

RADBOUD UNIVERSITY NIJMEGEN



LA FACULTAD DE LAS LETRAS

---

# Una interacción de animacidad, definitud y topicalidad

EL DESARROLLO HISTÓRICO DEL USO DE *a* EN EL DOM EN EL CASTELLANO

---

TRABAJO DE FIN DE GRADO PARA EL PROGRAMA DE LA FILOLOGÍA  
HISPÁNICA DE LA UNIVERSIDAD DE RADBOUD DE NIJMEGEN

*Autor:*

Nienke WESSEL  
s4598350

9320 palabras

*Supervisor:*

Dr. Gijs J. MULDER

*Segunda supervisora:*

Dr. Janine K. M. BERNS

31 de diciembre de 2022

## Resumen

El uso de *a* como marcador de los objetos directos en el español se ha estado investigando durante mucho tiempo, pero aún no está claro cómo interactúan los diferentes factores. Este estudio investiga cómo la animacidad y la definitud influyeron en el uso de *a* en el desarrollo del español mediante el análisis de un corpus histórico. Este análisis combina la anotación automática con la anotación manual. Encontramos que los textos bajo investigación generalmente se adhieren a las teorías discutidas, pero que los inanimados definidos a veces tienen *a* lo que los libros de gramática creen que es imposible. Además, la investigación encuentra que para los humanos definidos existe un aumento en el uso de *a* a lo largo del tiempo, aunque los datos limitados dificultan sacar conclusiones definitivas.

## Samenvatting

Het gebruik van *a* als markeerder van directe objecten in het Spaans wordt al langere tijd onderzocht, maar er is tot nu toe nog geen duidelijkheid over hoe verschillende factoren precies elkaar beïnvloeden. Het huidige onderzoek bekijkt de manier waarop animacy en definiteness van invloed zijn op het gebruik van *a* in de historische ontwikkeling van het Spaans door het analyseren van een historisch corpus. Deze analyse wordt gedaan door automatische annotatie aan te vullen met handmatige annotatie. De resultaten geven aan dat de onderzochte teksten voor het grootste deel overeenkomen met de theorieën, maar dat inanimate definite zelfstandig naamwoorden soms *a* gebruiken, iets wat onmogelijk is volgens de grammaticaboeken. Daarnaast laten de resultaten zien dat voor definite humans er inderdaad een toename te zien is in het gebruik van *a* door de tijd heen, hoewel de beperkte hoeveelheid data het moeilijk maakt definitieve conclusies te trekken.

# Índice

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>El marco teórico: desde un sistema de casos hasta <i>a</i></b>	<b>6</b>
2.1	El DOM . . . . .	6
2.2	El desarrollo de <i>a</i> en el castellano . . . . .	8
2.3	Las jerarquías . . . . .	10
2.4	Topicalidad . . . . .	15
2.5	El verbo . . . . .	16
2.6	Conclusión . . . . .	16
<b>3</b>	<b>La metodología: combinar análisis automático y manual</b>	<b>17</b>
3.1	El corpus . . . . .	17
3.2	Metódo de análisis . . . . .	18
3.3	Las oraciones analizadas . . . . .	19
3.4	Aspectos específicos del análisis . . . . .	19
3.5	Conclusión . . . . .	21
<b>4</b>	<b>Los resultados: la animacidad y definitud en la acción</b>	<b>22</b>
4.1	Pronombres . . . . .	22
4.2	Nombres de humanos . . . . .	25
4.3	Definidos humanos . . . . .	25
4.4	Indefinidos no específicos humanos . . . . .	27
4.5	Animales . . . . .	27
4.6	Definidos inanimados . . . . .	27
4.7	Conclusión . . . . .	28
<b>5</b>	<b>La discusión: hacia una visión global del DOM</b>	<b>29</b>
5.1	Implicaciones . . . . .	29

5.1.1	Teoria de Aissen . . . . .	30
5.1.2	Otros factores . . . . .	31
5.1.3	Pronombres . . . . .	32
5.2	Limitaciones . . . . .	33
5.3	Conclusión . . . . .	34
	<b>Bibliografía</b>	<b>35</b>

# Capítulo 1

## Introducción

El uso de la preposición *a* en castellano es difícil de describir. No solo se usa para denotar dirección (en: *voy a la biblioteca*), sino también para marcar los objetos directos e indirectos. Las reglas para marcar los objetos directos son complicadas. Las dos frases siguientes muestran que *a* solo es necesaria en una de las oraciones:

- (1) a. Veo la lámpara.  
b. Veo *a* mi amigo.

En general, es evidente que el tipo de sustantivo es importante para determinar si *a* es necesaria. Parece que se usa con los objetos humanos y no se usa con los objetos que refieran a cosas que no vivan. Sin embargo, investigaciones anteriores han demostrado que la situación es más compleja y también tiene que ver con otros factores. Las gramáticas de referencia intentan explicar los diferentes factores, pero no discuten los patrones generales (Batchelor & San José, 2010; Bosque & Demonte, 1999; Butt et al., 2019).

Si intentamos identificar patrones generales, nos damos cuenta de que el español no es la única lengua que marca ciertos objetos directos. En la lingüística, se conoce el fenómeno como el marcado diferencial de objetos (DOM, por sus siglas en inglés: *differential object marking*; algunos autores lo llaman ‘la marcación diferencial/diferenciada de objetos’). El término refiere al fenómeno que ciertos objetos gramaticales están más marcados con herramientas gramaticales que otros. En castellano, como hemos visto, significa que algunos objetos conllevan *a* y otros no.

Si bien ocurre en muchos idiomas, los lingüistas aún no han descubierto completamente cómo funciona el DOM precisamente. Es difícil formular reglas exactas y completas

sobre el uso del marcador. Además, sabemos que, en el castellano, el uso ha cambiado a lo largo del tiempo. No hubo tal marcador en latín; en el castellano medieval se empezó a usar y hoy en día el castellano lo usa con bastante frecuencia.

Por ello, es interesante investigar la evolución diacrónica del marcador. Podemos usar el castellano y su desarrollo histórico para comprobar teorías sobre el DOM. Cabe esperar que una teoría sobre el DOM se mantenga en todo momento ante cualquier cambio en el desarrollo del lenguaje. Aissen (2003) ha propuesto una teoría cross-lingüística que propone que el DOM se puede analizar en dos dimensiones: la definitud y la animacidad. La definitud tiene que ver con la especificidad y determinación del objeto. Por ejemplo, los pronombres se consideran más específicos que los nombres con artículos definidos, porque con un pronombre debe ser claro a qué refieren al usarlo. Además, los artículos definidos son más específicos que los artículos indefinidos. Así surge una jerarquía de definitud. La animacidad refiere al nivel de agencia y vitalidad que tiene el objeto. Hay tres niveles: seres humanos, animados e inanimados (objetos que no vivan). Por ejemplo, un ser humano tiene más animacidad que un animal, y un animal tiene más animacidad que un objeto sin vida.

La teoría de Aissen plantea que los objetos inesperados - es decir, los objetos que no son prototípicos porque son animados y específicos y, por lo tanto, se parecen más a los sujetos prototípicos - se marcan para enfatizar sus significados inesperados. Aissen (2003) sugiere que todas las lenguas utilizan estas dimensiones para determinar el uso del DOM.

Después, von Heusinger y Kaiser (2005) han ampliado la teoría de Aissen y argumentan que también necesitamos incluir la topicalidad del objeto para tener una teoría completa del DOM. La topicalidad de un constituyente depende de si se lo ha mencionado antes o no. Es una medida de la prominencia del constituyente en el discurso y la representación mental del discurso en los participantes.

Los tres conceptos gramaticales discutidos (definitud, animacidad y topicalidad) invocan reglas del uso del DOM. Como suponemos que cada etapa del desarrollo del lenguaje siga las reglas, se ha propuesto que, desde una perspectiva diacrónica, el cambio lingüístico del DOM debe respetar las jerarquías de animacidad, definitud y topicalidad.

Para el castellano histórico Aissen (2003) ha propuesto una teoría del DOM que argumenta que el DOM se desarrolló a lo largo de una dimensión de definitud y animacidad entre el castellano de *El Cantar de Mio Cid* (escrito entre 1140 y 1207) y el español moderno. Además, von Heusinger y Kaiser (2005) añade que también hay un tercer factor: la topicalidad. Si bien está claro que los tres factores juegan un papel en

el DOM, la manera exacta en que colaboran en el desarrollo del DOM en castellano aún no está clara.

En el presente estudio se examina la interacción entre las tres dimensiones desde una perspectiva diacrónica. Mientras que Aissen (2003) ha investigado dos momentos en el desarrollo del castellano, esta investigación intenta llenar los vacíos entre este período tanto en la escala de animacidad como en la de definitud. Si bien von Heusinger y Kaiser (2005) incluyen la topicalidad, no realizan una investigación de corpus para confirmar su teoría. von Heusinger (2008) ha investigado más periodos de tiempo pero solo de la medida en que el objeto se ve afectado por la acción del verbo. Al combinar los hallazgos y al utilizar las tecnologías de la investigación del procesamiento del lenguaje natural, investigamos la interacción entre los factores de manera más compleja.

Usamos el corpus de Sánchez-Martínez et al. (2013) que contiene textos del siglo XVI al XVIII. Automáticamente extraemos las oraciones que contienen un objeto directo y si utilizaban *a* o no. Las anotamos automáticamente para la dimensión de la definitud, y a mano para la dimensión de la animacidad. Así podemos determinar cómo se ha desarrollado el uso de *a* a lo largo del tiempo.

El segundo (siguiente) capítulo discute el marco teórico del estudio. Se discutirá el fenómeno del DOM y las investigaciones realizadas para el castellano. Este capítulo proporciona los fundamentos necesarios de los siguientes capítulos. También se discuten estudios anteriores y qué brecha de investigación está investigando este estudio.

El capítulo tres estará dedicado a la metodología; se describen el corpus y los pasos del análisis. Describimos la anotación utilizada, una combinación de anotación automática y manual.

Luego, discutimos los resultados de los análisis. En general, la teoría de Aissen cabe con los datos, pero hay objetos definidos inanimados que estaban marcados. Además, vemos que para los objetos definidos humanos hay un aumento en el uso de *a*, lo que era exactamente lo que se suponía. También se discuten algunos hallazgos más detallados.

