

2025

Soid Lazlo Ruiz
Ramírez

Lecturabilidad en manuscritos de investigación: un
análisis de correferencias guiado por corpus



Universidad Autónoma de Querétaro
Facultad de Lenguas y Letras

Lecturabilidad en manuscritos de investigación: un
análisis de correferencias guiado por corpus

Que como parte de los requisitos para obtener el título de:

Maestría en Lingüística

Presenta:

Soid Lazlo Ruiz Ramírez

Dirigida por:

Dr. Ignacio Rodríguez Sánchez

Querétaro, Qro., Julio, 2025

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciatario no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciatario.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



Universidad Autónoma de Querétaro

Facultad de lenguas y letras

Maestría en Lingüística

"Lecturabilidad en manuscritos de investigación: un análisis de correferencias guiado por corpus"

Tesis

Que como parte de los requisitos para obtener el Grado de Maestro en Lingüística

Presenta:

Soid Lazlo Ruiz Ramírez

Dirigida por:

Dr. Ignacio Rodríguez Sánchez

SINODALES

Dr. Ignacio Rodríguez Sánchez

Presidente

Dr. Antonio Reyes Pérez

Secretario

Dra. Eva Patricia Velásquez Upegui

Vocal

Dra. Carolina Urizar Ocampo

Suplente

Dra. Eva Margarita Godínez López

Suplente

Centro Universitario, Querétaro, Qro.

2025

México

Le dedico el esfuerzo de estos años de trabajo a mi pequeño Lev.

Contenido

Índice de figuras	7
Índice de tablas	7
Resumen	8
Abstract.....	8
1. Introducción.....	10
1.1. Justificación	12
1.2. Preguntas de investigación	14
1.3. Hipótesis	15
1.4. Objetivos.....	15
2. Marco teórico.....	16
2.1. Descripción del problema.....	16
2.2. Lecturabilidad	18
2.3. Cohesión	21
2.4. Correferencialidad	24
2.5. Encapsulación nominal vs no nominal	26
2.6. La forma pronominal <i>esto</i>	31
3. Antecedentes.....	33
3.1. Índices cuantitativos	33
3.2. Enfoques de la lingüística de corpus (LC).....	35
3.3. Características de un texto académico	36
3.4. Estudios previos relacionados	37
4. Metodología.....	39
4.1. Tipo de investigación	39
4.2. Población o unidad de análisis.....	42
4.3. Implicaciones éticas.....	43
4.4. Criterios de inclusión.....	44
4.5. Técnicas e instrumentos	44

4.6. Procedimientos	45
4.6.1. Pretratamiento de los datos.....	45
4.6.2. Etiquetado del corpus	46
4.6.3. Compilación del corpus	49
4.6.4. Identificación de los usos de <i>esto</i> como anáfora encapsuladora	52
4.6.5. Conteo de palabras por oración en los documentos	56
5. Resultados.....	57
5.1. Apariciones de <i>esto</i>	57
5.2. Longitud oracional.....	60
6. Conclusiones.....	68
Agradecimientos	69
Referencias	69

Índice de figuras

Figura 1. Las palabras comunidad, científica, encuentran y ello etiquetadas con los parámetros descritos	47
Figura 2. Área de trabajo de TagAnt	47
Figura 3. Etiquetas inadecuadas para los tokens “estrategias” y “resumen”. Se reconocen como nombres propios	48
Figura 4. Ventana de sustitución en Notepad++	49
Figura 5. Ventana inicial de AntConc	50
Figura 6. Gestor de corpus de AntConc	54
Figura 7. Creación del corpus en AntConc	52
Figura 8. Ventana de trabajo de AntConc: vista de líneas de concordancia	53
Figura 9. Ventana de trabajo de AntConc: vista del contexto amplio	53
Figura 10. Atributos de archivo. Se aprecia el conteo de palabras de una de las tesis de maestría incluidas en el análisis	56
Figura 11. Conteo de las relaciones de cohesión en el corpus general	58
Figura 12. Conteo de las relaciones de cohesión en los manuscritos de investigación inéditos	58
Figura 13. Conteo de las relaciones de cohesión en los artículos publicados	59
Figura 14. Conteo de las relaciones de cohesión en las tesis de maestría	59
Figura 15. Gráfico descriptivo de las longitudes oracionales por tipo de texto	61

Índice de tablas

Tabla 1. Características de cohesión, categorizaciones, descripciones y ejemplos en TAACO	34
Tabla 2. Repeticiones de <i>esto</i> por tipo de texto	57
Tabla 3. Estadísticos descriptivos entre los grupos	60
Tabla 4. ANOVA de las longitudes oracionales por tipo de texto	60

Resumen

La presente investigación muestra un análisis de la palabra *esto* como recurso de cohesión empleado en la redacción de textos académicos; para tal fin, se construyó un corpus a medida compuesto por manuscritos de investigación, tesis de maestría y artículos de investigación científica publicados. Como punto de partida, se plantea que la finalidad principal de la ciencia es la generación de conocimiento, pero que este se devalúa si no se comunica de forma clara y accesible. Por desgracia, en el ámbito académico, reina una preferencia por el contenido sobre la forma que puede perjudicar la lecturabilidad, es decir, la facilidad con que se comprende un escrito. Con base en el estudio lingüístico de estructuras sintácticamente anómalas, repeticiones léxicas y, especialmente, el uso ambiguo del pronominal *esto*, la intención de esta tesis es evidenciar los obstáculos que dificultan la comprensión lectora. Por su neutralidad morfológica, dicho pronombre tiene compatibilidad con múltiples antecedentes simultáneos, pero esta flexibilidad, más allá de ser una ventaja, a menudo incrementa la ambigüedad y dificulta la recuperación de referentes por parte del lector. En ese sentido, esta investigación se propone evaluar el impacto de *esto* en la claridad del texto, detectar los contextos en que genera mayor ambigüedad, medir su frecuencia en textos inéditos frente a publicados, clasificar el tipo de relaciones cohesivas que establece y cuantificar la longitud de las oraciones en que aparece. Como hipótesis, se plantea que el uso de *esto* tiende a disminuir la lecturabilidad cuando existen varios posibles referentes, es decir, cuando hay ambigüedad anafórica. La finalidad práctica es aportar conocimiento que enriquezca y facilite en alguna medida el trabajo editorial, en aras de una mejor efectividad comunicativa de la investigación científica en nuestro idioma y, en consecuencia, un progreso científico más fluido para nuestra sociedad.

Palabras clave: cohesión, encapsulación, lecturabilidad, redacción académica.

Abstract

This research presents an analysis of the word *esto* (“this”) as a cohesive device used in academic writing. To that end, a custom-built corpus was compiled, consisting of research manuscripts, master’s theses, and published scientific research articles. As a starting point, it is proposed that the main purpose of science is the generation of knowledge, but that such knowledge is devalued if it is not communicated clearly and accessibly. Unfortunately, in academic settings, there is often a preference for content over form, which can impair readability—that is, the ease with which a

text is understood. Based on a linguistic analysis of syntactically anomalous structures, lexical repetition, and especially the ambiguous use of the pronominal *esto*, the aim of this thesis is to highlight the obstacles that hinder reading comprehension. Due to its morphological neutrality, *esto* is compatible with multiple simultaneous antecedents, but rather than being an advantage, this flexibility often increases ambiguity and makes it difficult for readers to recover the intended reference. In this sense, the research aims to evaluate the impact of *esto* on textual clarity, identify the contexts in which it generates the most ambiguity, measure its frequency in unpublished versus published texts, classify the types of cohesive relations it establishes, and quantify the sentence length in which it appears. The working hypothesis is that the use of *esto* tends to reduce readability when there are several possible antecedents—i.e., when anaphoric ambiguity is present. The practical goal is to provide knowledge that enhances and facilitates editorial work, thus contributing to more effective scientific communication in Spanish and, consequently, a less hindered scientific progress for society.

