

RASAL

LINGÜÍSTICA

2023: 35-57

Recibido: 31.08.2022 | Aceptado: 03.03.2023

DOI: <https://doi.org/10.56683/rs232089>

UNA SINTAXIS GENERATIVA PARA EL MODELAMIENTO COMPUTACIONAL DE LA ELIPSIS NOMINAL DEL ESPAÑOL

A GENERATIVE SYNTAX FOR THE COMPUTATIONAL MODELING OF SPANISH NOMINAL ELLIPSIS

Hazel Barahona Gamboa
Universidad Nacional Arturo Prat

<https://orcid.org/0000-0001-9245-3645>

RESUMEN

En este trabajo se estudia la elipsis nominal del español desde los postulados generales de la Gramática Generativa (Chomsky, 1965, 1995) y la Teoría de la Identidad parcial (Saab, 2008, 2019) y, a partir de su explicación, se representa su alcance en una gramática computacional capaz no solo de identificar estas formaciones, sino también de reponer el correspondiente elemento nominal elidido. Si bien se muestra un componente computacional, el foco de atención es la estructura gramatical y cómo es un instrumento idóneo para resolver algunos de los problemas de la identificación automática de estas construcciones y otras similares. Ahora bien, la elipsis nominal no cuenta con un minucioso estudio como sí sucede con la elipsis verbal. Para comprobar la representación computacional de este fenómeno sintáctico se crearon en NooJ (Silberztein, 2005, 2016) dos herramientas: un diccionario electrónico y una gramática computacional. Ambas fueron evaluadas en un corpus de 423000 palabras pertenecientes al dominio médico. Es importante mencionar que esta gramática computacional no fue entrenada ni modificada en ningún corpus previo. En un barrido manual se identificaron 355 elipsis, de las cuales se logró detectar un total de 335 elipsis. Los resultados obtenidos fueron los siguientes: un 99,11 % de precisión, un 94,36 % de cobertura y un 96,68 % de medida f.

PALABRAS CLAVE: gramática; elipsis nominal; gramática generativa; Teoría Identidad Parcial; lingüística computacional.

ABSTRACT

In this paper, Spanish nominal ellipsis is studied from the general principles of Generative Grammar (Chomsky, 1965, 1995) and Partial Identity Theory (Saab, 2008, 2019). On the basis of

their explanation, nominal ellipsis is represented in a computational grammar capable of not only identifying this information, but also of replacing the respective nominal element that had been elided. Even though a computational component is shown, the main focus is on the grammatical structure and how this is a suitable tool to solve some problems of automatic identification of these constructions and similar ones. However, nominal ellipsis has not been as meticulously studied as verbal ellipsis has. To verify the computational representation of this syntactic phenomenon, two tools were created in NooJ (Silberztein, 2005, 2016): an electronic dictionary and a computational grammar. Both were evaluated in a 423,000-word corpus from the medical domain. It is important to mention that this computational grammar was not trained or previously modified in any other corpus. In a manual revision, 355 cases of ellipsis were identified, 335 of which were detected. The results were the following: 99.11% of accuracy, 94.36% of coverage and 96.68% of f-measure.

KEYWORDS: grammar; nominal ellipsis; generative grammar, Partial Identity Theory; computational linguistic.

1. Introducción

En este trabajo se estudia la elipsis nominal del español, entendida, inicialmente, como el “silenciamiento” del segundo elemento nominal bajo restricciones sintácticas particulares. Como por ejemplo en (1), en donde hay elipsis del segundo miembro coordinado *casa*. Se indica con el tachado la presencia de elipsis.

- (1) La casa de Juan y la ~~casa~~ de María.

Para describir este fenómeno se consideran los postulados teóricos de la Gramática Generativa (GG) y se modela su alcance descriptivo en una gramática computacional capaz no solo de identificar estas estructuras gramaticales, sino también de reponer el correspondiente elemento nominal elidido. Si bien se muestra un componente computacional, el foco de atención es la estructura gramatical y cómo esta es un instrumento idóneo para resolver algunos de los problemas de la identificación automática de estas construcciones y otras similares. Para esto, se empleó NooJ (Silberztein, 2005, 2016) un *software* computacional cuyo objetivo es describir exhaustivamente todas las oraciones de una lengua por medio del reconocimiento automático de textos escritos. Esta formalización se aplica en un corpus de textos escritos en lenguaje natural pertenecientes al dominio médico.

La elipsis ha sido abordada desde la sintaxis por Chomsky (1965, 1995), Sag (1976); Williams (1977), Fiengo y May (1994), Chung *et al.* (1995), Lobeck (1995), Lasnik (1999), Merchant (1999), Fox (2000) y Lasnik y Funakoshi (2019), desde la semántica por Hankamer y Sag (1984), Merchant (2001), van Craenenbroeck (2010a) y Aelbrecht (2010) o, a través de una combinación de ambas por Kehler (2002), Chung (2013) y Merchant (2013). Sin embargo, estos estudios se concentran exclusivamente en la elipsis verbal que consiste en el no pronunciamiento del verbo. No obstante, la elipsis nominal no cuenta con estudios minuciosos sobre su naturaleza sintáctica o sobre posibles contextos estructurales que permitan clasificar el contexto de aparición. Una de las razones podría deberse a su

aparición en contextos más amplios que su contraparte verbal. Sumado a esta complejidad estructural se encuentra que distintos fenómenos en los que suele omitirse alguna categoría gramatical se han denominado indistintamente elipsis nominal, como en Brucart (1987, 1999), la Real Academia Española (RAE, 2009), Brucart y Macdonald (2012) y Brucart y Fernández-Sánchez (2018) para quienes esta es un tipo de relación anafórica. Lo anterior también se replica en las investigaciones de corte computacional (Ferrández y Peral, 2000; Mel'čuk, 2006, Mitkov, 2001, 2002; Rello y Ilisei, 2009; Rello, 2010) que han denominado elípticas y nominales tanto a las categorías vacías, sujetos nulos con y sin flexión verbal como a los pronombres nulos expletivos, lo que ha ocasionado errores en su definición y descripción.

Para la representación computacional de este fenómeno sintáctico se crearon en NooJ (Silberztein, 2005, 2016) dos herramientas: un diccionario electrónico y una gramática computacional.¹ Por un lado, la función principal del diccionario electrónico es asociar una entrada léxica con su respectiva información morfosintáctica. Este funciona como un analizador morfosintáctico que etiqueta cada palabra con su correspondiente información morfosintáctica. Para este trabajo se empleó el diccionario del Proyecto FONDECYT 1171033 compuesto por 67862 entradas tomadas del Diccionario de la RAE.² Por otro lado, las gramáticas computacionales reconocen si una secuencia de texto pertenece o no a un lenguaje definido por una gramática. Estas secuencias gramaticales trabajan con la definición de variables –como las reglas de reescritura– que almacenan subsecuencias de elementos (rasgos morfosintácticos), con las que es posible agrupar, reubicar y ensamblar en distintos niveles de la gramática una misma descripción sintáctica. Ambos fueron evaluados en un corpus de 423000 palabras pertenecientes al dominio médico, también del proyecto FONDECYT. Es importante mencionar que esta gramática computacional no fue entrenada ni modificada previamente en ningún corpus previo. En un barrido manual se identificaron 355 elipsis, de las cuales se logró detectar un total de 335. Los resultados obtenidos fueron los siguientes: un 99,11 % de precisión, un 94,36 % de cobertura y un 96,68 % de medida f.

En lo que respecta a la organización de este artículo, en la primera sección se hace mención a algunos de los problemas sobre la definición de la elipsis nominal. Esto se aborda a través de un recorrido teórico que inicia en la GG y su diferencia con PRO y pro, para finalizar con la discusión de Saab (2008, 2019) y su propuesta. En la segunda, se presenta la *Teoría de la Identidad parcial* de la elipsis (TIP; Saab, 2008, 2019) cuyos postulados se adoptan en este estudio. En tercer lugar, se presenta cómo modelar una gramática computacional en NooJ (Silberztein, 2005, 2016). En cuarto lugar, se muestra la elaboración de las herramientas (el diccionario electrónico y la gramática computacional). En quinto lugar, la validación de la implementación y los resultados obtenidos. Luego, en sexto lugar, la discusión de los resultados. Aquí, se anotan algunos fenómenos observados tanto en las elipsis no detectadas, como en las detecciones erróneas y, finalmente, las conclusiones.