El trabajo termina con una discusión de los resultados. Este capítulo contesta la pregunta de investigación y analiza las implicaciones de los hallazgos para las teorías. Proponemos una pequeña modificación a la teoría de Aissen y profundizamos en lo que es necesario para proporcionar una investigación más completa sobre el DOM en español.

## Capítulo 2

# El marco teórico: desde un sistema de casos hasta *a*

En este capítulo se discute el desarrollo del DOM en castellano. Primero, profundizamos en el fenómeno del DOM en general con algunos ejemplos del castellano. Después se esboza cómo se desarrolló *a* en castellano. Luego, se discuten las teorías contemporáneas sobre el DOM en castellano. Por último, se discuten las predicciones de estas teorías. Por lo tanto, en este capítulo se presentan las teorías que evaluamos en el resto del estudio.

### 2.1 El DOM

El marcado diferencial de los objetos (DOM) es el fenómeno de que algunos objetos directos están marcados de una manera distinta a otras. Se hace marcando algunos factores sintácticos o semánticos. Es una forma específica del marcado diferencial de los argumentos (DAM, de *differential argument marking*). Cross-lingüísticamente, el DOM es un fenómeno común que se puede encontrar en muchas lenguas de distintas familias (ver p.e. los ejemplos de de Swart (2007) o Aissen (2003)).

Como el DOM se trata del marcado de los objetos, es solamente relevante estudiar oraciones transitivas (i.e. frases con un sujeto y un objeto directo). Los objetos pueden ser de distintos tipos<sup>1</sup>, pero gramaticalmente siempre son un sintagma nominal (SN o NP, por sus siglas en inglés: *Noun Phrase*).

El castellano tiene el DOM. Algunos objetos directos están marcados con *a* (o *al*, de *a* + *el*) mientras que otros no lo son. Por ejemplo, hay las siguientes reglas en el castellano

---

<sup>1</sup>Por ejemplo, pueden ser un pronombre, un nombre o un sustantivo sin o con artículos o adjetivos

moderno según las gramáticas de referencia (Bosque & Demonte, 1999, pp. 1780-1805):<sup>2</sup>

- (1)
- a. Encontré el/\*al libro en la biblioteca
  - b. Encontré \*(a) Marta en la biblioteca
  - c. Encontré \*el/al bibliotecario en la biblioteca
  - d. Encontré (a) un bibliotecario en la biblioteca

El ejemplo (1-a) muestra que es imposible utilizar *a* hablando de un objeto inanimado. Los ejemplos (1-b) y (1-c) muestran que es obligatorio utilizar *a* con objetos humanos definidos, i.e. personas específicas como *Marta* o un bibliotecario específico del que todos los interlocutores saben a quién se refiere. El ejemplo (1-d) muestra que es demasiado simple decir que se debe utilizar *a* con todos los objetos humanos; si el objeto refiere a una persona no específica a veces se puede utilizar *a* (Bosque & Demonte, 1999, pp. 1787-1789).

Las reglas para los objetos animales también son bastante complicados (Pensado, 1995).

- (2) Veo al/el perro

Esta frase es gramatical tanto con *a* como sin *a*; hay diferencias semánticas sutiles entre los dos. Por ejemplo, se usa *a* frecuentemente con los animales domésticos que con los animales salvajes. Batchelor y San José (2010, p. 376) proponen que se usa *a* con animales inteligentes o racionales. Butt et al. (2019, pp. 328-329) argumentan que depende del grado de humanización del animal.

Pasemos ahora a los pronombres. Son complicados porque el español distingue dos tipos de pronombres: los clíticos y los no clíticos. Véase los ejemplos siguientes:

<sup>2</sup>En los ejemplos, seguimos las normas lingüísticas de los juicios de gramaticalidad. Significa que tenemos los siguientes juicios posibles:

- \* *Yo gusta ir a la playa* esta frase se juzga inaceptable por hablantes nativos.
- *Me \*(gusta) ir a la playa* esta frase se juzga inaceptable **sin** el contenido dentro de los soportes pero es aceptable **con** el contenido dentro de los soportes.
- *Me gusta (\*verde) ir a la playa* esta frase se juzga inaceptable **con** el contenido dentro de los soportes pero es aceptable **sin** el contenido dentro de los soportes.
- *Me gusta ir a la playa (contigo)* la frase se juzga aceptable **sin y con** el contenido dentro de los soportes.

La frase es gramaticalmente aceptable para los hablantes nativos en todos los otros casos.

- (3) La veo a ella.

En este ejemplo, *la* es un pronombre clítico mientras que *ella* es un pronombre no clítico. Se usa *a* solo con los pronombres no clíticos (Batchelor & San José, 2010, pp. 378-379):

- (4) a. La encontré en la biblioteca  
b. La encontré \*(a) ella en la biblioteca

En la oración (4-a), no importa si *la* se refiere a un humano (por ejemplo, a *Marta*) o a un objeto femenino (por ejemplo, a *la revista*); no hay “a la encontré” o algo similar. La frase (4-b) muestra que es posible añadir el objeto dos veces; una vez como pronombre clítico (*la*) y otra vez con pronombre ‘normal’ (*ella*). La duplicación sucede principalmente por factores pragmáticos y se discutirá con más detalle más adelante.

Finalmente, tenemos unos casos interesantes (Bosque & Demonte, 1999, pp. 1796-1797):

- (5) a. Veo \*(a) alguien.  
b. No veo \*(a) nadie  
c. Veo \*(a) todos.  
d. Veo (\*a) otro.  
e. Veo (a) otro doctor.

Pues, *alguien*, *nadie* y *todos* usan obligatoriamente *a*. Son definidos, y por eso tiene sentido que use *a*. Por otro lado, *otro* nunca usa *a*. Obsérvese que en el último ejemplo ‘otro doctor’ es indefinido y, como vimos anteriormente, significa que *a* es opcional.

Cabe señalar que existen variaciones regionales en el uso contemporáneo de *a* (Tip-pets, 2011). Como este estudio solo analiza el uso histórico en España, no se centrará en variedades regionales. Tampoco nos enfocamos en la variedad regional dentro de España.

Antes de analizar planteamientos más rigurosos del DOM en castellano para determinar cuándo se usa *a*, primero discutimos la historia de *a* en castellano desde el latín hasta el castellano moderno. Nos ayudará a comprender mejor por qué han surgido estas complicadas reglas de uso.

## 2.2 El desarrollo de *a* en el castellano

El español ha evolucionado del latín vulgar. En latín las preposiciones tenían un papel mínimo. El sistema de casos era más importante, porque los casos ayudaban a distinguir entre los papeles de los diferentes elementos de la frase. El nominativo se usaba

para los sujetos gramaticales y el acusativo para los objetos gramaticales. Penny (2002, pp. 114-116) sostiene que en el habla la mayoría de los casos se parecían demasiado. Por eso, los hablantes empezaron a usar preposiciones para evitar ambigüedades. La preposición *a* viene de *ad* (*a* o *hacia*) del latín. Al inicio, solo reemplazó el caso de dativo para el objeto indirecto, pero después el uso se amplió al objeto directo en frases sin orden fijo. Según Penny (2002, pp. 115-116), en el castellano antiguo *a* ocurrió con frases con sintagmas nominales (SSNN) teniendo el mismo número gramatical para determinar cuál de estos SSNN era el sujeto y cuál el objeto. Mientras que otros idiomas se basan más en el orden de las palabras, el español antiguo tenía un orden de palabras relativamente libre. Podía provocar confusión y la preposición ayudó a remediar la confusión. La ambigüedad ocurría sobre todo cuando el objeto era humano o tenía mucha agencia<sup>3</sup>, ya que este tipo de sintagmas nominales generalmente son sujetos. Por lo tanto, la preposición se usó más que nada en esos contextos.

Antes de continuar, conviene discutir el compromiso entre la claridad y el uso económico del lenguaje (Ferreira, 2008; Haspelmath, 2008), lo que supone un importante compromiso en el DOM (Kurumada & Jaeger, 2015). Por un lado, las lenguas tienden a querer ser claras en su significado. Por eso, el marcado de los objetos era útil porque no puede haber confusión sobre el significado de una oración. Por otro lado, los idiomas tienden a ser económicos. No quieren usar demasiadas palabras innecesarias porque el lenguaje resulta verbosísimo. Por lo tanto, los objetos más típicos (es decir, las palabras de las que se espera que sean objetos porque suelen serlos) generalmente no están marcados, por lo que el lenguaje será menos prolijo. Por lo tanto, los idiomas necesitan encontrar un equilibrio entre la economía y la reducción de la ambigüedad. En algún momento, el compromiso empieza a codificarse en la gramática de una lengua. El español ha encontrado tal equilibrio después de la gramaticalización de *a*, pero, como veremos, el equilibrio aún está sujeto a cambio.

La gramaticalización es un proceso diacrónico por el cual una palabra de una categoría léxica pasa a ser parte de una categoría funcional. En el caso de *a*, pasó de ser una preposición solo utilizada para denotar dirección a una preposición utilizada para denotar una función gramatical específica. En las etapas iniciales del proceso de gramaticalización, las reglas que rodeaban el uso de esta preposición eran difusas. Con el paso del tiempo, las reglas se hicieron más fijas.

Penny (2002) concluye que al final del Siglo de Oro el proceso de la gramaticalización se finalizó y *a* era obligatorio en contextos específicos, también cuando no había posibili-

---

<sup>3</sup>En este estudio se entiende la *agencia* como la capacidad de emprender acciones.

dad de ambigüedad. Sin embargo, la función de eliminación de ambigüedades es todavía visible en algunas oraciones, como *mordió el perro al/?el gato*. Sin el *a* no se sabe cuál es el objeto y cuál el sujeto. En estos casos la ambigüedad es el factor importante para determinar si se necesita usar *a* o no.

