

Construcción y análisis preliminar de un corpus de artículos de investigación: aportes a la alfabetización académica

Design and preliminary analysis of a corpus of research articles: contributions to academic literacy

Picchio, Romina; Panza, Carolina; Gonzalez Vuletich, Laura; Scarafia, Paula

Romina Picchio

rpicchio@hum.unrc.edu.ar

Universidad Nacional de Río Cuarto, Argentina

Carolina Panza

Universidad Nacional de Río Cuarto, Argentina

Laura Gonzalez Vuletich

Universidad Nacional de Río Cuarto, Argentina

Paula Scarafia

Universidad Nacional de Río Cuarto, Argentina

Contextos de Educación

Universidad Nacional de Río Cuarto, Argentina

ISSN-e: 2314-3932

Periodicidad: Semestral

núm. 35, 2023

contextos@hum.unrc.edu.ar

Recepción: 14 Abril 2023

Aprobación: 11 Agosto 2023

URL: <http://portal.amelica.org/ameli/journal/693/6934646004/>

Autor de correspondencia: rpicchio@hum.unrc.edu.ar

Resumen: El presente trabajo tiene como objetivo describir el proceso de diseño, construcción y análisis preliminar de un corpus textual de artículos de investigación escritos en inglés de la disciplina Lingüística Aplicada. El corpus se construyó como una herramienta que aportará a los procesos de enseñanza y aprendizaje de la lectura y la escritura en el contexto del Profesorado y la Licenciatura en Inglés de la Universidad Nacional de Río Cuarto (UNRC). En los textos compilados, se apunta a estudiar el recurso nominalización, típico de la escritura académica, especialmente en cuanto a su potencial de expresar metáforas gramaticales ideacionales, tal como lo indica la Gramática Sistémico-Funcional. El trabajo describe la motivación que dio inicio a este estudio y resume brevemente el marco teórico en el que se enmarcan las distintas etapas del mismo. También se explicitan las decisiones que se tomaron sobre diferentes aspectos del diseño y construcción del corpus, el vínculo con el contexto al cual pretende aportar y las características de los textos que se incluyeron. Seguidamente, se presentan los datos que surgieron del primer análisis cuantitativo – a partir del uso de una herramienta de análisis computarizado –, el cual permitió conocer todos los sustantivos formados por sufijación derivados de adjetivos (deadjetivales) y de verbos (deverbales), con potencial de crear metáforas gramaticales. A modo de conclusión, se comparten algunas futuras líneas de acción para llevar adelante un análisis cualitativo del uso de la nominalización como metáfora gramatical, que permita aportar al proceso de alfabetización académica en nuestro contexto universitario de grado.

Palabras clave: alfabetización académica, corpus especializado, textos disciplinares, nominalización.

Abstract: This paper aims to describe the process of design, construction, and preliminary analysis of a corpus of Applied Linguistics research articles written in English. The corpus was created as a tool to facilitate the teaching and learning of academic reading and writing skills in two programs at the National University of Río Cuarto: Profesorado/Licenciatura en Inglés. The texts collected are aimed to be used to study a frequent resource in academic writing, nominalization, with a focus on its potential to encode ideational grammatical metaphors, as defined by Systemic Functional Grammar (SFG).

This paper describes the motivation for the study and provides a brief summary of the theoretical framework informing its different stages. It also elaborates on the decisions made about different aspects of corpus design, the relationship between the corpus and the academic context where it is intended to be used, and the linguistic features of the texts collected. The quantitative data emerging from a computer-assisted analysis is then presented in the form of a list of all the nouns formed by suffixation derived from adjectives (deadjectival) and verbs (deverbal) with the potential to create grammatical metaphors. To conclude, suggestions for future qualitative analysis are presented with the aim of contributing to the development of academic reading and writing skills in the aforementioned context.

Keywords: academic literacy, specialized corpus, disciplinary texts, nominalization.

1. INTRODUCCIÓN

El desarrollo de la lectura y la escritura en la universidad ha sido objeto de numerosos estudios con el objetivo de llevar adelante propuestas pedagógicas para fortalecer la alfabetización académica (Carlino, 2003, 2005, 2013). El proceso de alfabetización académica implica la adquisición de las herramientas que se necesitan tanto para acceder a material bibliográfico como para producir textos que circulan en una comunidad discursiva disciplinar determinada. En el Profesorado de Inglés y la Licenciatura en Inglés, carreras que ofrece la Universidad Nacional de Río Cuarto (UNRC), este proceso requiere, además, de la adquisición de un nivel avanzado en el idioma para comprender y producir textos de diversos géneros en inglés. En este contexto académico específico es relevante identificar los géneros textuales vinculados con la formación universitaria en estas carreras para estudiar sus características retóricas y lingüísticas típicas como apoyo a los procesos de comprensión y producción. En este trabajo se informa la primera etapa de una investigación que apunta a estudiar un recurso típico de los textos académicos, la nominalización, en un corpus de textos especializados en inglés. El estudio de la nominalización en este contexto está enmarcado en teorías del género (Bhatia, 1993, 2004, 2017; Swales, 1990, 2004, 2019) y en la Gramática Sistémico-Funcional, de aquí en adelante GSF (Halliday y Matthiessen, 2014; Thompson, 2013). El enfoque de corpus, por otra parte, está informado por los desarrollos teóricos de especialistas tales como Biber, et al (1998), Egbert et al. (2022), Hunston (2002), y Sinclair (1991).