2. Problemas sobre la definición de la elipsis nominal

Uno de los primeros registros sobre el tratamiento de la elipsis como fenómeno sintáctico con una función particular en la lengua fue en los estudios clásicos de la retórica y la oratoria. Sin embargo, no hay una explicación sistemática y formal de sus características y propiedades particulares, debido a que la omisión de algún segmento de la oración era considerada inadecuada y errónea, por lo que no formó parte de las figuras retóricas o de construcción que embellecían la lengua (Correas, 1967). No obstante, en los estudios contemporáneos la elipsis ha pasado a tener un papel destacado y se han formulado propuestas en las que este fenómeno marca un giro investigativo, en especial con la explicación de las categorías vacías (Chomsky, 1995) y en el desarrollo de la GG (Chomsky, 1965, 1995; Chung *et al.*, 1995; Fiengo y May, 1994; Fox, 2000; Lasnik, 1999, Lobeck, 1995; Merchant, 1999; Sag, 1976; Williams, 1977).

Un caso particular de interacción entre la forma y significado es la elipsis, pues como afirma Merchant (2001) representa un quiebre entre esa correspondencia, así como se ilustra en (2). Con *itálicas* se destaca el elemento antecedente.

- (2) María *trabaja* con Juan y Pedro *trabaja* con Antonio.

En (2) hay una elipsis verbal (*trabajar*) en el segundo miembro coordinado (*Pedro trabaja con Antonio*), esto es así porque es posible interpretar la información faltante (el verbo) a partir de su elemento antecedente. Inicialmente, se puede definir la elipsis como un proceso de eliminación del material fonológico que aparece citado previamente (Gallego, 2011, p.11). Brucart (1999) la define como un “mecanismo limitador de la redundancia léxica” (p. 2789), puesto que, si bien las oraciones de (3) son gramaticalmente adecuadas, también son innecesariamente reiterativas, ya que se repiten los elementos léxicos (*estudio, sabe inglés, presidente, hijo, se enamoró Antonio, compró y comprar pan*) mencionados.

- (3) a. Antonio *estudia* medicina y yo *estudio* lingüística.
 b. Juan *sabe* inglés, pero Antonio no *sabe inglés*.
 c. El *presidente* de Costa Rica visitó al *presidente* de Holanda.
 d. El *hijo* de Juan y el *hijo* de Pedro fueron al cine.
 e. *Antonio se enamoró*, pero no sé de quién *se enamoró Antonio*.
 f. María *compró* vino tinto y Antonio *compró* vino blanco.
 g. María quería *comprar pan*, pero Antonio no la dejó *comprar pan*.

En términos de Merchant (2001), los procesos elípticos permiten la economía en la expresión debido a la omisión de estructuras lingüísticas que ya cuentan con la información necesaria. Según este autor, la redundancia léxica es una propiedad de los sistemas biológicos en los que existe una competencia entre la economía de la expresión (aquellos principios basados en el mínimo esfuerzo) y el requerimiento de que una oración sea comprensible (es decir, interpretable).

Brucart (1999) explica que la posibilidad de recuperar un elemento elidido es la naturaleza de su antecedente, este último lo define como la unidad léxica que fija el valor que comparte la categoría elidida o los elementos elididos. Mientras que en la estructura elíptica el sitio que ocuparía el elemento elidido es el *gap* o hueco elíptico. Como puede verse en (3), este último concepto no se delimita solo al verbo (3a, b, f) o al nombre (3c, d), sino que también implica elipsis de estructuras más complejas (3e, g). Entonces, para generalizar estas diferencias, el tamaño del hueco elíptico puede incluir solo el núcleo del sintagma o bien, el núcleo del sintagma acompañado por algunos de sus complementos. Debido a lo anterior, se ha decidido acotar el dominio de este trabajo a la elipsis nominal.

Esta relación entre antecedente y hueco elíptico también se ha explicado desde los mecanismos anafóricos. No obstante, esta descripción ha ocasionado que diferentes procesos gramaticales en los que hay una omisión sean catalogados indistintamente como elípticos. Por ejemplo, Brucart (1987, 1999), la RAE (2009), Brucart y MacDonald (2012) y Brucart y Fernández-Sánchez (2018) tratan los sujetos tácitos pronominales como “elipsis del sujeto”. Los autores incluyen este tipo de elipsis en (4) dentro de las nominales, ya que identifican el pronombre personal *ellos* como elíptico, tal vez debido a la posición que ocupan.

- (4) Ø compraron pan.

La confusión también la comparten los trabajos sobre el reconocimiento automático de la elipsis nominal (Peral y Fernández, 2000; Mel’čuk, 2006; Mitkov, 2001, 2002; Rello y Ilisei, 2009; Rello, 2010), los cuales asumen que las categorías vacías, sujetos nulos con y sin flexión verbal y pronombres nulos expletivos son casos de elipsis nominal basados en el mecanismo de recuperación anafórica. Sin embargo, esto no es mayoritariamente así, puesto que la elipsis nominal tiene que ver con un antecedente cuya categoría principal de elisión es un nombre, el cual puede acompañarse de uno, dos o ningún complemento. Es decir, los pronombres no cumplen inicialmente con esta propiedad, ya que la suya es información que se obtiene de la concordancia del verbo. Así, se elide un núcleo nominal más un adjetivo (*blusa amarilla*), pero no una categoría vacía, pues suele ser un pronombre, como se muestra en (5) y (6).

- (5) La blusa amarilla de María es igual a la _____ de Josefa.

- (6) Ø nos trajeron confites.

Una segunda diferencia es que los casos de categorías vacías, sujeto nulo con y sin flexión verbal y pronombres nulos expletivos han recibido una explicación a partir de los conceptos de *pro* y *PRO*. El primer *pro* (en minúscula) actúa como un pronombre en algunas lenguas que tienen la capacidad de omitir el sujeto, condición denominada *pro-drop*. El español es un ejemplo, como se mostró en (4), en donde la flexión verbal permite reconocer el sujeto omitido. El segundo *PRO* tiene un contexto especial asociado a las construcciones de infinitivo, en las que se omite el sujeto preverbal. Compárese lo anterior con (7a) y (7b).

- (7) a. Antonio quiere PRO estudiar medicina.
 b. *Antonio quiere él estudiar medicina.³

Del contraste del par de oraciones en (7), la categoría PRO es un pronombre que no se realiza fonéticamente y en la que, en los casos que nos ocupan, hay una correspondencia con el sujeto principal. Así, se diferencia de *pro*, por la posición del sujeto en las oraciones de infinitivo en la que no hay rasgos de concordancia explícitos. Bosque y Gutiérrez (2016) afirman que el elemento nulo PRO tiene rasgos morfológicos, aunque no fonológicos; por lo que, los rasgos de género, número y persona de PRO permiten las relaciones de concordancia. En (8) hay una categoría vacía PRO con los rasgos [masculino] y [singular] que concuerdan con el adjetivo *solo*.

- (8) Antonio insiste en PRO trabajar solo.

Entre ambas categorías (*pro* y PRO) se ubican los pronombres nulos expletivos, los cuales según Bosque y Gutiérrez (2016) tienen una posición preverbal vacía, como sucede con los verbos meteorológicos (9) o incluso con otros verbos como los impersonales.

- (9) a. Ø Hace mucho calor en Quilpué.
 b. Ø Lloverá torrencialmente el fin de semana.

Una tercera diferencia es la localidad o, dicho de otro modo, la distancia sintáctica, dado que las categorías vacías, sujetos nulos con y sin flexión verbal y pronombres nulos poseen una posición limitada de la que suele aparecer la elipsis nominal. La elipsis responde a un mecanismo sintáctico específico que no necesariamente responde a los rasgos morfosintácticos del verbo, como sucede con las categorías pronominales. La extensión de la elipsis puede ser a nivel del sintagma o a nivel oracional, como se ejemplifica en (10a) y (10b), mientras que el pronombre o la categoría vacía irá lo más cerca posible del verbo (11).

- (10) a. El hijo de Juan y el ____ de Pedro fueron a acampar a Chiloé.
 b. Las embarazadas normotensas fueron comparadas con las ____ preclámpicas.

- (11) Antonio quiere PRO ser un médico exitoso.