La importancia de la prevención de confusiones ha sido atestiguada en estudios anteriores. Melis (1995) ha investigado el uso del objeto directo en *El Cantar de Mio Cid*, escrito entre 1140 y 1207. Nota que los objetos colocados detrás del verbo tienden a carecer de preposición mientras que los objetos antepuestos al verbo sí la llevan (pp. 133-134). Se explica observando que hay más confusión cuando el objeto está al principio de la oración que cuando no lo está. Al fin y al cabo, el español es una lengua SVO, lo que significa que según el orden de las palabras estándar en español (y latín, que es SOV) el sujeto se antepone al objeto. Melis también descubrió que era más probable que los objetos personales llevaran la preposición que los objetos inanimados y que importaba también la definitud .

Por lo tanto, la construcción sirvió principalmente para evitar interpretaciones erróneas cuando había dos elementos que podían ser sujetos (como ocurre con los nombres personales). Después la preposición se gramaticalizó y se formaron reglas relativamente estrictas. El DOM sirve para evitar malentendidos, pero su uso se rige por normas más estrictas (Fábregas, 2013, p. 5). Antes se usaba de forma más intuitiva.

## 2.3 Las jerarquías

La idea de que el DOM sirve para evitar ambigüedad no es exclusiva del castellano. En general, las teorías sobre el DOM proponen que se deben marcar los objetos que son similares a los sujetos. Por lo tanto, el DOM es una manera de reducir la ambigüedad. Conviene resaltar que, en otros idiomas, también es posible marcar sujetos inusuales en vez de marcar objetos inesperados<sup>4</sup>. El castellano solo usa el DOM, por lo que la discusión restante solo se centrará en el DOM.

Para determinar qué quiere decir ‘similitud al sujeto’, diferentes autores han propuesto distintas escalas y jerarquías para comparar sujetos y objetos. Silverstein (1986) propuso una jerarquía para entender la agencia de un SN:

<sup>4</sup>No hay una razón a priori por la que se debería marcar objetos inesperados en lugar de sujetos inesperados; ambos marcan lo inesperado y así llaman la atención y aclaran la estructura gramatical. Por lo tanto, ambos evitan la ambigüedad. Se conoce como marcado diferencial de sujetos (DSM, por sus siglas en inglés: *Differential Subject Marking*). Véase de Hoop y Malchukov (2008) para una discusión más elaborada sobre la relación entre el DOM y el DSM, y cómo y cuándo ocurre uno de ambos.

- (6) segunda persona (tú, vosotros/as) > primera persona gramatical (yo, nosotros/as) > tercera persona gramatical (ella, él, etc.) > nombre humano > sustantivo humano > sustantivo animal > sustantivo inanimado

De acuerdo con esta jerarquía, un argumento en primera persona tiene más agencia y, por lo tanto, es un sujeto más prototípico que un sustantivo animal. Por lo tanto, si el sustantivo animal es el sujeto y la primera persona el objeto, este objeto probablemente se marcaría para evitar confusiones. Sin embargo, si el sustantivo animal es el objeto y la primera persona el sujeto, no sucede nada inusual y no hay necesidad de resaltar nada. Por lo tanto no se marca nada en este caso.

Después de Silverstein (1986), distintos autores han propuesto otras jerarquías. Aissen (2003) introdujo dos jerarquías que separaban la animacidad de la definitud. De esta manera se resuelven casos difíciles que la jerarquía de Silverstein no tenía en cuenta (por ejemplo: nombres de animales) o no encajan con datos del mundo real (por ejemplo: en castellano es posible utilizar *a* con animales pero no con todos los sustantivos humanos, como hemos visto antes). Además, en su artículo Aissen señala que sus dos jerarquías brindan la precisión necesaria para adaptarse a una amplia gama de idiomas.

Su jerarquía de animacidad distingue tres niveles:

- (7) seres humanos > animados > inanimados

Intuitivamente, los humanos tienen más agencia por lo que es más probable que hagan cosas y sean sujetos. Los inanimados tienen poca agencia y, por lo tanto, es probable que sean objetos. Los animales están en el medio. El castellano sigue este patrón, ya que tiende a marcar a los humanos más que a los animales, los cuales a su vez se marcan con más frecuencia que los inanimados.

Como sugiere Aissen (2003, pp. 456-457), se pueden dividir las categorías principales de la animacidad en subcategorías. Por ejemplo, podemos distinguir entre diferentes tipos de animales, como los animales domésticos, los animales de granja, y los insectos. Estas jerarquías también parecen ser relevantes en español, aunque la naturaleza exacta sigue sin estar clara (Fábregas, 2013, pp. 13-14). Sin embargo, para mantener el alcance de esta investigación limitado, no se desglosará más la jerarquía.

La segunda jerarquía de Aissen es la jerarquía de definitud.

- (8) Pronombres > nombres > definidos > indefinidos específicos > indefinidos no específicos

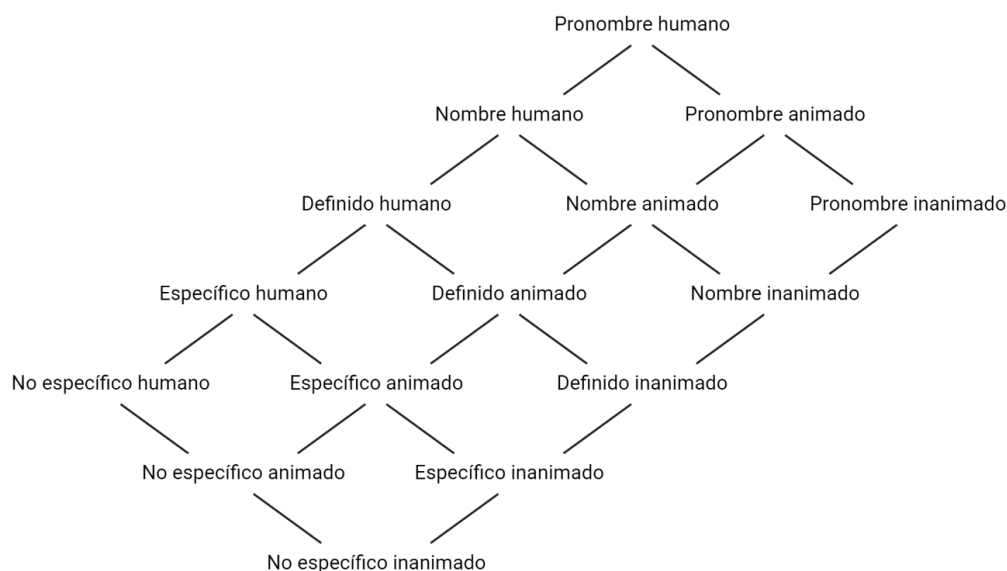


Figura 2.1: La combinación de las jerarquías propuestas por Aissen (2003). Los SSNN de arriba son prototípicamente sujetos y no objetos, mientras que los SSNN de debajo son prototípicamente objetos. Adaptada de Aissen (2003, p. 459) y figura creada por Maaïke Wessel.

Aissen define la definitud como la claridad y determinación con la que se introduce un referente discursivo por el SN (p. 444). Los pronombres de la primera y segunda persona están fijados en el discurso. La tercera persona solo puede referirse a una entidad prominente pero no participante en el discurso. Los nombres están fijos por convención. Los definidos están definidos por conversación previa; los interactuantes deben saber a qué o a quién se refieren. Los indefinidos no están fijos. Los indefinidos específicos ofrecen algunas opciones (*una de los dos niñas*) mientras que los indefinidos no específicos no imponen ningún límite en el conjunto de referentes (*un niño*).

En general, a los hablantes de un idioma les gusta usar sujetos definidos y objetos indefinidos (p. 445). Por tanto, objetos definidos o sujetos indefinidos son inesperados. Debido a esto, las lenguas tienden a marcar estos objetos (o los sujetos, véase la nota anterior). Como vimos anteriormente, para los objetos humanos, el castellano se ajusta a este patrón. Los objetos humanos indefinidos no están marcados, mientras que otros tipos de objetos humanos sí lo están.

Aissen (2003) combina las dos jerarquías para dar cuenta del patrón que encontramos en español y en otros idiomas. Forma la red reproducida en la figura 2.1.

Si tenemos un SN desde la parte superior de la red, suponemos que este SN es un

sujeto, combinando el conocimiento de las jerarquías de definitud y animacidad. Si, por otro lado, tenemos un SN del fondo de la red, suponemos que es un objeto. Por lo tanto, si nos encontramos con un SN de la parte superior de la jerarquía que es un objeto, es inesperado. Las lenguas tienden a marcar estos objetos, que es precisamente lo que vemos en el castellano.

Ahora bien, la hipótesis de Aissen es que todos los idiomas siguen este patrón. Aissen utiliza la teoría de la optimidad (*Optimality Theory*) para hacer plausible esta afirmación. Si bien la teoría de la optimidad es una teoría elegante e interesante, está más allá del alcance del presente estudio. Más bien, me centraré en la medida en que la afirmación de Aissen es válida para el castellano (histórico), es decir hasta qué punto el castellano (histórico) sigue el patrón.

“Seguir el patrón” quiere decir que un idioma dibuja una línea en algún lugar de la red. Por encima de la línea el lenguaje marca los objetos y por debajo de la línea no. En otras palabras, creemos que es imposible encontrar un idioma que solo marque, por ejemplo, sustantivos definidos humanos, pero no nombres propios humanos o pronombres. Aissen argumenta que también es posible introducir una zona gris. Por ejemplo, en un idioma puede ser opcional marcar algunos objetos, mientras que es obligatorio marcar o no marcar otros.

La teoría de Aissen es interesante porque aplica también a los idiomas que no utilizan ninguna marca. Estos idiomas simplemente trazan una línea en la parte superior de la red, sobre el “pronombre humano”. Todos los elementos por encima de la línea (que no hay ninguno) están marcados, mientras que todos los elementos por debajo de la línea (todos los elementos) no lo están. De manera similar, las lenguas que marcan todos los objetos trazan la línea debajo de la red. De esta manera, todos los objetos por encima de la línea (todos los objetos) están marcados, mientras que todos los objetos por debajo de la línea (ninguno) no lo están.

Aissen (2003) ha colocado tanto el español moderno como el español utilizado en *El Cantar de mio Cid* en la red. El resultado está visualizado en la figura 2.2. Por encima de la línea superior, el marcado de objetos es obligatorio, mientras que por debajo de la línea inferior está prohibido. Entre las dos líneas marcar es opcional. Las líneas verdes muestran el límite en español de *El mio Cid* mientras que las líneas azules muestran los límites en español moderno.

