el de iconicidad, pues en la situación anómala o excepcional de un sujeto y un objeto con rasgos semejantes, la iconicidad dictamina la marcación del objeto para distinguirlo del otro participante en la construcción transitiva.

Adicionalmente, Costanza (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.

(5) Juan rasca a Mairena.

(6) 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).

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

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

En el presente estudio se tuvieron en cuenta los rasgos de *aspecto*, el cual se refiere a la telicidad de la acción, es decir, si es un evento delimitado o completo; *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); *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); *afirmación*, es decir, si la acción es afirmativa o negativa; *modo*, el cual se refiere a la distinción entre *realis* e *irrealis*, y

*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.

### **3. Metodología**

#### *Recolección de datos*

##### *El corpus*

Jimenez et al. (2018) recolectaron un corpus de manera semiautomática en la página web de Twitter, compuesto por 217'928.418 tuits de 333 ciudades con más de 100.000 habitantes, en 21 países en los que se habla español. Dicho corpus contiene tuits generados en su mayoría por usuarios individuales o institucionales de dicha red social así como contenido generado por cuentas automatizadas. Estas cuentas automatizadas se caracterizan por producir grandes cantidades de contenido. En una inspección manual del corpus, entre los 50 usuarios más prolíficos, 32 resultaron ser cuentas automatizadas para reporte de tendencias, clima, etc. En general el contenido generado por estas cuentas es una plantilla textual como “<#hashtag> es ahora una tendencia en <lugar>” y “<#hashtag> acaba de convertirse en TT ocupando la <número><sup>a</sup> posición en <lugar>”. Inspeccionando las plantillas utilizadas por estas 32 cuentas automatizadas, no encontramos apariciones de la MDO, por lo tanto consideramos que no es necesario remover este contenido para el análisis a realizar. Este 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.

##### *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:

Método para encontrar los casos de **marcado** del OD en el corpus de tuits en Colombia.

1. Se tokenizaron (segmentaron) todos los tuits con el método TweetTokenizer de la librería NLTK1 para Python, con el fin de identificar unidades lingüísticas, en este caso, palabras, signos de puntuación, emoticones, etc.
2. Se identificaron todas las apariciones de los tokens a y al.
3. Para cada aparición, se recolectaron contextos de cinco unidades: dos tokens a la izquierda y dos a la derecha de los tokens a y al, adicionales a estos (por ejemplo, [...] que destruyen a una persona).
4. 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.
5. De manera manual se identificaron los verbos que aparecieran una o dos posiciones antes de a o al.
6. 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 (por ejemplo, tenerte a

---

<sup>1</sup> <https://www.nltk.org/api/nltk.tokenize.html>

mi lado). En general, se mantuvieron todos los 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 conjugaciones 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:

7. De manera manual, se creó un listado de todas las conjugaciones de los 129 verbos encontrados.
8. Con un programa en Python se buscaron contextos de seis unidades que consistieron en cada una de las conjugaciones de los verbos más dos tokens 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.
9. 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.
10. 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.
11. 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.

12. 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}.
13. Finalmente, se obtuvieron 4.967 pares los cuales se utilizaron en el análisis estadístico.

### *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 2. 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 pequeños<sup>2</sup>. 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 2, con el fin de identificar en qué medida estas 4.967

---

<sup>2</sup> 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.

combinaciones encontradas eran utilizadas en los demás países. En la Tabla 3 se presentan las cantidades de pares que aparecen y los que no aparecen en el corpus de cada país.

**Tabla 2** — Apariciones de combinaciones {verbo conjugado, objeto} en las variedades del español por país

<b>País</b>	<b>ISO</b>	<b>Apariciones de pares en corpus</b>	<b>Pares que no aparecen</b>
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
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

*Sistematización 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 2.1.2. y se establecieron los siguientes rasgos binarios:

- Rasgos del OD<sup>3</sup>: ±humano, ±animado, ±definido, ±nombre propio, ±singular, y ±sustantivo escueto.
- Rasgos del verbo: ±modo, ±afirmativo, ±telicidad, ±volicionalidad, ±afectación, ±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 3 se presentan algunas consideraciones importantes para la anotación de estas propiedades.