La semejanza entre la elipsis y las categorías vacías, sujetos nulos con y sin flexión verbal y pronombres nulos expletivos es su no realización fonética. No obstante, su sintaxis es diferente. En este trabajo se concuerda con Saab (2008), quien afirma que la necesidad de describir dos mecanismos diferentes bajo una misma teoría obedece a la explicación unificada a la que querían llegar sus primeros exponentes (Brucart, 1987; Lobeck, 1995) en el marco de las teorías del *pro* (principio de Ligamento). Lo expuesto anteriormente lo refuerza Panagiotidis (2002, 2003) cuando menciona que estas “variaciones anafóricas” pertenecen a lo que denomina *nombres nulos*, pues son elementos vacíos de contenido denotativo.

Hankamer y Sag (1976) estudian los distintos procesos anafóricos y muestran que tanto los procesos señalados como los elípticos son producto de dos anáforas distintas, a saber, las superficiales y las profundas. La elipsis que se estudia aquí se correspondería con las primeras, ya que serían estructuras sintácticas sujetas a operaciones transformacionales (en especial las de borrado). Mientras que las segundas son pronombres nulos que no cuentan con una estructura interna. No obstante, Saab (2008) señala que esa diferencia no se ajusta al español.

Para Saab (2008) la elipsis del español, en especial la verbal (SV) y la del sintagma tiempo (ST), son anáforas superficiales no locales según la clasificación de Hankamer y Sag, pues entre el sitio (hueco) elíptico y su antecedente no hay una relación de localidad estricta (como sí lo tienen PRO y PRO). Hankamer (2003) señala una serie de características particulares para la anáfora no local; no obstante, para demarcar el propósito de esta sección, se considera el requisito del centinela (un antecedente), el cual legitima la elipsis.

Brucart (1987) afirma que tanto (12) como (13) son producto de las mismas construcciones, es decir, ambas son categorías PRO.

(12) El padre de Juan y el PRO de María nunca se conocieron.

(13) El PRO que quiera vendrá conmigo.

Saab (2008) plantea que (12) no puede ser analizada transformacionalmente en términos de las anáforas superficiales, ya que no hay un nombre que esté sintácticamente presente en la Forma Fonológica, por lo que una categoría vacía es más adecuada. Sin embargo, en (13) sí hay un antecedente lingüístico *padre*. A partir de estas respuestas, ajustar dos procesos sintácticos a un mismo orden no parece lo más explicativo. Para Saab (2008), (13) es una construcción con estructura interna sujeta a lo que él define como condición de *Identidad* y, por ende, un caso propio de elipsis nominal. O, dicho de otro modo, en (12) hay un centinela que licencia la elipsis; mientras que en (13) lo único reconocible es un posible rasgo [+humano] de ese “alguien” que vendrá.

Siguiendo a Saab (2008), el rasgo principal de la elipsis es el antecedente y las dependencias sintácticas de su remanente. Por tanto, una relación elíptica catafórica no es posible, ya que “las dependencias sintácticas entre el constituyente remanente y el sitio elíptico se obtienen del mismo modo que en las oraciones no elípticas” (Saab, 2008, p. 498). Por ende, (12) no puede ser un caso de elipsis porque PRO es una construcción nula que posee una interpretación fija independiente de cualquier antecedente. En (14) se transcriben los ejemplos de (24) de Saab (2008, p. 499).

(14) El tonto / el del frente / el que quiera vendrá conmigo.

En resumen, para esta investigación, se entiende la elipsis como un mecanismo propio de la gramática. En el que se lleva a cabo un proceso de recuperación del elemento elidido posible de identificar por los mismos mecanismos de construcción de su estructura antecedente. Así, la elipsis se caracteriza por tener una estructura sintáctica que se encuentra delimitada (licenciada, restringida) por la misma sintaxis del español.

3. Teoría de la Identidad parcial de la elipsis

Para los fines de este trabajo, se entenderá la elipsis nominal como un mecanismo sintáctico que se genera a partir de los mismos ensamblajes sintácticos de su estructura antecedente, pero que sufre el silenciamiento de un elemento nominal bajo circunstancias gramaticales específicas. Como esta investigación cuenta con un componente de formalización computacional que se requiere comprobar empíricamente, se consideró la Teoría de la Identidad parcial de la elipsis de Saab (2008, 2019), ya que se plantean evidencias, ejemplos y contraejemplos del fenómeno. Dicho autor menciona que una teoría de la elipsis es una teoría de la interfaz entre la sintaxis y la Forma Fonológica. La TIP, al inscribirse en la postura generativista, se ciñe a la adecuación descriptiva y explicativa y al supuesto de la Tesis minimalista fuerte. Bajo este marco, la TIP intenta responder por qué la identidad tiene efectos en la sintaxis y por qué se silencian los elementos nominales implicados.

Para dar una respuesta a lo anterior, el autor enmarca su análisis de la elipsis en el modelo de la Morfología Distribuida (MD) (Halle y Marantz, 1993; Harley y Noyer, 1999; Embick, 2000; Embick y Noyer, 2001; Embick y Halle, 2004). La MD plantea que la sintaxis opera con nodos desprovistos de información fonológica y que este contenido se agrega a partir de lo que se denomina *Reglas de inserción de vocabulario*. La inserción tardía de exponentes fonológicos tiene un rol fundamental en la TIP de Saab.

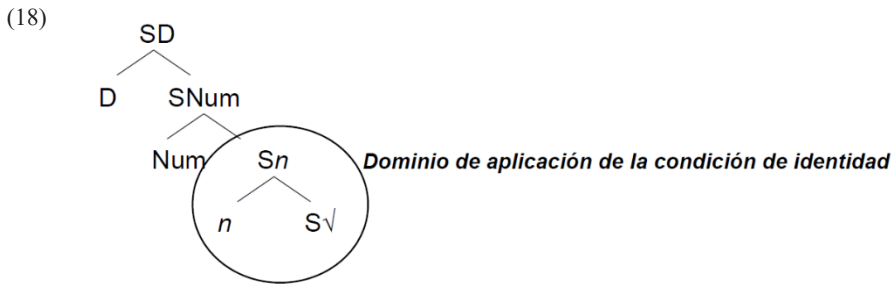
El inventario básico de terminales sintácticos con el que trabaja la MD está compuesto por morfemas abstractos y raíces. Los morfemas abstractos son aquellos que están compuestos exclusivamente por rasgos sintáctico-semánticos como [pasado], [plural], [masculino], [definido], [específico], mientras que las raíces son complejos de rasgos fonéticos que se identifican con el símbolo $\sqrt{\quad}$ e incluyen ítems como $\sqrt{\text{MES-}}$, $\sqrt{\text{AMIG-}}$, $\sqrt{\text{CANT-}}$, $\sqrt{\text{AYER}}$, $\sqrt{\text{LIND-}}$. Estas raíces aparecen “desnudas”, lo que quiere decir que no pertenecen a una clase de palabra en particular, sino que son categorizadas por un núcleo funcional definidor de categoría llamado *categorizador* y con el que debe compartir localidad. A este proceso se le denomina *Supuesto de categorización*. Embick y Marantz (2008, p. 5), siguiendo los planteamientos de Marantz (1995, 1997), lo definen de la siguiente manera:

Las raíces no pueden aparecer (no pueden pronunciarse o interpretarse) si no están categorizadas. Las raíces se categorizan a partir del ensamble sintáctico con núcleos funcionales que definen categorías. (La traducción es de Saab, 2008, p. 25)

Ahora bien, ¿qué sucede entonces con la sintaxis? Aquí es donde interviene la *Identidad*, pues en la elipsis nominal existe un contexto de no inserción léxica bajo identidad sintáctica. Esto quiere decir que, para que exista una verdadera elipsis nominal se deben compartir los rasgos que aún no han entrado en el resto de la derivación sintáctica. Recuérdese el caso de (14), en el que no es posible hablar de una elipsis nominal, ya que el único rasgo con el que eventualmente puede llenarse un terminal es con el rasgo [+humano]. Añádase a esto que la elipsis nominal comparte con su elemento antecedente una identidad del rasgo de género.

- (15) El hijo de Juan y el **hijo** de María fueron al cine
- (16) El hijo de Juan y los **hijos** de María fueron al cine.
- (17) *El hijo de Juan y la **hija** de María fueron al cine.