Aissen se basa en datos de Melis (1995) del *Mio Cid*. Dicha lingüista ha encontrado sustantivos humanos definidos sin y con *a* en el *Cantar*. Por ejemplo (ejemplos reproducidos de Aissen (2003, p. 462)).

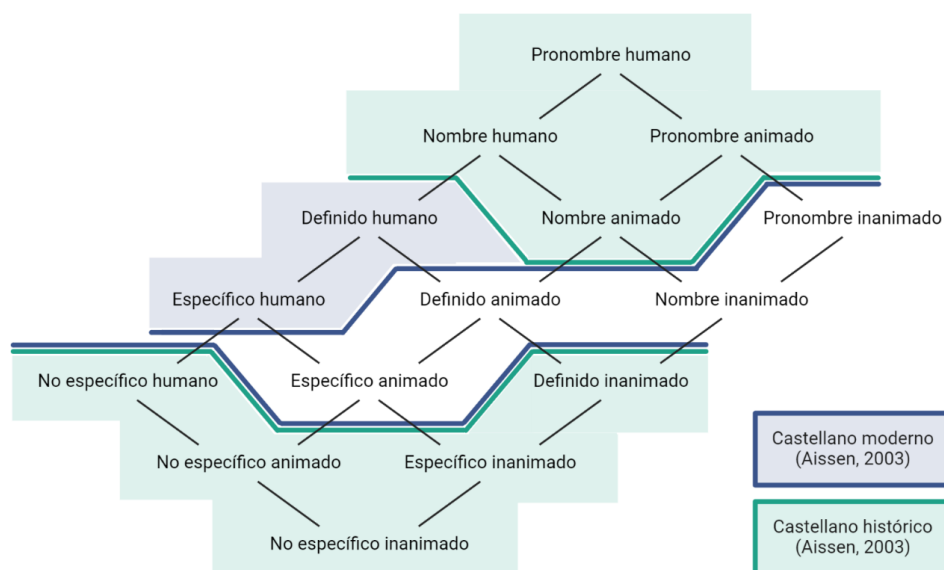


Figura 2.2: El cambio en el DOM en castellano según Aissen (2003). Las líneas azules refieren al español moderno mientras que las líneas verdes refieren al español del *Cid*. Figura adaptada de Aissen (2003, p. 472) y creada por Maaike Wessel.

- (9) a. ... rreciba *a mios yernos* como él pudier mejor.  
 b. ... quando dexaron *mis fijas* en el rrobredo de Corpes ...

Aissen concluye que en el caso de humanos definidos marcar es opcional. Sin embargo, en el español moderno ya no es el caso como se discutió anteriormente. Los hallazgos de Aissen para el español están en línea con lo que Penny (2002) señaló anteriormente: que el español antiguo permitía una mayor variación en los objetos marcados con *a* que el español moderno, en el que esta preposición está completamente gramaticalizada.

Como se supone que esta red es válida para todos los idiomas, también suponemos que es válida para todas las etapas intermedias de un idioma. Por lo tanto, no solo debería ser válida para el español del *Mio Cid* o el español moderno, sino también para cada paso del español en el medio. Entonces, Aissen (2003) predice que los cambios en las situaciones donde el marcado de objetos es obligatorio solo pueden cambiar junto con estas jerarquías. Además, suponemos un cambio gradual entre las dimensiones discutidas anteriormente. El cambio de idioma no ocurre de la noche a la mañana, sino que generalmente lleva años o incluso siglos. Esta investigación tiene como objetivo investigar si de hecho el desarrollo del español ha seguido el patrón discutido.

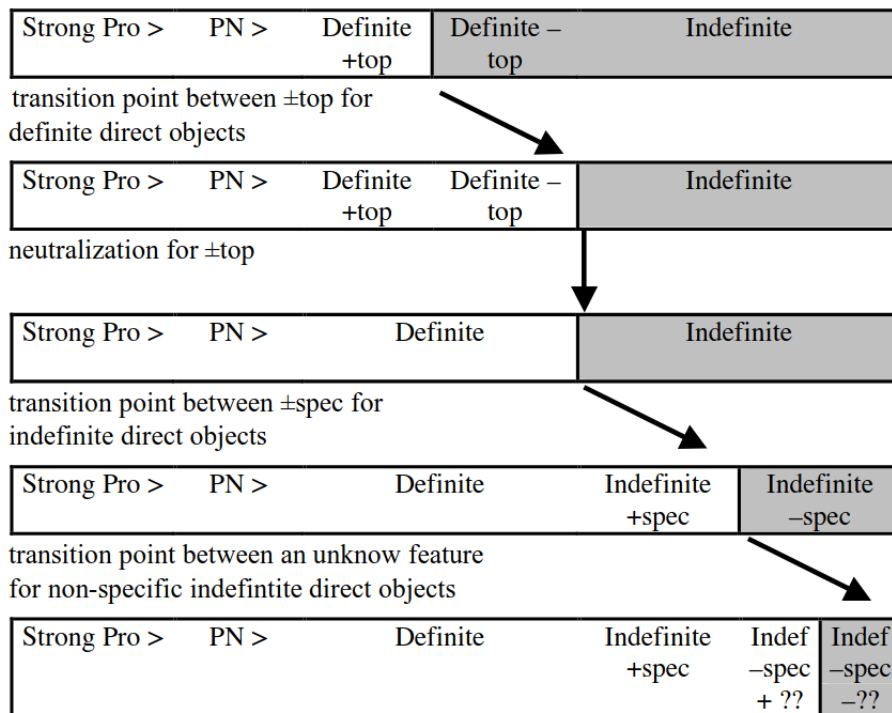


Figura 2.3: La evolución del DOM desde el castellano antiguo hasta el castellano moderno para los objetos animados (incluyendo los objetos humanos) según von Heusinger y Kaiser (2005). Las áreas blancas designan los casos en los que un es obligatorio y las áreas grises los casos en los que está prohibido. *PN* refiere al ‘nombre’ por sus siglas en inglés (*proper name*). Tomada de von Heusinger y Kaiser (2005, p. 45).

## 2.4 Topicalidad

Tras haber discutido la propuesta de Aissen, ahora es el momento de pasar a otros factores que podrían influir en la medida en que se usa la preposición *a*.

Según von Heusinger y Kaiser (2005) necesitamos añadir la topicalidad como dimensión del DOM. La topicalidad se refiere a la influencia de la estructura de información del discurso. Por ejemplo:

- (10) A la bibliotecaria la encontré en la biblioteca, al bibliotecario no

En esta oración, se agregan ‘a la bibliotecaria’ y ‘al bibliotecario’ para enfatizar el contraste. Por lo tanto, von Heusinger y Kaiser (2005, p. 44) argumentan que es más probable que se agregue *a* a una oración definida en español antiguo si el SN es tópico que cuando no lo es, pero que este es solo el caso cuando el SN es definido.

A partir de ahí, averiguan cómo el español antiguo pudo haberse convertido en el español moderno, que se muestra en la figura 2.3. Observamos que, según von Heusinger y Kaiser, la topicalidad jugó un papel en la clase definida en el español antiguo y luego dejó de tener un papel a medida que el uso de *a* comenzó a extenderse. Al diferencia de Aissen, esta propuesta solo se refiere a los objetos animados. Además, von Heusinger y Kaiser dividen algunas de las clases propuestas por Aissen en +top (tiene topicalidad) y -top (no tiene topicalidad). Así von Heusinger y Kaiser (2005) intentan evitar la opcionalidad que contiene la teoría de Aissen; von Heusinger y Kaiser buscan una teoría del DOM que explique cada uso de *a*.

En un etapa posterior (fila inferior de la figura 2.3), von Heusinger y Kaiser (2005) observan que el uso para marcar los objetos animados indefinidos es más opcional. Suponen que hay un factor subyacente aún por identificar que separa la categoría indefinida, pero von Heusinger y Kaiser no se profundizan en ellos.

## 2.5 El verbo

También se han identificado otros factores que pueden haber desempeñado un papel en la medida en que *a* se usa en casos opcionales. Por ejemplo, el verbo que invoca el objeto directo puede ser de influencia; von Heusinger (2008) investiga el mismo texto bíblico de diferentes períodos de tiempo. Compara diferentes clases de verbos y descubre que algunos tipos de verbos tienden a invocar la preposición y otros no. A esta propiedad se llama ‘affectedness’ (afectación) de los verbos. El verbo ‘matar’ tiende a invocar *a* porque el objeto está afectado por ‘matar’ mientras que el verbo ‘tomar’ no lo hace. Por lo tanto, también se debe tener en cuenta el tipo de verbo.

## 2.6 Conclusión

En este capítulo se ha discutido el fenómeno del DOM en general y concretamente el DOM en el castellano histórico y moderno. Se han discutido dos propuestas sobre su desarrollo en castellano. El resto de este estudio se enfoca en afirmar estas teorías con un corpus histórico del español. El siguiente capítulo explica el método utilizado para hacerlo.

## Capítulo 3

# La metodología: combinar análisis automático y manual

En este capítulo se discute el método utilizado. Utilizamos una combinación de anotación automática y manual de forma complementaria, porque es más eficaz preprocesar todos los datos de forma automática. Nos permite analizar más datos en menos tiempo. Como no existen métodos suficientemente avanzados disponibles para analizar la animacidad automáticamente con suficiente precisión, también utilizamos anotación manual para complementar la anotación automática y corregirla cuando fuera necesario. En este capítulo se describen ambos métodos con más detalle.

### 3.1 El corpus

Usamos el corpus de Sánchez-Martínez et al. (2013) que contiene textos de los años 1480 hasta los años 1650. Como ha demostrado von Heusinger (2008, p. 21), el mayor cambio en el uso de *a* ocurrió entre el siglo XVI y XX, y, como argumenta Penny (1995), en su mayoría en el siglo de oro (1492-1680). Por lo tanto, los textos de dicho época son los más interesantes. El corpus contiene textos de este período.

Este corpus es *open source* y consiste de dos partes: una parte llamada “GT” y otra llamada “BVC”. La parte BVC está completamente anotada, lo que significa que incluye información sobre el tipo de constituyente y una variante ortográfica moderna de palabras españolas antiguas. Las ortografías modernas nos permiten aplicar herramientas modernas de procesamiento de lenguaje natural a los textos. No usamos la parte GT por falta de anotación. Para una lista completa de los textos remitimos al artículo de

Sánchez-Martínez et al. (2013, p. 18).