Keywords: cohesion, encapsulation, readability, academic writing.

1. Introducción

Para las redes de investigación, resulta imprescindible que exista una comunicación eficiente entre los académicos dedicados a expandir el conocimiento, pues el contacto interdisciplinar es la vía por la cual la ciencia progresá de manera más rápida y efectiva. Gracias al esfuerzo colectivo de los investigadores, la sociedad en su conjunto es capaz de comprender el mundo y transformarlo a conveniencia y de manera sostenible. En esta línea, el acervo científico debe ser accesible y comprensible, de modo que permita a la civilización enfrentar los problemas sanitarios y medioambientales para mejorar la calidad de vida de las personas; asimismo, debe habilitar la toma de decisiones responsables en materia de políticas públicas, empresariales e individuales. Por ese motivo existen programas de investigación y cuerpos editoriales dedicados a mostrar al mundo en publicaciones periódicas los hallazgos científicos fruto de la investigación experimental y teórica, para ampliar las fronteras del conocimiento.

El presente trabajo surge de la observación de textos académicos dedicados a la investigación que se redactan en pos de salir a la luz en publicaciones científicas. Como cabe suponer, son necesarios procesos de revisión editorial por medio de los cuales se dota a los materiales de los requisitos para ser considerados dignos de aparecer en una publicación de renombre. Así, a partir del acceso directo a manuscritos sometidos a los procesos de revisión, es posible identificar los errores de redacción, deficiencias estructurales, redundancias y otros vicios textuales que suponen obstáculos para la comprensión lectora.

En el caso de la investigación presentada en esta tesis, la atención se centra en los recursos textuales de cohesión dentro de los trabajos de investigación científica, en específico, la encapsulación anafórica mediante unidades pronominales. Cabe mencionar que la encapsulación se trata de una herramienta totalmente legítima, pero que puede cobrar en los textos un talante vicioso, principalmente porque impone al lector una tarea adicional a la lectura: la resolución de ambigüedades. En ese sentido, ataña a los lectores entablar un compromiso profundo con el texto si es que quieren ser capaces de desentrañarlo. No basta ya con decodificar el léxico, la semántica y la estructura sintáctica con que el escrito se presenta; hace falta también emprender búsquedas constantes, ejercer un esfuerzo por hallar un referente específico, oculto en las líneas precedentes —y a veces ni siquiera allá. Una encomienda de tal magnitud puede rápidamente disuadir al lector de su interés por la obra, obstaculizando la transmisión del conocimiento y entorpeciendo el avance de la ciencia.

Con base en estas ideas, el presente escrito se propone realizar una descripción de manuscritos de investigación, artículos publicados y tesis de nivel maestría, todos en el ámbito de la agronomía, a fin de dilucidar cuán eficientes son para transmitir el conocimiento en función de su lecturabilidad. Se eligieron estos tipos de textos a partir del supuesto colectivo de que los profesionales de las ciencias exactas presentan más carencias en redacción que los dedicados a las ciencias sociales y humanidades. Al respecto, se citan dos estudios cuyas conclusiones respaldan en alguna medida tal suposición.

Romero y Álvarez (2020) se centraron en caracterizar la escritura académica de los estudiantes del grado en Lenguas Modernas de la Universidad de Deusto, España. Con ese propósito, recopilaron 96 reseñas producidas *ad hoc* por los alumnos de los cuatro cursos impartidos en dicha institución; posteriormente, analizaron el corpus teniendo en cuenta la estructura, la voz del autor y la corrección lingüística. De su trabajo, constataron mejorías en la selección de la información y la inclusión del lector en el texto, pero no ocurrió lo mismo en lo que concierne a la argumentación e, incluso, se manifiesta cada vez mayor despreocupación por el posicionamiento crítico del autor, la ortografía y la puntuación.

Por su parte, Flores Aguilar (2018) contribuyó a esclarecer la percepción de los cuerpos estudiantiles y docentes de dos carreras de ingeniería respecto de las prácticas de escritura académica en el aula; incluyeron además el análisis de su retroalimentación y la evaluación de competencias de escritura según cada disciplina. Se adaptó y aplicó la “Encuesta Europea sobre la Escritura Académica (EEEA)” a una muestra de 86 estudiantes y 15 de sus profesores. Tanto unos como otros coincidieron en la necesidad de mayor retroalimentación de los textos escritos y los estudiantes refirieron sentirse inseguros de su propio dominio ante las competencias de escritura académica.

En los apartados sucesivos se ofrece una justificación del presente estudio, enfatizando la relevancia de las relaciones correferenciales y la formación de la cohesión en los documentos de investigación científica. Asimismo, se destaca la necesidad de un examen minucioso de estos procesos lingüísticos, específicamente en la construcción de textos científicos en español, pues la investigación al respecto es limitada, al menos en términos de género y registro. Así pues, se recolectó una muestra de manuscritos de investigación inéditos, artículos científicos publicados y tesis de maestría, todos relacionados con la temática de la agronomía. La selección del campo como

ámbito de estudio responde a razones circunstanciales, dado el acceso a documentos de investigación inéditos en una revista científica.

Posteriormente, se describen las preguntas y objetivos de la investigación. Una vez establecidos los propósitos, se acota el presente estudio en un marco teórico compuesto por los conceptos de lecturabilidad, correferencialidad, cohesión y encapsulación, y se puntualizan los fundamentos de cada uno, al igual que su conveniencia dentro de este análisis.

En el subtítulo de Metodología se explicitan las características de la investigación y el corpus analizado en términos de su origen, constitución y pertinencia, así como los procedimientos necesarios y las herramientas informáticas con las cuales se llevaron a cabo. Seguidamente, en Resultados se exhiben algunas tendencias generales del fenómeno, las apariciones de la forma *esto* según el tipo de texto examinado y las longitudes oracionales en términos estadísticos. Asimismo, se ofrece una disección detallada de los ejemplos más representativos detectados en el corpus de estudio. Uno a uno, se analizan en función de las implicaciones de la estructura sintáctica en la lecturabilidad. Finalmente, en la sección de 6. Conclusiones, se explicitan las consecuencias de los hallazgos en la lecturabilidad de textos científicos y académicos, y se plantea la importancia de cuidar los medios de transmisión del conocimiento. Este aporte de la lingüística aplicada al ámbito científico que puede inspirar investigaciones que profundicen el análisis de otros fenómenos a fin de refinar la calidad de los artículos y otros tipos de textos investigativos.