En la tríada de ejemplos anteriores, se observa que tanto (15) como (16) son casos de elipsis nominal, pues antecedente y núcleo nominal elidido establecen una relación de identidad amparada en la no inserción de rasgos. Lo anterior se sostiene en el principio de disyuntividad de rasgos de Embick (2000). Dicho autor menciona que los rasgos fonológicos o las propiedades arbitrarias de los ítems de vocabulario no están presentes en la sintaxis y que los rasgos sintáctico-semánticos no se insertan en la morfología. El dominio de la elipsis nominal en español es como se reproduce en (18) de acuerdo con Saab (2003).



En (18) se observa que los rasgos de número quedan fuera de la condición de identidad, por lo que, casos como los de (15) y (16) son válidos, ya que antecedente y nombre elidido comparten el rasgo de identidad de la elipsis. Mientras que los de número entran posterior a la asignación de los rasgos de género y cotejan sus respectivos rasgos con los del determinante.

Para los fines expositivos de este trabajo en particular, se parte de que el dominio de identificación tanto del antecedente como del lugar de la elipsis es el sintagma determinante, como se muestra en (18). Así, se adoptan las dos operaciones con las que se caracteriza esta estructura en las lenguas latinas (Cinque, 1993; Longobardi, 1994 y, para el español, Bosque y Picallo, 1996). En primer lugar, el no ascenso de los nombres a SNUM y, en segundo lugar, la operación post-sintáctica de copiado de rasgos (Halle y Marantz, 1993; Saab, 2008) vía disociación de rasgos, en especial los de género y número en el determinante.

Dado que la teoría es adecuada descriptivamente y responde mayoritariamente al fenómeno de la elipsis nominal en español, su representación debiera permitir construir una metodología de identificación automática aplicable en textos escritos. Esto es lo que se detalla a continuación.

4. La gramática generativa como herramienta computacional: NooJ

Una de las críticas hacia la gramática generativa es la procedencia de sus datos y cómo estos pueden dar evidencia sobre las posibilidades combinatorias del lenguaje. NooJ intenta ser un programa que, basado en formalismos gramaticales, intenta dar cuenta de cómo emplearlos como mecanismo de búsqueda de información y descripción de un fenómeno particular de la lengua. Por ende, puede afirmarse que es una herramienta valiosa para construir corpus específicos con los cual dar evidencia sobre la adecuación y explicación descriptiva.

Trabajos como los de Hardt (1993, 1997), Lappin y Leass (1994), Lappin y Shih (1996), Hardt y Rambow (2001), Nielsen (2004), Lappin (2005), McShane y Babkin (2016) identifican solamente el antecedente, debido a que las categorías vacías, pronombres, sujetos nulos y relaciones anafóricas son tratadas como fenómenos similares y sintácticamente equivalente, porque comparten el mecanismo de búsqueda hacia atrás. Dicho de otro modo, el objetivo es identificar en el contexto previo, es decir, el elemento “omitido”. Si bien los resultados son prometedores, el fin es optimizar el porcentaje de rendimiento. La mayoría de estos estudios tiene como base aprendizajes automatizados, etiquetadores morfosintácticos previamente establecidos y listas de mejores candidatos a antecedente elíptico. Otros estudios como los de Peral y Ferrández (2000), Mitkov (2002), Han (2004), Pastor (2008), Hu (2008), Rello y Ilisei (2009) y Rello (2010) comparten propósitos similares.

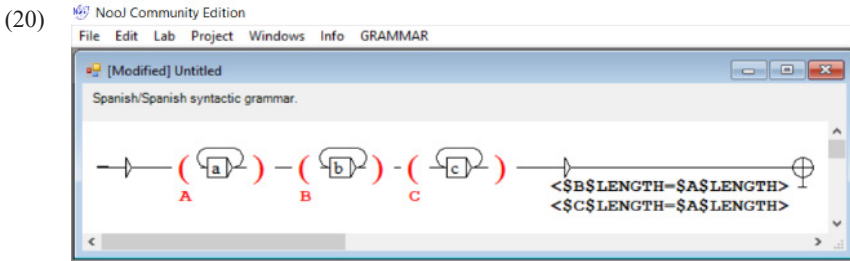
Debido al contexto que rodea a estas estructuras en el dominio computacional, se decidió trabajar con NooJ (Silberztein, 2005, 2016). Esta es una herramienta computacional cuyo objetivo es describir exhaustivamente todas las oraciones de una lengua, por medio del reconocimiento automático en textos escritos. Silberztein (2005, 2018) señala que algunos problemas de aplicación del Procesamiento del Lenguaje Natural pueden resolverse de manera económica a través de diccionarios electrónicos, gramáticas morfológicas y sintácticas basadas en la Jerarquía de gramáticas formales de Chomsky-Schützenberger (Chomsky, 1957), con la que es posible tomar de los distintos niveles de análisis formal aquellas propiedades que mejor se ajustan a la descripción sintáctica.

NooJ permite crear varias herramientas que le dan autonomía al análisis. La primera es la creación de diccionarios electrónicos cuya función principal consiste en asociar una entrada léxica con su respectiva información morfosintáctica. Este funciona como un analizador morfosintáctico que etiqueta cada palabra con su correspondiente información morfosintáctica. Para crear este diccionario se toma un ítem (entiéndase *una palabra*) y se le asocia un modelo de flexión que responde a los rasgos de concordancia, como se muestra en (19).

(19) abeja, N+sing+fem+FLX=CASA

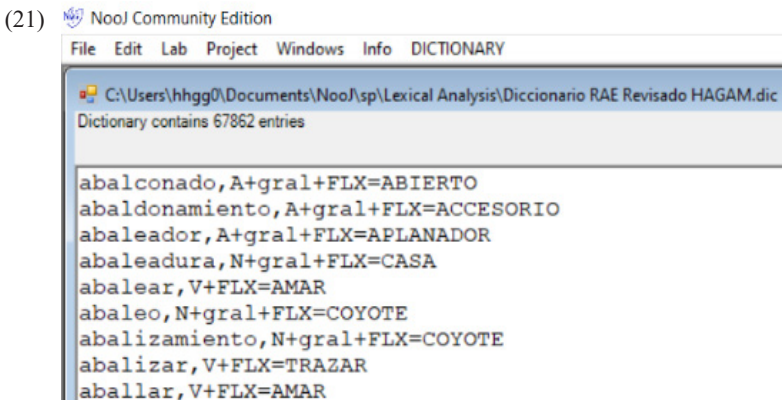
En (19) las siglas FLX representan el modelo de flexión y ejecuta los rasgos *nombre femenino singular y plural*. Este modelo se puede aplicar a todos los ítems que se flexionen del mismo modo.

La segunda herramienta son las gramáticas computacionales que reconocen si una secuencia de texto pertenece o no a un lenguaje. Estas gramáticas trabajan con la definición de variables –al modo de las reglas de reescritura– las cuales almacenan subsecuencias de elementos (rasgos morfosintácticos), con las que es posible agrupar, reubicar y ensamblar en distintos niveles de la gramática una misma descripción sintáctica. Un sencillo ejemplo es el de (20) en donde hay tres variables \$A, \$B y \$C dentro de las que se almacenan secuencias de a, b y c. Esta gramática reconoce secuencias como “aaabbbccc” pero no “aaabccc”, debido a la regla de restricción de contexto indicada con paréntesis angulares (<>) que especifica la misma cantidad de a, b y c.



5. Modelamiento de la Teoría de la Identidad parcial: elaboración de herramientas

El diccionario electrónico empleado en esta investigación se creó en el Proyecto FONDECyT 1171033 y se tomaron todas las entradas léxicas del Diccionario de la RAE. Este diccionario cuenta con 67862 entradas, todas anotadas con su respectivo modelo de flexión. El resultado se observa en (21).



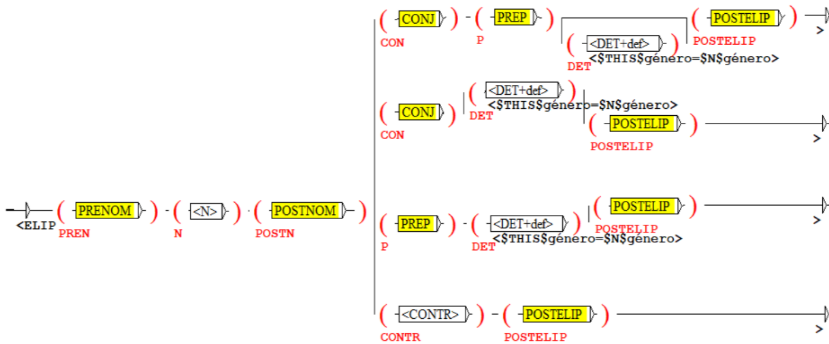
El diccionario de (21) permite identificar las distintas categorías léxicas en las que puede aparecer una palabra.