## 3.2 Método de análisis

En esta sección primero se explican los pasos de anotación en general para dar una visión general. Después, se describen con más detalle algunas de las decisiones tomadas.

El corpus se analizó extrayendo todas las oraciones y luego anotándolas con etiquetas Part-of-Speech y dependencias sintácticas utilizando la biblioteca Python `es_core_news_sm` de spaCy. El modelo estadístico ha sido capacitado en textos periodísticos en español. Los textos de dicha investigación son literarios (poesía y prosa). Si bien estos textos son diferentes, un estudio piloto ha demostrado que el analizador funciona bastante bien en las versiones de palabras actualizadas que están disponibles en el corpus, pero comete más errores que en los textos modernos. En consecuencia, los resultados de los siguientes pasos se han verificado a mano. El código utilizado está disponible en Github<sup>1</sup>.

Tuvimos que limpiar algunos de los datos del corpus para hacerlos más adecuados para el analizador. Eliminamos palabras que consistían solo de mayúsculas; generalmente se trababa de los nombres de los personajes en obras de teatro. Además, resultó que en el corpus no se había modernizado toda la ortografía. Por lo tanto cambiamos algunas ortografías antiguas que ocurren con frecuencia a sus variantes modernas (como *vn* a *un* y *auia* a *había*).

Después, se extrajeron automáticamente todas las oraciones con objetos, ya que estas son las frases de interés. En el siguiente paso, se anotaban automáticamente las frases con spaCy. De estas frases, extrajimos después las oraciones subordinadas que contenían los objetos. De estos objetos, determinamos si iban con *a* y dónde figuraban en la jerarquía de definitud. También anotamos el verbo del objeto. Todos estos datos se pusieron en una hoja de cálculo que se anotó a mano. Durante la anotación manual se comprobaron los datos generados automáticos y se corrigieron o se complementaron si era necesario. Además, se anotó la animacidad y el orden del verbo, objeto y, si estaba presente, el sujeto. Dado que el proceso de anotación manual requiere mucho trabajo, solo se anotó una parte limitada de los datos.

Resumiendo, se anotaban las siguientes propiedades por oración que contiene un objeto directo.

- Uso de *a* o no

---

<sup>1</sup><https://github.com/NienkeWessel/ThesisSpanish>

- Animacidad
- Orden del sujeto, objeto y verbo
- Definitud
- El verbo que ha invocado el objeto

El lugar del objeto (antes o después del verbo) es una medida indirecta de la topicalidad. Es difícil medir la topicalidad por completo, ya que depende en gran medida del contexto. Por lo tanto, anotamos el orden del sujeto, objeto y verbo.

Después de la anotación, los textos se dividieron en períodos de tiempo de cuarenta años cada uno: 1480-1519, 1520-1559, 1560-1599, 1600-1649. Se ha elegido un periodo de cuarenta años para dividir los textos disponibles en cuatro categorías. En cada período de tiempo, para cada combinación de animacidad y definitud, se calculó el porcentaje de  $a$  usado. Además, en algunos casos interesantes (vease la sección de los resultados) se calculó si el verbo venía después del objeto o antes, para las oraciones con  $a$  y las sin  $a$ . No se usó el tipo de verbo en ningún análisis cuantitativo, sino que solo se examinó para el análisis cualitativo.

### 3.3 Las oraciones analizadas

Al final, analizamos 10 textos diferentes de los cuales 637 oraciones. En el cuadro 3.1 se resumen los datos utilizados en el estudio.

Los textos anotados se eligieron en parte al azar, pero con algunas limitaciones. En primer lugar, nos aseguramos de tener aproximadamente la misma cantidad de textos y frases de cada período de tiempo. En segundo lugar, hemos elegido textos de diferentes autores en lugar de elegir varios textos del mismo autor, para aumentar la representatividad y la posibilidad de generalización; un autor puede tener cierta preferencia, pero estamos interesados en crear una imagen más general. La elección de diferentes autores del mismo período de tiempo ayuda a proporcionar esta imagen general.

### 3.4 Aspectos específicos del análisis

Lo más difícil de analizar automáticamente un corpus como este, es la cantidad de tipos de oraciones, que posiblemente contengan una o varias cláusulas subordinadas y diferentes construcciones. Más que nada era difícil determinar el verbo principal. El

Período	# de frases	Año	Texto	Autor	NBD	# de frases
1480-1519	165	1481	Laberinto de Fortuna	Juan de Mena	138	79
		1499	La Celestina	Fernando Rojas	90	40
		1507	Representación sobre el poder del amor	Juan del Encina	8648	46
1520-1559	150	1544	Dialogo de mujeres	Cristóbal de Castillejo	1010	50
		1554	Farsa de la muerte	Diego Sánchez de Badajoz	8108	100
1560-1599	160	1585	Dialogo de la dignidad del hombre	Fernán Pérez de Oliva	164	47
		1590	Comedia del Príncipe Ynocente	Lope de Vega	10107	113
1600-1649	162	1613	Novela de la Ilustre Fregona	Miguel de Cervantes Saavedra	8784	59
		1628	La cueva de Salamanca	Juan Ruiz de Alarcón y Mendoza	449	62
		1647	Oráculo manual y arte de prudencia	Baltasar Gracián	23	41

Cuadro 3.1: La información sobre las oraciones analizadas. NBD refiere al número en el base de datos del corpus de Sánchez-Martínez et al. (2013).

anotador automático asumió que el primer verbo era el principal. Durante la corrección manual, esto resultó ser en su mayoría una suposición correcta, aunque a veces tuvo que ser ajustada.

Durante la corrección manual, muchas oraciones tuvieron que ser eliminadas porque el analizador las consideró oraciones con objeto mientras que no eran relevantes para nuestra investigación. Por ejemplo, el analizador no distingue entre objetos directos e indirectos, sino que simplemente marca ambos como *obj*. Por lo tanto, muchas oraciones tuvieron que ser eliminadas por contener solo un objeto indirecto. Además, las oraciones con solo pronombres reflexivos o *se* impersonal también se marcaron como oraciones de objeto, mientras que no eran relevantes para el estudio. Además, también se excluyeron los verbos que siempre usan una determinada preposición, ya que no permiten el uso de una preposición adicional *a*, por ejemplo *hablar de*. Finalmente, también hubo algunas oraciones que contenían estructuras subordinadas muy complejas de las cuales era difícil zanjar cómo debían interpretarse. Esas oraciones también fueron excluidas.

También hubo algunos casos infrecuentes. Un nombre con artículo definido se codificó como nombre y no como sustantivo definido si el artículo siempre iba con el nombre/puede considerarse parte del nombre. Por ejemplo: *Las Filipinas* se consideraba un nombre. Si la misma oración aparecía dos veces (como ocurre a veces en la poesía) solo se incluía una vez por texto para evitar que una determinada formulación tuviera demasiada influencia en los resultados. Gramaticalmente, ‘Dios’ y el diablo fueron anotados como ‘humanos’, porque su agencia está al nivel de los humanos (o posiblemente mayor que la de los humanos). Si había un conflicto en la clasificación porque podía pertenecer a múltiples clases de definitud, se clasificaba como la más alta de esas clases. Por ejemplo, ‘mi Elicia’ se puede clasificar como definido debido a ‘mi’ o como un nombre porque Elicia es un nombre propio. En este trabajo, se clasifica como nombre, porque los nombres están más arriba en la jerarquía que los sustantivos definidos. La lógica detrás de la jerarquía de definitud es que pertenece a la medida en que el referente está fijado en el discurso. El escalón más alto de la jerarquía determina hasta qué punto se fija. En este caso, el nombre lo fija ‘más’ que el ‘mi’.

### 3.5 Conclusión

A base de una combinación de anotación automática y manual, se analizaron más de 600 oraciones de diferentes períodos de tiempo. Los resultados de este análisis se discuten en el siguiente capítulo.

## Capítulo 4

# Los resultados: la animacidad y definitud en la acción

En este capítulo se analizan los resultados de la investigación. Se hace principalmente utilizando las dimensiones propuestas por Aissen (2003). Donde sea relevante, se discuten otros factores.

Este capítulo se centrará en torno a la figura 4.1 que muestra los resultados divididos entre las dimensiones propuestas por Aissen. En cada caso se da el porcentaje del uso de *a*. El texto comparará los porcentajes con el porcentaje esperado a base de las teorías discutidas en los capítulos anteriores. En las siguientes secciones se profundizará en algunas partes de la figura. Primero hablaremos de arriba abajo de los humanos (la primera columna), luego de los animados (la columna en el medio) y finalmente de los inanimados (la columna de la derecha).

### 4.1 Pronombres

Si bien los pronombres son los más definidos, el gráfico muestra que usan relativamente poco *a*. A primera vista, es sorprendente, pero cuando excluimos todos los pronombres clíticos surge otra imagen. Los resultados para los pronombres no clíticos se muestran en la figura 4.2. Vemos que todos los pronombres usan la preposición, lo cual está en línea con las teorías.

Ocasionalmente se dice que es más común doblar con un pronombre no clítico si el objeto está preverbal, pero no es lo que encontramos. De hecho, encontramos más casos de objetos posverbiales que de objetos preverbiales, como se muestra en la tabla 4.1.