La nominalización es una característica de los discursos asociados a las ciencias por su potencial para contribuir a la condensación de información y a la organización textual. (Halliday, 1998, 2004). Numerosos estudios previos han investigado el uso de la nominalización en diferentes géneros textuales asociados con distintas disciplinas y niveles académicos. En esta línea, resulta relevante considerar el análisis del uso de la nominalización en libros de texto de nivel universitario de biología (Hao y Humphrey, 2019) y de física y lingüística aplicada (Jalilifar et al., 2017), y en manuales de biología de nivel secundario (Martin, 1993). Otras investigaciones han analizado el uso de la nominalización como recurso expresivo en artículos de investigación en inglés vinculados con diferentes contextos disciplinares (Arizavi et al, 2015; Bello, 2016; Frigal y Mustafa, 2017; He y Yang, 2018; Holtz, 2009; 2011; Kaidan et al., 2021). Algunos de los estudios

NOTAS DE AUTOR

rpicchio@hum.unrc.edu.ar

aquí citados utilizaron un enfoque de corpus para el análisis lingüístico, mientras que otros obtuvieron resultados a partir del análisis manual y en profundidad de textos individuales. Muchos de ellos, además, analizan el uso de la nominalización como metáfora gramatical ideacional, desde la perspectiva de la GSF. A partir de sus hallazgos, las diferentes investigaciones revelan características del uso de la nominalización según el género textual y la disciplina y señalan la importancia de estudiar su uso para posteriormente diseñar intervenciones didácticas para la enseñanza de la lectura y escritura académicas.

El presente trabajo¹ describe las primeras etapas de una investigación que pretende estudiar el uso de la nominalización como metáfora gramatical ideacional en textos asociados con el proceso de alfabetización académica en el nivel universitario de grado para las carreras Profesorado de Inglés y Licenciatura en Inglés. En primer lugar, desarrollamos de manera sintética el marco teórico que informa esta investigación. Seguidamente, centramos el desarrollo del trabajo en la etapa de construcción del corpus textual como base de datos fundamental para una descripción precisa del uso del recurso en textos académicos relevantes al contexto de enseñanza mencionado. En este sentido, describimos el corpus en cuanto a su vínculo con la disciplina que informa, el género que representa y los criterios de diseño y compilación de los textos que lo componen. Finalmente, presentamos algunos resultados preliminares del uso de la nominalización en el corpus textual y proponemos posibles usos futuros del corpus para explorar en profundidad el recurso estudiado e informar el desarrollo de propuestas pedagógicas.

2. MARCO TEÓRICO

El marco teórico de este estudio integra tres líneas teóricas que informan las diferentes etapas de la investigación realizada: la escuela de Inglés con Fines Específicos (Bathia, 1993, 2004, 2017; Swales, 1990, 2004, 2019), el análisis lingüístico basado en corpus (Biber, Conrad y Reppen, 1998; Egbert, Biber y Gray, 2022; Gray, 2015; Hunston, 2002; Sinclair, 1991) y la Gramática Sistémico-Funcional (Halliday y Matthiessen, 2014; Thompson, 2013).

2.1. Inglés con Fines Específicos

Desde la literatura anglosajona sobre los estudios de géneros textuales, se pueden reconocer diferentes escuelas, tales como la Nueva Retórica (Berkenkotter y Huckin, 1985; Freedman y Medway, 1994; Miller, 1984, 2015), la Lingüística Sistémico-Funcional (Martin, 1997; Martin y Rose, 2008) e Inglés con Fines Específicos (Bhatia, 1993, 2004, 2012, 2017; Swales, 1990, 2004, 2019). Ésta última, la escuela de análisis del género Inglés con Fines Específicos define a los géneros textuales como “instancias de logros exitosos de propósitos comunicativos específicos usando conocimiento convencionalizado de recursos lingüísticos y discursivos” (Bathia, 1993, p. 16, nuestra traducción). En este enfoque, el análisis, descripción y enseñanza de los géneros textuales se vincula principalmente con sus patrones retóricos y características lingüísticas. La finalidad de esta escuela es identificar las convenciones estructurales y características lingüísticas de los géneros textuales que circulan en ámbitos académicos y profesionales para poder utilizarlos con fines pedagógicos.

Por su atención a un contexto académico específico y al estudio de las convenciones lexicogramaticales de un género textual particular, esta línea teórica es, entonces, la más adecuada para dar marco a la investigación que aquí se describe. Esta teoría permite reconocer al artículo de investigación como un género textual relevante al ámbito académico en los últimos años de formación universitaria y posterior práctica docente. Asimismo, sustenta la identificación de recursos lingüísticos típicos tales como la nominalización, la cual es el foco central de análisis en este estudio. La escuela de Inglés con Fines Específicos se nutre del análisis

lingüístico de otras teorías y metodologías, entre las cuales se pueden mencionar el enfoque de corpus y la Gramática Sistémico-Funcional.

2.2. Enfoque de corpus: pautas para su diseño y construcción

Este trabajo aplica el enfoque metodológico de corpus, el cual permite estudiar aspectos del lenguaje en una gran cantidad de textos, a partir de un análisis cuantitativo, cualitativo, o mixto. El término *corpus textual* ha sido definido por autores tales como Biber, et al (1998), Hunston (2002), Sinclair (1991; 2005). En línea con estos especialistas, tomamos la definición más reciente de Egbert et al. (2022), quienes definen corpus como “una gran muestra de textos reunida conforme a principios y diseñada para representar un dominio particular del uso del lenguaje” (p.7, nuestra traducción). Esta definición del término contiene conceptos que resultan centrales al momento de definir criterios de construcción de un corpus con un objetivo de análisis particular: tipos de texto, tamaño, representatividad, dominio y distribución.