1.1. Justificación

El objetivo básico de la investigación científica es la generación de conocimiento, pero de nada sirve el saber acumulado si no se pone a disposición y servicio de la sociedad; entonces, todo investigador debe contar con la facultad de transmitir el conocimiento que cosecha tras largas jornadas de trabajo intelectual. Por desgracia, a menudo dicha capacidad flaquea cuando se trata de la expresión en página, quizás en gran medida porque se prioriza el contenido sobre la forma; es decir, los investigadores suelen focalizar su atención en la importancia de los hallazgos en sí, relegando a un segundo plano el proceso de comunicarlos. Por supuesto, la omisión es hasta cierto punto comprensible, después de todo, no se puede exigir a nadie que se especialice en cuestiones lingüísticas en paralelo con su formación en cualquier otro campo. No obstante, es innegable que una expresión escrita carente de claridad poco valor aporta a cualquier ámbito de estudio.

A grandes rasgos, los manuscritos de investigación son una fuente rica de fenómenos lingüísticos analizables. Abundan, por ejemplo, las construcciones anómalas sintácticamente, en el sentido de que se intenta articular ideas despojadas de verbos principales, como se exhibe a continuación en (1).

- (1) Si bien el uso de internet ha crecido exponencialmente en la última década y existe una gran cantidad de aplicaciones y sistemas de tecnología digital disponibles en el mercado. Aunado a todo esto se requiere que los profesionales de la educación estén a la vanguardia para adaptarse a los procesos de aprendizaje de las ciencias y buscar estrategias de enseñanza con objetivos claros y precisos... (Corpus propio, 2024)

Se observa en el ejemplo de arriba cómo la primera oración, delimitada por el punto y seguido, carece en realidad de un verbo principal que aporte conclusión a la cláusula iniciada en *si bien*. La estructura predecible tendría la forma de *si bien X, Y*, donde *X* y *Y* representan proposiciones que se perciben como opuestas en cierta medida. En tales instancias, concierne al lector —o en este caso especial, al corrector de estilo— encontrar cuál es la proposición central de un segmento y reconstruirla a partir de los elementos que sí están disponibles en la página; con frecuencia esta labor implica un esfuerzo colaborativo entre el equipo editorial y el autor del texto, lo cual a su vez supone un retraso en el proceso de edición.

Otro yerro común es la aparición de palabras repetidas que, aunque no ostenta la misma gravedad, otorga al texto una propensión al tedio y, peor aún, refleja una carencia léxica que juega en contra del prestigio del autor. El párrafo desplegado en seguida ejemplifica esta clase de fenómeno (2):

- (2) El potencial de hidrógeno o pH es una variable del suelo que solo ha tenido **importancia** en el sector agrícola debido a que es una propiedad química de mucha **importancia** al indicar que tan ácida o alcalina es la solución del suelo, que es donde las raíces y los microorganismos toman sus nutrientes, sin embargo, es **importante** conocer que ocurre con el comportamiento geotécnico del suelo si se le modifica su pH natural. (Corpus propio, 2024)

De entre todos los obstáculos que yacen en la página, el recurso de cohesión (Halliday y Hasan, 1976) basado en la anáfora construida con el pronominal *esto* llama la atención del presente estudio. Dado que su versatilidad lo convierte en un instrumento fácil de abusar, es necesario tener plena conciencia del impacto que tendrá en la lecturabilidad general. Al respecto, como se trata de una unidad de cohesión anafórica, remite al lector hacia los segmentos anteriores, pero su carácter

neutro puede generar ambigüedad, puesto que a menudo es poco claro a cuál elemento previo hace referencia: puede referir a un sujeto anterior, a un objeto directo o, con frecuencia, a una proposición completa.

- (3) Desde Venezuela, Jesús Núñez, explica el abordaje de los saberes “procesos y productos de la creación humana” vinculándoles con las distintas generaciones: abuelos, jóvenes, adultos o niños, quienes poseen dispares grados de dominio, **esto** es susceptible de entenderse desde la pedagogía de los saberes, **esto** es una iniciativa que busca impulsar una educación acorde a las características y condiciones de los escenarios rurales, **esto** es en la cotidaneidad, **esto** descansa en los señalamientos de Paulo Freire, aspecto que Nuñez lo adapta y da como resultado el reconocimiento de tipos de saberes, a saber aquellos salvaguardados, hibridados, sustituidos y emergentes.

Por último, la presente investigación se desarrolla con la intención no solo de describir teóricamente un fenómeno lingüístico en la construcción de textos académicos, sino también de aportar algún beneficio práctico para los profesionales de la edición. Estar conscientes de dicho tema y sus implicaciones prácticas representa una ventaja en el intrincado proceso de producir, editar y publicar trabajos de investigación.

Hay que mencionarlo, los problemas de lecturabilidad tienen un impacto que supera el ámbito de la gramática; la labor de enfrentarlos no surge de un enfoque prescriptivista del lenguaje académico, sino de un compromiso con el acceso a la información, el avance de la ciencia en nuestro idioma y la claridad en la comunicación entre instituciones de investigación. Una baja lecturabilidad obstaculiza la transferencia del conocimiento entre los académicos e investigadores, impacta el ámbito de la educación y contraviene la construcción de una sociedad mejor informada. También es preciso señalar la importancia de la lecturabilidad para los trabajos de traducción, cuya calidad depende del material de origen. Así, la lengua en que se genera el saber científico no debe obstaculizar el progreso de la sociedad, la industria y la tecnología.

1.2. Preguntas de investigación

- ¿Cómo contribuye la anáfora pronominal con *esto* a la lecturabilidad en textos académicos?
- ¿En qué contextos genera mayor ambigüedad?
- ¿Es característica de un texto inédito o también aparece en publicaciones revisadas?
- ¿Qué tipo de relaciones de cohesión entabla la anáfora pronominal con *esto*?

- ¿Se respeta el límite de 35 palabras por oración que supuestamente favorece la lecturabilidad?

1.3. Hipótesis

- Se plantea que la introducción de “esto” como recurso anafórico de cohesión contribuye a incrementar la ambigüedad textual y, por consecuencia, perjudica la comprensión, es decir, disminuye la lecturabilidad.
- Generará mayor ambigüedad en función de cuántos referentes posibles satisfagan la conexión anafórica.
- Aparecerá con mayor frecuencia en los textos inéditos que los publicados.
- Las oraciones excederán con frecuencia el límite de 35 palabras.

1.4. Objetivos

General:

- Llevar a cabo un análisis enfocado en el uso de la unidad lingüística pronominal *esto*, a fin de describir las implicaciones que tiene en la lecturabilidad.

Específicos:

- Compilar un corpus de estudio compuesto por artículos de investigación, manuscritos de investigación y tesis de maestría en temas afines.
- Revisar a profundidad los contextos de aparición de la partícula en términos del tipo de relación cohesiva establecida por la anáfora con *esto*, grado de anidación de la cláusula y longitud oracional.
- Describir el comportamiento de la anáfora con *esto* en los casos donde sea más prominente su impacto en la lecturabilidad.

2. Marco teórico

2.1. Descripción del problema

Existe una oportunidad de investigación en el análisis lingüístico de la escritura académica que corresponde a los textos inéditos, es decir, aquellas versiones que aún no se someten a procesos editoriales, tales como revisiones de pares, corrección de estilo y lecturas de pruebas finales. En ese sentido, la disponibilidad de manuscritos “de puño y letra del autor” supone una oportunidad única de adentrarse en los procesos lingüísticos que los académicos manifiestan en su labor textual. La presente investigación pretende aprovechar esa oportunidad, surgida de la labor editorial en el Despacho de Publicaciones de la Facultad de Ingeniería, para poner de manifiesto las características de dichos escritos.