**Tabla 3** — Anotación de rasgos en las combinaciones {verbo conjugado, objeto} obtenidas en el corpus

Rasgo <sup>4</sup>	Método de anotación de combinaciones que sí cumplen con el rasgo	Ejemplo del corpus con rasgo binario +
1. Humano (H <sub>o</sub> )	OD de persona	mis hijos, alguien
2. Animado (AN <sub>o</sub> )	OD de entidades vivas o con movimiento	animales, la selección
3. Definido (D <sub>o</sub> )	OD con sustantivos acompañados de artículos definidos	los recuerdos
4. Nombre propio (NP <sub>o</sub> )	OD con sustantivos propios de personas, animales, películas, entre otros	Caperucita, Cartagena
5. Singular (S <sub>o</sub> )	OD con sustantivos singulares	tu casa, Dora
6. Sustantivo escueto (SE <sub>o</sub> )	OD con sustantivos sin determinantes que no sean nombres propios	rosas, vida
7. Modo ( <i>Realis</i> e <i>Irrealis</i> ) (M <sub>v</sub> )	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

<sup>3</sup> 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 ±específico.

<sup>4</sup> El subíndice en paréntesis en la nomenclatura de cada rasgo indica si el rasgo pertenece al objeto (o) o al verbo (v).

8. Afirmativo (AFI <sub>v</sub> )	Combinaciones {verbo conjugado, objeto} que no estuvieran acompañados de formas como <i>no</i> o <i>nunca</i>	cambiar, quiero
9. Telicidad (T <sub>v</sub> )	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 <sub>v</sub> )	Acciones que se realicen con intención o que impliquen una decisión por parte del agente	destruir, elegir
11. Afectación (AFE <sub>v</sub> )	Acciones que causan que el OD experimente un cambio de estado	herimos, lastimes
12. Puntualidad (P <sub>v</sub> )	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 4. De esta manera fue posible recopilar la información sobre las características de los verbos y OD obtenidos en la búsqueda en el corpus, así como la cantidad de veces que el OD apareció marcado y no marcado.

**Tabla 4** — Ejemplos de combinaciones {verbo conjugado, objeto} anotados y cantidad de apariciones en los datos de Colombia

{verbo conjugado, objeto}	H <sub>o</sub>	AN <sub>o</sub>	D <sub>o</sub>	NP <sub>o</sub>	S <sub>o</sub>	SE <sub>o</sub>	M <sub>v</sub>	AFI <sub>v</sub>	T <sub>v</sub>	V <sub>v</sub>	AFE <sub>v</sub>	P <sub>v</sub>	#	# no marc
{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
{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

## *Normalización de los datos*

### *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 afecten las estadísticas. El problema de 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 3.5) 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 naturalmente 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 ilustramos el método de normalización descrito. El país en cuyo corpus se encontró el menor número de apariciones fue Guatemala, con ( $T'=18,372$ ) casos de marcado de la MDO. En comparación, Colombia y México tuvieron en total  $T_{col} = 103,570$  y  $T_{mex} = 155,051$  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  $\{elegir, alguien\}$  del cual este objeto fue marcado en Guatemala 10 veces,

$M_{\{elegir,alguien\}} = 58$  en Colombia, y  $M_{\{elegir,alguien\}} = 73$  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.

*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 aparición 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. Además, el uso del porcentaje de marcado simplifica el análisis a una sola variable en lugar de tener que analizar las frecuencias de marcación y omisión por separado.

