

**Detección automática de rasgos de despersonalización
en artículos de investigación científica en español.
Un análisis comparativo acerca de la frecuencia
en diferentes áreas disciplinares**

**Automatic Detection of Depersonalization Traits
in Spanish-Language Scientific Research Articles.
A Comparative Analysis of their
Frequency across Disciplines**

Carolina Paola Tramallino
Consejo Nacional de Investigaciones científicas y Técnicas
Universidad Nacional de Rosario, Argentina
tramallino@irice-conicet.gov.ar
ORCID: <https://orcid.org/0000-0003-4760-7005>

Romina Paola Arnal
Universidad Nacional de Rosario, Argentina
arnal_romina_paola@santafevirtual.edu.ar
ORCID: <https://orcid.org/0009-0005-4814-5991>

Celina Beltrán
Facultad de Ciencias Agrarias
Universidad Nacional de Rosario, Argentina
cbeltran2510@gmail.com
ORCID: <https://orcid.org/0000-0001-8529-894X>

Resumen

En este trabajo nos proponemos emplear un software de acceso abierto para detectar automáticamente y contabilizar determinadas estructuras sintácticas propias de artículos de investigación científica escritos en español pertenecientes a tres áreas disciplinares disímiles: ingeniería y ciencias exactas; medicina,

biología y ciencias de la salud; ciencias sociales y humanidades. Particularmente, nos referimos a oraciones en voz pasiva y construcciones compuestas por un nombre y un participio verbal. Cabe aclarar que este estudio forma parte de un proyecto de investigación sobre herramientas digitales colaborativas en la construcción de textos académicos en lengua española. Los objetivos específicos son: realizar una descripción y una medición de cada una de estas construcciones halladas en los tres subcorpus de artículos científicos escritos por hablantes nativos de español. Asimismo, se contabilizan las formas verbales flexionadas en primera persona plural para efectuar un análisis comparativo. Por último, se propone contrastar la frecuencia de ambas estructuras en las tres áreas de conocimiento para extraer conclusiones que puedan aportar al campo de la didáctica de la escritura académica. La metodología abarca un enfoque mixto; en primer lugar, se realiza un análisis descriptivo cualitativo; en segundo lugar, se aplican técnicas estadísticas para medir los resultados. Entre los hallazgos, se obtiene que el grupo de artículos perteneciente a ingeniería presenta una mayor cantidad de estructuras que indican despersonalización. En lo que respecta a la presencia de verbos en primera persona plural, se hallan diferencias significativas en los diferentes subcorpus.

PALABRAS CLAVE: detección automática, artículos de investigación científica, escritura académica, disciplinas científicas.

Abstract

This study, part of a research project on collaborative digital tools for the elaboration of academic texts in Spanish, uses an open access software program to automatically detect and count certain syntactic structures typical of scientific research articles written in Spanish in three dissimilar areas: engineering and exact sciences; biology, health sciences and medicine; and humanities and social sciences. The syntactic structures of interest are sentences in the passive voice and constructions composed of a noun and a verbal participle. The specific objective is to carry out a description and measurement of each of these constructions found in three subcorpora of scientific articles written by native Spanish speakers. The inflected verb forms in the first person plural are also counted for purposes of comparative analysis. Finally, it compares the frequency of each structure in the three areas of knowledge to draw conclusions that may contribute to the teaching of academic writing. The methodology includes a mixed approach: a qualitative descriptive analysis followed by statistical analysis. Among the findings are that engineering articles use a larger number of impersonal structures in certain types of syntactic constructions. Significant differences are found among the three disciplines in the frequency of verbs used in the first-person plural.

KEYWORDS: automatic detection, scientific research articles, academic writing, scientific disciplines

FECHA DE RECEPCIÓN: 29-01-2025

FECHA DE ACEPTACIÓN: 08-05-2025



1. Introducción

Está claro que el artículo de investigación científica (AIC) no sólo se propone patrocinar y comunicar la existencia de una investigación, sino también, validar la legitimación de su autor como investigador ante la comunidad científica de un determinado campo disciplinario (Meza, 2016). De las muchas aristas que posee la escritura de este género textual, numerosas investigaciones han puesto el foco en el proceso alfabetizador (Swales, 1990; Carlino, 2005; Codina y Lopezosa, 2022); en la dificultad que entraña confeccionarlo, debido al desconocimiento que exhiben estudiantes de posgrado acerca de la superestructura textual o de la intencionalidad comunicativa propia de sus secciones (Swales, 2004; Hyland, 2000, 2005a, 2005b; Doury y Plantin, 2016).

En lo que refiere a su arquitectura lingüística, se lo ha abordado desde el estudio de las movidas retóricas que lo componen (Hyland, 2000; Swales, 2004). Para Hyland (2000), las organizaciones retóricas son unidades discursivas que cumplen funciones comunicativas coherentes dentro de un discurso escrito, por ejemplo: establecer la importancia del tema, exponer la metodología, presentar los resultados, discutir la relevancia de la investigación. Por su parte, Swales (2004) define a las movidas retóricas como las estrategias empleadas por los autores para construir textos que satisfagan las expectativas de la comunidad académica.

Por otro lado, Sabaj (2012) examina los propósitos comunicativos que se textualizan en diferentes patrones léxico-gramaticales de artículos científicos y realiza un recorrido por estudios que han abordado tres cuestiones: en primer lugar, las movidas retóricas, en segundo lugar, los rasgos lingüísticos y, por último, investigaciones que integran ambas. En cuanto a los rasgos lingüísticos, se han examinado las marcas de atenuación, el uso de nominalizaciones, las cláusulas reportativas y las marcas de autoría, entre otros.

De igual manera, se ha puesto el foco en el mecanismo de despersonalización del discurso científico (Álvarez López, 2012, 2013; Lamprea Abril et al., 2023), el cual se construye mediante numerosos recursos lingüísticos que inciden o impactan en mayor o menor medida para lograr este efecto.

El presente estudio tiene como objetivo específico detectar automáticamente, mediante un software de acceso libre, determinadas construcciones sintácticas pertenecientes a la formulación escrita del género textual artículo de investigación científica (AIC) en idioma español, que son tomadas como rasgos de la despersonalización. De la totalidad de mecanismos, abordamos las construcciones formadas por nombre seguido de participio; el uso de la voz pasiva en dos de sus formas: la llamada perifrástica y la que se construye con *se*; las oraciones impersonales cuya estructura sintáctica coincide con esta última clase de pasiva y los verbos conjugados en primera persona plural.

Es importante mencionar que el estudio se vincula con un proyecto de investigación encuadrado en el campo de la lingüística computacional que se propone examinar herramientas digitales que colaboran con el proceso de escritura académica para la confección de artículos científicos en español (Tramallino y San Martín, 2023). Por lo tanto, el interés radica en indagar acerca de las características textuales del género mencionado con la finalidad de establecer si los rasgos lingüísticos considerados apare-

cen y se distribuyen de la misma manera en artículos científicos pertenecientes a tres áreas disciplinares diferentes. Los campos disciplinares son: ingeniería y ciencias exactas; medicina, biología y ciencias de la salud y, por último, humanidades y ciencias sociales.

2. Antecedentes

El artículo de investigación científica se constituye como uno de los géneros académicos más investigados en las últimas dos décadas (Sabaj, 2009, 2012; Rodríguez Hernández y García Valero, 2015; Meza, 2016, 2017) debido a su impacto en la transmisión del conocimiento especializado, en beneficio del avance de la ciencia. Está claro que las publicaciones científicas constituyen un indicador de la actividad de investigación, por lo cual este acto de escribir y publicar reviste un carácter estratégico-persuasivo (Meza, 2016). Por lo tanto, es importante considerar los aspectos socio-discursivos del género y reparar en el hecho de que los autores recurren a diversas estrategias retórico-discursivas para persuadir a sus lectores.

Dentro de estas estrategias o recursos podemos señalar el mecanismo de despersonalización, definido como “fenómeno discursivo que consiste en la acción de evitar, mediante un conjunto de mecanismos lingüísticos, marcas que señalen la presencia explícita y directa de las personas o participantes del discurso” (Álvarez López, 2012, p. 27). La autora identifica diferentes grados de despersonalización de acuerdo al recurso lingüístico empleado: el uso de la primera persona plural indica un grado menor de despersonalización; el de las oraciones pasivas perifrásticas con complemento agente portan un grado medio, mientras que las pasivas con *se* y las oraciones impersonales detentan un grado alto de despersonalización. Asimismo, el uso de la primera persona singular funciona como marca explícita del autor.

Con respecto al empleo de infinitivos, gerundios y participios, estos son potencialmente elementos de despersonalización siempre y cuando no estén asociados a verbos auxiliares o principales y puedan suplir la referencia a un sujeto; en otras palabras, si restringen la asociación a una persona gramatical concreta. En efecto, el participio se distingue de las otras formas no personales por su carácter pasivo y por presentar un comportamiento paralelo al de las construcciones pasivas ya que habilita la dependencia de un complemento agente (Álvarez López, 2013).

Hyland (2012) investiga en la escritura académica los rasgos lingüísticos mediante los cuales se afirma el posicionamiento del autor y se trasluce la identidad. En este sentido, le otorga a la comunidad académica en la que se inscribe el autor la facultad de constituir una identidad disciplinaria. Por lo tanto, afirma que las disciplinas se corresponden con comunidades culturales y no responden, solamente, a afiliaciones institucionales. Trabaja las formas discursivas que son indicios de una identidad individual y grupal. Asimismo, Lamprea Abril et al. (2023) investigan acerca de los posicionamientos enunciativos frente a la actividad de investigar en corpus de artículos científicos sobre el lenguaje y para ello consideran tanto los mecanismos lingüísticos empleados por los autores para generar “efectos de responsabilidad explícita frente a lo que se enuncia” como los que privilegian marcas de distanciamiento con el fin de objetivar los enunciados planteados en sus textos, a los que llaman efectos de despersonalización. Para ello, examinan artículos de investigación pertenecientes al área de las ciencias sociales y humanidades y se detienen en las formas verbales empleadas con mayor frecuencia; consideran el modo verbal, la tercera persona y las oraciones en voz pasiva. Concluyen que la modalización entendida como la actitud del enunciador con respecto al destinatario y al contenido de su enunciado (tomada de Charaudeau y Maingueneau, 2002) “opera en la construcción de un discurso de quien se considera

experto a través de formas despersonalizadas que configurarían una retórica de demostración y de evidencia” (Lamprea Abril et al., 2023, p. 481).