En ese sentido, se propone que existe una deficiente calidad en la redacción de los textos en general, pero particularmente en los artículos de investigación. Aquí cabe señalar las características que se entienden por norma general como determinantes de una buena redacción: el texto debe ser claro, conciso y libre de ambigüedades; las ideas deben estar organizadas de manera lógica y secuencial; deben emplearse conectores y palabras de transición para mantener la coherencia y facilitar la comprensión al lector. Además, la estructura ha de estar constituida por párrafos que organicen la información en unidades coherentes, y estos a su vez por ideas contenidas en unidades lingüísticas llamadas *oraciones*. Las oraciones se componen generalmente de un sujeto y un predicado, y se conectan entre sí mediante marcadores discursivos que establecen el modo en que las ideas del párrafo evolucionan y se interrelacionan; a este proceso de interconexión se le denomina *cohesión*, y se detallará más adelante.

Como unidades provistas de sentido y constituyentes principales del párrafo, las oraciones juegan un papel determinante en la comprensión del texto; a todo esto, su longitud (medida en número de palabras por oración) podría ser un factor definitivo de la facilidad o dificultad de un escrito. Björnsson (1983) llevó a cabo un experimento donde se cuantificaba la lecturabilidad de textos en función de la longitud media de oración y palabra (su estudio se detalla en la sección de antecedentes); no obstante, como el adagio popular reza, “para muestra, un botón”, a continuación, se presentan algunos ejemplos de oraciones cuya longitud presenta en cierta medida un desafío para el lector. Los ejemplos (4)-(6) se extrajeron del corpus construido a medida para el presente trabajo.

- (4) Sin embargo, estas ideas sólo podrían implicar que existe una clara falta de integración del dispositivo dentro de las aulas y dentro de las estrategias pedagógicas, así como la falta de empatía con las características de las nuevas generaciones y falta de aprovechamiento de las oportunidades que el acceso rápido a la información pueden llegar a brindar, esto debido a que el uso de estas tecnologías no debe catalogarse como una herramienta para sustituir el aprendizaje o algún proceso, sino, por el contrario, se debería considerar como un elemento de apoyo en el proceso de la enseñanza-aprendizaje o un facilitador para el mismo fin. (Corpus propio, 2024)
- (5) El diagnóstico de trastornos de atención como el Trastorno por Déficit de Atención e Hiperactividad (TDAH) es el que más se realiza en niños de edad escolar, sin embargo, las cifras varían dependiendo del país de residencia de los niños y en algunos casos como en México no se tienen cifras oficiales de cuantos niños lo padecen, esto se cree que es debido a que los instrumentos de diagnóstico actuales no toman en cuenta todos los factores que influyen en los mecanismos de la atención, por ejemplo el concepto que se tiene de la niñez en la sociedad podría estar influyendo al creciente número de casos de TDAH, entre otros factores. (Corpus propio, 2024)
- (6) Sin embargo, la llegada de la tecnología ha cambiado los métodos con los cuales se trata de obtener una medición precisa, la principal herramienta tecnológica usada para el diagnóstico de trastornos de atención y que generalmente derivan en un diagnóstico de TDAH es el uso de Electroencefalogramas (del inglés electroencephalography, EEG), que es una forma de monitorear la actividad eléctrica del cerebro, en específico, la fluctuación de electricidad de las neuronas [10], se ha utilizado EEG para el diagnóstico de problemas de atención principalmente para TDAH desde hace 75 años, hasta la actualidad. (Corpus propio, 2024)

La memoria de trabajo, también llamada funcional o mediata, “es la que guarda y procesa durante breve tiempo la información que viene de los registros sensoriales” (Etchepareborda *et al.*, 2005). Tal capacidad cuenta con una duración limitada y es susceptible de interferencias; por lo tanto, se satura conforme las oraciones se extienden, y la comprensión lectora se ve afectada. Por su parte, en *Manual de estilo para la redacción de textos científicos y profesionales*, el bioquímico Xavier Fuentes Arderiu apunta que “para que sean claras, las oraciones se deben construir con un máximo de 25 palabras que expresen una sola idea, sin elementos retóricos que no aporten ningún contenido al significado de la oración” (2003, p. 4). Si bien su aseveración resulta irónica (33 palabras, bastantes más que su límite autoimpuesto), queda de manifiesto que incluso los profesionales de

las ciencias ajenas a la lingüística están conscientes de la importancia de mantener un estilo de redacción sucinto.

Kintsch y Monk (1972) demostraron que la complejidad sintáctica puede ralentizar la velocidad de lectura, incluso si el contenido proposicional permanece intacto. Por lo tanto, aumentar el número de proposiciones de una frase manteniendo constante el número de palabras va necesariamente acompañado de un aumento de la complejidad sintáctica. Posteriormente, Kintsch y Keenan (1973) confirmaron de manera experimental que el tiempo de procesamiento por proposición depende de la longitud del texto leído: para las frases más breves, los sujetos necesitaron poco tiempo de lectura por proposición, mientras que para los párrafos de 50 palabras se requirió aproximadamente el triple de tiempo de procesamiento por unidad textual.

En los ejemplos (4)-(6), puede observarse que se trata de párrafos enteros, compuestos por ideas distintas pero constituidos por una sola oración. Asimismo, aunque es posible discernir con relativa precisión dónde una finaliza y la siguiente inicia, lo innegable es que se trata de secuencias exageradamente largas. Además, resalta el uso de un tipo de “encapsulamiento”: parece ser que los autores se sirven del pronombre demostrativo *esto* (en el tercer ejemplo, el relativizador *que*) como un modo de encadenar las proposiciones siguientes. No obstante, el recurso termina extendiendo la longitud y forzando la memoria de trabajo del lector. Allí yace el foco de este trabajo: se pretende examinar las oraciones en esa dimensión y determinar cómo la práctica de encapsulación pronominal afecta a la comprensión lectora.

2.2. Lecturabilidad

Se denomina lecturabilidad a la facilidad o dificultad de un texto para ser comprendido y leído (Campos *et al.*, 2014), y debe distinguirse categóricamente de la legibilidad, la cual se detallará más adelante en la sección 3. Antecedentes. La complejidad de lectura se encuentra contenida en aspectos lingüísticos, como la diversidad del vocabulario y la longitud de las oraciones; esta última característica es fundamental para el presente estudio, puesto que la lecturabilidad se ve comprometida en la medida en que se incrementa la extensión oracional (ponderada en número de palabras).

Un concepto estrechamente relacionado con la lecturabilidad es la comprensión lectora. Se trata de un proceso cognitivo de alto nivel, el cual, según Kintsch (1998), se compone de dos fases primordiales: la construcción y la integración. En orden, durante la primera etapa, el lector realiza

la percepción, el reconocimiento y el procesamiento sintáctico del texto, es decir, se lleva a cabo la aproximación sensorial y el procesamiento a nivel superficial de la información. Durante la etapa posterior, se entablan las vinculaciones en representaciones de la memoria de largo plazo; es entonces cuando se teje y se dota de significado al entramado global de proposiciones contenidas en cada oración.