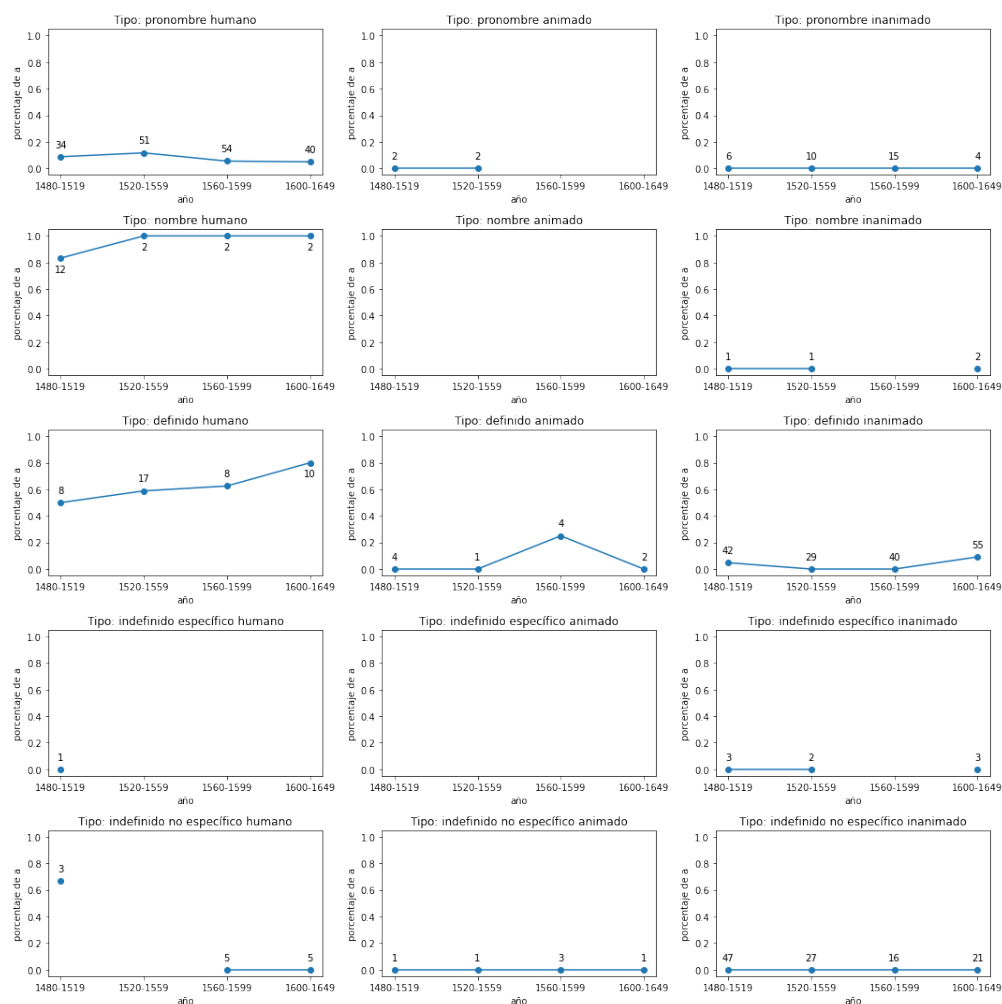


Figura 4.1: El porcentaje de casos de objetos directos que contienen *a* del total número de objetos directos, clasificados según las diferentes dimensiones de Aissen (2003). Horizontalmente se pueden ver los resultados de la dimensión de animacidad y verticalmente de la dimensión de definitud. La cantidad presentada en cada franja de años indica el número de oraciones en la categoría.

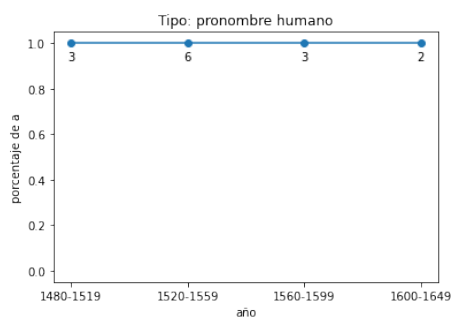


Figura 4.2: Los resultados para los pronombres humanos, excluyendo los resultados para los pronombres clíticos.

Pronombre doble	No doble	Doble	Total
Orden			
OV	2	3	5
VO	2	7	9
Total	4	10	14

Cuadro 4.1: La cantidad de oraciones que tienen pronombres no clíticos. ‘Doble’ se refiere a que la oración también tenía un pronombre clítico, por ejemplo ‘a mí me ve’. ‘El orden’ se refiere al orden del verbo y el objeto marcado con *a* (el pronombre clítico, si está presente, se ignora al determinar el orden; se hizo porque se rige por reglas bastante estrictas, por lo que deja poco margen de maniobra).

Es interesante señalar que los objetos de pronombres no clíticos nunca aparezcan en las clases animadas o inanimadas.

## 4.2 Nombres de humanos

En dos de nuestros tres textos más antiguos, tenemos objetos con nombres humanos que no usan la preposición. Es algo digno de mención porque no está en consonancia con ninguna de las teorías descritas anteriormente, que afirman que en la época del Cid (unos 300 años antes de estos textos) ya estaba fijado que los nombres propios de los humanos llevaran la preposición. Por eso queremos discutir estas frases en más detalle. Son las siguientes:

- (1) a. ... por ver mi Elicia ... (*La Celestina*, Fernando Rojas 1499)
- b. ... quando librava en el siglo los calidones del bravo vestiglo ... (*Laberinto de Fortuna*, Juan de Mena, 1481)

La primera frase de *La Celestina* tiene un nombre y un pronombre posesivo. Recordando la definición de la noción de la definitud, la caracterizamos como muy definido ya que el referente está bastante fijo en el discurso. Por lo tanto es extraño que no lleve *a*. Tal vez deberíamos analizarlo como un sintagma nominal definido (por la *mi*) en lugar de un nombre. Para SSNN definidos, *a* es opcional según las teorías discutidas (véase también la siguiente sección). Podría explicar los hallazgos extraños de este tipo.

*Los calidones* es aún más difícil de entender; pueden referir a un pueblo o una nación por lo que es difícil de clasificar en las tres categorías de animacidad. Si creemos que se usa en este contexto para referirse a una nación, se vuelve inanimado y, por lo tanto, debe trasladarse a otra categoría.

En conclusión, encontramos aquí dos casos extraños, pero ambos pueden ser reinterpretados.

## 4.3 Definidos humanos

Que conste que en el caso de los humanos definidos el uso de *a* es opcional en todos los períodos de tiempo. Los siguientes ejemplos muestran algunas oraciones sin *a*.

- (2) a. ... crió Dios la primera muger ... (*Dialogo de mujeres*, Cristóbal de Castillejo, 1544)

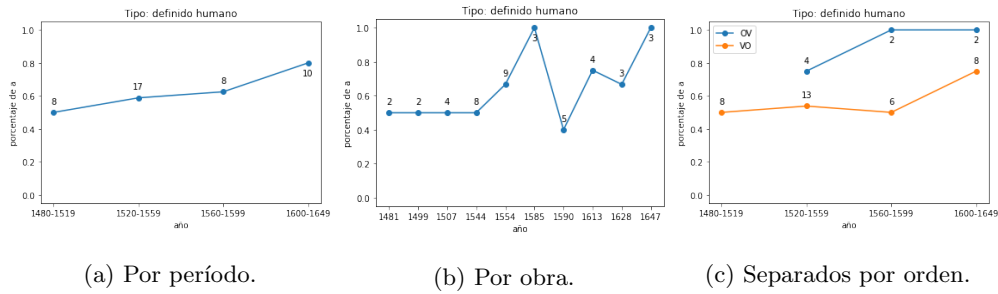


Figura 4.3: Los resultados para los definidos humanos.

- b. ... haz el diablo la presa en grandes porque se ciegan y en chicos porque reñegan de verse hechos su mesa.! (*Farsa de la muerte*, Diego Sánchez de Badajoz, 1554)
- c. ... querer los hijos de Adán sin sudor comer el pan ... (*Farsa de la muerte*, Diego Sánchez de Badajoz, 1554)

En la figura vemos, como era de esperar, un pequeño aumento en el uso de  $a$  durante el período de tiempo investigado. Hubo una correlación de  $r = 0,22$ . Es una correlación débil que también se basa en datos limitados, por lo que debe interpretarse con precaución.

Se realizaron algunos análisis adicionales para investigar esta categoría de manera más detallada. La figura 4.3 muestra los resultados más detallados para los seres humanos definidos. La figura 4.3a muestra de nuevo el gráfico de la figura 4.1 a efectos comparativos. La figura 4.3b muestra la cantidad de  $a$  utilizada por texto. Es más detallado que por período y demuestra que hay bastante variación entre los autores. Una vez más, sin embargo, debe tenerse en cuenta que hay pocos puntos de datos y es complicado generalizar los datos. La figura 4.3c muestra los resultados cuando separamos el orden de las palabras y observamos si el objeto va antes o después del verbo. Parece que es menos probable que el orden verbo-objeto lleve  $a$  que el orden objeto-verbo.

Estos resultados están en línea con la teoría de Heusinger y Kaiser (2005), mostrada en la figura 2.3 del capítulo anterior. Si el objeto está delante del verbo, suele ir precedido por  $a$ . El objeto es más tópico si viene antes del verbo, por lo que podemos argumentar que es más probable que los objetos más tópicos estén precedidos por  $a$ . Además, vemos que cambia a lo largo del tiempo, lo que hace probable que, de hecho, los SSNN definidos humanas más tópicos se marcaron primero con  $a$ , después de lo cual los SSNN definidos humanas menos tópicos comenzaron a marcarse. Nuevamente, sin

embargo, estos hallazgos están limitados por la cantidad limitada de datos.

## 4.4 Indefinidos no específicos humanos

Encontramos pocos casos de objetos sustantivos humanos indefinidos no específicos. De hecho, encontramos dos casos que tienen *a*. Ambos son de *La Celestina*. Escribe dos veces ‘a otro’ como objeto.

- (3) a. ... tenía a otro consigo ...  
 b. Torpe cosa es mentir el que enseña a otro ...

Un análisis de las partes sin anotaciones de los otros textos encontró que otros textos investigados no usan esta forma como objeto (a veces ocurre para indicar dirección (“de aquel lugar a otro”), pero no como objeto directo). Debido a esto, es probable que sea una peculiaridad del autor más que un caso representativo para el período de tiempo.

## 4.5 Animales

Lamentablemente, ha resultado difícil encontrar datos suficientes para los objetos animales. Significa que la columna central de la figura 4.1 está casi vacía. Los datos disponibles muestran que *a* se usaba una vez. No contradice hallazgos previos, que afirman que *a* es opcional en español moderno, pero tampoco soporta la noción de ‘opcionalidad’. En otras palabras, es difícil sacar conclusiones concretas de los datos limitados.