2.1 *Voz pasiva y oraciones impersonales en español*

La literatura especializada en el análisis de estructuras en voz pasiva plantea que existe la tendencia universal en las lenguas de construir en pasiva cuando no hay agente o cuando se pretende efectuar una afirmación acerca del objeto directo de la oración activa (González Calvo, 1992).

En español, la estructura pasiva que se formula con el auxiliar *ser* conjugado y el participio de la forma verbal habilita la presencia del agente como complemento agente, por ejemplo, “Los datos termopluviométricos fueron suministrados por la Estación Meteorológica de INTA Cerro Azul” (Texto 1 corpus Ingeniería), sin embargo, el objeto semántico funciona obligatoriamente como el sujeto sintáctico.

La otra estructura pasiva que se denomina refleja conserva como sujeto sintáctico el objeto semántico pero no permite la presencia del agente, es decir, no admite el complemento encabezado por la preposición *por*, salvo que se interprete como un adjunto de causa. Además, se ve impedida en los casos en que el OD es animado (De Miguel, 2004), por ejemplo: “se produce una capa carbonizada” (Texto 1 corpus Humanidades)

Carranza (2023) define a la pasivización en español como la variante pasiva de las cláusulas activas con objeto directo, en la cual el argumento funciona como sujeto. Examina, además, la posibilidad que admiten ciertos verbos ditransitivos del español para construir oraciones pasivas con argumentos meta o beneficiarios.

Maldonado (1999) reserva el término pasiva para la llamada pasiva perifrástica, ya que considera que esta construcción es la que contrasta con la activa en español. En esta, el verbo no está

empleado en forma activa y la estructura depende de una formación de participio pasado con la cual se invierten los términos de prominencia entre participantes, por lo que el paciente se constituye como figura principal. Al mismo tiempo, el participio cancela la visión procesal del evento para dar prominencia a la función terminal. Entonces lo primero que se observa es el cambio de estado que sufre el tema y luego se identifica la fuente energética que lo indujo.

En esta clase de construcciones, el inductor es siempre un agente humano y ocupa un lugar secundario. Este inductor se codifica en forma oblicua (en una frase agentiva con *por*) o puede quedar presupuesto marcado morfológicamente en la concordancia verbal. La escena es perfectiva, pero no puntual; pone el perfil en el proceso de cambio de estado con prominencia terminal, pero no excluye la inducción y el desarrollo del evento.

En cuanto a las pasivas perifrásticas, Maldonado (1999) expone que la intencionalidad del agente es sin duda más prominente que en las construcciones con *se* y coincide con Fernández Ramírez (1986), quien señala en estas un grado más alto de dinamismo; en otras palabras, los momentos más sensibles (posición y movimiento) y las intenciones (reales o metafóricas) se manifiestan de una manera más prominente y eficiente.

El requisito de perfectividad en estas construcciones ha sido explicado también por De Miguel (2004), quien propone que, para que este tipo de oraciones pueda construirse, el verbo debe ser perfectivo, es decir, pertenecer a la clase aspectual de las realizaciones o bien a la de los logros que fuera propuesta por Vendler (1967). Las realizaciones son eventos dinámicos con duración y con un final, por ejemplo: *escribir una tesis*. Por su parte, los logros poseen un final pero sin duración, por ejemplo: *encontró la partida de nacimiento de su hija*.

Takagaki (2008) y De Miguel (2004) le otorgan un papel determinante al aspecto léxico del verbo en la formación de la

pasiva en español, al tiempo que ratifican la hipótesis a partir del análisis de corpus de textos contemporáneos, según la cual la pasiva perifrástica se emplea con menos frecuencia porque posee más restricciones aspectuales que la pasiva con el pronombre *se*.

Álvarez López (2013) observa un paralelismo entre las pasivas perifrásticas y las construcciones de N + participio, que pueden denominarse construcciones absolutas internas siguiendo a la RAE (1973, 2009) o de participios adjuntos libres, como las llaman Hernanz y Suñer Gratacós (1999, p. 2543), cuando el participio funciona como complemento de un N o SN externo a la construcción de participio, y *construcciones de participio absoluto* cuando el SN es interno. En estas estructuras, el participio presenta un aspecto perfectivo-resultativo y el significado pasivo que presentan se debe a que el N o SN con el que concuerda el participio funciona de manera análoga a la del sujeto gramatical que semánticamente se interpreta como tema o paciente. Además, permiten un complemento agente encabezado por la preposición *por*.

Ambas estructuras de N + participio funcionan para la autora como mecanismos de despersonalización que evitan las marcas del sujeto o las trasladan, si presentan un complemento agente explícito, de la misma manera en que lo hacen las pasivas perifrásticas.

Respecto de las construcciones con *se* y verbo conjugado, Maldonado (1999) realiza un estudio exhaustivo. Propone que la forma *se* tiene la doble función de focalizar la posición final del evento y de dejar la fuerza inductiva en forma esquemática. La presencia de *se* implica que la causa iniciativa pierde peso en favor de la focalización del cambio que sufre un elemento temático.

A raíz de esta afirmación, el autor postula una organización gradual de eventos, cuyas formas polares serían: la de ocurren-

cia espontánea, en la que no hay energía inductora (*El cinturón se rompió*) y la construcción impersonal en la que el inductor es necesariamente un humano no especificado, en este caso la energía es humana y volitiva, aunque esquemática (*Se prohíbe fumar*) (Maldonado, 1999, p. 272). Los grados de manifestación de la fuerza inductiva varían de acuerdo con factores semántico-sintácticos relacionados con el requerimiento de la agentividad humana del verbo, el aspecto (léxico y morfológico) y el grado de definición del elemento temático.

Para el autor, si bien es cierto que las “pasivas”¹ pueden aparecer en cualquier tiempo, hay una serie de restricciones aspectuales de importancia. Los que demarcan la terminación puntual de un evento son los morfemas perfectivos. Esta perfectividad combina armoniosamente con la prominencia terminal. En cambio, la forma imperfectiva proyecta el cambio como repetitivo o habitual. Los verbos imperfectivos oscurecen la función focal del *se* y, en consecuencia, gana prominencia la fuerza inductiva del cambio, por lo que se obtiene una lectura genérica.

Además, el factor fundamental a tener en cuenta es el de la composición semántica del verbo, ya que si este requiere de un agente humano como fuerza inductora emergerá una construcción impersonal. Por lo tanto, la motivación inicial para distinguir una “pasiva” con *se* de una construcción impersonal es que, en esta última, hay un agente humano obligatoriamente implícito, mientras que en la primera el agente está ausente y el evento es visto como una ocurrencia espontánea. Si bien la similitud que presentan es que no cuentan con un agente identificado que induce la acción, ya que como se indicó anteriormente, el inductor es esquemático, en las impersonales la acción del agente es

¹ Maldonado (1999) utiliza las comillas para designar como “pasivas” a las pasivas sintéticas y a las construcciones con *se* a las cuales denomina “medias de prominencia terminal” (MPT).

observable, mientras que en la “pasiva” hay un bajo nivel de prominencia de las fuerzas inductoras y del desarrollo de la acción, en contraste con la prominencia terminal en la que se focaliza el cambio de estado o el resultado de una acción.

Con respecto al tratamiento de estos dos tipos de estructuras sintácticas que hemos considerado (pasiva perifrástica, y *se* + verbo conjugado, ya sea “pasiva” o impersonal), como rasgos de despersonalización, Álvarez López (2013) define a la pasiva con *se* como un mecanismo que omite las marcas de la persona del discurso, mientras que considera que el mecanismo de la pasiva perifrástica evita las marcas y las traslada a una posición poco relevante en la oración. En coincidencia con Maldonado (1999), plantea que la pasiva perifrástica tiene un carácter intencional debido a que muestra una mayor presencia del agente y puede hacerlo explícito mediante un sintagma preposicional con *por*, para ello se basa en el estudio emprendido por Mendikoetxea (1999) (Álvarez López, 2013, p. 370).

En cuanto a la distinción entre las pasivas con *se* y las impersonales, coincide con Gutiérrez Ordóñez (2002) en afirmar que ambas evidencian un proceso de impersonalización efectuado por el elemento *se*, que denota la suspensión del sujeto, cuya función es la más preeminente de la oración (Álvarez López, 2013, p. 375).

Por lo tanto, consideramos que las oraciones de voz pasiva perifrásticas junto con las estructuras de nombre más participio portan un grado menor de despersonalización que las estructuras formadas por *se* más verbo conjugado.

2.2 *La primera persona plural en artículos de investigación científica*

En lo que refiere a manifestaciones de la subjetividad en artículos de investigación científica, García Negroni (2008) analiza textos

de cuatro disciplinas diferentes con el objeto de demostrar que el locutor del discurso científico-académico deja huellas de su presencia aun en las ciencias llamadas duras o exactas. Pone en discusión la característica tradicional de la sintaxis académica caracterizada por “la ausencia o baja frecuencia de marcas de 1ª persona coincidente con el número del/los autor/es” (García Negroni, 2008, p. 12) Para ello, contabiliza las formas despersonalizadas en las que incluye a las nominalizaciones y metonimias y las oraciones pasivas “con *ser* y con *se*” en las que pone el acento por considerar que permiten dejar indeterminado a quien se considera responsable de la acción. De forma contrastiva, mide las formas personalizadas, estas incluyen el empleo de la primera persona singular y la primera plural.