A su vez, Riffó y Véliz (2011) proponen una jerarquización de la comprensión lectora en función de la profundidad con que un lector es capaz de asimilar un determinado escrito. A saber, plantea una estratificación triple, en la superficie de la cual se encuentra la comprensión de nivel textual; debajo de esta faz se dispone la comprensión pragmática, y al fondo del constructo se asienta el nivel de la comprensión crítica. Como cabe suponer, cada estrato del modelo requiere un esfuerzo cognitivo más extenuante por parte de la audiencia.

La comprensión de nivel textual refiere al proceso de reconocer las estructuras léxicas, proposicionales, sintácticas, etcétera; es decir, las operaciones básicas de reconocimiento de las características lingüísticas del texto. Por estos motivos, aunque no exento de complejidad, se considera un proceso superficial. Aquí se engloban las tareas de identificar una palabra, reconocer su significado, establecer la relación con otros términos dentro del mismo párrafo; más adelante, se reconocen los límites y secuencias oracionales, y se vinculan los hechos contenidos en las sentencias a través de relaciones funcionales, de causalidad o condicionalidad, por ejemplo. En suma, se establece el significado global del texto: ¿qué está diciendo?

El nivel subyacente, el pragmático, se centra en identificar la relación que media entre el texto y el contexto. Es decir, el lector ya no solamente debe descifrar la información contenida en lo que lee, sino también reconocer el marco temporal, espacial e incluso cultural en que dicho material fue producido. De tal manera, debe prestar atención a quiénes son los participantes del circuito comunicativo establecido por el texto y cuáles son las intenciones detrás de dicha comunicación. Puesto en términos simples, ¿quién escribe?, ¿para quién? y ¿por qué?

En la roca madre de este modelo se encuentra el nivel de comprensión crítica. A esta profundidad, ya no basta con reconocer el contenido, la forma y el contexto de un escrito. Es necesario además aplicar una serie de criterios —de ahí el nombre— para evaluarlo. A partir de dicha valoración, el individuo será capaz de reconocer las operaciones retóricas, cotejar el conocimiento adquirido con su propio acervo cultural o sistema de valores, y en última instancia, instrumentalizarlo en la realización de tareas o la resolución de problemas.

Existe una clasificación adicional de la complejidad textual, formada por los apartados léxico, sintáctico y semántico. Dentro del apartado léxico se encuentra a su vez la densidad, que refiere a la proporción entre palabras nacionales y gramaticales de un discurso: cuando las primeras sobrepasan a las segundas con un margen considerable, se dice que hay una densidad alta; como puede esperarse, cualquier desajuste en el balance de la densidad léxica impactará la lecturabilidad, ya que sería difícil transmitir con claridad una idea sin partículas gramaticales, más aún sin palabras ricas en significado léxico. La diversidad léxica, en cambio, refiere a la relación entre tipos y casos; se denomina *tipo* al conjunto de formas que puede adoptar una palabra, y *caso* a cada aparición individual estas formas. Para comprender esta clasificación de manera breve, las palabras *mujer*, *mujercita* y *mujeres* constituyen tres casos del tipo *mujer*. De manera sencilla, la relación tipo-caso (también conocida como *type-token ratio*) es una medida cuantitativa de la repetitividad. Entre más se aproxima a la unidad, refleja una mayor variedad de vocabulario, en contraste, según se acerca a cero, se infiere un nivel alto de repetición. Al igual que con el aspecto anterior, tampoco es preferible tender hacia los extremos, la moderación es clave para la facilidad de comprensión. El último apartado léxico corresponde a la frecuencia, pero no debe confundirse con la repetitividad a la que alude la diversidad léxica; en su lugar, la frecuencia refiere al uso efectivo de las palabras dentro de una comunidad establecida, en el sentido de que los individuos sean capaces de reconocer el vocabulario y su pertinencia contextual. Así, de nada sirve un lexicón abundante si la audiencia no puede reconocer el significado de las unidades, o si considera que su uso es inapropiado en una situación comunicativa específica.

Por añadidura, Véliz (1999) alude a la complejidad sintáctica en términos de la organización de un texto. En ese sentido, argumenta que “La tendencia a construir oraciones estructuralmente más complejas no va siempre asociada a la escritura de textos más coherentes, especialmente cuando se trata de textos elaborados por escolares o escritores inexpertos”. La autora expone también que hay un proceso de lenta evolución hacia lo que denomina “la madurez sintáctica”. Dicho estado se alcanza cuando el redactor es capaz emplear a conciencia y con discriminación los patrones sintácticos disponibles en su sistema lingüístico convencional. Pero además debe ejercer discreción según la situación comunicativa en la que se desenvuelva, de modo tal que no se expresará del mismo modo durante una narración que al exponer una argumentación, por dar un ejemplo. Así, los sujetos a menudo transicionan por etapas en que construyen oraciones de longitud exagerada sin necesariamente comunicarse de manera exitosa. Existe entonces un punto crítico, un equilibrio

que debe establecerse a partir de la medida en la subordinación clausal: aunque no exista una ley tallada en piedra que prescriba cuántas cláusulas secundarias pueden anidarse dentro de una principal, es definitivo que abusar de dicha práctica solo puede perjudicar la lecturabilidad. En otras palabras, la estructura sintáctica puede llegar a ser tan densa que compromete la eficacia comunicativa.

La complejidad semántica es un aspecto difícil de medir, pues supone adoptar una perspectiva cognitiva del lenguaje. A tal efecto, se considera como medida de complejidad el contenido proposicional (Campos, 2014); cada proposición se compone de un predicado y uno o más argumentos (el número de estos depende del predicado en cuestión). Verbos, adjetivos y adverbios son algunos ejemplos de unidades lingüísticas que comunican predicados, es decir, que codifican propiedades, relaciones, estados y eventos. En cambio, los argumentos pueden ser expresiones referenciales, conceptos o incluso otras proposiciones, y desempeñan lo que se denomina *roles semánticos*, el tipo de participación dentro de un predicado. Introducidos por Fillmore (1968), los roles semánticos pueden tomar la forma de un agente (cuando se trata de una entidad que realiza una acción volitiva), paciente (cuando se recibe una acción), tema (un objeto del cual se predica posesión, localización o transferencia), localización (el escenario donde se lleva a cabo el predicado), entre otras. Ahora bien, las proposiciones gozan de relevancia para la comprensión toda vez que constituyen unidades de medida del procesamiento. Como ejemplo, cada proposición ofrece un empaquetado de información que favorece el registro en la memoria; “las palabras se recuerdan mejor cuando están dentro de una misma proposición que cuando están fuera de ella” (Campos, 2014). En síntesis, a medida que crece la cantidad de proposiciones, se incrementa la complejidad del texto y, por tanto, se dificulta su comprensión.

Con base en el modelo de compresión descrito, Campos centra su esfuerzo en encontrar los factores que inciden en la lecturabilidad; tras la ejecución de un diseño experimental orientado al rendimiento de estudiantes, concluyó que, además de la capacidad individual, la complejidad estructural del texto influye considerablemente en la facilidad con que se entiende un escrito.

2.3. Cohesión

El presente trabajo adopta el término *cohesión* según como describen Halliday y Hasan (1976): la cohesión constituye un mecanismo por medio del cual la interpretación de un elemento del discurso se logra a partir de la decodificación de otro. Es decir, cuando un elemento del discurso hace

referencia a algún otro elemento del mismo, hay cohesión (y correferencialidad, que se detallará más adelante). El concepto describe cómo se tejen las relaciones semánticas esenciales que dotan a los pasajes individuales del discurso o de la escritura con las características de un texto como tal.