#### *Función de densidad de probabilidad Beta*

La distribución Beta 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 tal como el “porcentaje de marcado”, en el cual cada caso está delimitado entre el 0% y el 100%. La función de densidad de probabilidad de la distribución Beta se define por:

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

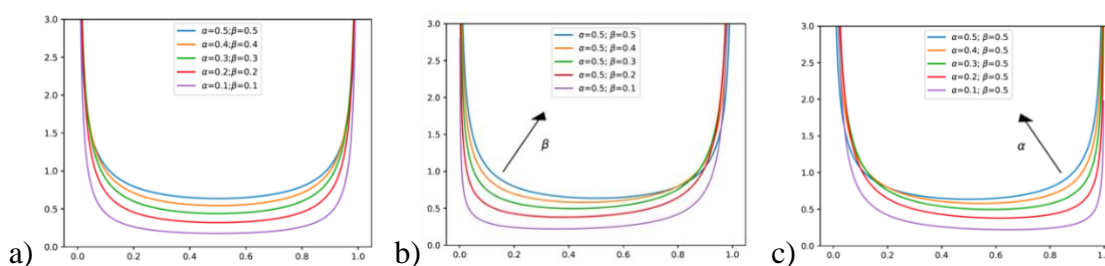
en donde  $K$  es una constante de normalización según la siguiente expresión:

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

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. Más adelante veremos que el “porcentaje de mercado” se comporta de esa manera.

**Figura 1** — Parámetros de la distribución Beta en forma de U a) caso simétrico, b) caso asimétrico a la izquierda controlado por  $\beta$ , y c) caso asimétrico a la derecha controlado por  $\alpha$



Como se aprecia en la Figura 1, los parámetros  $\alpha$  y  $\beta$  establecen la forma de la función de la distribución Beta. Esto es análogo a el caso de la distribución normal donde los parámetros  $\mu$  y  $\sigma$  dan forma a la campana invertida de la función de distribución. El parámetro  $\mu$  (media), que se estima con el promedio de la población, establece la posición horizontal de la campana invertida, y  $\sigma^2$  (varianza) estimada con la desviación estándar, controla que tan empinada o aplanada es la campana invertida. De manera similar  $\alpha$  y  $\beta$  controlan la forma de la función de distribución Beta y el método para estimar sus valores a partir de una muestra de la población se muestra más adelante. En el caso en el que  $\alpha$  es igual a  $\beta$  (Figura 1a) la forma de la “U” es simétrica y si los valores de  $\alpha$  y  $\beta$  aumentan, la “U” sube, aumentando la probabilidad de los valores intermedios y disminuyendo la probabilidad de los extremos. En contraste, cuando  $\alpha$  y  $\beta$  disminuyen, la “U” baja, reduciendo la probabilidad de los valores intermedios y aumentando la de los extremos.

Cuando  $\alpha$  y  $\beta$  son diferentes, la forma de la “U” es asimétrica. Las Figuras 1b y 1c muestran que el codo izquierdo de la “U” es controlado por  $\beta$  y el derecho por  $\alpha$ . A medida

que  $\beta$  aumenta, el “codo” se hace más redondeado y se desplaza hacia arriba haciendo que la probabilidad de los valores cercanos a 0 aumente.

Más adelante veremos que el “porcentaje de marcado” se distribuye según Beta  $\alpha < 1, \beta < 1$  y  $\alpha < \beta$ .

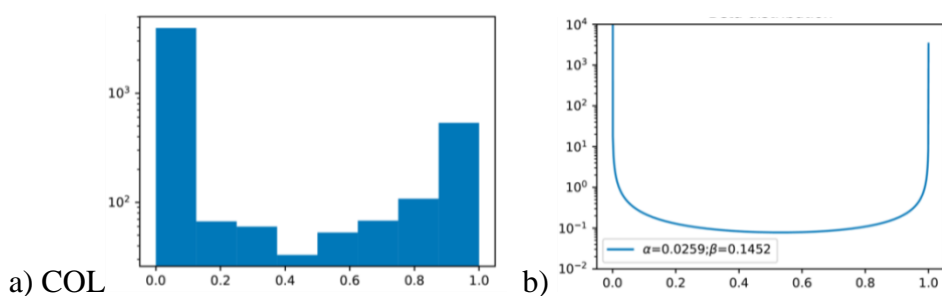
La estimación de los parámetros de la distribución Beta se hace por el método de los momentos (Owen, 2008. Ecuaciones 2.19 y 2.20) con las expresiones

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

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

En la Figura 2a se muestra un histograma en forma de “U” de frecuencias de la variable *porcentaje de marcado* correspondiente a los datos de Colombia. En la Figura 2b, se muestra la forma de la distribución Beta con los valores estimados de  $\alpha$  y  $\beta$  para los datos de Colombia. En todas las figuras se usó una escala logarítmica en la ordenada 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



#### 4.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 hallazgos respecto a la marcación opcional de OD en los países observados.