En primer lugar, es necesario determinar qué textos – en términos de tipología y registro – conformarán el corpus. La construcción de un corpus requiere, además, determinar el tamaño adecuado para el análisis de ciertos patrones lingüísticos para obtener resultados cuantitativos confiables. En este sentido, Egbert et al. señalan que las decisiones respecto del tamaño del corpus se vinculan con el problema de la representatividad del mismo, la cual debe ser evaluada según el dominio de los textos que conforman la muestra. Estos autores agregan que la representatividad del corpus debería ser considerada en términos de la potencialidad del mismo para responder las preguntas de investigación que motivan su construcción, lo cual involucra atender a cuestiones de dominio (composición del corpus en cuanto a tipos de texto) y de distribución (la cantidad de textos a incluir para poder estudiar aspectos lingüísticos y arribar a datos confiables). Estos principios fueron tenidos en cuenta al momento de tomar decisiones para la construcción del corpus que constituye la base de datos del presente trabajo, tal como intentamos reflejar en la sección Metodología de este escrito.

2.3. La nominalización según la Gramática Sistémico-Funcional

En términos simples, la nominalización puede definirse como un proceso que permite la formación de sustantivos a partir de palabras provenientes de otra categoría gramatical, típicamente verbos o adjetivos. Para comprender el uso de la nominalización en discursos específicos, por ejemplo, los asociados a las ciencias, es preciso partir de la idea de que los significados ideacionales pueden expresarse de más de una forma a través de las configuraciones de transitividad de la cláusula (Taverniers, 2003). En este sentido, desde la perspectiva de la GSF, se distinguen dos tipos de realizaciones entre la gramática y la semántica, una que se considera natural o congruente y otra que se considera no natural o incongruente (Martin, 1993; Halliday, 1985; 1994; 2004). Por medio de las realizaciones congruentes las personas, lugares y cosas son expresadas mediante sustantivos, las acciones mediante verbos y las relaciones lógicas a través de conjunciones. Sin embargo, también puede ocurrir que las acciones se representen a través de sustantivos y las relaciones lógicas a través de verbos. Esta última forma no natural o realización incongruente (Martin, 1993; Halliday, 2004) se lleva a cabo mediante la reconstrucción metafórica de procesos (verbos) y cualidades (adjetivos) como sustantivos.

Debido a que las realizaciones incongruentes son frecuentemente expresadas a través de la nominalización, este recurso se ha identificado como la forma más recurrente para crear metáforas gramaticales ideacionales tal como lo indican Halliday y Matthiessen:

nominalizar es el recurso individual más poderoso para crear metáforas gramaticales. A través de este mecanismo, los procesos (congruentemente realizados a través de verbos) y las características (congruentemente realizadas a través de adjetivos) son expresadas metafóricamente como sustantivos; en lugar de funcionar en la cláusula como Proceso o Atributo, funcionan como entidad en el grupo nominal. (2014, p. 656, nuestra traducción)

Es por esto que la investigación que se presenta en este estudio reporta la identificación de las nominalizaciones en un corpus de lingüística aplicada como primer paso para un futuro análisis cualitativo manual que identifique instancias de metáfora gramatical ideacional.

A continuación, se presentan dos ejemplos de nominalizaciones extraídos del corpus que ilustran cómo la nominalización permite la reconstrucción de significados creando metáforas gramaticales ideacionales:

1. *The present study differs in two aspects from the earlier study which examined the effects of the digital game*, English Extras In Business with A, An, and The on the acquisition of proper English article usages by students (Reynolds & Kao, 2019). *The first major difference* lies in the treatments received by the participants. (S-8)2
2. In contrast, *when we are extrinsically motivated*, we engage in an activity because of some outcomes that are contingent upon it. *Extrinsic motivation* is further categorized from the least self-determined (i.e., external regulation) to the most self-determined (i.e., integrated and identified regulations), with one in between (i.e., introjected regulation). (SLLT-1)

En el Ejemplo (1) se observa la reconstrucción metafórica del verbo *difiere* (*differs*, en inglés) como un sustantivo, *diferencia* (*difference*, en inglés). En el Ejemplo (2), se ilustra la creación de una metáfora ideacional a través de la nominalización del adjetivo motivado (*motivated*, en inglés) que se reconfigura como un sustantivo, *motivación*, (*motivation*, en inglés).

La metáfora gramatical ideacional particularmente centrada en la nominalización es una característica de los discursos asociados a las ciencias (Halliday, 1998, 2004; Halliday y Matthiessen, 2014) y su evolución en el tiempo ha sido documentada en varios estudios (Banks, 2005; Biber y Gray, 2010, entre otros). Su uso permite incrementar el potencial de creación de significados ya que cuando un verbo o adjetivo es reconstruido como un sustantivo, puede ser pre y post modificado casi infinitamente, permitiendo la condensación de información en un grupo nominal (Halliday, 2004). Este proceso permite lograr altos grados de abstracción y tecnicización lingüística que son característicos de los textos académicos (Martin, 1993). Asimismo, la metáfora gramatical ideacional permite organizar la información en un texto, contribuyendo de manera central con la cohesión y la progresión temática (Baratta, 2010; Halliday y Matthiessen, 2014; Liardét, 2016; Ravelli, 2003; Susinskiene, 2012; Taverniers, 2006; Thompson, 2013).

3. METODOLOGÍA

A continuación, se describen las decisiones que informaron la construcción del corpus textual, los pasos que se siguieron para construirlo, y los criterios que se tomaron para el análisis preliminar de los datos obtenidos.