Para la representación de la sintaxis, se definieron dos formalizaciones: una para la producción en la que se caracteriza la estructura sintáctica completa de la elipsis, con sus respectivas restricciones combinatorias; y otra, para la formalización sobre la comprensión para la identificación de los remanentes del sitio de la elipsis. Para la primera formalización, se trabajó con las definiciones y organización de variables. Estas se delimitaron considerando la distribución, las características y las combinaciones sintácticas presentes en la estructura antecedente y en las de los remanentes. Las combinaciones que preceden al N principal son recogidas en PRENOM y las que le siguen POSTNOM. Para reconocer el elemento antecedente y la estructura elíptica, se definieron variables a partir de la construcción que las une, como se menciona a continuación: (i) para las coordinadas CONJ intervienen elementos como *y, ni, e, o, u*; (ii) para las verbales, SV, ya que media un verbo o se inserta una subordinación; y en (iii) las comparativas, COMP para las posibilidades *menor, mayor, peor, mejor*. Para las variables de los remanentes, se sigue el mismo criterio, DET para los determinantes y POSELIP para todas aquellas que continúan al sitio de la elipsis nominal. Algunos casos identificados en el corpus de textos de dominio médico son las que se observan (22).

- (22) a. La tasa de escolaridad y las tasas de vacunación afectaron la salud de los niños.
- b. El informe patológico definitivo fue el informe de un tumor de Sértoli-Leydig

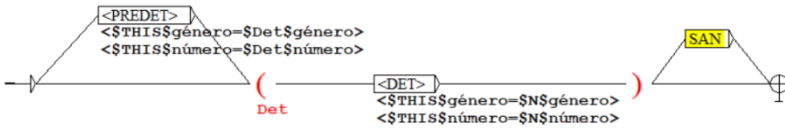
A partir de lo anterior, se modeló la gramática computacional de la siguiente manera.

(23)



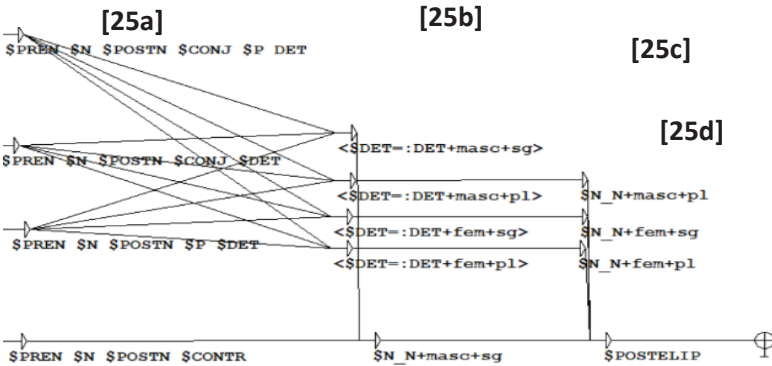
La gramática computacional de (23) identifica y restringe los posibles contextos de aparición de los elementos antecedentes de la elipsis nominal y reconoce casos como los de (22). Cada una de las variables en (23) contiene a su vez gramáticas incrustadas que aparecen en amarillo. Estas últimas tienen la particularidad de contener reglas autorreferenciales; es decir, reglas que se definen a partir de las mismas reglas. En lenguaje formal se le denomina *recursión*. En PREN hay una gramática incrustada denominada PRENOM cuyas combinaciones son las que se muestran en (24).

(24)



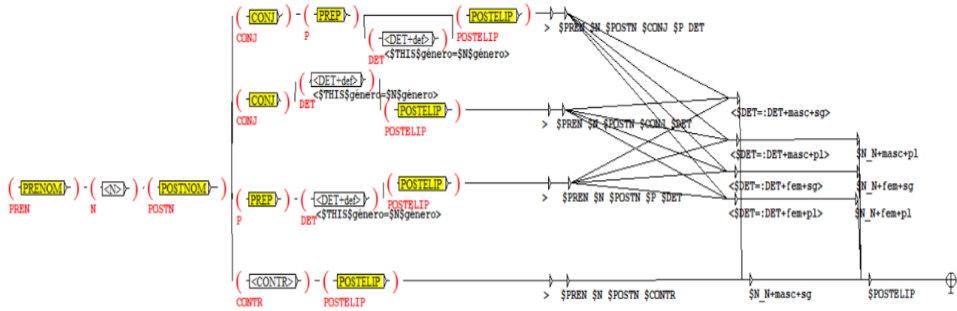
Tanto en (23) como en (24), DET respeta la condición de Identidad y de no inserción del N elidido, considerando que el género debe coincidir con el N antecedente en todo momento. Ahora bien, si verdaderamente hay una elipsis nominal será posible recuperar el elemento nominal elidido. En (25) se observa la salida de la gramática de (23), la cual comprueba que es posible recuperar el elemento elidido bajo la sintaxis que generó la estructura antecedente.

(25)



En (25a) se indica el *input* que es la gramática de (23). En (25b) se definen los rasgos de género y número de la variable rectora DET. Si el determinante antecedente es femenino o masculino singular o plural cabe la posibilidad de que la elección del N elidido sea singular o plural (25c). En (25c) se observa cómo actúa la Identidad (Saab, 2008, 2019), pues se coteja el rasgo de género, pero no la variación del número. Finalmente, (25d) es la última variable que completa la reposición de la elipsis. La gramática completa se observa en (26).

(26)



6. Resultados: validación de la gramática computacional

La formalización gramatical de (26) se aplicó en un corpus de 423.000 palabras. Esta gramática no fue entrenada ni modificada previamente. El corpus de validación pertenece al proyecto FONDECYT 1171033 compuesto por artículos de revistas científicas publicadas en internet sobre el dominio médico. En un barrido manual se identificaron 355 elipsis, de las cuales se logró detectar un total de 335 elipsis. La cuantificación de los resultados se muestra en el Cuadro 1.

Total de elipsis	Elipsis detectadas	Elipsis no detectadas	Detecciones erróneas	Precisión	Cobertura	Medida f
355	335	20	3	99,11 %	94,36 %	96,68 %

Cuadro 1. Medidas de rendimiento obtenidas de la aplicación de la descripción sintáctica en el modelamiento computacional

A partir de los resultados del Cuadro 1 se puede concluir que tanto las gramáticas formales como la TIP dan cuenta del fenómeno de la elipsis nominal en español y son útiles para estudiar y describir su sintaxis. Como es evidente, la gramática computacional de (26) detecta construcciones coordinadas en las que hay elipsis en el sintagma determinante. Dentro de las elipsis no detectadas se incluyen otras formaciones como en las verbales, subordinadas y comparativas que no fueron incluidas en la representación y que también pueden ser elípticas como, por ejemplo, las siguientes:

- (27) a. Disminuyendo los valores de colesterol de baja densidad y elevando los valores de colesterol de alta densidad.
- b. El porcentaje fue mayor en el grupo de derecho comparado con el grupo de medicina.

Las elipsis no detectadas obedecieron a alguno de los siguientes aspectos: (i) la ausencia del determinante en la estructura antecedente, como en (28a); (ii) elipsis de núcleos más su complemento; en (28b) se muestra en *itálicas* el núcleo nominal y el adjetivo que lo acompaña; (iii) estructuras complejas que no fueron incluidas en construcciones comparativas, pues su extensión superaba la concatenación de dos sintagmas adicionales; en (28c) se resalta la subordinación que interviene entre la estructura antecedente y el sitio de la elipsis; (iv) elipsis denominadas en cascada entendidas como aquellos casos de múltiples elipsis, en el que un elemento elidido depende a su vez de otro elemento elidido. En (28d) con *itálicas*, se reproduce la repetición del elemento elidido; (v) palabras desconocidas que no aparecen en el diccionario electrónico, como nombres de enfermedades, términos de especialidad médica y extranjerismos; y (vi) fallos en la puntuación por parte de los autores, como excesivo uso de comas.