### *Distribución de probabilidad del porcentaje de mercado*

La Tabla 5 muestra los resultados de las estimaciones de los parámetros  $\alpha$  y  $\beta$  para cada uno de los países analizados. En esta se puede observar que, para todos los países,  $\alpha$  y  $\beta$  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 ( $\underline{x}$ ) y la varianza ( $\underline{y}$ ) como referencia, dado que con estos valores se obtienen las estimaciones de  $\alpha$  y  $\beta$ , 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 (Mayorga, 2004), 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 a la distribución Beta<sup>5</sup>.

**Tabla 5** — Estadísticas de la función de densidad de probabilidad del porcentaje de mercado en las variedades del español

ISO	$\hat{\alpha}$	$\hat{\beta}$	Test	
			KS	$p^*$

<sup>5</sup> 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>)

AR	0,022	0,125	0,018	0,115
G	8	0	7	5
	0,023	0,125	0,012	0,572
CHL	5	5	2	9
	0,025	0,145	0,012	0,466
COL	9	2	2	1
	0,021	0,116	0,009	0,915
CRI	7	7	1	8
	0,029	0,145	0,011	0,538
ECU	2	0	9	4
GT	0,024	0,115	0,010	0,679
M	0	5	9	8
ME	0,023	0,113	0,014	0,323
X	2	2	4	3
	0,029	0,149	0,015	0,229
NSH	5	4	5	9
	0,025	0,127	0,010	0,769
PAN	5	3	0	7
	0,026	0,128	0,010	0,740
PRY	0	2	4	3
	0,022	0,111	0,009	0,853
PER	7	9	3	9
	0,028	0,152	0,017	0,176
PRI	1	8	1	8
DO	0,029	0,165	0,012	0,476
M	3	5	7	8
	0,023	0,120	0,014	0,393
ESP	9	5	1	5
	0,032	0,170	0,015	0,202
USA	5	2	9	7
UR	0,026	0,144	0,012	0,609
Y	1	0	4	7
VE	0,027	0,144	0,015	0,233
N	5	8	2	6

\*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  $\alpha$  y  $\beta$ . Según lo comentado para la Figura 1 en la sección 3.5,  $\hat{\alpha}$  y  $\hat{\beta}$  se pueden interpretar como fuerzas que dan forma a la función de distribución de probabilidad. En la Tabla 6 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.