3.1. Diseño y construcción del corpus textual

Para diseñar el corpus textual se tuvieron en cuenta las características del contexto de enseñanza y aprendizaje en el cual se planifica aplicar los resultados con fines pedagógicos, es decir, el Profesorado de Inglés y la Licenciatura en Inglés. En primera instancia, para considerar el contexto de formación docente universitaria, se revisaron documentos oficiales acerca de los profesados universitarios en general y para la formación de profesores y licenciados en lengua extranjera en particular. Este paso involucró la lectura de dos documentos: la Resolución N° 1166/2016 del Comité de Asuntos Académicos del Consejo Interuniversitario Nacional (<https://www.cin.edu.ar/archivo.php>) que establece los lineamientos generales de la formación docente comunes a los profesados universitarios y el Proyecto de mejora para la formación inicial de profesores para el nivel secundario del Ministerio de Educación de la Nación (<http://www.bnm.me.gov.ar/giga1/documentos/EL005898.pdf>). En este análisis, se advirtió la mención de géneros textuales relevantes a la formación

y capacitación continua de docentes y licenciados en lenguas extranjeras. En un segundo momento, para considerar el contexto local de ambas carreras en la UNRC, se analizaron los programas de las asignaturas que incluyen el desarrollo de la alfabetización académica en inglés dentro de sus objetivos: Lengua Inglesa I, II, III, IV, Lengua Inglesa Instrumental, Lengua Inglesa Académica y Lengua Inglesa para la Comunicación Social. Este paso permitió la identificación de los géneros textuales a partir de los cuales la enseñanza de la lectura y la escritura se lleva a cabo en estas asignaturas. El análisis realizado mostró que el artículo de investigación se incluye tanto en documentos oficiales (entre los géneros textuales característicos de la práctica institucional de la formación docente) y en los programas analíticos de asignaturas que desarrollan la lectura y escritura en inglés a nivel avanzado (como texto meta o *target text*). Por lo tanto, se consideró pertinente construir un corpus de artículos de investigación en inglés que permitiera el estudio del uso de un recurso típico de los textos académicos, la nominalización, para eventualmente aportar información relevante acerca de la lectura y la escritura académicas en inglés en el contexto de enseñanza y aprendizaje mencionado.

Para responder a las necesidades contextuales, se delimitó la selección a artículos de investigación correspondientes a la disciplina lingüística aplicada, ya que ésta se vincula con las necesidades en cuanto a la lectura y la escritura en el contexto académico y profesional que da origen a este estudio. Según Schmitt (2010), la Lingüística Aplicada se interesa por conocer “(a) el lenguaje, (b) cómo este se aprende y (c) cómo éste se usa para lograr un objetivo o resolver algún problema en la vida real” (2010, p. 1, nuestra traducción). En esta línea la Asociación Norteamericana de Lingüística Aplicada agrega que la disciplina “recurre a una amplia variedad de enfoques teóricos y metodológicos de varias disciplinas (...) para desarrollar conocimientos acerca del lenguaje, sus usuarios y usos y sus condiciones sociales y materiales subyacentes” (www.aal.org, nuestra traducción). Se estimó que sería relevante, entonces, seleccionar artículos de investigación que representarían las temáticas y abordajes metodológicos de interés para el Profesorado de Inglés y la Licenciatura en Inglés: los textos y sus características, el proceso de enseñanza y aprendizaje de la lengua extranjera/segunda lengua, y el vínculo entre la producción de la lengua y el contexto social en el que se lleva a cabo.

El diseño del corpus textual fue informado también por el estudio de líneas teóricas ya mencionadas en el marco teórico de este trabajo (Biber et al 1998; Hunston, 2002; Sinclair, 1991). Como primer criterio de búsqueda de fuentes para obtener los artículos de investigación a incluir en el corpus, se decidió consultar al portal Scimago (<https://www.scimagojr.com/>), que sugiere una serie de indicadores acerca de la calidad y el impacto de las publicaciones científicas. La decisión de considerar la información publicada en este portal tuvo como objetivo que el corpus textual a conformar incluyera instancias de discurso académico consideradas de importancia en el campo de conocimiento de la disciplina en cuestión (Lake y Cortés, 2020). En el portal Scimago, se realizó una búsqueda de artículos de investigación bajo el rubro “Language and Linguistics”, propuesto por el sitio. Esta acción se realizó en el mes de marzo de 2021. A partir de la lista de las primeras 50 revistas presentadas en orden de prestigio por el sitio web, se seleccionaron aquellas que cumplieran con los siguientes criterios:

1- Disponibilidad: la posibilidad de acceder a los artículos completos (revista de acceso libre o incluida en la biblioteca electrónica de MinCyT, vinculada a la Biblioteca Central de la UNRC).

2- Relevancia disciplinar: la indicación, dentro de la descripción de contenidos publicada por cada revista, del foco de los artículos en uno de los siguientes aspectos: (a) enseñanza y aprendizaje de la lengua y (b) lengua extranjera/segunda lengua y/o inglés como lengua extranjera/inglés como segunda lengua.

Estos criterios permitieron seleccionar las siguientes revistas científicas para incluir en el corpus, que se denominó RAAL (Research Articles of Applied Linguistics). El detalle de los textos incluidos en el corpus se encuentra sintetizado en la Tabla 3.

TABLA 3
Textos incluidos en el corpus

| Revista | Volumen | N° artículos | Codificación en el corpus |
|--|----------------------------------|--------------|---------------------------|
| Modern Language Journal | Vol. 104, núm. 4, Invierno 2020 | 9 | MLJ |
| Language Learning | Vol. 70, núm. 4 Diciembre, 2020 | 6 | LL |
| Language, Learning and Technology | Vol. 24, núm. 3 Octubre, 2020 | 8 | LLT |
| Journal of Second Language Writing | Vol. 50, Diciembre 2020 | 8 | SLW |
| Journal of English for Academic Purposes | Vol. 48, Noviembre, 2020 | 10 | EAP |
| System | Vol. 95, Diciembre, 2020 | 14 | S |
| Assessing Writing | Vol. 46, Octubre, 2020 | 7 | AW |
| Studies in Second Language Learning and Teaching | Vol. 10, núm. 4, Diciembre, 2020 | 5 | SSLT |
| English for Specific Purposes | Vol. 60, Octubre, 2020 | 13 | ESP |
| Journal of Phonetics | Vol. 83, Noviembre, 2020 | 3 | JP |