En ese sentido, la idea del *nexo* habilita el análisis de un texto en función de sus propiedades cohesivas y, además, da cuenta sistemáticamente de sus patrones, de su “textura”. Todas las unidades gramaticales —oraciones, cláusulas, sintagmas y palabras— gozan de cohesión interna por el simple hecho de estar estructuradas. Así, la cohesión incorpora una relación general de formación del texto, tanto individual como colectivamente. Al mismo tiempo, algunas de estas unidades están sujetas a determinadas restricciones cuando se incorporan a la estructura de una frase, sin duda porque la condición gramatical de ser una frase garantiza que las partes sinergicen para conformar un texto.

En términos simples, la cohesión se sirve de la correferencialidad con el fin de tender un hilo que entreteje las partes del texto o el discurso como un solo lienzo de información. Para comprender dicha propiedad, primero es necesario reconocer que en toda lengua existen ítems léxicos que ostentan la capacidad de referenciar: su significado trae a la conciencia de los interlocutores entidades abstractas o concretas y las deja a disposición para construir el discurso alrededor de ellas. Con esto en cuenta, se sigue que la correferencialidad alude a la propiedad con la cual dos o más ítems hacen referencia a la misma entidad. Su uso resulta versátil en tanto que permite mantener un referente activo sin necesidad de nombrarlo en cada cláusula subsecuente donde aparezca. En la teoría de Givón (2001) sobre la distancia referencial, ofrece un método de acortar dicha travesía al mismo tiempo que se observa la economía del lenguaje y se reserva la repetición como un mecanismo dedicado específicamente a las funciones pragmáticas.

Una descripción de los dispositivos de cohesión no estaría completa si pasara por alto las conjunciones consecutivas y los conectores ilativos (García Izquierdo, 1998). Aunque ambas categorías se asemejan en virtud de que se componen por elementos remáticos —que proporcionan un anclaje desde el cual se ejecuta una transición temática entre oraciones—, los ilativos presentan una característica adicional: no están limitados a su naturaleza semántica de consecutividad, sino que además realizan una función pragmática al orientar el discurso y dirigirlo hacia un fin comunicativo específico. Conviene también registrar que se manifiestan diferencias adicionales en términos de la escala; los ilativos se desenvuelven al nivel interoracional, de modo que contribuyen en mayor medida a establecer la estructura del texto, respecto de las conjunciones consecutivas,

cuyo dominio es el intraoracional. A continuación, se propone una categorización no exhaustiva de los conectores ilativos a fin de ilustrar sus funciones.

- Aditivos (añaden información o ideas similares): *además, también, asimismo, igualmente, de igual modo, incluso, encima, por añadidura*.
- Contrastivos (expresan diferencias o contradicciones entre ideas): *pero, sin embargo, no obstante, aunque, a pesar de, en cambio, por el contrario, mientras que, al contrario*.
- Causales (indican la causa o motivo de algo): *porque, ya que, puesto que, debido a que, gracias a que, como, por motivo de, a causa de*.
- Consecuenciales (muestran el resultado o efecto de algo): *por lo tanto, así que, por consiguiente, en consecuencia, de modo que, por eso, de ahí que, como resultado*.
- De finalidad (expresan el propósito u objetivo de una acción): *para, a fin de, con el propósito de, con la intención de, con el fin de, con el objetivo de*.

En la redacción de artículos de investigación a menudo se prescinde de las herramientas descritas en favor de los recursos pronominales, lo cual puede perjudicar la claridad del texto. Es cierto que los recursos pronominales, tales como los pronombres relativos o demostrativos, son instrumentos valiosos; brindan medios para reducir la repetitividad de un escrito y darle fluidez, pero si se emplean con exceso o de manera inadecuada, repercuten negativamente en la comprensión al dar cabida a la ambigüedad. Peor aún, combinados con otros vicios de la redacción, pueden obstaculizar o directamente imposibilitar la compresión lectora. De nueva cuenta es menester preservar el equilibrio; en este caso, debe aprovecharse el contrapeso que aportan los conectores ilativos, pues no solo refinan la precisión semántica y la variedad léxica, sino que trazan un camino a través de la estructura argumentativa del discurso.

Por otro lado, el valor de los conectores discursivos trasciende el aspecto gramatical; más allá de su función textual, son un medio estratégico para construir el discurso académico. Ya sea a través de la codificación explícita de causas y consecuencias, o de la descripción de contrastes y similitudes, dan cuenta de las relaciones lógicas entre las proposiciones planteadas. De modo que no solo la cohesión del texto se beneficia, también su capacidad persuasiva y rigor científico. En este sentido, los investigadores y redactores académicos deben armarse con un arsenal de conectores ilativos, pero también asegurarse de saber cómo blandirlos, procurando que se ajusten

al tono y propósito del texto. De ese modo, al alternar estratégicamente entre diferentes tipos de conectores se mantiene la claridad, la variedad y el dinamismo de la escritura.

Finalmente, el dominio de los mecanismos de cohesión es una habilidad que se refina con la práctica y la revisión constante. Resulta imprescindible para todo escritor dedicar tiempo al escrutinio y la edición de sus propios textos, en aras de optimizar la coherencia y la cohesión del discurso. Tal empeño se refleja en la calidad, la facilidad de comprensión y el seguimiento por parte de los lectores de la línea argumentativa.

2.4. Correferencialidad

Un concepto vinculado con la cohesión textual es la *correferencialidad*, cuyo estudio comenzó con los trabajos de Halliday y Hasan (1976). Si bien los autores no acuñaron el término, su análisis de los mecanismos anafóricos y catafóricos es fundamental para comprender cómo se interrelacionan los elementos textuales. Poco después, Van Dijk (1977, 1980) y Givon (1980) exploraron tales conexiones en términos semánticos y pragmáticos, y de la continuidad temática en el discurso como un instrumento para mantener activos los referentes en la mente del lector a lo largo del texto.

Concretamente, el concepto de correferencia denota la vinculación que hay entre dos o más unidades lingüísticas que comparten un mismo referente (RAE, s. f.). En español, las relaciones correferenciales pueden darse de tres maneras:

- Entre sintagmas nominales
- Entre pronombres
- Entre sintagmas nominales y pronombres

Aunque no se trata de una característica imprescindible del lenguaje, pues existen contextos donde resulta innecesaria, como en la oración imperativa “abre la puerta”, sí que es prominente en contextos comunicativos complejos, porque dispone un método para mantener la persistencia temática a medida que se extiende una serie de proposiciones. A modo de ejemplo, se despliega en el siguiente párrafo un fragmento extraído del corpus de estudio del presente trabajo, con las anotaciones respectivas de correferencialidad. Se han añadido manualmente los subíndices literales **i, j, k, l, m y n** en las frases nominales y pronominales a fin de indicar cuáles son correferenciales,

(7) [Las plantas]_i al ser [organismos sésiles]_i recurren al reconocimiento específico de [patrones moleculares provenientes de microorganismos]_j para defenderse ante el ataque de patógenos u

otras situaciones de estrés. Dichos [patrones moleculares]_j son también conocidos como [elicitores o estresores biológicos]_j, y son definidos como [moléculas de origen biogénico]_j ([proteínas, carbohidratos, lípidos u otros]_j), capaces de desencadenar [una respuesta fisiológica o morfológica]_k para la activación de [la respuesta inmune vegetal]_k. También, existen [otros tipos de estresores]_l, como [los físicos]_{l, m} ([temperatura, radiación electromagnética, y campos eléctricos]_m) y [químicos]_{l, n} ([nanopartículas, pesticidas, sales, y elementos esenciales]_n) que pueden provocar [respuestas favorables]_k en [las plantas]_i cuando son aplicados en ciertos tiempos, dosis y etapa fenológica, [los cuales]_i no serán tratados en la presente revisión; sin embargo, información relevante de [estos]_i puede encontrarse en [...]. Por lo tanto, en esta revisión el término [elictor]_j será utilizado para hacer referencia a [moléculas biológicas]_j.