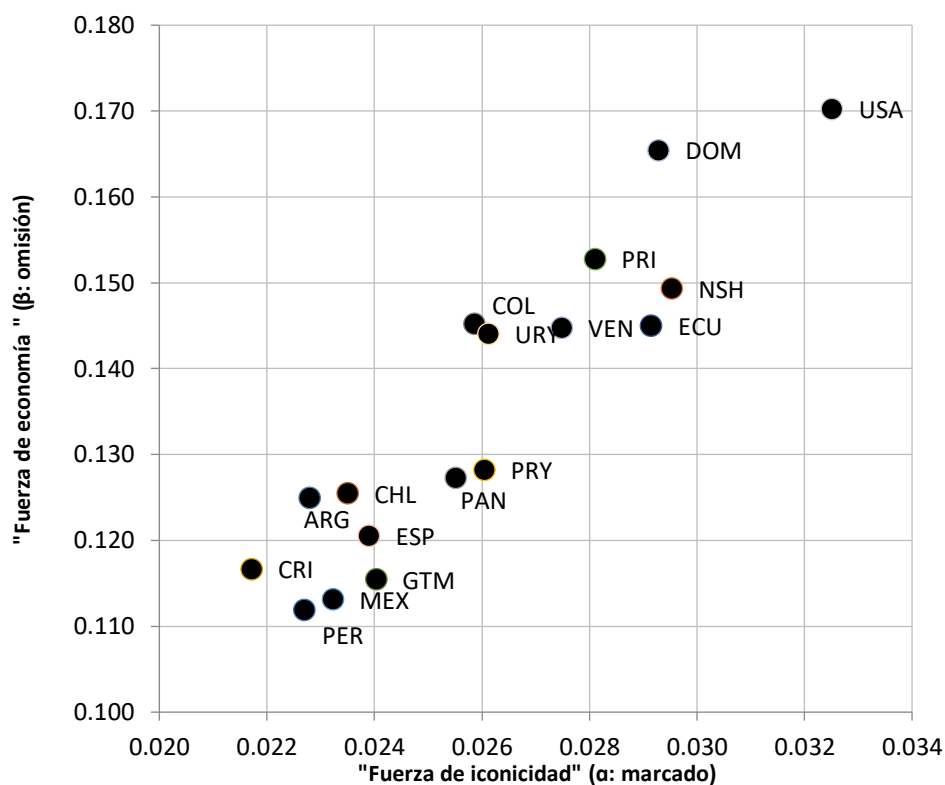
De igual manera, como se indicó en la sección 2.1.1.2, la teoría actual sugiere que el fenómeno de la MDO está sujeto a las fuerzas opuestas de iconicidad y economía. Dado que  $\alpha$  da forma al codo derecho de la función de distribución de probabilidad, cuyos valores están cercanos al valor máximo (100% del porcentaje de mercado), es pertinente relacionar a  $\alpha$  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  $\beta$ , 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 3 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  $\alpha$  y la de economía representada por  $\beta$ . 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 3** — Variación de la “Fuerza de iconicidad” representada por la estadística  $\alpha$  versus la “Fuerza de economía” representada por  $\beta$ , 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 mayor intensidad de esta fuerza, los extremos de total marcación y de total omisión se reconcilian, por tanto, es posible nombrar esta fuerza como **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:

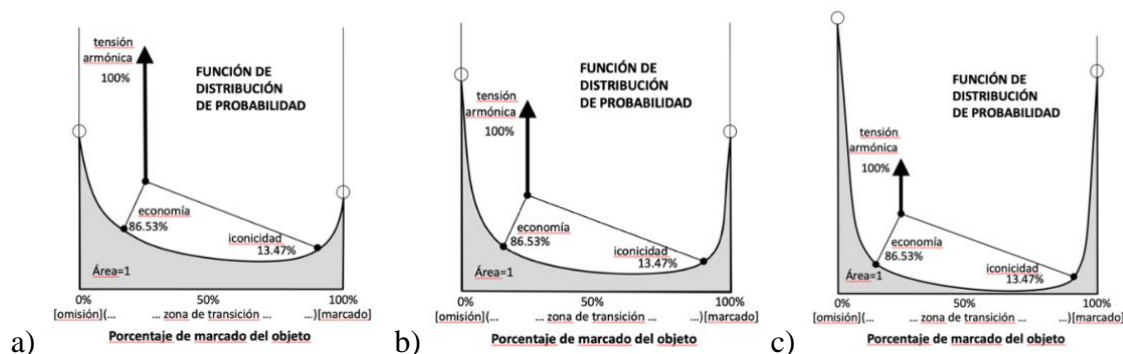
$$\text{economía} = 0,8653H - 0,0042$$

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

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 evidencia un pequeño sesgo de 0,0042 que se remueve de la economía y favorece la iconicidad.

La Figura 4 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 4** — 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 proporcionalmente 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 es siempre igual a 1. Cuando la tensión armónica es fuerte la función es halada hacia arriba haciendo que los codos de la función sean redondeados, mientras que, los extremos de porcentaje de marcado descienden.

Al disminuir la tensión, la función desciende, sus codos se hacen más agudos y los extremos ascienden. Este comportamiento se da por la restricción del área constante bajo la curva, lo que se hace más evidente cuando la tensión es débil. 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 "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.

*Jerarquías de rasgos asociados a la MDO*

Con el fin de comprender 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. 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 entre los valores reales del porcentaje de marcado y las predicciones del modelo lineal. 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 (estudio de ablación). 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 estudio 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 6, 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 ( $V_v$ ) resultó considerable.