De las revistas seleccionadas, se obtuvieron los 83 artículos de investigación incluidos en el último volumen publicado en el año 2020. Los artículos fueron preparados para ser analizados con el software lingüístico AntConc 3.5.8 (Anthony, 2019). Dicho proceso implicó eliminar encabezados y pies de página, información acerca de los autores, tablas y figuras, imágenes, apéndices, agradecimientos, notas, referencias, ecuaciones, extractos de entrevistas y ejemplos extensos. Por último, también se eliminaron los resúmenes de los artículos (abstracts) para considerar la construcción futura de otro corpus con estos textos, lo que permitiría comparar el uso de la nominalización entre ambas bases de datos (Holtz, 2009).

3.2. El contexto situacional de los textos del corpus RAAL

Para profundizar la descripción de la base de datos de este estudio, se utilizó una adaptación de criterios propuestos por Gray (2015). Con este objetivo, se llevó a cabo un análisis situacional de cada uno de los textos que conforman el corpus, atendiendo a:

- sus temáticas: a partir del registro de las palabras clave del artículo.
- su(s) objetivo(s): tal como se encontraban expresado(s) en la sección Introducción: enunciado(s), pregunta(s) de investigación o ambas formas.
- la naturaleza de los datos que se obtenían a partir de la investigación: provenientes (a) del análisis de textos/documentos, (b) de la observación a participantes, (c) de a y b.

- sus formas organización textual: si respetaban la organización IMRD (Introducción, Método, Resultados, Discusión), o si presentaban variaciones a esta estructura, agregando secciones o incluyendo subtítulos para organizar cada sección.
- los tipos de metodología: cuantitativa, cualitativa o mixta.
- la descripción del diseño metodológico: claridad en la expresión del objetivo, nivel de detalle de la explicación del diseño metodológico, y la inclusión de evidencia de los datos (figuras, tablas, pasajes textuales).

Entre los aspectos más destacables de esta mirada al contexto situacional de los textos se puede identificar que los artículos de investigación que conforman el corpus realizan aportes vinculados con el análisis, exploración, evaluación y descripción de la enseñanza y el aprendizaje de lenguas extranjeras según diferentes contextos, participantes, habilidades, procedimientos y géneros textuales. Otro aspecto interesante es que en el 41% de los artículos se usan datos que provienen del análisis de textos, en un 36% de los artículos se analizan datos que provienen de textos y participantes, y en el 23% restante se analizan datos provenientes únicamente de participantes. En este sentido, puede decirse que las temáticas identificadas en los textos del corpus, así como también la naturaleza de los datos de las investigaciones que describen, se vinculan estrechamente con aquellas incluidas en la descripción de la disciplina lingüística aplicada (ver Sección 3.1 en este artículo). Además, el análisis situacional de los artículos del corpus reveló que casi la mitad de los textos (48%) describen investigaciones que siguen una metodología cuantitativa, mientras que la otra mitad aplican métodos cualitativos (17%) y mixtos (35%). También se observó que la metodología se describe siempre en detalle, expresando claramente el objetivo del trabajo, en todos los casos por medio de una proposición y, en el 78% de los artículos se incluyen, además, preguntas de investigación. Asimismo, en todos los artículos se describen los procedimientos paso a paso y se muestran de manera explícita los instrumentos (cuestionarios, entrevistas, tests) y datos (respuestas, fragmentos de textos, gráficos y tablas). En cuanto a la organización textual, se identificó variabilidad en cuanto a extensión, ya que se pudo saber que en promedio los artículos de investigación del corpus tienen 19 páginas de extensión (con un rango de entre 9 y 45 páginas). Por otra parte, el 80 % presenta una subdivisión interna que sigue la estructura típica IMRD (Introducción, Metodología, Resultados y Discusión), incluyendo frecuentemente (72%) una sección denominada Conclusión, y subtítulos conceptuales (85%), especialmente dentro de las secciones Introducción y Metodología.

Es importante señalar que el objetivo de este análisis fue profundizar conocimientos respecto de los textos que componen el corpus, para reconocer similitudes y diferencias situacionales entre los textos, y comprender la conexión entre la base de datos y la disciplina con la cual se pretende vincular. La inclusión de esta sección, entonces, no debe interpretarse como un resultado principal de la investigación, sino como un paso previo a la descripción del aspecto lingüístico de interés de este estudio. Se espera que este análisis situacional de los textos aporte a la validez de los datos obtenidos a partir del corpus y resulte de utilidad al momento de relacionar los resultados con las características de los textos en etapas futuras de análisis.

3.3. Criterios para el análisis preliminar de los datos

Para analizar los datos del corpus, se utilizó el programa AntConc en su versión 3.5.8 (Anthony, 2019), en particular, la herramienta *Wordlist* del programa para obtener la cantidad de palabras (*tokens*) y la cantidad de tipos de palabras (*types*). En la lista obtenida, se identificaron los sustantivos formados por sufijación (Biber, et al. 1999; Quirk, et al. 1980) que derivan de adjetivos (deadjetivales) y que derivan de verbos (deverbales). Seguidamente, se eliminaron de manera manual aquellas nominalizaciones que no tenían el potencial de crear metáforas gramaticales. Se reconoce que este primer análisis no toma en cuenta las nominalizaciones formadas por conversión (Quirk, et al. 1980), es decir cuando la misma palabra funciona como verbo y

sustantivo (por ejemplo, *use*) y que otras palabras que no son nominalizaciones pueden también servir para crear metáforas gramaticales. Sin embargo, se toma esta decisión metodológica como paso inicial ya que permite identificar una gran cantidad de datos para comenzar la investigación cualitativa.