En términos discursivos, Meza (2016, 2017) propone el Posicionamiento Estratégico del Autor que tiene su fundamento en el carácter argumentativo que detenta y del que se sirve el sujeto que escribe para atraer la atención de los lectores y para convencerlos de la relevancia de los datos y/o resultados que expone. La finalidad de situarse estratégicamente responde a la necesidad de que dichos resultados o conclusiones sean considerados en el campo de conocimiento en el que se inscriben. Sabemos, incluso, que cada sección del AIC contiene recursos retóricos (Swales, 2004) para persuadir a la audiencia y lograr la aceptación de esos otros investigadores/científicos que tomarán, posteriormente, en sus propios trabajos o trayectos investigativos el conocimiento obtenido de esa lectura.

Es importante mencionar los estudios de Hyland (2012), quien indaga sobre las formas de representación del autor en la escritura científica en ocho disciplinas diferentes y determina que su figura está presente en todas las áreas. Asimismo, Benveniste (1971), quien analiza el aparato formal de la enunciación y a los pronombres personales como elementos portadores de la subjetividad del enunciador en el lenguaje, afirma que estos no se

refieren a entidades léxicas, sino a entidades puramente lingüísticas. En cuanto a la primera persona plural, se trata de un nosotros de autoría que no implica una multiplicación del yo, sino una amplificación difusa de la primera persona, es decir, “expresa una comprensión más indefinida o una generalización vaga del yo” (Benveniste, 1971, p. 68). En este sentido, González de Requena Farré (2020) investiga los diferentes usos de la primera persona y determina que su empleo en el discurso académico se traduce en una forma de atenuación del enunciador.

3. Metodología

La metodología se llevó a cabo a través de diferentes etapas; la primera de ellas consistió en conformar un corpus de artículos de investigación científica, para ello se empleó la técnica de selección aleatoria simple (Hurtado y Toro, 2005). Este procedimiento permite elegir elementos de forma tal que cada uno tenga la misma probabilidad de ser seleccionado, sin sesgos ni preferencias. Por lo tanto, los textos que forman parte del análisis se escogieron al azar de un conjunto mayor, en este caso, la totalidad de artículos disponibles en los números publicados por las revistas escogidas durante los años 2021 y 2022. Así, se garantiza un corpus representativo y no influenciado por criterios del sujeto que investiga; además, posibilita que los resultados del análisis puedan generalizarse o reflejar características típicas del corpus completo para asegurar la fiabilidad del estudio.

La estrategia metodológica consiste en adherir a la lingüística de corpus (LC) como método válido para la investigación de las lenguas y del lenguaje que permite llevar a cabo investigaciones empíricas en contextos auténticos. Por lo tanto, se puede indagar el uso lingüístico, tal como es producido, comunicado y comprendido entre hablantes/escribientes y oyentes/lectores reales (Parodi, 2008).

Para este estudio, se conforman tres muestras de artículos de investigación que responden a diferentes áreas científicas y se eligen publicaciones periódicas de centros de estudios pertenecientes a universidades de distintos países hispanoparlantes. Asimismo, dichas revistas se encuentran indizadas en los portales Scielo y Scopus. La tabla que sigue exhibe los nombres de las revistas y los países de edición de las publicaciones periódicas.

ÁREA	INGENIERÍA	MEDICINA, BIOLOGÍA Y CS. DE LA SALUD	HUMANIDADES Y CS. SOCIALES
Revista 1	<i>Revista Facultad de Ingeniería</i> (Colombia)	<i>Revista Ciencias de la Salud</i> (Colombia)	<i>Calidad en la educación</i> (Chile)
Revista 2	<i>Revista de Ciencias y Tecnología</i> (Argentina)	<i>Acta Bioquímica Clínica Latinoamericana</i> (Argentina)	<i>Estudios de Filosofía Práctica e Historia de las Ideas</i> (Argentina)
Revista 3	<i>Revista Tecnia</i> (Perú)	<i>Revista peruana de Biología</i> (Perú)	<i>Anuario de Letras. Lingüística y Filología</i> (México)

Tabla 1. Revistas seleccionadas para la conformación del corpus

A continuación, se confeccionó un corpus digital de artículos científicos (AC) escritos en español pertenecientes a tres campos disciplinares diferentes. Por cada área disciplinar se seleccionaron treinta artículos científicos (AC) de tres publicaciones periódicas diferentes, en total se reunió un corpus de noventa AC extraídos de nueve revistas de alto impacto publicados entre los años 2021 y 2022. En cuanto a los países editores se encuentran Argentina, Chile, Colombia, Perú y México.

En la siguiente etapa, se analizó a través de la herramienta informática NooJ cada uno de los textos con el fin de extraer la cantidad de estructuras sintácticas en voz pasiva, de estructuras con *se* + verbo conjugado y de las construcciones de nombres

seguidos por participios. Además, se contabilizaron los verbos conjugados en primera persona plural. Cabe aclarar que se extrajeron resultados por cada texto en particular, ya que cada artículo fue guardado en el software y se le asignó un nombre de acuerdo al orden cardinal y al área científica correspondiente. De esta forma, se conservaron en el sistema NooJ treinta artículos de ingeniería y ciencias exactas, clasificados del T1 I al T30 I; treinta artículos del área de medicina, biología y ciencias de la salud (T1 B al T30 B); y, por último, treinta textos del área humanidades y ciencias sociales (T1 H al T30 H).

Para finalizar, se recurrió al análisis cuantitativo mediante la aplicación de técnicas estadísticas para determinar si existían diferencias significativas en el uso de dichas estructuras sintácticas y formas verbales entre los diferentes subcorpus según la disciplina científica.

4. Análisis de los datos

Como ya se mencionó, para la detección automática se utilizó el sistema NooJ (Silberztein, 2016), este software de acceso libre está disponible para el análisis morfológico, sintáctico y semántico de más de veinte lenguas naturales. Asimismo, ofrece para el análisis textual un procesamiento básico de tokenización y segmentación y listas de concordancia léxica basada en corpus y en frecuencias de palabras. Para llevar a cabo el reconocimiento automático se empleó el módulo *sp (spanish)* Argentina (Tramallino, 2013; Rodrigo y Bonino, 2019).

El programa cuenta con variados recursos; entre estos, se hallan los diccionarios que son listas de palabras ingresadas en el sistema como entradas léxicas y acompañadas por diversos tipos de información lingüística, como por ejemplo, la categoría gramatical asignada mediante una etiqueta morfosintáctica. Además, las propiedades correspondientes al sistema lingüístico de

la lengua con la que opera el programa están consignadas en un tipo de archivo denominado propiedades definidas. Este reúne una información categorial completa para cada ítem lexical, por ejemplo: V (verbo) y sus rasgos morfológicos: tiempo, modo, conjugación, persona, número, entre otros.

El programa posee la posibilidad de crear tres tipos de gramáticas informáticas; por un lado, las gramáticas morfológicas y derivaciones que incluyen los modelos de flexión y de derivación correspondientes de acuerdo a la categoría gramatical de la palabra. Por otro lado, las gramáticas productivas que se basan en sistemas de grafos para poder detectar y analizar cadenas con caracteres. La tercera de las gramáticas es la que corresponde a gramáticas sintácticas, las cuales son útiles para formalizar relaciones entre palabras que constituyen sintagmas. Estas permiten detectar y analizar cadenas formadas por dos o más unidades léxicas. Pueden formalizarse de dos maneras: por un lado, a partir de determinados ítems léxicos y por el otro, mediante etiquetas asignadas por los diccionarios correspondientes. Con esta última gramática se elaboran gramáticas más complejas que son capaces de reconocer secuencias de categorías.

Para este estudio, se recurre a una gramática sintáctica de grafos que permite localizar las oraciones en voz pasiva formadas por el auxiliar *ser* más el participio de la forma verbal. A su vez, esta funciona a partir de la selección adecuada de diccionarios, gramáticas y propiedades definidas (Tramallino, 2013; Rodrigo y Bonino, 2019).

La captura que sigue muestra la gramática de grafos realizada para la detección automática de oraciones en voz pasiva construidas mediante el auxiliar *ser* conjugado más el participio correspondiente que concuerda en género y número con el núcleo del sujeto, es decir, las llamadas pasivas perifrásticas.

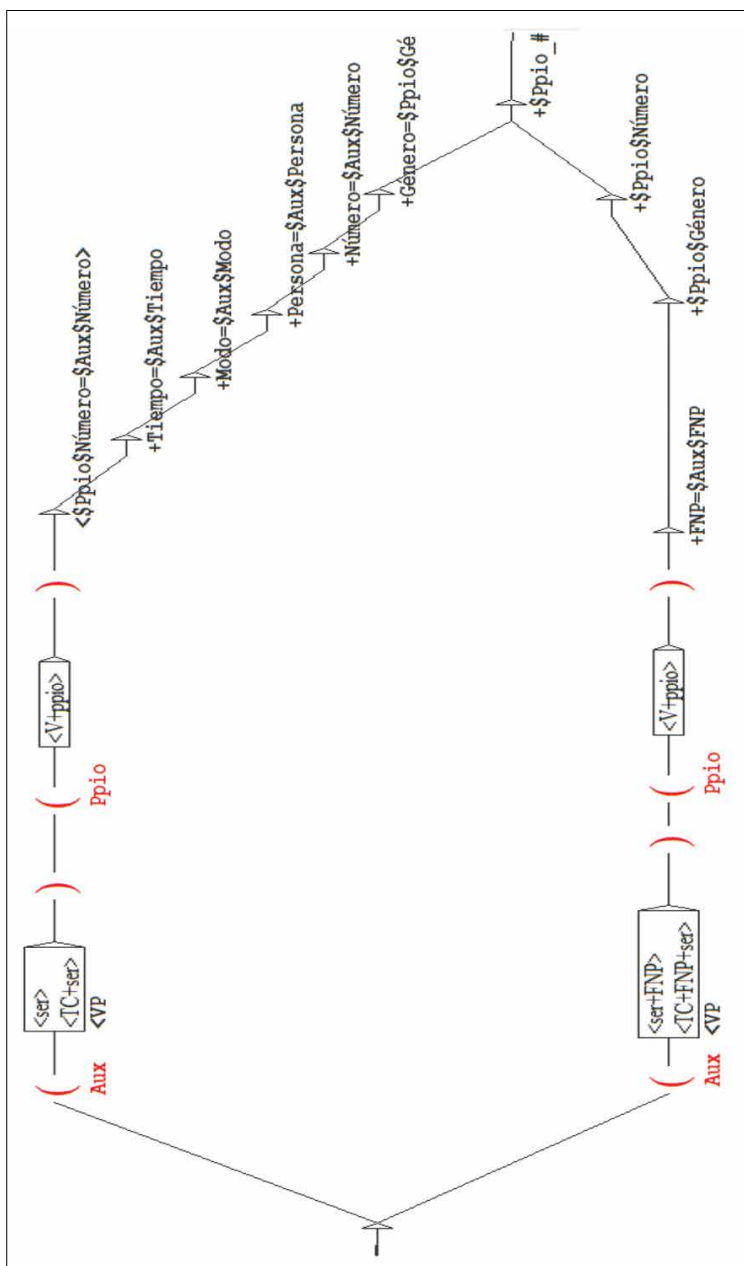


Imagen 1. Gramática de grafos para voz pasiva

Con esta gramática se detectan construcciones como las siguientes:

- (1) En este trabajo, el criterio que *es adoptado* para determinar el número total de escenarios que *serán simulados*, ... (T24 I)
- (2) Durante las simulaciones *es asumido* un factor de potencia (T24 I)
- (3) ... lo cual *será analizado* en esta revisión (T3 B)
- (4) ...el triptófano disponible *es internalizado* en la célula (T3 B)

Por otra parte, el software presenta el recurso Locate, a partir del cual se ejecutan diferentes tipos de búsquedas en los textos; desde palabras, categorías de palabras, secuencias de palabras, sufijos o terminaciones, hasta construcciones sintácticas, sinónimos y términos pertenecientes al mismo campo semántico. El programa muestra no solo la lista de ocurrencias (Seq) y la cantidad total (Query), sino también la porción del texto en el que se encuentra el ítem o los ítems léxicos junto con las cinco palabras anteriores (Before) y las cinco posteriores (After), como se observa en la siguiente imagen.

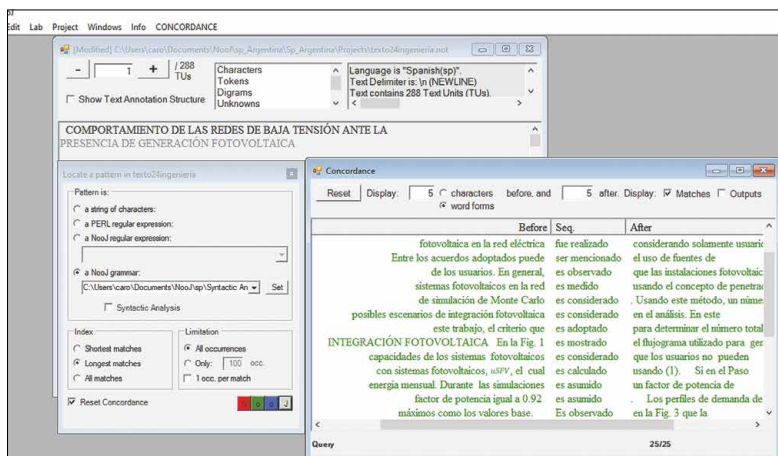


Imagen 2. Localización gramática de grafos para estructuras en voz pasiva en el texto núm. 24 del corpus de ingeniería

De esta forma, se detectaron las estructuras sintácticas que se componen de un nombre seguido de participio, a partir de las categorías léxicas establecidas en las propiedades definidas del programa y sus rasgos, con la correspondiente fórmula $\langle N \rangle \langle V + \text{ppio} \rangle$. Las categorías se registran en mayúscula y se localizan mediante paréntesis angulares, en el caso planteado $\langle N \rangle$ y $\langle V \rangle$. Además, puede especificarse la búsqueda añadiendo los rasgos pertinentes, que se colocan en minúscula precedidos del signo +. Por lo tanto, a la categoría V (verbo) se le agregó el rasgo +ppio (participio).

Así, se pudieron identificar construcciones como las que siguen:

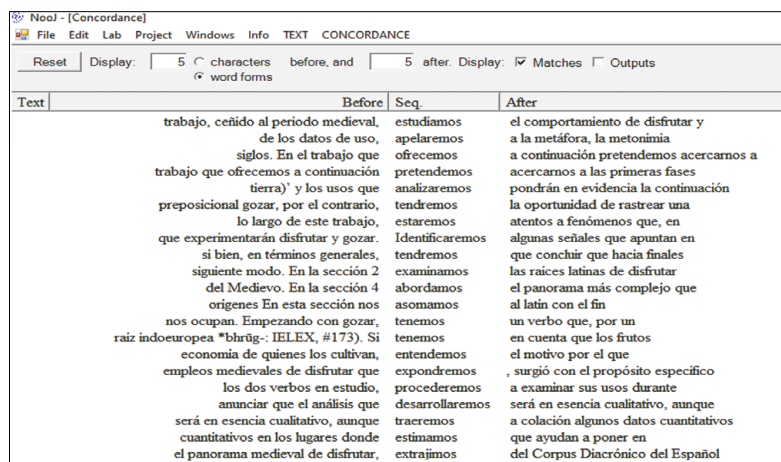
- (5) ... que es el *desecho generado* en la operación... (T6 I)
- (6) El *algoritmo propuesto* es... (T26 I)
- (7) Como puede observarse en el *ejemplo dispuesto* en la tabla (T17 H)
- (8) Es importante destacar un *estudio realizado* en Buenos Aires (T3 B)

También, con esta función es posible reconocer las estructuras sintácticas que se componen de la partícula *se* más el verbo conjugado, mediante una búsqueda combinada de palabra y categoría léxica que se consigna de la siguiente forma: $\text{se} \langle V + 3a \rangle$ para poder localizar estructuras como las siguientes:

- (9) En los que no *se detectaron* variables químicas (T3 B)
- (10) En 2016, *se define* un Plan de Renovación (T10 H)
- (11) ...un paciente en el que *se aisló* esta bacteria (T3 B)

Sin embargo, el sistema no efectúa una distinción semántica de estas estructuras para poder diferenciar las pasivas con *se* de las oraciones impersonales. No obstante, como ambas construcciones son consideradas rasgos con un alto grado de despersonalización, cuestión teórica que ya hemos explicado en el apartado anterior, esta limitación no afecta el objetivo propuesto.

En tanto, este mismo procedimiento se utiliza para localizar los verbos en primera persona plural. Entre signos angulares se indica que busque con la etiqueta V (verbo) y se especifican los rasgos 1a (primera persona) y pl (plural). Como resultados obtenemos las siguientes concordancias que corresponden al texto núm. 3 del corpus correspondiente a humanidades:



The screenshot shows the NooJ Concordance window with the following settings: Display: 5 characters before, and 5 after. Display: ☒ Matches, ☐ Outputs. The search results are displayed in a table with three columns: Text, Before, Seq., and After.

Text	Before	Seq.	After
trabajo, cenido al periodo medieval, de los datos de uso, siglos. En el trabajo que trabajo que ofrecemos a continuación tierra') y los usos que preposicional gozar, por el contrario, lo largo de este trabajo, que experimentarán disfrutar y gozar. si bien, en términos generales, siguiente modo. En la sección 2 del Medioevo. En la sección 4 orígenes En esta sección nos nos ocupan. Empezando con gozar, raíz indoeuropea *bhr̥tǵ-: IELEX, #173). Si economía de quienes los cultivan, empleos medievales de disfrutar que los dos verbos en estudio, anunciar que el análisis que será en esencia cualitativo, aunque cuantitativos en los lugares donde el panorama medieval de disfrutar,	estudiamos apelaremos ofrecemos pretendemos analizaremos tendremos estaremos Identificaremos tendremos examinamos abordamos asomamos tenemos tenemos entendemos expondremos procederemos desarrollaremos traeremos estimamos extraíjamos		el comportamiento de disfrutar y a la metáfora, la metonimia a continuación pretendemos acercarnos a acercarnos a las primeras fases pondrán en evidencia la continuación la oportunidad de rastrear una atentos a fenómenos que, en algunas señales que apuntan en que concluir que hacia finales las raíces latinas de disfrutar el panorama más complejo que al latín con el fin un verbo que, por un en cuenta que los frutos el motivo por el que , surgió con el propósito específico a examinar sus usos durante será en esencia cualitativo, aunque a colación algunos datos cuantitativos que ayudan a poner en del Corpus Diacrónico del Español

Imagen 3. Resultados de la localización de verbos conjugados en primera persona plural en el texto núm. 3 del corpus de humanidades y ciencias sociales

Una vez contabilizadas las ocurrencias halladas en cada uno de los textos que componen los tres subcorpus, se confeccionaron tablas con la cantidad de las localizaciones para poder efectuar el posterior análisis estadístico y obtener resultados de acuerdo al campo disciplinar.

5. Resultados

La herramienta informática utilizada arrojó los resultados con una gran cobertura, sin embargo, en cuanto a la localización de estructuras compuestas por *se* más el verbo conjugado, se observó

una limitación del programa en lo que se refiere al análisis cualitativo de los diferentes tipos de estructuras (pasivas, impersonales, medias, etc.) que coinciden en la construcción sintáctica.

Para el análisis cuantitativo de los datos se emplearon dos técnicas estadísticas: la prueba no paramétrica de Kruskal-Wallis y el test de Wilcoxon para muestras independientes aplicado entre cada par de grupos con corrección de significancia cuando en el primero se hallaron diferencias entre grupos.