**Tabla 6** — Importancia relativa de los rasgos en el modelo predictivo lineal para el porcentaje de marcado

<b>Rasgos</b>											
<b>H<sub>o</sub></b>	<b>AN<sub>o</sub></b>	<b>V<sub>v</sub></b>	<b>D<sub>o</sub></b>	<b>SE<sub>o</sub></b>	<b>AFI<sub>v</sub></b>	<b>NP<sub>o</sub></b>	<b>S<sub>o</sub></b>	<b>T<sub>v</sub></b>	<b>AFE<sub>v</sub></b>	<b>M<sub>v</sub></b>	<b>P<sub>v</sub></b>
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 después de analizar los rasgos con el estudio de ablación es la que se observa en (9).

- (9) Humano > Animado > Volicionalidad > Definitud > Sustantivo Escueto > 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 la 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* 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 escueto* rasgo asociado a la definitud como se indicó en 2.1.1.3, 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 2.1.2) 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 esto, 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 Costanza (2016) referentes a las hipótesis de la transitividad o de Næss (2007), de máxima distinción de argumentos (ver sección 2.1.2), los cuales indican que, a mayor cantidad de rasgos asociados a la transitividad, mayor es la posibilidad de que un OD aparezca marcado.

## 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 que se distribuye acorde a la función de distribución de probabilidad Beta en forma de “U”, y que los parámetros de esta distribución,  $\alpha$  y  $\beta$ , modelan adecuadamente las nociones teóricas de iconicidad y economía. Esta metodología se recomienda para analizar otros fenómenos de marcado binarios.

Asimismo, los resultados obtenidos en relación con las tendencias de marcado en el español de los 17 países corroboran las intuiciones de estudios previos que afirman 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 **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 este fenómeno en español.

Respecto a los rasgos de los OD y de los verbos, se logró corroborar empíricamente que la jerarquía propuesta coincide con las propuestas de estudios previos; además, se adicionaron rasgos que antes no habían sido considerados. 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 y la de máxima distinción de argumentos, según las cuales el objeto 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.

### **Conflicto de intereses**

*(X) Los autores declaran no tener ningún conflicto de intereses.*

### **Contribución de los autores**

*Identificamos el aporte de los autores del artículo dentro de las siguientes categorías: conceptualización del estudio (Stephanie Puentes (SP), Daniel Chaves (DC) y Sergio Jimenez (SJ)), metodología (SP), diseño del estudio (SP, DC, SJ), análisis formal de los datos (SP), análisis estadístico de los datos (SP), supervisión del proyecto (DC, SJ), recopilación de datos (SP, SJ), generación de datos (SP), validación de datos (SP, DC), edición y visualización de datos (SP), redacción y edición final del artículo (SP). Al final, se debe insertar una confirmación de que todos los autores aprueban la versión final del manuscrito y son responsables de todos los aspectos, incluida la garantía de su veracidad e integridad.*

## Referencias

- Aissen, J. (2003). Differential object marking: iconicity vs. economy. *Natural Language and Linguistic Theory*, 21 (3), 435–483. <http://dx.doi.org/10.1023/A:1024109008573>
- Balasz, S. (2011). Factors determining Spanish differential object marking within its domain of variation. In J. Michnowicz & R. Dodsworth (Eds.), *Selected proceedings of the 5th Workshop on Spanish Sociolinguistics* (pp. 113–124). Cascadilla Press.
- Bosque, I., & Gutiérrez-Rexach, J. (2016). *Fundamentos de sintaxis formal*. Akal.
- Bosson, 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* (pp. 143–170). John Benjamins.
- Bybee, J. (2011). Markedness: Iconicity, Economy, and Frequency. In Jung Song, J. (Ed.) *The Oxford handbook of Linguistic Typology* (pp. 131–147). Oxford University Press.
- Camacho, J. (2018). *Introducción a la sintaxis del español*. Cambridge University Press.
- Costanza, Mario Andrés Della. (2016). *La marcación diferencial del objeto (DOM) en español - ¿Una construcción con varios significados?* [Unpublished doctoral dissertation]. Universidad of Zurich. [https://www.zora.uzh.ch/id/eprint/130530/1/della\\_costanza.pdf](https://www.zora.uzh.ch/id/eprint/130530/1/della_costanza.pdf) (accessed 03 August, 2021)
- Croft, W. (2003). *Typology and Universals*. Cambridge University Press.
- Garachana, M. (2008). Gramática y pragmática en el empleo del artículo en español. *Revista de Didáctica Español Lengua Extranjera*, 7, 1–18. <https://marcoele.com/descargas/7/garachana.pdf>

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of Language* (pp. 73–113). MIT Press.