RESULTADOS Y DISCUSIÓN

Un análisis global del corpus utilizando la herramienta Wordlist del programa AntConc (Anthony, 2019) mostró que el corpus RAAL contiene un total de 626.715 palabras y 18.200 tipos de palabras diferentes. El análisis manual de la lista de tipos de palabras permitió identificar 1466 nominalizaciones contando formas plurales y singulares como un solo tipo de palabra. Un segundo paso en el análisis manual permitió identificar y eliminar aquellas nominalizaciones formadas por sufijación que no tenían el potencial de crear metáforas gramaticales por constituir expresiones congruentes de significado. Por ejemplo, los sustantivos *learner*, *teacher* y *participant*, que derivan de los verbos *learn*, *teach* y *participate*, son nominalizaciones, pero expresan significados congruentes típicos de los sustantivos, es decir que refieren a entidades (Halliday y Matthiessen, 2014).

Este análisis dio como resultado una lista de 1175 sustantivos formados por sufijación que tienen potencial de crear metáforas gramaticales. Estos sustantivos ocurren 30,290 veces en el corpus, lo que constituye un 4.8% del total de las palabras. Es decir, en el corpus RAAL hay aproximadamente una nominalización con potencial de crear metáfora gramatical cada 20 palabras. La cantidad de palabras para cada sufijo analizado y su frecuencia en el corpus RAAL se muestran en la Tabla 4.1:

TABLA 4.1
Cantidad de tipos diferentes de palabras por sufijo y su frecuencia en el corpus RAAL

| Sufijos | Cantidad de tipos de palabras (types) | Cantidad de palabras (tokens) |
|------------|---------------------------------------|-------------------------------|
| -tion/sion | 542 | 15577 |
| -ty | 245 | 3759 |
| -ness | 99 | 879 |
| -ance/ence | 92 | 3149 |
| -ment | 82 | 3327 |
| -cy | 50 | 2498 |
| -ism | 25 | 245 |
| -al | 16 | 241 |
| -age | 9 | 129 |
| -y/ry | 6 | 62 |
| -ure | 5 | 358 |
| -hood | 1 | 44 |
| -dom | 3 | 22 |
| Total | 1175 | 30290 |

Como se muestra en la Tabla 4.1, los sufijos *-tion/sion*, *-ty*, *-ness* fueron identificados como los más productivos en el corpus RAAL. Estos resultados concuerdan con los datos obtenidos por Biber et al (1999) en un corpus de registro académico de 5 millones de palabras que contiene extractos de libros y artículos de

investigación de 15 disciplinas diferentes. Otros dos sufijos altamente productivos en el corpus RAAL son *-ance/ence* y *-ment*, los cuales fueron identificados como productivos y estudiados en profundidad por otros investigadores en diferentes géneros académicos (Baratta, 2010; He y Yang, 2018; Holtz, 2011).

TABLA 4.2
Nominalizaciones por sufijación de mayor frecuencia en el Corpus RAAL

| Sufijos | Palabras más frecuentes |
|------------|---|
| -tion/sion | information - instruction - communication - interaction - motivation - education - comprehension - discussion - addition - production - attention - intervention - acquisition - recognition - relation - comparison - correlation - variation - revision - introduction |
| -ty | ability - complexity - variety - reliability - intertextuality - authenticity - majority - validity - anxiety - similarity - difficulty - sensitivity - comprehensibility - subjectivity - diversity - opportunity - possibility - capacity - familiarity - personality |
| -ness | awareness - effectiveness - goodness - willingness - accentedness - preparedness - usefulness - openness - politeness - agreeableness - nativeness - appropriateness - conscientiousness - skewness - readiness - nativelikeness - consciousness - conciseness - indirectness - technicalness |
| -ance/ence | performance - difference - reference - variance - competence - importance - preference - significance - relevance - significance - distance - presence - confidence - invariance absence - occurrence - consequence - utterance - coherence - confidence |
| -ment | development - agreement - assessment - engagement - assignment - argument - improvement - entrainment - measurement - achievement - statement - treatment - management - judgment - movement - involvement - embodiment - government - enjoyment - requirement |
| -cy | proficiency - frequency - fluency - proficiency - accuracy - efficacy - literacy - adequacy - tendency - consistency - fluency - legitimacy - competency - transparency - sufficiency - intimacy - primacy - saliency - efficiency - inconsistency |

El análisis de frecuencia reveló que los sufijos *-tion/sion* y *-ty* no solo son los más productivos, sino que presentan una mayor ocurrencia en el corpus, tal es el caso de las palabras terminadas con sufijos *-tion/sion* que totalizan 15.577. Estos resultados también coinciden con los datos obtenidos por Biber et al. (1999). La Tabla 4.2 muestra las veinte palabras más frecuentes formadas con los seis sufijos más productivos en el Corpus RAAL.

5. CONCLUSIÓN

Este trabajo resumió las primeras etapas de una investigación acerca del uso de la nominalización por sufijación con potencial de expresar significados metafóricos en textos académicos vinculados a la disciplina Lingüística Aplicada. A partir de un estudio teórico de temas vinculados al desarrollo del proyecto y de la consideración del contexto de enseñanza y aprendizaje al que pretende aportar, se delimitaron criterios para el diseño, construcción y análisis preliminar de un corpus de artículos de investigación. Este análisis

preliminar permitió identificar sustantivos derivados de verbos y adjetivos, de los cuales se seleccionaron aquellos casos con potencial de expresar metáforas gramaticales ideacionales. Esta mirada reveló la cantidad de nominalizaciones con respecto a la cantidad total de palabras del corpus y los sufijos que más frecuentemente conforman las nominalizaciones en la base de datos diseñada. Este primer análisis representa el paso previo a un futuro análisis cualitativo manual de la nominalización para identificar instancias de metáfora gramatical ideacional.