A partir del texto anterior, es posible efectuar un análisis ilustrativo de los tres tipos de relación correferencial:

- (8) [Las plantas]_i al ser [organismos sésiles]_i recurren al reconocimiento específico de patrones moleculares provenientes de microorganismos.

En (8) se observa una relación de correferencia de identidad entre sintagmas nominales: *las plantas* y *organismos sésiles*, establecida de manera explícita y directa por el verbo copulativo *ser*. En términos sintácticos, *las plantas* funciona como el sujeto de la oración, mientras que *organismos sésiles* hace las veces del complemento atributivo.

- (9) [...] existen [otros tipos de estresores]_i [...] [los cuales]_i no serán tratados en la presente revisión [...]

El vínculo que se establece por correferencialidad entre *otros tipos de estresores* y *los cuales* identifica ahora a una frase nominal con una forma pronominal. En este caso, se presenta además una concordancia de género y número, la cual guía al lector hacia el antecedente correcto, favoreciendo la cohesión en el texto.

- (10) [...] [los cuales]_i no serán tratados en la presente revisión; sin embargo, información relevante de [estos]_i puede encontrarse en [...].

Se despliega en (10) una correspondencia entre dos formas pronominales; las expresiones *los cuales* y *estos*, en este caso, refieren a la misma entidad mencionada en (9), *otros tipos de estresores*. La concordancia de género y número aporta nuevamente una pista para resolver el antecedente de ambas expresiones.

- (11) También, existen [otros tipos de estresores]_l, como [los físicos]_{l, m} ([temperatura, radiación electromagnética, y campos eléctricos]_m) y [(los) químicos]_{l, n} ([nanopartículas, pesticidas, sales, y elementos esenciales]_n)

Al examinar (11), se observa el establecimiento de una jerarquía correferencial, donde un antecedente general se desglosa en dos subtipos más específicos; las estructuras deductivas como la mostrada son comunes en el discurso científico y favorecen la manifestación de una propiedad lingüística relevante en este tema: el fenómeno de que un mismo antecedente puede ser correferencial con dos expresiones que de hecho no lo son entre ellas. La locución genérica *otros tipos de estresores* se complementa con los sintagmas *los físicos* y *(los) químicos*, que a su vez se identifican con sus respectivas descripciones, donde se enlistan los elementos que conforman cada clase de estresor. Ambas numeraciones entre paréntesis describen dos categorías que se presentan como mutuamente excluyentes y, por tanto, es imposible que refieran a la misma entidad, de ese modo, se erigen fronteras claras entre los conceptos; como última observación, el entramado que el autor construye en (11) demuestra que la compartición de referentes no se limita a las relaciones de identidad, sino que alcanza un nivel de clasificación y especificación.

2.5. Encapsulación nominal vs no nominal

Fue a finales de 1995 que Schmid acuñó el término en inglés *shell noun* para describir un tipo específico de sustantivos: los encapsuladores (en esta sección se usarán indiscriminadamente los términos *shell noun* y *encapsulador*). No obstante, cabe aclarar que el estatus de estos nominales como encapsuladores es contingente. La noción se define en términos funcionales, no es que un nominal cuente con rasgos inherentes que lo conviertan en encapsulador; en su lugar, se trata de un conjunto de funciones que se desempeñan en un discurso dado, ya sea la señalización, la cohesión o la caracterización (Schmid, 2018). Schmid propone, por ejemplo, que las palabras *fact*, *idea*, *problem*, *aim* y *plan* se asemejan —para respetar su metáfora— a un cascarón dentro del cual se resguarda cierta información.

En nuestro idioma, Urizar y Rodríguez (2022) llevaron a cabo un estudio donde consideran los nominales *hecho*, *idea*, *tema* y *problema* en el Corpus del Lenguaje Académico del Español (CLAE). Aunque los autores sostienen que el uso de los encapsuladores difiere con el nivel de lengua y el contexto de aprendizaje de quien los produce, las diferencias no implican una mayor o menor corrección de las producciones escritas; en su lugar, se atribuyen al estilo y la experiencia en la

redacción. Hay que destacar que no cualquier sustantivo cuenta con las características necesarias y suficientes para constituir un encapsulador; es más, ni siquiera puede afirmarse que cualquier sustantivo abstracto pueda cumplir con tal función. En específico, debe haber un espacio semántico que llenar con la información desperdigada en el contexto lingüístico; sirva de ejemplo el contraste de (12) con (13):

- (12) Otro aspecto a considerar, y que afecta directamente al fundamento de la técnica, es que no está claro si el VIH-1 infecta a los espermatozoides. En este sentido hay controversia, ya que los distintos estudios que han abordado el **tema** presentan resultados dispares. (RAE, s. f., corpus CREA)
- (13) Así, se puede ir contabilizando de una forma sistemática y moderna todas las labores, de modo que permita llevar los datos a la mano, todos los días, todo el tiempo, ¿en qué estamos errando?, ¿qué estamos haciendo hoy día? para ir corrigiendo sobre todo el **tema** de costos. (RAE, s. f., corpus CREA)

Nótese cómo en (12), el *tema* al que se hace referencia alude a la proposición *no está claro si el VIH-1 infecta a los espermatozoides*, en virtud de lo cual la frase nominal se comporta como encapsulador. No se da el mismo fenómeno en (13), donde el *tema* se manifiesta únicamente como sustantivo abstracto; de hecho, debe complementarse con la frase preposicional *de costos* para que pueda dotarse de sentido.

Las características funcionales que distinguen a la clase de los *shell nouns* se agrupan en tres niveles descriptivos: a nivel cognitivo, los encapsuladores contribuyen a la formación temporal de entidades con cualidades definidas, a la reificación de unidades conceptuales complejas. Al tiempo que acotan los límites de dichas nociones, generan contrastes con otros sustantivos. Un rasgo distintivo del *shell noun* es que su alcance es local y depende del contexto; entonces, aquello que se caracteriza como *problema*, *desafío* u *objetivo* varía de un uso a otro y depende en gran medida de la proposición encapsulada.

El siguiente nivel es semántico; aquí es donde se desenvuelven las propiedades caracterizadoras del encapsulador al atribuir a las proposiciones codificadas en el contexto lingüístico los rasgos asociados al *shell noun*. Sirva de ejemplo la proposición “sólo es posible lidiar con un número limitado de aspectos”; si se acepta que dicha idea puede encapsularse en el sustantivo *reto*, se describe entonces como “una tarea difícil”. A diferencia de la capacidad de encapsular, la caracterización no es contingente, sino inherente; se encuentra imbuida en el significado léxico

más que en el contextual. Además, se distribuye sobre un gradiente que va desde lo más genérico (*cosa, hecho, caso, situación o evento*) a lo más específico (*desventaja o peculiaridad*).