- (28) a. Se comprometen DET valores como los de responsabilidad y lealtad médica.
 b. La *mortalidad perinatal* en Chile y la de Estados Unidos.
 c. La contaminación por células maternas *debido a que el índice mitótico de las células deciduales* es mucho más bajo que la de las células en división activa.
 d. Se describen dos técnicas: la *técnica* de una sola aguja y la *técnica* de dos agujas.

Ahora bien, las detecciones erróneas obedecieron a la ambigüedad sintáctica entre elementos que son tanto nombres como adjetivos. En el caso de (29), si bien el primer N *embarazadas* es el antecedente y portadora su adjetivo, ambos son adjetivos y nombre, como se ejemplifica en (29).

- (29) a. La mujer embarazada fue atendida por la UCI y la otra fue operada de emergencia.
 b. El portador del pasaporte y el de la cédula de identidad se perdieron.

En el caso de (29a) la formalización implementada logra identificar una sola estructura antecedente pero dos posibilidades en cuanto a la reposición del elemento elidido, como también se muestra en (30).

(30)

Todas las embarazadas portadoras y las con factores de riesgo<ELIP>Todas las embarazadas portadoras y las portadoras con factores de riesgo
 Todas las embarazadas portadoras y las con factores de riesgo<ELIP>Todas las embarazadas portadoras y las embarazadas con factores de riesgo

Ambas posibilidades en (30) si bien son gramaticales dado el cotejo, la verificación de los remanentes y el sitio de la elipsis coincide en que ambas entradas son femeninas y concuerdan en plural, solo la segunda es un verdadero caso de elipsis nominal, pues hay Identidad. No obstante, considerar las portadoras implica que aparte de las *embarazadas portadoras* hay otro grupo, de mujeres, pero no *embarazadas*.

7. Discusión de los resultados

Estudios como este validan la descripción de una teoría, en este caso la TIP, sobre el funcionamiento del sistema gramatical elíptico. Además, es posible ejemplificarla con casos extraídos de corpus escritos en lenguaje natural. Dicho de otro modo, es posible dotarle un carácter experimental al contexto en el que se produce y emplea la gramática sin que intervenga directamente el especialista. En cierta medida –y sin afán controversial–, se logra describir cómo actúa la competencia, pues se sabe que tanto autor como editores realizan una revisión del artículo médico antes de su publicación –aunque claro, aún pueden quedar algunas estructuras consideradas como agramaticales–.

En este apartado se muestran algunas construcciones sintácticas detectadas por la gramática computacional similares a otras que fueron estudiadas por su particularidad estructural y que requieren un estudio adicional y pormenorizado. Además, se sugieren algunas respuestas o abordajes teóricos a partir de otras propuestas teóricas sobre fenómenos similares. En primer lugar, la evaluación de la presencia-ausencia del determinante en la estructura antecedente, en donde algunos contextos bajo predicación no requieren la presencia del determinante (31). En ambos, tal dicotomía no es motivo de agramaticalidad.

- (31) a. Con teorías como la de Bandura en Kaufman y la de Schön.
 b. Con las teorías como la teoría de Bandura en Kaufman y la de Schön

Laca (1999) describe este mismo fenómeno y afirma que “la aparición de sustantivos o grupos nominales sin determinantes está más o menos severamente restringida” (p. 893). La autora menciona que la ausencia del determinante es posible según el tipo de predicado, el sujeto gramatical, presencia de dativos, posición del complemento directo, con complementos de régimen verbal con *de*, sintagmas preposicionales incluidos en el sintagma nominal, en el predicado de las oraciones copulativas, en los complementos predicativos y en las estructuras informativas. No obstante, la elipsis no se registra como una posibilidad para explicar y describir este fenómeno. Esta observación apunta no a la naturaleza de los remanentes en la elipsis nominal que tienen sus propias restricciones y deben ser modificadores o complementos del nombre; sino como afirma Saab (2008) la observación refiere al licenciamiento o tal vez tipos de licenciamiento, entre elipsis y antecedente. Así, este problema se relaciona con lo que Ackema y Szendroi (2002), Arregi y Centeno (2005) y Centeno (2012) denominan *determiner sharing* (determinante partido) que aparece en construcciones coordinadas elípticas en inglés, en las que se puede omitir el determinante siempre y cuando haya identidad entre sí. En (33) se reproducen dos ejemplos de Ackema y Szendroi (2002. p. 190).

- (32) a. Too many Irish setters are named Kelly, ~~too many~~ German shepherds ~~are named~~ Fritz, and too many huskies ~~are named~~ Nanook.
 b. The duck is dry and ~~the~~ mussels are tough.

Una segunda observación es la elipsis de unidades complejas, como se muestra en los siguientes ejemplos.

- (33) a. Las *malformaciones congénitas* de origen multifactorial y las ~~malformaciones congénitas~~ de origen monogénico.
 b. Las *tasas de infección* por Chlamydia se elevan y la ~~tasa de infección~~ de vaginosis bacteriana disminuyen.

Las unidades complejas a las que se refieren los casos de (34) podrían definirse como la combinación de un nombre (+ preposición) + adjetivo o por nombre + nombre, a partir de los cuales surgen algunos inconvenientes que a nivel sintáctico son un desafío para el modelamiento sintáctico y también para NooJ, ya que es necesario delimitar el espacio del sitio elíptico a partir de la selección pertinente de los elementos antecesores. A esto se añade que NooJ posee un fuerte componente léxico, por lo que habrá que modelar una gramática que lo resuelva desde la sintaxis. En (34) y (35) se ejemplifica esta distinción.

(34) La *fiesta de cumpleaños* de Pedro y la ~~fiesta de cumpleaños~~ de Juan fueron el mismo día.

(35) La *fiesta de cumpleaños* de Pedro y la ~~fiesta de~~ matrimonio de Juan fueron el mismo día.

Casos como los anteriores podrían tener alguna respuesta desde lo que Kornfeld (2012) trata como una clase natural que se generaría en un nivel particular denominado Sintaxis temprana producto de la fusión directa (Contreras y Masullo, 2000). El punto clave que trata Kornfeld (2012) es la numeración, es decir, la selección de los ítems que se van a combinar en la sintaxis; puesto que el cotejo de los rasgos de género, número y caso se realiza sobre una sola unidad. Así, en los ejemplos de (36) se identifica y repone un nombre compuesto por: (i) un nombre + adjetivo y (ii) un nombre+preposición+nombre. Este mismo fenómeno se ha explicado para el inglés. Ntelitheos (2004) afirma que la elipsis nominal puede incluir el adjetivo, una oración relativa y complementos preposicionales siempre que haya un antecedente realizado fonéticamente. En (36) se reproducen los ejemplos citados por el autor.

- (36) a. Ten students attended the play but six students left disappointed.
 b. I like Bill's yellow shirt, but not Max's yellow shirt.

The students that Peter invited attended the play but most/some/all ~~students that Peter invited~~ went home disappointed.

Ntelitheos (2004) concluye que los casos de (36) involucran un movimiento sintáctico del sintagma nominal interno a una posición de tópico. Este movimiento es precedido por una eliminación fonológica debido a la elipsis nominal. Otros estudios como los de Pereltsvaig (2013), Ruda (2016), Gruet-Skrabalova (2016), Zdravkovska-Adamova (2017) y Raposo (2018) estudian también casos de elipsis de UC en lenguas como el macedonio, el portugués, el polaco, el húngaro y lenguas eslavas, respectivamente.

Finalmente, la tercera observación obedece a las construcciones comparativas. Las oraciones que no fueron detectadas mostraron una variación tanto de las características sintácticas de la estructura antecedente como en la elíptica. Según la gramática descriptiva (GDLE, 1999; RAE, 2009) las formaciones comparativas son aquellas que cuentan con

elementos comparativos como *mayor, menor, peor, mejor que*. Gutiérrez y Bosque (2009) adiciona *más/menos/tanto* introducidos por *que/de* o *como*. Formaciones que incluyan esos elementos se diferencian del ejemplo citado por Kornfeld y Saab (2005) y Saab (2008) en su tipología de anáforas de dominio nominal, pues en las construcciones que señala la RAE no hay remanente (un determinante) que resguarde los rasgos de número (37).

(37) El estudiante de física es más inteligente que el ~~estudiante~~ de matemáticas.