Estos procedimientos son adecuados cuando no se satisfacen las condiciones para poder aplicar un modelo clásico basado en la distribución normal de la variable en estudio, como sucede en este caso y se sustentan en el uso de los rangos asignados a las observaciones (Beltrán y Barbona, 2016).

Por lo tanto, en una primera instancia se realizó una comparación de las disciplinas respecto al total de palabras por texto. Dado que el número de palabras por texto difería en cada grupo definido por la disciplina ($p < 0.0001$, resultando el número de palabras por texto significativamente mayor en los pertenecientes a humanidades), se realiza la comparación en términos relativos al número de palabras, es decir, sobre la cantidad de ocurrencias (V+participio y *se+v* conjugado) por cada 100 palabras. Esto es:

$$\begin{array}{lcl} \text{VS} & & \text{SE} \\ +_i P \times 100P = \frac{\text{cantidad de VS + PARTICPIO}}{\text{total de palabras}} \cdot 100_i & & +_i VC \times 100P = \frac{\text{cantidad de SE + V CONJUGADO}}{\text{total de palabras}} \cdot 100_i \end{array}$$

En promedio, los textos de ingeniería contienen 3075 palabras; mientras que aquellos de ciencias de la salud y biológicas alcanzan 4215 palabras y los de humanidades presentan un promedio mayor de 7430 palabras; por lo tanto, estas diferencias resultaron estadísticamente significativas. Al examinar las medianas, se observa que el 50% de los textos de ingeniería y ciencias

exactas posee menos de 3034 palabras, cifra que asciende a 3266 palabras en medicina, biología y ciencias de la salud y a 7738 en humanidades y ciencias sociales. Además, el coeficiente de variación indica que el corpus de medicina, biología y ciencias de la salud presenta la mayor variabilidad en la extensión de sus textos.

La tabla siguiente muestra las medidas descriptivas del total de palabras según la disciplina:

ÁREA DEL CONOCIMIENTO	PROMEDIO	COEFICIENTE DE VARIACIÓN	MEDIANA
Ingeniería	3075	20.03	3034
Medicina, biología y ciencias de la salud	4215	67.4	3266
Humanidades	7430	25.3	7738

Tabla 2. Medidas descriptivas del total de palabras por texto según disciplina

5.1 *Pasivas perifrásticas*

Las tablas que se exhiben a continuación muestran las medidas descriptivas, luego de efectuado el análisis cuantitativo, de acuerdo a cada una de las estructuras sintácticas halladas en los tres grupos de artículos científicos de acuerdo al área del conocimiento al que pertenecen. En primer lugar, contabilizamos las ocurrencias de voz pasiva perifrástica, formadas por *ser* más el participio, por ejemplo “...es internalizado” (T24 I) y obtuvimos las medidas dispuestas en la tabla 3.

ÁREA DEL CONOCIMIENTO	PROMEDIO	COEFICIENTE DE VARIACIÓN	MEDIANA
Ingeniería	0.1813	165.5	0.0829
Medicina, biología y ciencias de la salud	0.0795	100.6	0.0701
Humanidades	0.0844	106.6	0.0541

Tabla 3. Medidas descriptivas de la cantidad de ocurrencias verbo + participio por cada 100 palabras por texto según disciplina

Respecto a la cantidad de ocurrencias verbo *ser* + participio, el promedio mayor lo presenta el corpus de ingeniería con 0.1813 mientras que los restantes promedios están próximos a 0.08. Asimismo, el 50% de los textos de ingeniería alcanza el valor para este índice de 0.0829 a diferencia del subcorpus de medicina, biología y ciencias de la salud, y el de humanidades y ciencias sociales cuyos valores son de 0.0701 y 0.0541, respectivamente. Por lo tanto, se evidencia una mayor diferencia en promedio y no así en mediana. La variabilidad es alta en los tres corpus, mostrando coeficientes de variación superiores al 100%.

En síntesis, se obtuvo que de la comparación de la cantidad resultante de estructuras denominadas pasivas perifrásticas por cada 100 palabras, no se hallaron diferencias significativas ($p = 0.1928$) indicando que no hay evidencia para pensar que la frecuencia de estructuras de este tipo en voz pasiva difiere en las disciplinas consideradas.

5.2 Estructuras con *se* y verbo conjugado

Con respecto a la comparación de la estructura sintáctica con *se* + verbo conjugado sí se encuentran diferencias significativas entre las áreas ($p = 0.0004$). El promedio de ocurrencias de este tipo de estructuras por cada 100 palabras para el corpus de ingeniería fue de 2 mientras que para medicina, biología y ciencias de la salud, y para humanidades resultó menor a 1.6 en ambos subgrupos. Mientras que el 50% de los textos de ingeniería alcanza una ocurrencia por cada 100 palabras de 1.8, en los corpus de medicina y humanidades esta cifra desciende a 1.7 y 1.4, respectivamente.

Por lo tanto, en la comparación de pares se ha encontrado que los textos pertenecientes a la disciplina ingeniería presentan una frecuencia significativamente mayor de este tipo de estructura que los restantes ($p < 0.05$), mientras que las diferencias

halladas en las disciplinas medicina y humanidades no muestran relevancia estadística ($p = 0.1635$). El coeficiente de variación revela similar variabilidad en los tres corpus.

A continuación, se expone la tabla 4 con las medidas descriptivas de este tipo de estructuras.

DISCIPLINA	PROMEDIO	COEFICIENTE DE VARIACIÓN	MEDIANA
Ingeniería	2.0004	36.5	1.814
Medicina, biología y ciencias de la salud	1.5446	34.3	1.6603
Humanidades	1.3889	38.9	1.3881

Tabla 4. Medidas descriptivas de la cantidad de ocurrencias SE + VC por cada 100 palabras por texto según disciplina

5.3 Estructuras formadas por un nombre y un participio

En la comparación de pares respecto a nombre + participio se ha encontrado que los textos pertenecientes a la disciplina ingeniería presentan una frecuencia significativamente mayor de este tipo de estructura que los restantes ($p < 0.01$), mientras que las diferencias halladas en las disciplinas medicina, biología y cs. de la salud y humanidades no muestran relevancia estadística ($p=0.72$).

La tabla 5 describe las medidas halladas para esta variable.

DISCIPLINA	PROMEDIO	COEFICIENTE DE DESVIACIÓN	MEDIANA
Ingeniería	0.3555	66.6	0.3106
Ciencias biológicas	0.1633	65.6	0.1384
Humanidades	0.1548	41.9	0.1447

Tabla 5. Medidas descriptivas de la cantidad de ocurrencias de las estructuras de nombre seguido de participio por cada 100 palabras por texto según disciplina

En el corpus de ingeniería, el promedio de ocurrencias de las estructuras fue de 0.3555 por cada 100 palabras, mientras que en los otros dos corpus la frecuencia fue menor, con valores de 0.1633 y 0.1548, respectivamente. Al analizar las medianas, se observa que el 50% de los textos de ingeniería registran al menos 0.31 ocurrencias por cada 100 palabras, mientras que en medicina, biología y ciencias de la salud esta cifra disminuye a 0.1384, y a 0.1447 en humanidades y ciencias sociales. En relación con la variabilidad, esta es alta en los tres subcorpus, supera el coeficiente de variación el 40%.

5.4 Verbos en primera persona plural

En la comparación de los verbos conjugados en primera persona plural, resulta significativamente más alto el promedio de ocurrencias en el corpus de artículos referidos al área de las humanidades y ciencias sociales como se muestra en la tabla 6.

DISCIPLINA	PROMEDIO	COEFICIENTE DE VARIACIÓN	MEDIANA
Ingeniería	0.0372	200	0
Ciencias biológicas	0.0499	209.02	0
Humanidades	0.2790	91.0	0.2102

Tabla 6. Distribución del número de ocurrencias de verbos en primera persona plural por cada 100 palabras según disciplina

El 50% de los textos de humanidades presenta 0.21 ocurrencias de este tipo por cada 100 palabras mientras que en los otros dos subcorpus, por lo menos el 50% de los textos no presenta dicha estructura. Por lo tanto, los artículos de investigación con mayor heterogeneidad para la ocurrencia de verbos en primera persona plural son los que pertenecen a ingeniería, y medicina, biología y ciencias de la salud. Entre las otras dos áreas disciplina-

res la que menor cantidad exhibe es la de ingeniería (ver gráficos 7 y 8 en Anexos).

Para los fines de indagar sobre la coincidencia entre el empleo de este *nosotros* y la cantidad real de autores de las investigaciones, extrajimos los datos autoriales de cada artículo, de acuerdo a la clasificación: *a)* autor único y *b)* dos o más autores.

Resultó que en el subcorpus de humanidades y ciencias sociales hay 19 artículos con un solo autor, por lo tanto, supera al 60%; a diferencia del subcorpus del área de medicina, biología y ciencias de la salud, en la que solo un artículo es de autor individual (3%), cinco de ellos poseen dos autores (17%) y 24 (el 80%) están escritos por grupos de más de tres autores hasta ocho. En tanto, en el grupo de ingeniería y ciencias exactas, el 100% es presentado por más de dos autores (el 20% posee dos autores y el 80% entre tres y cinco).

6. Discusiones

En la comparación de la cantidad de estructuras de voz pasiva perifrástica por cada cien palabras en cada uno de los subcorpus, no encontramos diferencias representativas, lo cual indica que no hay evidencia para pensar que la frecuencia de este tipo de estructuras en voz pasiva difiere en las disciplinas consideradas.

Sin embargo, al establecer la comparación con la otra estructura pasiva que porta un grado mayor de despersonalización, construida con *se* más el verbo conjugado, sí hallamos diferencias reveladoras entre las tres áreas. En efecto, descubrimos que los textos pertenecientes a la disciplina ingeniería y ciencias exactas presentan una frecuencia significativamente mayor que el resto. Por otra parte, los corpus que provienen de medicina, biología, ciencias de la salud, y humanidades y ciencias sociales no muestran discrepancias que impliquen relevancia estadística.