A partir de estos datos, se pueden establecer futuras líneas de acción que implican tanto la profundización del estudio como la transferencia de los resultados a situaciones de enseñanza vinculadas a la alfabetización académica. En primer lugar, se realizarán los pasos necesarios para completar los resultados de la identificación de nominalizaciones por sufijación, identificando también casos de nominalización por conversión (Quirk, et al. 1980). En segundo lugar, se llevará a cabo un estudio de las nominalizaciones más frecuentes identificando casos de expresión de metáfora gramatical ideacional y analizando la contribución de este recurso a la cohesión textual en una muestra de textos del corpus (Susinskiene, 2012; Liardét, 2016). Se estima que, a partir de los resultados de este análisis más profundo y focalizado de las nominalizaciones frecuentes se podrían diseñar actividades que propendan a mejorar la interpretación de la metáfora gramatical ideacional de los textos académicos que se leen en el contexto disciplinar descripto. Asimismo, sería posible proponer tareas de producción escrita en las que el uso de la nominalización como recurso expresivo sea apropiado y aporte a la condensación de información y cohesión textual. Se utilizaría un abordaje que parta del reconocimiento de la nominalización como uno de los indicadores típicos (no el único) de la presencia de metáforas gramaticales. Se desarrollarían actividades que permitan conectar formas congruentes y no congruentes de expresión y se discutiría la pertinencia del uso de una u otra forma según variables contextuales. De esta manera, se espera que los resultados obtenidos aporten a la alfabetización académica de los estudiantes del Profesorado y la Licenciatura en inglés de nuestro contexto de enseñanza universitaria.

REFERENCIAS

- Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Arizavi, S., Namdari, N. y Mousavi, S. (2015). Nominalization in the Research Article Discussion Sections of Local and International Journals of Applied Linguistics. *Iranian Journal of Applied Linguistics (IJAL)*, 18(1) pp. 1-28.
- Banks, D. (2005). On the historical origins of nominalized process in scientific text. *English for Specific Purposes*, 24(3), pp. 347-357. <https://doi.org/10.1016/j.esp.2004.08.002>.
- Baratta, A. (2010). Nominalization development across an undergraduate academic degree program. *Journal of Pragmatics*, 42, pp. 1017–1036. <https://doi.org/10.1016/j.pragma.2009.08.007>
- Bello, I. (2016). Cognitive implications of nominalizations in the advancement of scientific discourse. *IJES*, 16 (2), pp. 1–23. <https://doi.org/10.6018/ijes/2016/2/262921>
- Berkenkotter, C., y Huckin, T. (1995). Rethinking genre from a sociocognitive perspective. En C. Berkenkotter, y T. Huckin, *Genre knowledge in disciplinary communication. Cognition/ culture/ power* (pp. 1-25). Lawrence Erlbaum.
- Bhatia, V. K. (1993). *Analysing Genre: Language Use in Professional Settings*. Longman.
- Bhatia, V. K. (2004). *Worlds of Written Discourse: A Genre-Based View*. Continuum International.
- Bhatia, V. K. (2012). Critical reflections on genre analysis. *Ibérica*, 24, pp. 17-28.
- Bhatia, V. K. (2017). *Critical genre analysis. Investigating interdiscursive performance in professional practice*. Routledge.
- Biber, D., Conrad, S., y Reppen, R. (1998). *Register variation and English for specific purposes. Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.