Los estudios de Lechner (2004, 2019) debaten esto, pues en las construcciones comparativas elípticas no habría uno sino dos procesos de elipsis diferentes. El primero actuaría a nivel de la derivación denominado *Comparative Deletion* (eliminación comparativa) y planteado por Bresnan (1973), pero profundizada por Lechner (2004, 2019). El segundo se ejecuta a nivel superficial conocido como *Comparative Ellipsis* (elipsis comparativa). Según Lechner (2004, 2019) ambos intervienen en las oraciones comparativas, solo que el primero es exclusivo de operaciones de eliminación como *gapping, VP-ellipsis, Pseudogapping, stripping, across-the-board movement (ATB)* y *Right-Node-Raising (RNR)*.

8. Conclusiones

Este trabajo tomó la descripción y el alcance teórico de la Teoría de la Identidad parcial no solo para describir las particularidades que definen la elipsis nominal del español, sino que también se evidenció que sus formalizaciones son óptimas para la representación computacional de una gramática que identifique, coteje y reponga el elemento nominal elidido basadas en los rasgos gramaticales de la lengua.

Una implementación computacional que considere el accionar de la gramática logra aumentar el rendimiento en las tareas de búsquedas y reposición de elementos elididos. A pesar de los prometedores resultados de Hardt (1993, 1997), Lappin y Leass (1994), Hardt y Rambow (2001), Nielsen (2004) y McShane y Babkin (2015) aún no es posible identificar adecuadamente la elipsis nominal. El empleo de un diccionario electrónico anotado con información morfosintáctica proporciona un óptimo etiquetaje automatizado y la aplicación de una misma configuración gramatical para el reconocimiento y reposición del elemento elidido en otros dominios y corpus.

Los estudios teóricos expuestos en este trabajo evidencian que la elipsis, al ser un mecanismo propio de la lengua, es transversal a distintas operaciones sintácticas. Esto ha provocado que no se diferencien formaciones en donde interviene la elipsis (sujeto nulo, categorías vacías, sujeto pronominal, sujeto cero) y las construcciones legítimas de elipsis (elipsis nominal, elipsis comparativa).

Algunos de los inconvenientes de trabajar con corpus de textos escritos en lenguaje natural es que las diferentes combinaciones sintácticas poseen elementos extralingüísticos, dentro de los cuales se pueden mencionar los siguientes:

1. Existencia de muchas palabras desconocidas, pues a pesar de que se incorporaron algunas de ellas en el diccionario electrónico, no fue posible (ni viable) registrarlas todas porque al ser un corpus de dominio médico, siempre habrá términos nuevos y especializados, extranjerismos, nombres de sustancias y moléculas, nombres de instituciones médicas, nombres de enfermedades, instrumentos médicos y sus respectivos procedimientos.
2. La gran cantidad de fórmulas, abreviaturas, ecuaciones y datos estadísticos que aparecían entre paréntesis que nos hicieron descartar algunas construcciones elípticas.
3. Casos de estructuras complejas y de excesiva subordinación, los cuales no fueron contempladas al momento de elaborar las reglas, debido a que su recursión excedía más de dos sintagmas. Esto habría complejizado las gramáticas computacionales tanto para el análisis automatizado como en su aplicación en otros textos.
4. Gran cantidad de errores ortográficos y de puntuación de los escritores, especialmente en cuanto a la concordancia y al excesivo uso de las comas que impedían el reconocimiento automático.

Para finalizar, la gramática computacional desarrollada en este estudio puede aplicarse a otros textos y géneros discursivos, puesto que se contribuyó con base en la gramática. Así se lograría y ampliar el registro de este fenómeno. Además, puede sugerirse analizar la elipsis nominal en segundas lenguas y en comparación con el español. Esto podría revelar resultados novedosos con respecto a la adquisición y dominio de la lengua meta. Otra de las investigaciones pendientes es en corpus paralelos para mejorar las traducciones automáticas; y, a nivel de la interpretación y traducción de lenguas, el estudio sobre la elipsis nominal puede mejorar el rendimiento en la ejecución y la simultaneidad entre los dominios especializados y terminológicos de estas disciplinas.

Notas

- ¹ Herramientas como esta y para otras lenguas se encuentran disponibles en <https://atishs.univ-fcomte.fr/nooj/>
- ² Este diccionario se puede consultar en <https://www.walterkoza.com/traduccion-automatizada-del-conocimiento/>
- ³ En algunos contextos es posible este tipo de construcciones, conocidas como foco contrastivo y que forman parte de las categorías pragmáticas (Gutiérrez, 2008). Sin embargo, en esta investigación de corte gramatical, se omiten aquellos aspectos que no sean categorías sintácticas.

Referencias

- Ackema, P., y Szendroi, K. (2002). Determiner sharing as an instance of dependent ellipsis. *Journal of Comparative Germanic Linguistics* 5, 3–34.
- Aelbrecht, L. (2010). *The syntactic licensing of ellipsis*. John Benjamins Publishing Company.
- Arregi, K., y Centeno, N. (2005). Determiner sharing and cyclicity in Wh-movement. En R. Gess y E. J. Rubin (Eds.), *Theoretical and experimental approaches to Romance linguistics* (pp. 1–19). John Benjamins.
- Bosque, I. y Demonte, V. (Eds.). (1999). *Gramática descriptiva de la lengua española*. Espasa.
- Bosque, I., y Gutiérrez, J. (2016). *Fundamentos de Sintaxis Formal*. Akal.
- Bresnan, J. (1973). Syntax of the Comparative Clause Construction in English. *Linguistic Inquiry* 4 (3), 275-343.
- Brucart, J. (1987). *La elisión sintáctica en español*. Bellaterra.
- Brucart, J. (1999). La elipsis. En I. Bosque y V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 2787-2863). Espasa.
- Brucart, J. M., y MacDonald, J. E. (2012). Empty Categories and Ellipsis. En *The Handbook of Hispanic Linguistics* (pp. 579-601). Wiley Blackwell.
- Brucart, J. M., y Fernández-Sánchez, J. (2018). *Ellipsis (in Romance languages)* [Archivo PDF]. http://docs.wixstatic.com/ugd/d72a87_e042fdaabcb64f09a137aba5763a16cc.pdf
- Centeno, N. (2012). Determiner sharing in Spanish: DET y DETFoc. En R. Campos-Astorkiza y J. Franco (eds.), *Papers in linguistics by the BIDE generation, Anuario del Seminario de Filología Vasca “Julio de Urquijo” XLVI-1 (2012)*, 109-123 [Archivo PDF]. <https://addi.ehu.es/bitstream/handle/10810/49505/12597-46273-1PB.pdf?sequence=1>
- Chomsky, N. (1957). *Syntactic Structure*. Mouton.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Chomsky, N. (1995). *The Minimalist Program*. MIT Press.
- Chung, S., Ladusaw, W., y McCloskey, J. (1995). Sluicing and Logical Form. *Natural Language Semantics*, 3, 239-282.
- Chung, S. (2013). Syntactic identity in sluicing: How much and why. *Linguistics Inquiry* 44(1), 1-44.
- Cinque G. (1993). On the Evidence for Partial N-movement in the Romance DP. University of Venice *Working Papers in Linguistics* 3 (2), 21-40.
- Contreras, H. y Masullo, P. (2000). Motivating Merge. En M. Leonetti., O. Fernández y V. Escandell Vidal (Eds.) *Current Issues in Generative Grammar*. Universidad de Alcalá, Madrid.
- Correas, G. (1967). *Arte Kastellana*. Universidad de Santiago de Compostela.
- Embick, D. (2000). Features, Syntax, and Categories in the Latin Perfect. *Linguistic Inquiry* 31(2), 185-230.
- Embick, D. (2007). Linearization and Local Dislocation: Derivational Mechanics and Interactions. *Linguistic Analysis* 33(3-4), 2-35.
- Embick, D., y Noyer, R. (2001). Movement Operations after Syntax. *Linguistic Inquiry* 32(4), 555-595.
- Embick, D., y Halle, M. (2004). *Word Formation: Aspects of the Latin Conjugation in Distributed Morphology*. Mouton de Gruyter.
- Embick, D y Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry* 39(1), 1-53.
- Fiengo, R. (1977). On trace theory. *Linguistic Inquiry* 8, 35-61.
- Fiengo, R., y May, R. (1994). *Indices and Identity*. MIT Press.
- Fox, D. (2000). *Economy and Semantic Interpretation*. MIT Press.