Incluso, en la comparación de pares con las construcciones de nombre + participio obtuvimos que los artículos de ingeniería y ciencias exactas detentan una frecuencia significativamente mayor que los restantes, mientras que las disimilitudes halladas en los textos de medicina, biología y ciencias de la salud, y humanidades y ciencias sociales no manifiestan importancia estadística.

Por consiguiente, determinamos que en los artículos científicos provenientes del área de ingeniería y ciencias exactas se configura una retórica de demostración y evidencia, siguiendo a Lamprea Abril et al. (2023), debido a un mayor empleo del recurso de despersonalización del discurso con el que consiguen crear un efecto de veracidad para poder convencer a la comunidad lectora “probablemente” experta.

En lo que atañe a la comparación de pares con respecto al porcentaje de verbos conjugados en primera persona plural, se ha obtenido que los artículos científicos del área de las humanidades y ciencias sociales ostentan una frecuencia significativamente mayor que el resto, mientras que en las diferencias detectadas en los artículos de ingeniería, ciencias exactas, y medicina, biología y ciencias de la salud no se manifiesta relevancia estadística.

Ahora bien, al contrastar la cantidad de formas verbales en primera persona plural halladas en cada grupo con la cantidad real de autor o autores de los artículos, resultó que, llamativamente, humanidades y ciencias sociales, que ostenta la mayoría de artículos de autoría individual, sin embargo, es el área que se distingue de las otras por un uso significativo de verbos en primera persona plural. En consecuencia, podemos establecer que este uso de *nosotros*, que en la mayoría de los artículos se corresponde a un solo enunciador, obedece a una estrategia del autor en cuanto mecanismo que le permite asumir un “plural de modestia”, que González de Requena Farré (2020) conceptualiza como una forma de desfocalización del sujeto que investiga.

Estos resultados son coincidentes, además, con los obtenidos por Álvarez López (2012), quien luego de analizar las marcas de despersonalización en textos científicos de ocho disciplinas diferentes, halla un porcentaje elevado en el uso de la primera persona plural en textos pertenecientes a lingüística, mientras que la disciplina derecho es la que exhibe una menor cantidad.

En relación con esta cuestión, García Negroni (2008), quien considera el empleo de verbos en primera persona plural como marca de personalización, concluye que a pesar de que las marcas de persona son menos frecuentes que las de despersonalización, estas también se emplean para hacer referencia al sujeto de la enunciación con diferente regularidad de acuerdo a la disciplina. Obtiene que la mayor ocurrencia de marcas de personalización se manifiesta en el corpus de lingüística. Por su parte, en el área de medicina, las formas de primera persona del plural muestran “un grado de responsabilidad enunciativa explícita” (García Negroni, 2008, p. 16). Sin embargo, en este mismo grupo halla una alta frecuencia de estrategias de despersonalización, lo cual da cuenta de un efecto mayor de “objetividad y neutralidad” asociado a las disciplinas llamadas “duras” (García Negroni, 2008, p. 28).

Por lo dicho anteriormente, podemos determinar que en los subcorpus de ingeniería y de medicina, el empleo de verbos en primera persona plural no responde a una estrategia del autor sino que, por el contrario, coincide con la autoría colectiva de los artículos; por lo tanto, la utilización de *nosotros* puede considerarse como una marca de personalización, incluso, si tenemos en cuenta que ingeniería es el área que detenta una mayor cantidad de estructuras de despersonalización.

En este sentido, Hyland (2012) afirma que es imposible borrar la presencia del autor, aun en las disciplinas en las que no se explicitan interpretaciones personales, ya que esta se visibiliza en las modalizaciones, en la primera persona o en auto-

referencias. Concluye que en el ámbito de las ciencias sociales los escritores creen pertinente la explicitación de sus posiciones personales, mientras que, en las ciencias exactas, como es el caso del área de ingeniería, prevalecen los recursos de demostración y generalización, pero se evitan las interpretaciones personales. En la misma dirección, Antolí Martínez (2019) declara que la búsqueda de una claridad expositiva y de un estilo de redacción directo y asertivo es lo que lleva a emplear la primera persona gramatical en la escritura científica, incluso en el ámbito de las disciplinas con escasa manifestación del investigador en el texto.

7. Conclusiones

Este artículo cumplió con el objetivo de indagar acerca de los elementos lingüísticos que se configuran como rasgos de despersonalización en artículos de investigación científica en tres áreas disciplinares diferentes. Para ello fueron reconocidas automáticamente y contabilizadas determinadas estructuras sintácticas. Se obtuvo, como resultados, que estas se utilizan, mayormente, en la escritura científica propia del área de ingeniería y ciencias exactas.

Por otra parte, contabilizamos el empleo de la primera persona plural y su correspondencia con la cantidad real de autores de los artículos, así hallamos que los correspondientes a disciplinas como lingüística, filosofía y ciencias de la educación son las que presentan una diferencia significativa no solo en su utilización sino también en la coincidencia con el número de autores, ya que la mayoría de los artículos que componen este subcorpus son de autor único.

En cuanto a las proyecciones, se espera continuar con la elaboración de recursos lingüísticos dentro del módulo español del software empleado, en lo que respecta a diccionarios semánticos de verbos, para que el programa distinga, por ejemplo, los verbos que son agentivos de aquellos que no lo son, con el objeto

de obtener avances en el análisis cualitativo de estructuras sintácticas de despersonalización. Además, se prevé ampliar no solo la muestra de artículos científicos, sino también la diversidad de áreas disciplinares y realizar una contabilización pormenorizada de estas estructuras por cada una de las secciones que integran dicho género académico.

En síntesis, consideramos que este estudio significa un incipiente aporte al ámbito de la alfabetización académica en el nivel superior en lo que atañe a las características enunciativas de artículos de investigación en idioma español, de acuerdo a las formas particulares de comunicar la ciencia que adquiere cada área disciplinar. En efecto, comprobamos que los autores se representan a sí mismos de diferentes maneras según el campo científico al que pertenecen (Hyland, 2012) y, precisamente, son los rasgos lingüísticos los que materializan estas representaciones. Por lo tanto, afirmamos que cada campo disciplinar construye sus formas de decir de acuerdo a determinadas herramientas discursivas propias.

Anexo

Tablas

La siguiente tabla muestra los resultados correspondientes al área de humanidades y ciencias sociales:

NÚM. TEXTO	TOTAL PALABRAS	V SER + PARTICIPIO	SE + VERBO	NOMBRE + PARTICIPIO	V 1ERA P. PL
1	1751	0	12	2	0
2	9815	6	170	15	10
3	6930	1	77	5	14
4	8014	10	117	15	17
5	8720	4	132	26	33
6	8742	10	93	19	1
7	4354	6	53	10	1

8	7866	4	123	8	2
9	5696	0	82	8	2
10	7732	12	108	18	8
11	7554	2	112	11	65
12	10196	1	154	15	22
13	6412	27	73	9	16
14	7745	5	228	13	8
15	7609	6	160	19	42
16	8159	3	88	8	13
17	6685	1	103	10	12
18	7981	1	184	15	0
19	8345	2	189	9	38
20	7276	3	108	6	16
21	8166	3	97	7	17
22	5859	7	66	16	23
23	9731	20	90	14	67
24	5598	2	38	4	22
25	7322	16	101	20	48
26	9063	20	96	11	29
27	8538	10	67	10	29
28	4158	1	12	5	7
29	6418	4	130	4	9
30	10456	6	123	16	102

Tabla 1. Resultados de localizaciones en el corpus del área de humanidades

La tabla que se encuentra a continuación exhibe los resultados de los artículos correspondientes al área de ingeniería y ciencias exactas.

NÚM. TEXTO	CANTIDAD DE PALABRAS	VERBO SER + PARTICIPIO	SE + V	NOMBRE+ PARTICIPIO	V EN 1ERA P. PL
1	3429	2	54	6	0
2	2541	1	77	9	0
3	2908	1	84	6	5
4	2612	0	49	13	0

5	1969	2	63	15	1
6	3026	7	45	14	0
7	2300	4	60	23	2
8	3426	2	88	4	0
9	3478	5	71	8	0
10	3891	10	117	22	0
11	1850	3	50	4	0
12	3224	3	83	12	2
13	2332	6	55	4	0
14	3373	11	56	5	0
15	3326	2	58	16	0
16	3913	1	65	6	0
17	3436	4	81	35	0
18	3627	1	66	14	1
19	3041	0	81	7	2
20	2597	1	39	11	0
21	3482	7	65	9	0
22	4494	2	66	10	0
23	4119	3	62	12	12
24	2974	25	11	11	0
25	2831	43	8	4	0
26	2727	4	72	12	7
27	3414	1	76	6	1
28	2556	8	29	11	1
29	2425	0	41	8	0
30	2930	2	44	1	1

Tabla 2. Resultados de localizaciones en el corpus del área de ingeniería

Por último, la tabla que sigue presenta los resultados del subcorpus correspondiente al área de medicina, ciencias biológicas y de la salud.

TEXTO	TOTAL PALABRAS	V SER + PARTICIPIO	SE + VERBO	NOMBRE + PARTICIPIO	VERBOS EN 1ERA P.PL.
1	4585	4	50	14	0
2	15277	11	207	26	1
3	1466	1	31	1	0
4	5711	0	89	5	0
5	2275	3	29	12	0
6	1706	0	40	4	0
7	4943	1	90	9	0
8	2676	1	48	4	0
9	3267	2	65	4	0
10	4392	8	92	6	0
11	4380	7	73	6	0
12	3711	9	41	3	12
13	2859	0	51	4	1
14	2920	3	20	2	0
15	9146	7	53	10	31
16	3116	1	35	4	3
17	4471	4	55	7	3
18	8654	8	167	6	0
19	5332	5	64	6	6
20	1632	5	14	2	0
21	3114	6	45	6	1
22	3265	3	54	15	0
23	5409	2	56	10	1
24	1081	0	24	1	4
25	2041	0	36	1	1
26	2764	0	56	4	0
27	6752	10	31	15	0
28	3999	1	75	6	2
29	2687	0	65	6	0
30	2821	1	52	2	0

Tabla 3. Resultados de localizaciones en el corpus del área de ciencias biológicas y de la salud

Gráficos

Los gráficos siguientes muestran los diagramas de caja e histograma según la disciplina a la que pertenece el artículo científico de acuerdo a las cuatro localizaciones automáticas efectuadas.