- Biber, D., y Gray, B. (2013). Nominalizing the verb phrase in academic science writing. En B. Aarts, J. Close, G. Leech, & S. Wallis (Eds.), *The Verb Phrase in English: Investigating Recent Language Change with Corpora* (pp. 99-132). Cambridge University Press. <https://doi.org/10.1017/CBO9781139060998.006>
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- Carlino, P. (2003). Leer textos científicos y académicos en la educación superior: obstáculos y bienvenidas a una cultura nueva. *UNI-PLURI/VERSIDAD*, 3(2), Versión Digital. Facultad de Educación. Universidad de Antioquia. Medellín, Colombia.
- Carlino, P. (2005). *Escribir, leer y aprender en la universidad. Una introducción a la alfabetización académica*. Fondo de Cultura Económica.
- Carlino, P. (2013). Alfabetización académica. Diez años después. *Revista Mexicana de Investigación Educativa*, 18(57), pp. 355-381.
- Egbert, J, Biber, D. y Gray, B. (2022). *Designing and evaluating Language Corpora. A Practical Framework for Corpus Representativeness*. Cambridge University Press.
- Freedman, A., y Medway, P. (1994). Locating genre studies: antecedents and prospects. En A. Freedman, y P. Medway (Eds.), *Genre and the new rhetoric* (pp. 1-20). Taylor and Francis.
- Friginal, E. y Mustafa, S. (2017). A comparison of U.S.-based and Iraqi English research article abstracts using corpora. *Journal of English for Academic Purposes*, 25, pp. 45-57. <https://doi.org/10.1016/j.jeap.2016.11.004>
- Gray, B. (2015). *Linguistic Variation in Research Articles. When discipline tells only part of the story*. John Benjamins Publishing Company.
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar* (First Edition). Edward Arnold.
- Halliday, M.A.K. (1994). *Introduction to Functional Grammar* (Third Edition). Edward Arnold.
- Halliday, M.A.K. (1993). Some grammatical problems in scientific English. En M.A.K Halliday y J. R. Martin (Eds.), *Writing Science: Literacy and Discursive Power* (pp. 76-94). Routledge.
- Halliday, M. A. K. (1998). Things and Relations: Regrammaticizing Experience as Technical Knowledge. En J. Martin y R. Veel (Eds.), *Reading Science: Critical and Functional Perspectives on Discourses of Science* (pp. 185-235). Routledge.
- Halliday, M. A. K. (2004). *The Language of Science*. Continuum.
- Halliday, M.A.K. y Matthiessen, C. (2014). *Halliday's Introduction to Functional Grammar* (Fourth Edition). Routledge.
- Hao, J. y Humphrey, S. (2019). Reading nominalizations in senior science. *Journal of English for Academic Purposes*, 42, pp. 1-15. <https://doi.org/10.1016/j.jeap.2019.100793>
- He, Q. y Yang B. (2018). A corpus-based study of the correlation between text technicality and ideational metaphor in English. *Lingua*, 203, pp. 51-65. <https://doi.org/10.1016/j.lingua.2017.10.005>
- Holtz, M. (2011). *Lexico-grammatical properties of abstracts and research articles. A corpus-based study of scientific discourse from multiple disciplines*. [Tesis Doctoral]. Disponible en URL: <http://tuprints.ulb.tu-darmstadt.de/2638>.
- Holtz, M. (2009). Nominalisation in scientific discourse. A corpus-based study of abstracts and research articles. Trabajo presentado en la Jornadas *Corpus Linguistics 2009*. Disponible en: <http://ucrel.lancs.ac.uk/publications/cl2009/>.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Jalilifar, A., White, P. y Malekizadeh, N. (2017). Exploring nominalization in scientific textbooks: A cross-disciplinary study of hard and soft sciences. *IJES*, 17(2), pp. 1-20. <https://doi.org/10.6018/ijes/2017/2/272781>
- Kaidan, Z., Jalilifar, A. y Don, A. (2021). On the significance of disciplinary variation in research articles: Perspectives from nominalization. *Cogent Education*, 8(1), pp. 1-18. <https://doi.org/10.1080/2331186X.2021.1890872>
- Lake, W. y Cortés, V. (2020). Spanish and English psychology Methods sections. En Friginal, E. y Hardy, J. (Eds.), *The Routledge Handbook of Corpus Approaches to Discourse Analysis* (pp. 334-353). Routledge.

- Liardét, C. (2016). Grammatical metaphor: Distinguishing success. *Journal of English for Academic Purposes*, 22, pp. 109 – 118. <https://doi.org/10.1016/j.jeap.2016.01.009>
- Martin, J. R. (1993). Life as a Noun: Arresting the Universe in Science and Humanities. En M.A.K Halliday y J. R. Martin (Eds.), *Writing Science: Literacy and Discursive Power* (pp. 242-293). Routledge.
- Martin, J. R. (1997). Analyzing genre. Functional parameters. En F. Christie, y J.R. Martin (Eds.), *Genre and institutions. Social processes in the workplace and school* (pp. 1-39). Continuum.
- Martin, J. R., y Rose, D. (2008). *Genre relations. Mapping culture*. Equinox.
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech*, 70, pp. 151-167.
- Miller, C. (2015). Genre as social action (1984), revisited 30 years later (2014). *Letras & Letras*, 31(3), pp. 56-72.
- Quirk, R. Greenbaum, S. Leech G. y Svartvik, J.(1980). *A Grammar of Contemporary English*. Longman.
- Ravelli, L. (2003). Renewal of connection. Integrating theory and practice in an understanding of grammatical metaphor. En A. Simon-Vanderbergern, M. Taverniers, y L. Ravelli (Eds), *Grammatical Metaphor. Views from Systemic Functional Linguistics* (pp. 37-64). John Benjamins Publishing Company.
- Schmitt, N. (2010). *An Introduction to Applied Linguistics*. Hodder Education.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sinclair, J. (2005). Corpus and text. Basic principles. En M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp.1-16). Oxbow Books. Retrieved from <http://ahds.ac.uk/linguistic-corpora>
- Susinskiene, S. (2012). Nominalization as a lexico-grammatical cohesive device in science popular texts. *Filologija*, 17, pp. 133-144.
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.
- Swales, J. (2004). *Research Genres. Explorations and Applications*. Cambridge University Press.
- Swales, J. (2019). *The Futures of EAP Genre Studies: A Personal Viewpoint*. *Journal of English for Academic Purposes*, 38, pp. 75-82. <http://dx.doi.org/10.1016/j.jeap.2019.01.003>
- Taverniers, M. (2003). Grammatical Metaphor in SFL: A Historiography of the Introduction and Initial Study of the Term. En A. M. Simon-Vandenbergen, M. Taverniers, y L. J. Ravelli (Eds.), *Grammatical Metaphor: Views from Systemic Functional Linguistics* (pp. 5-33). John Benjamins.
- Taverniers, M. (2006). Grammatical Metaphor and Lexical Metaphor: Different Perspectives on Semantic Variation. *Neophilologus*, 90, pp. 321–332. <https://doi.org/10.1007/s11061-005-0531-y>
- Thompson, G. (2013). *Introducing Functional Grammar*. Third Edition. Routledge.

NOTAS

- 1 Asociado al proyecto de investigación Análisis de la nominalización en textos académicos en inglés para aplicar a la enseñanza de la lectura y la escritura en el nivel superior (SeCyT-UNRC 2020-2023), y al proyecto de beca de investigación: Diseño, construcción y análisis de un corpus de textos en inglés para el estudio de la nominalización en la disciplina lingüística aplicada (SeCyT-UNRC 2021), Becaria: Paula Scarafia Directora: Romina Picchio – Codirectora: Laura González Vuletich.
2. La codificación de los artículos del corpus se describe en la sección 3.1.