- Gallego, A. (2011). *Sobre la elipsis*. Madrid: Arco.
- Grimshaw, J. (1990). *Argument Structure*. MIT Press.
- Gruet-Skrabalova, H. (2016). Ellipsis in sequences of multiple wh-words. *Linguistica Brunensia*, Masarykova univerzita, Brno 64(1), 57-69.
- Gutiérrez, J., y Bosque, I. (2009). *Fundamentos de sintaxis formal*. Akal.
- Halle, M. (1997). Distributed Morphology: Impoverishment and Fission. *MIT Working Papers in Linguistics* 30, 425-449.
- Halle, M., y Marantz, A. (1993). Distributed Morphology and the Pieces of Inflection. En K. Hale y S. Keyser (Eds.). *The view from Building 20* (pp. 111-176). MIT Press.
- Hankamer, J., y Sag, I. (1976). Deep and Surface Anaphora. *Linguistic Inquiry* 3 (7): 391-426.
- Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning and processing* [Tesis doctoral, University of Pennsylvania].
- Hardt, D. (1997). An empirical approach to VP ellipsis. *Computational Linguistics*, 23(4), 525-541.
- Hardt, D., y Rambow, O. (2001). Generation of VP ellipsis: A corpus-based approach. En *Proceedings of ACL*.
- Harley, H., y Noyer, R. (1999). Distributed Morphology. *Glott International* 4(4), 3-9.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford: CDLI Publications.
- Kester, P. M., y Sleeman, P. (2002). N-Ellipsis in Spanish. En H. Broekhuis, y P. Fikkert (Eds.), *Linguistics in the Netherlands 2002* (pp. 107-116). Benjamins.
- Kornfeld, L. (2012). *Núcleos complejos y división del trabajo lingüístico* [Tesis de Maestría, Universidad Nacional de Comahue].
- Kornfeld, L., y Saab, A. (18-20 de agosto, 2005). Hacia una tipología de las anáforas nominales en español. III Encuentro de Gramática Generativa, Neuquén.
- Laca, B. (1999). Presencia y ausencia del determinante. En *Gramática descriptiva de la lengua española*. Ignacio Bosque y Violeta Demonte (Eds.). 3 tomos. Espasa-Calpe.
- Lappin, S. (2005). A sequenced model of anaphora and ellipsis resolution. En A. Branco., T. McEnery y R. Mitkov (Eds.), *Anaphora processing: Linguistic, cognitive, and computational modelling*, (pp 3-16). Ámsterdam: John Benjamins.
- Lappin, S., y Leass, H. (1994). A Syntactically Based Algorithm for Pronominal Anaphora Resolution. *Computational Linguistics*, 20, 535-561.
- Lappin, S. y Shih, H. (1996). A generalized reconstruction algorithm for ellipsis resolution, in *Proceedings of COLING-96*, University of Copenhagen, pp. 687-692.
- Lasnik, H. (1999). A Note on Pseudogapping. En *Minimalist Analysis* (pp. 151-174). Malden MA: Blackwell.
- Lechner, W. (2019). Comparative deletion. En J. van Craenenbroeck y T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis* (pp. 19-45). Oxford.
- Lechner, W. (2004). *Ellipsis in Comparatives*. Mouton de Gruyter.
- Lobeck, A. (1995). *Ellipsis: Functional Heads, Licensing and Identification*. Oxford University Press.
- Longobardi, G. (1994). Reference and Proper Names: A Theory of N-movement in Syntax and Logical Form. *Linguistic Inquiry* 25(4), 609-665.
- Marantz, A. (1997). No Escape from Syntax: Don't try Morphological Analysis in the Privacy of Your Own Lexicon. En Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark, y Alexander Williams (eds.), *Proceedings of the 21st Penn Linguistics Colloquium*. Philadelphia: UPenn Working Papers in Linguistics, 201-225.

- Marantz, A. (1995). *Cat'as a Phrasal Idiom: Consequences of Late Insertion in Distributed Morphology*. MIT.
- McShane, M., y Babkin, P. (2016). Detection and Resolution of Verb Phrase Ellipsis. *Linguistic Issues in Language Technology*, 13(1), 1–34.
- Mel'čuk, I. (2006). *Aspects of the Theory of Morphology*. Mouton de Gruyter.
- Merchant, J. (1999). *The Syntax of Silence: Sluicing, Islands and Identity in Ellipsis* [Tesis de Doctorado, Universidad de Santa Cruz].
- Merchant, J. (2001). *The syntax of silence: sluicing, islands and theory of ellipsis*. Oxford University Press.
- Merchant, J. (2013). Ellipsis phenomena. Jeroen van Craenenbroeck and Jason Merchant. In Marcel den Dikken (ed.), *The Cambridge Handbook of Generative Syntax*, 701-745. Cambridge University Press: Cambridge.
- Mitkov, R. (2001). Outstanding issues in anaphora resolution. En Al. Gelbukh (Ed.) *Computational Linguistics and Intelligent Text Processing* (pp. 110–125). Springer.
- Mitkov, R. (2002). *Anaphora Resolution*. Longman.
- Nielsen, L. (2004). Verb phrase ellipsis detection using automatically parsed text. En *COLING '04: Proceedings of the 20th international conference on Computational Linguistics* (pp. 1093–1099). Ginebra.
- Ntelitheos, D. (2004). *Syntax of elliptical and discontinuous nominals* [Tesis de Maestría, UCLA].
- Panagiotidis, P. (2002). *Pronouns, clitics and empty nouns*. John Benjamins.
- Panagiotidis, P. (2003). Empty Nouns. *Natural Language and Linguistic Theory* 21:381-432.
- Peral J., Ferrández, A. (2000). Generation of Spanish Zero-Pronouns into English. En D.N Christodoulakis (Ed.), *Natural Language Processing* (pp. 252–260). Springer.
- Pereltsvaig, A. (2013). Noun phrase structure in article-less Slavic languages: DP or not DP? *Language and Linguistics Compass* 7(3), 201–219.
- Raposo, E. (2018). Nominal ellipsis and prepositional modifiers in Portuguese: A phase-theoretic approach. *Probus* 30(2), 277-303.
- Real Academia Española (2009). *Nueva Gramática de la Lengua Española*. Espasa.
- Rello, L. (2010). *Elliphant: A machine learning method for identifying subject ellipsis and impersonal constructions in Spanish*. [Tesis de Maestría, University of Wolverhampton y Universitat Autònoma de Barcelona].
- Rello, L., y Ilisei, I. (2009). A Rule-Based Approach to the Identification of Spanish Zero Pronouns. En *Proceedings of Student Research Workshop, RANLP* (pp. 60–65).
- Ruda, M. (2016). NP Ellipsis (Effects) in Polish and Hungarian: FFs on Fs, Agree, and Chain Reduction. *The Linguistic Review* 33, 649–677.
- Saab, A. (7-9 de agosto, 2003). Identidad Morfológica Estricta e Inserción Tardía. II Encuentro de Gramática Generativa, Instituto de Lenguas Vivas Juan Ramón Fernández.
- Saab, A. (2008). *Hacia una teoría de la identidad parcial en la elipsis* [Tesis de Doctorado, Universidad Nacional de Buenos Aires].
- Saab, A. (2012). On the notion of partial pro-drop in Romance languages and beyond. En M. Kato y F. Ordóñez (Eds.), *The Morphosyntax of Portuguese and Spanish in Latin America* (pp. 49-77). Oxford University Press.
- Saab, A. (2019). Nominal Ellipsis. En J. van Craenenbroeck y T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis*. Oxford University Press

- Sag, I. (1976). *Deletion and Logical Form* [Tesis de Doctorado, MIT].
- Silberztein, M. (2005). NooJ: A Linguistic Annotation System for Corpus Processing. Demo. En *Proceedings of the HLT/EMNLP2005 Conference*.
- Silberztein, M. (2016). *Formalizing Natural Languages: The Nooj Approach*. Wiley-Iste.
- van Craenenbroeck, J. (2010a). *The syntax of ellipsis: Evidence from Dutch dialects*. Oxford University Press.
- van Craenenbroeck, J. (2010b). Invisible last resort: A note on clefts as the underlying source for sluicing. *Lingua*, 120, 1714–1726.
- Williams, E. (1977). Discourse and Logical Form. *Linguistic Inquiry*, 8(1), 101-139.
- Zdravkovska-Adamova, B. (2017). Ellipsis in the Macedonian noun phrase. *Seeu Review* 12 (2),82-107.