## 4.6 Definidos inanimados

Con base en la teoría de Aissen, no esperamos encontrar ningún caso en que *a* se use con sustantivos inanimados definidos. Sin embargo, los gráficos muestran que encontramos algunos tanto en el primer como en el último período de tiempo. Justifica un análisis más detallada. A continuación se enumeran todas los casos de sustantivos inanimados definidos que usan *a*.

- (4) a. ... por no hacer agravio a la grande amistad que professauan. (*Novela de la Illustre Fregona*, Miguel de Cervantes Saavedra, 1613)  
 b. ... el que viere a la justicia ... (*La cueva de Salamanca*, Juan Ruiz de Alarcón y Mendoza, 1628)

- c. ... pero no se ha de regar al exceso de callar para que yerre ... (*Oráculo manual y arte de prudencia*, Baltasar Gracián, 1647)
- d. Es robarle a la esperança cortés fiar del agradecimiento villano ... (*Oráculo manual y arte de prudencia*, Baltasar Gracián, 1647)
- e. ... porque no se oponen a las regalías del rey ... (*Oráculo manual y arte de prudencia*, Baltasar Gracián, 1647)
- f. ... encaran a los ojos luego enclavan las entrañas (*Representación sobre el poder del amor*, Juan del Encina, 1507)
- g. ... pero si el Austro comueve al tridente ... (*Laberinto de Fortuna*, Juan de Mena, 1481)

Nótese que también encontramos muchos casos sin *a*. Sin embargo, como Aissen nos dice que los casos con *a* no deberían ocurrir, nos centramos en los casos con *a*. Algunos de estos casos podrían explicarse analizando *a* como una preposición direccional en lugar de un marcador de objeto. Además, en algunos casos, como la frase (4-b), podríamos reinterpretar el objeto como una personificación, convirtiéndose en un objeto humano. En el castellano moderno a menudo se usa *a* con sustantivos personificados (Batchelor & San José, 2010, pp. 377-378). No obstante, algunos de estos casos no dejan de extrañarnos. El castellano moderno generalmente no usa *a* con estos verbos. Podría ser que en el pasado, algunos de estos verbos llevaran la preposición como un *phrasal verb*, pero parece poco probable en el caso de, por ejemplo, ‘co(n)mover’. No está claro por qué este verbo necesitaría una preposición direccional. En cualquier caso, exige más investigaciones sobre los usos exactos de estos verbos en el español histórico para entenderlo completamente. Algunos casos tienen otras interpretaciones posibles, pero siguen difíciles de explicar en las teorías discutidas.

## 4.7 Conclusión

En este capítulo hemos analizado los hallazgos de nuestros datos por las dimensiones introducidas por Aissen. En el próximo capítulo, veremos cómo se relacionan nuestros resultados con la literatura y qué son sus consecuencias para futuras investigaciones.

## Capítulo 5

# La discusión: hacia una visión global del DOM

Al discutir los resultados, surgen dos hallazgos principales. En primer lugar, la mayoría de los datos confirman las hipótesis de Aissen (2003); en algunas categorías no hay marcado, en otras hay marcado obligatorio y además hay categorías con marcado opcional. Sin embargo, también encontramos algunas excepciones interesantes. Por ejemplo, la falta de marcas con nombres de humanos y la presencia de marcas con objetos inanimados definidos. Se han propuesto algunas posibles explicaciones, pero sigue siendo difícil explicar en detalle estas anomalías.

En segundo lugar, al examinar la categoría de los sustantivos definidos humanos, descubrimos que usar marcas gramaticales era, de hecho, opcional. Además, observamos a lo largo del tiempo un aumento aparente en el uso de marcas, lo que apoya la teoría de Aissen. También encontramos alguna evidencia de que el orden de las palabras podría ser relevante, pero necesitamos más investigación al respecto.

En resumen, llegamos a la conclusión de que el DOM en español se basa en la animacidad y la definitud. Además, es probable que el orden de las palabras también importe. A lo largo del tiempo, los casos en los que se aplica el DOM se mueven en función de estos tres criterios.

### 5.1 Implicaciones

En esta sección discutimos cómo los resultados se relacionan con la teoría de Aissen y los otros autores. Primero observamos la teoría de Aissen, luego ampliamos nuestra

perspectiva a otras teorías y finalmente volvemos brevemente a los pronombres.

### 5.1.1 Teoría de Aissen

En general, concluimos que los resultados respaldan la teoría de Aissen, salvo las excepciones mencionadas anteriormente.

Especialmente en el caso de los sustantivos inanimados indefinidos podemos preguntarnos si es necesaria una modificación de la teoría de Aissen. Hay dos maneras posibles de abordar esta incoherencia. Primero, podemos argumentar que los casos encontrados que usan *a* no son realmente casos de *a* como marcador de objeto sino, por ejemplo, una preposición de dirección. Ya se discutió brevemente en el capítulo anterior. Si podemos volver a analizar todos los casos de esta manera, la teoría de Aissen no exige cambios. Sin embargo, como algunos de los ejemplos usan verbos que normalmente no llevan *a*, parece poco probable. Si creemos que no podemos explicar todos estos casos de otra manera, deberíamos volver a examinar críticamente la teoría de Aissen. Así que supongamos por ahora que los casos discutidos en el capítulo anterior son evidencia convincente de que *a* también puede usarse en el caso de inanimados definidos. Si volvemos a la figura 2.2, vemos que los elementos inanimados definidos caen justo debajo de la línea inferior del lado derecho de la figura. Por lo tanto, podríamos argumentar que la teoría general de Aissen es válida y que la jerarquía combinada de la figura es correcta, pero que las líneas están dibujadas de manera incorrecta. En ese caso, propondríamos volver a dibujar las líneas según la figura 5.1. En esta figura, las líneas verdes claras denotan la teoría modificada para el español, mientras que las líneas azules y verdes oscuros muestran el dibujo original de Aissen. En la teoría adaptada, el marcado de inanimados indefinidos es opcional en lugar de prohibido, lo que está en línea con los datos del capítulo anterior. Obviamente, la teoría modificada necesita ser comprobada en futuras investigaciones. Además, si realmente creemos que esto es cierto, necesitamos más datos históricos para rastrear el desarrollo del marcado de sustantivos inanimados indefinidos. Melis (1995), que sirvió de base para Aissen (2003), no analiza los sustantivos definidos inanimados, por lo que no está claro cuál sería la mejor línea de dibujar en el castellano español antiguo. Podría ser un tema para futuras investigaciones.

Mirando algunas investigaciones recientes, vemos que para el español moderno también se ha atestiguado que los objetos directos definidos inanimados pueden estar marcados. García García (2007), por ejemplo, discute muchos ejemplos y ha investigado en qué circunstancias sucede. Otorga un papel destacado para el tipo de verbo. Futuras investigaciones podrían, a su vez, confirmar en qué medida estos hallazgos están en

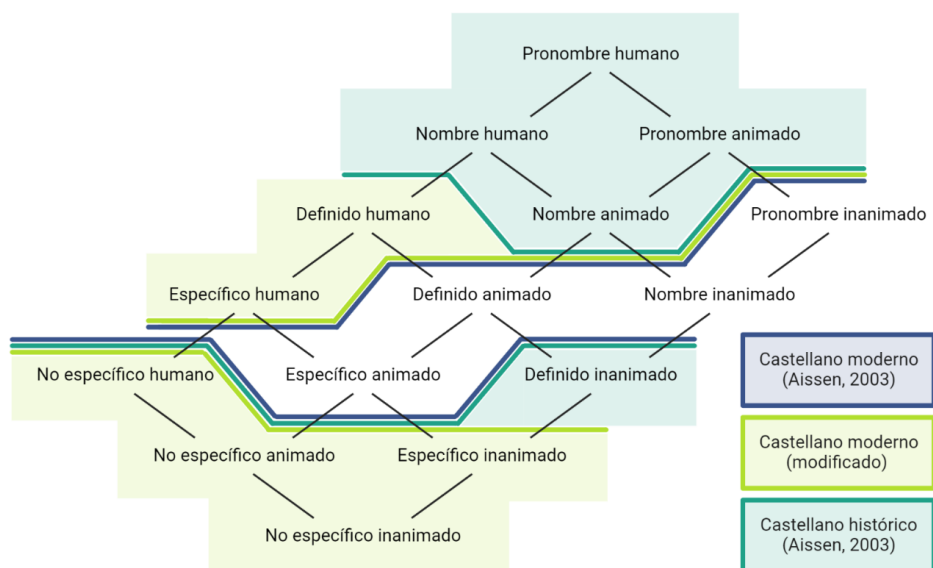


Figura 5.1: La teoría de Aissen modificada al incluir la posibilidad de marcado opcional para los indefinidos inanimados. Las líneas rojas denotan la teoría modificada para el español, mientras que la línea rosa muestra la línea original de Aissen. Figura producida por Maaïke Wessel.

consonancia con los textos de los corpus históricos.

Nos lleva a un inconveniente importante de la teoría de Aissen; aunque elegante, proporciona una gran área en la que el marcado es 'opcional'. No es muy satisfactorio y esperaríamos que algunas o muchas otras variables influyan y mitiguen esta opcionalidad. Los factores pueden ser: topicalidad, tipo de verbo, diferencias dialectales, número y persona del sujeto y objeto, etc. (Lizárraga Navarro & Mora-Bustos, 2010). Por lo tanto pasamos ahora a los otros factores identificados en investigaciones previas.

### 5.1.2 Otros factores

Encontramos que el orden de las palabras/topicalidad parecía tener influencia en el caso de los sustantivos definidos humanos, pero los datos limitados dificultan la evaluación adecuada. Parece que este no es el único factor, ya que sigue habiendo opcionalidad de *a* incluso cuando se tenía en cuenta el orden de las palabras, especialmente el orden verbo-objeto. Cualquiera de los otros factores mencionados o un factor aún por descubrir podría tener un papel importante en este caso. En cualquier caso, parece que la teoría de Aissen proporciona un buen punto de partida que incluye gran parte de la variedad

que se encuentra en el uso del DOM y concretamente en el DOM en español, pero aún no explica todo. Las teorías de von Heusinger y Kaiser (2005), von Heusinger (2008) y García García (2007) ayudan explicar la variedad.