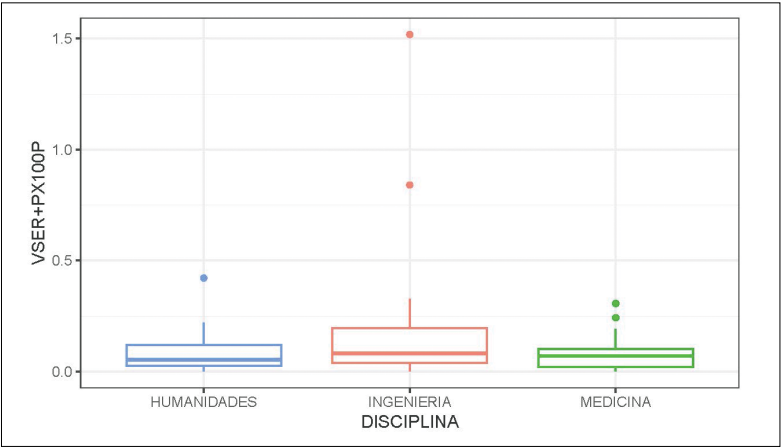


Gráfico 1. Diagramas de caja de la distribución del número de ocurrencias de voz pasiva verbo *ser* + participio por cada 100 palabras según disciplina.

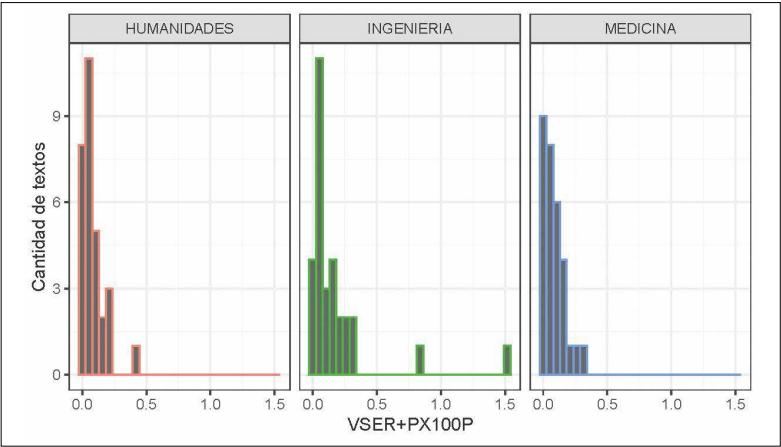


Gráfico 2. Histograma de la distribución del número de ocurrencias de voz pasiva verbo *ser* + participio por cada 100 palabras según disciplina

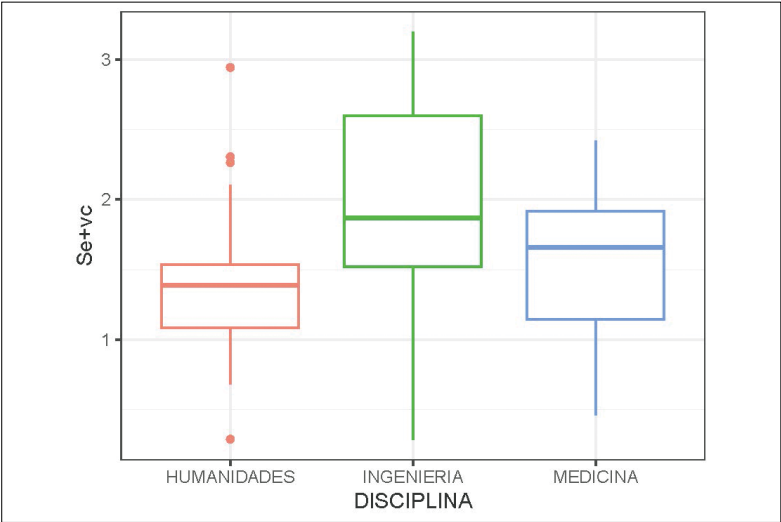


Gráfico 3. Diagramas de caja de la distribución del número de ocurrencias de voz pasiva pronombre *se* + verbo conjugado por cada 100 palabras según disciplina

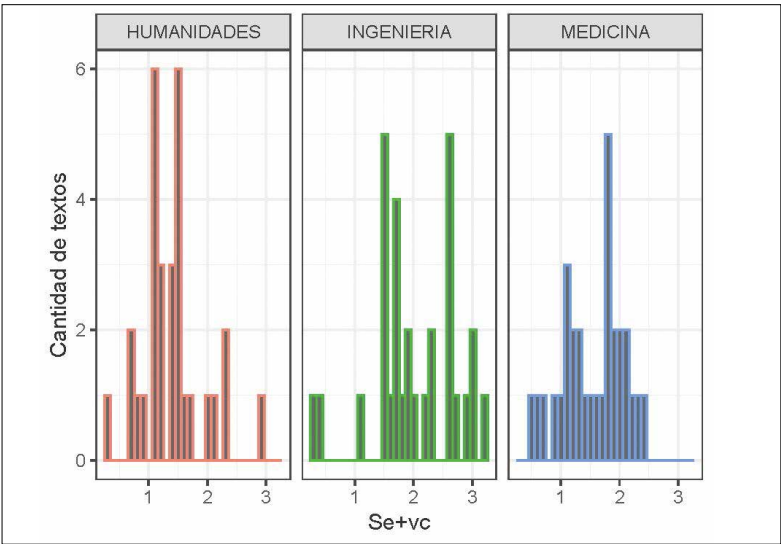


Gráfico 4. Histograma de la distribución del número de ocurrencias de voz pasiva pronombre *se* + verbo conjugado por cada 100 palabras según disciplina

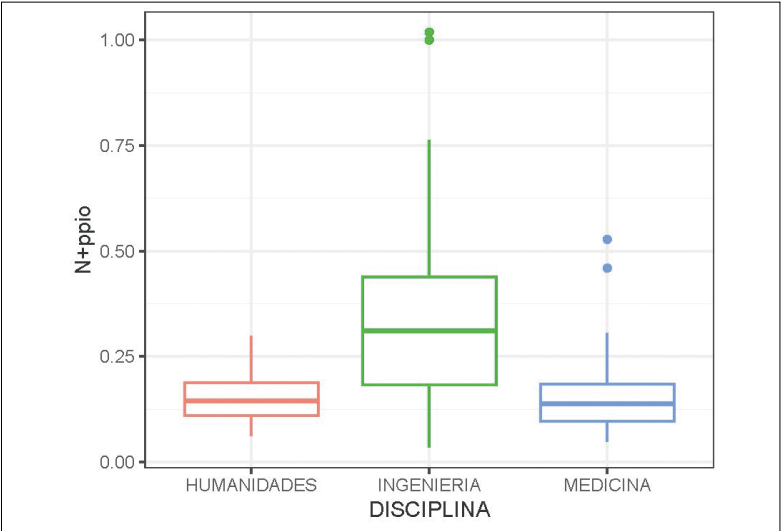


Gráfico 5. Diagramas de caja de la distribución del número de ocurrencias de estructuras de nombre seguido de participio por cada 100 palabras según disciplina

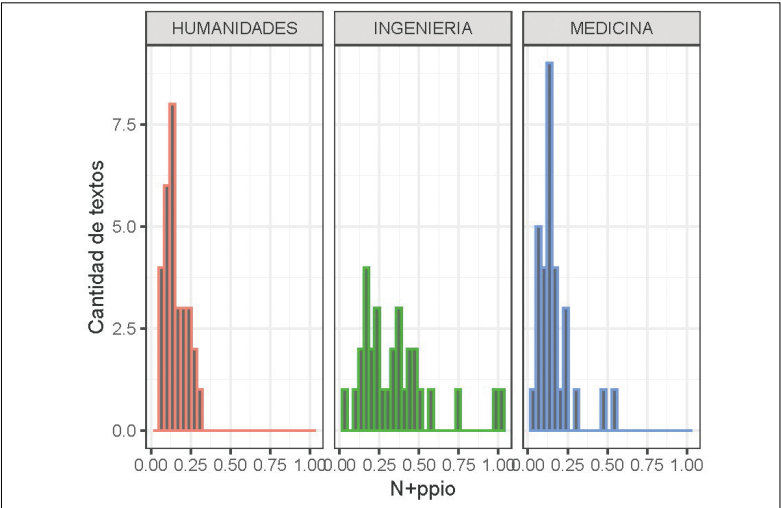


Gráfico 6. Histograma de la distribución del número de ocurrencias de estructuras de nombre seguido de participio por cada 100 palabras según disciplina

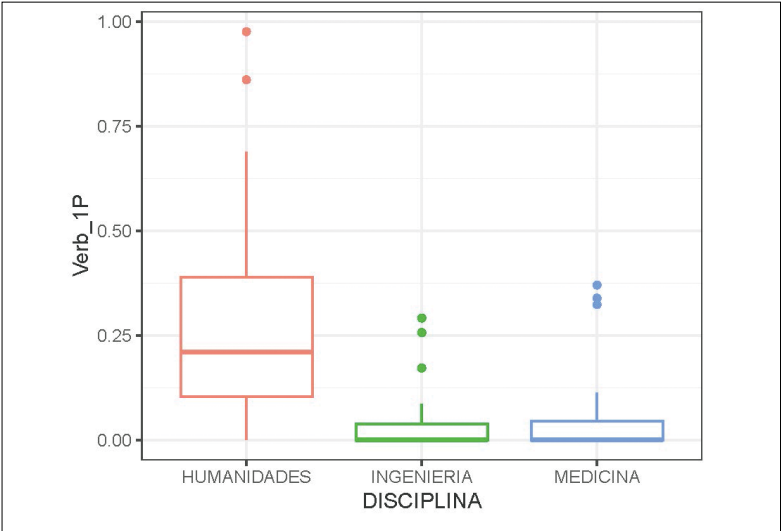


Gráfico 7. Diagramas de caja de la distribución del número de ocurrencias de verbos conjugados en primera persona plural