Hopper, P., y Thompson, P.J. (1980). Transitivity in Grammar and Discourse. *Language*, 56, 251–299. <https://doi.org/10.2307/413757>

Jimenez S., Dueñas G., Gelbukh A., Rodriguez-Diaz C.A., & Mancera S. (2018). Automatic Detection of Regional Words for Pan-Hispanic Spanish on Twitter. In: Simari G., Fermé E., Gutiérrez Segura F., & Rodríguez Melquiades J. (Eds.) *Advances in Artificial Intelligence - IBERAMIA 2018. Lecture Notes in Computer Science, vol 11238* (pp. 404–416). Springer. [https://doi.org/10.1007/978-3-030-03928-8\\_33](https://doi.org/10.1007/978-3-030-03928-8_33)

Lizárraga, G. y Mora-Bustos, A. (2010). Variación en la marcación diferenciada de objeto en español. *Forma y Función*, 23(1), 9–38.

Laca, B. (1995). Sobre el uso del acusativo preposicional en español. In C. Pensado (Ed.), *El complemento directo preposicional* (pp. 61–91). Visor.

Laca, B. (2006). El objeto directo. La marcación preposicional. In C. Company (Dir.), *Sintaxis histórica de la lengua española. Parte I. La frase verbal*. (pp. 423–479). Fondo de Cultura Económica.

Levinsohn, S. y Tandioy, F. (2000). Inga. In: González de Pérez, M. & Rodríguez de Montes, M. (Eds.), *Lenguas indígenas de Colombia: una visión descriptiva* (pp. 121–134). Instituto Caro y Cuervo.

López, L. (2012). *Indefinite objects. Scrambling, choice functions and differential marking*. MIT Press.

Mayorga, J. (2004). *Inferencia Estadística* (p. 268). 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.

Nesset, T. (2000). Iconicity and prototypes: A new perspective on Russian verbs of motion. *Scando-Slavica*, 46(1), 105–119. <https://doi.org/10.1080/00806760008601154>

Næss, A. (2007). *Prototypical transitivity*. John Benjamins Publishing.

Odria, A. (2014). Differential Object Marking and the nature of dative Case in Basque varieties. *Linguistic Variation* 14(2), 289–317. <https://doi.org/10.1075/lv.14.2.03odr>

Owen, C. E. B. (2008). *Parameter estimation for the beta distribution* [Unpublished master's thesis]. Brigham Young University.

Real Academia Española [RAE] y Asociación de Academias de la Lengua Española [ASALE] (2009). *Nueva gramática de la lengua española*. Espasa.

Rodríguez-Mondoñedo, M. (2007). *The syntax of objects: agree and Differential Object Marking* [Unpublished doctoral dissertation]. University of Connecticut.

Silverstein, M. (1976). Hierarchy of Features and Ergativity. In R. M.W. Dixon (Ed.), *Grammatical Categories in Australian Languages* (pp. 112–171). Australian Institute of Aboriginal Studies.

Seifart, F. (2005). *The structure and use of shape-based noun classes in Miraña (North West Amazon)* [Unpublished doctoral dissertation]. Max Planck Institute for Psycholinguistics.

Torrego, E. (1998). *The dependencies of objects*. MIT Press.

Tigău, M. (2010). Towards an Account of Differential Object Marking in Romanian. *Bucharest Working Papers in Linguistics* 12(1), 137–159.

This preprint was submitted under the following conditions:

- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The deposited manuscript is in PDF format.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors agree that if the manuscript is accepted and posted on the SciELO Preprints server, it will be withdrawn upon retraction.