Para obtener una mejor comprensión del DOM en español, la investigación deberá explorar la combinación de diferentes factores. Este tipo de investigación requiere una gran cantidad de datos; idealmente, necesitaríamos unas docenas de oraciones para cada combinación de factores. En la investigación actual, ya hemos visto que era difícil encontrar suficientes datos para las quince categorías de animacidad y definitud combinadas. Si queremos tener al menos 20 puntos por combinación de factores, necesitaríamos  $3 \times 5 \times 4 \times 20 = 1200$  puntos/oraciones. Incluir factores como el orden de las palabras o los tipos de verbos significa que se necesitarán aún más datos. Por supuesto, la dimensión histórica complica aún más la discusión, ya que se necesitan suficientes datos de todos los períodos bajo investigación. Por lo tanto, un estudio exhaustivo que analice todos los factores combinados sería imposible hasta que encontremos formas eficientes de manejar todos estos datos. Por lo tanto, los estudios anteriores se han centrado en uno, dos o tres factores. Los sistemas del procesamiento del lenguaje natural podrían constituir una solución (parcial) a este problema, pero actualmente, como ya señalábamos, no tienen la calidad suficiente.

### 5.1.3 Pronombres

Finalmente, encontramos que los pronombres no siguen ninguna teoría debido a la complicada situación de los pronombres en el español.

En los resultados observamos que los pronombres clíticos no usan *a*. Por alguna razón, en investigaciones anteriores nunca se han discutido en relación con la teoría de Aissen. Podemos preguntarnos por qué los pronombres clíticos no usan la preposición. Si creemos en las teorías del DOM que se basan en la reducción de la ambigüedad, notamos que los pronombres conservan un sistema de casos similar al que se usaba en latín. Significa que no hay posibilidad de confusión sobre cuál es el objeto y cuál es el sujeto; ‘yo’ es siempre sujeto mientras que ‘me’ es siempre objeto. Explicaría por qué no es necesaria la preposición en el caso de los pronombres clíticos. La pregunta entonces es por qué sí se usa en el caso de pronombres no clíticos, que todos usan *a* (véase figura 4.2 del capítulo anterior). Generalmente, este tipo de pronombre no proclítico se usa para enfatizar en ciertos órdenes de palabras. Por lo tanto, podría ser que la *a* proporcione un énfasis adicional. Sin embargo, no encontramos evidencia de que el orden de las palabras influyera en la duplicación o no. Además, el hecho de que no se

usen los pronombres no clíticos para objetos animados o inanimados nos confirma que la animicidad probablemente es de importancia. Sin embargo, exige más investigación para examinar con más detalle el surgimiento de este tipo de pronombres. Se han realizado algunos estudios sobre clíticos y el DOM (vea, por ejemplo, Leonetti (2008)), pero, hasta donde sepamos, no en combinación con los factores investigados en este estudio.

## 5.2 Limitaciones

Esta investigación está limitada en algunos aspectos. El más importante es la limitada cantidad de datos analizados. Anotar a mano es un trabajo engorroso que requiere mucho tiempo. Significa que no se anotaron todos los textos y oraciones disponibles. Plantea cuestiones de generalización. No fue posible hacer pruebas estadísticas en muchos casos por la cantidad limitada.

Además, este estudio ha demostrado que, si bien las herramientas modernas de procesamiento del lenguaje natural son muy útiles, se necesita más desarrollo técnico antes de que puedan funcionar sin supervisión humana. El anotador que se utilizó alcanza un 99 % de precisión en los textos de los periódicos modernos, pero obtuvo una precisión muy inferior en los textos en español antiguo<sup>1</sup>. Por lo tanto, es necesario que el campo del procesamiento del lenguaje natural comience a poner a disposición un anotador para variaciones más antiguas de idiomas así como sus variedades modernas. Si dicho sistema estuviera disponible, se podría combinar con clasificadores automáticos de animicidad para automatizar completamente el procedimiento utilizado en esta investigación. Ya se han dado pasos significativos en la clasificación automática de la animicidad (Bloem & Bouma, 2013; Bowman & Chopra, 2012), lo que podría ser de interés en investigaciones como la presente.

Además, la selección de textos en esta investigación no fue la ideal. Los géneros difieren entre las épocas investigadas. Sobre todo, la poesía no sigue de forma estricta las reglas de la lengua y por eso podría haber utilizado *a* de otra manera que los textos del género de prosa. Es difícil evitar porque los textos antiguos son complicados de encontrar en ciertos géneros. Sin embargo, sigue siendo algo a tener en cuenta.

Los datos también muestran que el cambio de opcional a obligatorio aún no se había completado al final del periodo investigado. Significa que investigaciones futuras podrían analizar un periodo más amplio, especialmente hacia el español moderno.

---

<sup>1</sup>No calculamos la precisión exacta, pero la cantidad de correcciones que se hicieron en la anotación manual claramente lograron una tasa de error del 1%.

Finalmente, sería interesante para investigaciones futuras examinar de manera sistemática y cuantitativa hasta qué punto el número del objeto y del sujeto influyen el uso de *a*. Con todo, si el sujeto y el objeto tienen un número diferente, el acuerdo verbal evita cualquier confusión. Por lo tanto, esperaríamos que la preposición fuera menos necesaria. Debería ser visible en el período anterior a la estandarización, donde el hablante tiene libertad de acción sobre cuándo y cómo use la preposición. Las investigaciones anteriores han prestado relativamente poca atención a esto.

### 5.3 Conclusión

En resumen, aún hay muchas incógnitas en el funcionamiento del DOM español y sobre los factores que rigen exactamente su uso, tanto en el español actual como en el español histórico. Esta investigación ha dado algunos pasos para completar nuestros conocimientos del tema. El mayor desafío será encontrar formas eficientes de analizar grandes cantidades de datos necesarias para comprender completamente el fenómeno en cuestión.

# Bibliografía

- Aissen, J. (2003). Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory*, 21 (3), 435-483. <https://doi.org/10.1023/A:1024109008573>
- Batchelor, R. E. & San José, M. A. (2010). *A reference grammar of Spanish* [OCLC: ocn502037201]. Cambridge University Press.
- Bloem, J. & Bouma, G. (2013). Automatic animacy classification for Dutch. *Computational Linguistics in the Netherlands Journal*, 3, 82-102. Consultado el 13 de agosto de 2022, desde <https://clinjournal.org/clinj/article/view/27>
- Bosque, I. & Demonte, V. (Eds.). (1999). *Gramática descriptiva de la lengua española*. Espasa.
- Bowman, S. R. & Chopra, H. (2012). Automatic Animacy Classification. *Proceedings of the NAACL HLT 2012 Student Research Workshop*, 4.
- Butt, J., Benjamin, C. & Rodríguez, A. M. (2019). *A new reference grammar of modern Spanish* (6th edition). Routledge.
- de Hoop, H. & Malchukov, A. L. (2008). Case-Marking Strategies. *Linguistic Inquiry*, 39(4), 565-587.
- de Swart, P. (2007). *Cross-linguistic Variation in Object Marking* (PhD Thesis). Radboud University. Nijmegen.
- Fábregas, A. (2013). Differential Object Marking in Spanish: state of the art [Number: 2]. *Borealis – An International Journal of Hispanic Linguistics*, 2(2), 1-80. <https://doi.org/10.7557/1.2.2.2603>
- Ferreira, V. S. (2008). Ambiguity, Accessibility, and a Division of Labor for Communicative Success. En B. H. Ross (Ed.), *Psychology of Learning and Motivation* (pp. 209-246). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)00006-6](https://doi.org/10.1016/S0079-7421(08)00006-6)
- García García, M. (2007). Differential Object Marking with inanimate objects. En G. A. Kaiser & M. Leonetti (Eds.), *Proceedings of the Workshop „Definiteness, Specifi-*

- city and Animacy in Ibero-Romance Languages*". Konstanzer Online-Publikations-System.
- Haspelmath, M. (2008). 8 Creating Economical Morphosyntactic Patterns in Language Change. En J. Good (Ed.), *Linguistic Universals and Language Change* (pp. 185-214). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199298495.003.0008>
- Kurumada, C. & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, 83, 152-178. <https://doi.org/10.1016/j.jml.2015.03.003>
- Leonetti, M. (2008). Specificity in Clitic Doubling and in Differential Object Marking [Publisher: De Gruyter Mouton Section: Probus], 20(1), 33-66. <https://doi.org/10.1515/PROBUS.2008.002>
- Lizárraga Navarro, G. Z. & Mora-Bustos, A. (2010). Variación en la marcación diferenciada de objeto en español. *Forma y Función*, 23(1), 9-38.
- Melis, C. (1995). El objeto directo personal en el Cantar de Mio Cid: Estudio sintáctico-pragmático. En C. Pensado (Ed.), *El complemento directo preposicional* (pp. 133-163). Viso Libros.
- Penny, R. J. (2002). *A history of the Spanish language* (2nd ed). Cambridge University Press.
- Pensado, C. (1995). El Complemento Directo Preposicional: Estado de la Cuestión y Bibliografía Comentada. En C. Pensado (Ed.), *El complemento directo preposicional*. Viso Libros.
- Sánchez-Martínez, F., Martínez-Sempere, I., Ivars-Ribes, X. & Carrasco, R. (2013). An open diachronic corpus of historical Spanish. *Language Resources and Evaluation*, 47. <https://doi.org/10.1007/s10579-013-9239-y>
- Silverstein, M. (1986). Hierarchy of features and ergativity. En P. Muysken & H. v. Riemsdijk (Eds.), *Features and Projections* (pp. 163-232). De Gruyter. <https://doi.org/10.1515/9783110871661>
- Tippets, I. (2011). Differential Object Marking: Quantitative Evidence for Underlying Hierarchical Constraints across Spanish Dialects. *Selected Proceedings of the 13th Hispanic Linguistics Symposium*, 12.
- von Heusinger, K. (2008). Verbal semantics and the diachronic development of DOM in Spanish [Publisher: De Gruyter Mouton Section: Probus], 20(1), 1-31. <https://doi.org/10.1515/PROBUS.2008.001>

- von Heusinger, K. & Kaiser, G. A. (2005). The evolution of differential object marking in Spanish. *Proceedings of the Workshop "Specificity and the Evolution / Emergence of Nominal Determination Systems in Romance"*, 33-69.