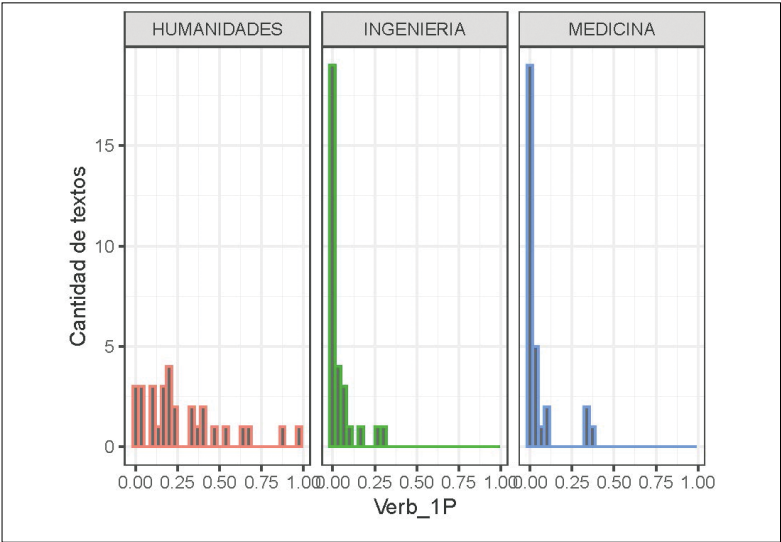


Gráfico 8. Histograma de la distribución del número de ocurrencias de verbos en primera persona plural por cada disciplina

Referencias

- ÁLVAREZ LÓPEZ, F. (2012). ¿ Presento, presentamos o se presenta? La autoría en el español académico escrito de CC. Económicas y Empresariales. En “El español de las profesiones”. Artículos seleccionados del IV Congreso Internacional de Español para Fines Específicos (CIEFE) (p. 27).
- ÁLVAREZ LÓPEZ, F. (2013). La despersonalización en el discurso académico escrito (Tesis doctoral). Universidad de Alcalá. <https://dialnet.unirioja.es/servlet/tesis?codigo=96949>
- ANTOLÍ MARTÍNEZ, J. M. (2019). El uso de la primera persona en el discurso académico. AulaMagna 2.0. [Blog]. <https://cuedespyd.hypotheses.org/6505>
- BELTRÁN, C., & BARBONA, I. (2016). *La estadística en la investigación. Elementos básicos y aplicaciones*. Ediciones del Revés.
- BENVENISTE, E. (1971). *Problemas de lingüística general*. Siglo XXI.
- CARLINO, P. (2005). *Escribir, leer y aprender en la universidad. Una introducción a la alfabetización académica*. Fondo de Cultura Económica.
- CARRANZA, F. M. (2023). Sobre las pasivas con argumentos meta o beneficiario en español. *Lexis*, 47(1), 69-98. <https://dx.doi.org/10.18800/lexis.202301.003>
- CHARAUDEAU, P., & MAINGUENEAU, D. (2002). *Dictionnaire d'analyse du discours*. Éditions du Seuil.
- CODINA, L., & LOPEZOSA, C. (2022). *Escritura de artículos científicos: estructura, redacción, fases y publicación*. Universitat Pompeu Fabra. https://repositori.upf.edu/bitstream/handle/10230/55062/Codina_dig_reda.pdf?sequence=1&isAllowed=y
- DE MIGUEL, E. (2004). La formación de pasivas en español: análisis en términos de la Estructura de Qualia y la Estructura Eventiva. *Vebra Hispánica*, 12(1), 107-129.
- DOURY, M., & PLANTIN, C. (2016). Un enfoque lingüístico e interaccional de la argumentación. *Traslaciones. Revista Latinoamericana de Lectura y Escritura*, 3(6), 11-46. <https://dialnet.unirioja.es/servlet/articulo?codigo=6895516>
- FERNÁNDEZ RAMÍREZ, S. (1986) *Gramática española: el verbo y la oración* (Vol. 4). Arco Libros.

- GARCÍA NEGRONI, M. (2008). Subjetividad y discurso científico-académico. Acerca de algunas manifestaciones de la subjetividad en el artículo de investigación en español. *Revista signos*, 41(66), 9-31.
- GONZÁLEZ CALVO, J. M. (1992). Notas sobre las estructuras llamadas pasivas con “ser” en español. *Anuario de Estudios Filológicos*, 15, 107-123. <http://hdl.handle.net/10662/2766>
- GONZÁLEZ DE REQUENA FARRÉ, J. A. G. (2020). Los otros en *nosotros* y la gramática de la primera persona del plural. *Revista de filología y lingüística de la Universidad de Costa Rica*, 46(1), 195-217. <https://doi.org/10.15517/rfl.v46i1.41116>
- GUTIÉRREZ ORDÓÑEZ, S. (2002). *Forma y sentido en sintaxis*. Arco Libros.
- HERNANZ, M. L. Y SUÑER GRATACÓS, A. (1999). La predicación no copulativa. Las construcciones absolutas. En I. Bosque y V. Demonte (Dir.), *Gramática descriptiva de la lengua española* (Vol. II, pp. 2525-2560). Espasa-Calpe.
- HYLAND, K. (2000). *Disciplinary Discourses. Social Interaction in Academic Writing*. Longman.
- HYLAND, K. (2005a). *Metadiscourse: Exploring Interaction in Writing*. Continuum.
- HYLAND, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- HYLAND, K. (2012). *Disciplinary Identities. Individuality and Community in Academic Discourse*. Cambridge University Press.
- HURTADO, I., & TORO, J. (2005). *Paradigmas y métodos de investigación en tiempos de cambio*. CEC.
- LAMPREA ABRIL, N., TORRES PERDIGÓN, A., & MORENO RODRÍGUEZ, D. C. (2023). ¿Qué nos dicen los verbos acerca de lo que consideramos como investigación? Estudio de la modalización verbal en artículos de investigación relacionados con el lenguaje. *Revista Signos. Estudios de Lingüística*, 56(113). <https://revistasignos.cl/index.php/signos/article/view/627>
- MALDONADO, R. (1999). *A media voz. Problemas conceptuales del clítico se*. Universidad Nacional Autónoma de México.
- MENDIKOETXEA, A. (1999). Construcciones con *se*: medias, pasivas e impersonales. En I. Bosque y V. Demonte (Dir.), *Gramática descriptiva de la lengua española* (Vol. II, pp. 1631-1721). Espasa-Calpe.

- MEZA, P. (2016). El posicionamiento estratégico del autor en artículos de investigación: una propuesta para su estudio. *Forma y Función*, 29(2), 111-134. <https://doi.org/10.15446/fyf.v29n2.60191>
- MEZA, P. (2017). El posicionamiento estratégico del autor en artículos de investigación: un modelo empíricamente fundado. *Logos. Revista de Lingüística, Filosofía y Literatura*, 27(1), 152-164. <https://doi.org/10.15443/RL2711>
- PARODI, G. (2008). Lingüística de corpus: una introducción al ámbito. *RLA. Revista de lingüística teórica y aplicada*, 46(1), 93-119.
- REAL ACADEMIA ESPAÑOLA (RAE) (1973). *Esbozo de una nueva gramática de la lengua española*. Espasa-Calpe.
- REAL ACADEMIA ESPAÑOLA Y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (RAE-ASALE) (2009). *Nueva gramática de la lengua española*. Espasa-Calpe.
- RODRIGO, A., & BONINO, R. (2019). *Aprendo con NooJ: de la lingüística computacional a la enseñanza de la lengua*. Ciudad Gótica. <http://hdl.handle.net/2133/>
- RODRÍGUEZ HERNÁNDEZ, M. Y GARCÍA VALERO, M. (2015). Escritura de textos académicos: dificultades experimentadas por escritores noveles y sugerencias de apoyo. *CPU-e, Revista de Investigación Educativa*, (20), 249-265. <https://doi.org/10.25009/cpue.v0i20.1332>
- SABAJ, MERAUNE O. (2009). Descubriendo algunos problemas en la redacción de Artículos de Investigación Científica (AIC) de alumnos de postgrado. *Revista Signos. Estudios de Lingüística*, 42(69), 107-127.
- SABAJ MERUANE, O. (2012). Uso de movidas retóricas y patrones léxico-gramaticales en artículos de investigación en español: Implicancias para la enseñanza de la escritura científica. *Boletín de filología*, 47(1), 165-186. <https://dx.doi.org/10.4067/S0718-93032012000100007>
- SILBERZTEIN, M. (2016). *Formalizing Natural Languages. The NooJ approach*. ISTE Wiley.
- SWALES, J. (1990). *Genre Analysis: English in Academic and Research Settings*. M. H. Long & J. C. Richards (Eds.). Cambridge University Press.
- SWALES, J. (2004). *Research Genres: Explorations and Applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139524827>

- TAKAGAKI, T. (2008). On the productivity of the Spanish passive constructions. En *Corpus-Based Approaches to Sentence Structures* (pp. 289-309). John Benjamins.
- TRAMALLINO, C. P. (2013). Análisis morfológico con herramientas informáticas: reconocimiento de nombres en textos de español con el sistema Nooj. *Lingüística y Literatura*, (64), 33-48.
- TRAMALLINO, C. P., & SAN MARTÍN, P. S. (2023). Hacia la co-construcción de un dispositivo hipermedial dinámico para el proceso de escritura de artículos científicos en español de estudiantes de postgrado. *Lengua y Sociedad. Revista de lingüística teórica y aplicada*, 22(2), 261-282.
- VENDLER, Z. (1967). Verbs and Times. En *Linguistics in Philosophy* (pp. 97-121). Cornell University Press.