En términos discursivos, las funciones correferenciales y cohesivas de la encapsulación indican al lector o interlocutor que debe relacionar la caracterización semántica del sustantivo con el contenido codificado en su interior. Luego, los encapsuladores comparten las características deícticas con las formas pronominales en el sentido de que son capaces de apuntar a referentes contextuales; pero, al contrario de lo que ocurre con los pronombres o adjetivos demostrativos, el encapsulador añade una capa de simbolización sobre la simple aplicación referencial. Por tal motivo, un hablante puede recurrir a la construcción “este problema es relevante” en lugar de “esto es relevante”, y de ese modo aprovechar las implicaciones discursivas del *shell noun* para transmitir con énfasis su mensaje a la audiencia.

Ahora bien, la anáfora discursiva es el mecanismo mediante el cual una palabra remite a elementos textuales conocidos (introducidos previamente) en el discurso. La correferencialidad a través de las formas pronominales es una manera productiva de entablar anáforas, aunque sean desprovistas de las ventajas de su contraparte léxica. Así, un pronominal, como *esto*, se encuentra dotado de un significado meramente procedimental; en esencia, se trata de una instrucción: envía al lector a escudriñar el discurso precedente en búsqueda de una construcción oracional que cumpla la función de referente. Como se mencionó antes, se trata de un recurso sumamente productivo, flexible y, por tanto, fácil de abusar. Cualquier descuido en la construcción del contexto lingüístico puede dar pie a la aparición de ambigüedades; por esa razón, la encapsulación nominal se considera una característica de los textos altamente planeados, como es el caso de los académicos y científicos (Nadal Sanchis, 2024). Vale la pena explicitar a continuación algunas instancias que ilustran el contraste entre ambos tipos de encapsulación.

- (14) Los vehículos oficiales de cargos públicos podrán circular igualmente, si bien se restringirá su estacionamiento. **Esta actuación** permitirá la plena peatonalización de la “almendra” del casco viejo y cerrar el programa iniciado en 1980, cuando se restringió el tráfico rodado en la rúa Nova, do Villar, Xelmírez, Fonseca, Obradoiro y Faxeira. (RAE, s. f., corpus CREA)

- (15) La campaña comenzará con una “carta abierta” a los padres sobre la marihuana que será publicada en más de 300 diarios y luego seguirá con avisos gráficos y televisivos. “Cada vez más adolescentes están fumando marihuana a edades más tempranas; necesitamos detener esa tendencia”, indica el

inspector general de Sanidad, Richard Carmona. Y asegura que, como padres, líderes y miembros de la comunidad, necesitamos enfrentar **este problema**. (RAE, s. f., corpus CREA)

En los ejemplos (14) y (15), se observa la implementación de los encapsuladores *actuación* y *problema*. Ambas partículas refieren a proposiciones en la cláusula adyacente anterior de sus respectivos contextos; a saber, la actuación a que alude (14) refiere concretamente a “los vehículos oficiales de cargos públicos podrán circular igualmente, si bien se restringirá su estacionamiento”. En ese caso se caracteriza dicha proposición como una acción implementada, seguramente por alguna entidad gubernamental; se trata entonces de una *actuación* en el sentido de un ejercicio de poder normativo. En analogía, (15) ilustra un ejemplo de un encapsulador léxicamente menos ligero: la palabra *problema* refiere inequívocamente a una situación indeseable, una condición que conviene evitar; por consiguiente, se atribuyen tales características a la proposición contenida en “Cada vez más adolescentes están fumando marihuana a edades más tempranas”. Así, queda clara la postura del autor en cuanto a los eventos que relata. Empero, hay una particularidad adicional que destacar entre ambos encapsuladores: primero, es cierto que recuperan con éxito la porción conocida del discurso y la constituyen como un suelo firme para cimentar la información nueva; sin embargo, ocupan posiciones distintas en el espectro de lo genérico a lo específico.

En efecto, la elección de *actuación* en (14) refleja una neutralidad discursiva, libre de cualquier juicio de valor positivo o negativo. La adopción de tal estilo goza de una ventaja, pues permite presentar los datos de modo objetivo sin influir (supuestamente) en la perspectiva del lector. En forma opuesta, el *problema* al que se hace alusión en (15) lleva una carga valorativa que sugiere una interpretación negativa de los hechos. Al describir la situación como un problema, el autor realiza varias acciones simultáneas, trae de vuelta determinada información previa, la enmarca en un tenor de urgencia y preocupación, pero también declara su postura respecto de las circunstancias. En el mismo ejemplo, aparece la palabra *tendencia*, la cual es correferencial con *problema*; pese a ello, al igual que ocurre en el ejemplo de *actuación*, *tendencia* carece de la carga connotativa que acarrea el *problema*. Quizás por ese motivo el autor precisa redactar una cláusula adicional donde emplea dicho recurso.

Como contraste, se proponen en (16) y (17) los mismos ejemplos, pero con una ligera alteración: se han eliminado los encapsuladores nominales para colocar en su lugar la forma pronominal *esto*. Tal modificación permite observar que la ausencia de especificidad léxica que aportaban los

encapsuladores en los ejemplos anteriores puede afectar la precisión semántica y comprometer la intención discursiva.

- (16) Los vehículos oficiales de cargos públicos podrán circular igualmente, si bien se restringirá su estacionamiento. **Esto** permitirá la plena peatonalización de la “almendra” del casco viejo y cerrar el programa iniciado en 1980, cuando se restringió el tráfico rodado en la rúa Nova, do Villar, Xelmírez, Fonseca, Obradoiro y Faxeira.
- (17) La campaña comenzará con una “carta abierta” a los padres sobre la marihuana que será publicada en más de 300 diarios y luego seguirá con avisos gráficos y televisivos. “Cada vez más adolescentes están fumando marihuana a edades más tempranas; necesitamos detener esa tendencia”, indica el inspector general de Sanidad, Richard Carmona. Y asegura que, como padres, líderes y miembros de la comunidad, necesitamos enfrentar **esto**.

En (16), la sustitución por la partícula *esto* desemboca en una pérdida de información semántica, puesto que no es claro si la proposición precedente describe una acción, política o evento. Algo similar ocurre en (17), donde *esto* suplanta la palabra *problema*, pero deja de lado la interpretación valorativa del hecho como un suceso indeseable, diluyendo hacia la neutralidad la postura del autor.

La diferencia que surge entre ambos pares de ejemplos manifiesta cómo el tipo de encapsulación tiene implicaciones pragmáticas. Entre autor y lector hay una relación unidireccional que el primero puede manipular por medio de la encapsulación nominal; de hecho, al codificar sustantivos, traza un camino sobre el cual el lector puede transitar hacia la interpretación pretendida del texto. En contrapartida, la encapsulación no nominal abre el espacio para la elucidación personal, con ciertos costos involucrados. Al igual que en cualquier otro ámbito, para cada decisión hay un costo de oportunidad, pero uno debe ser consciente del precio que debe pagar por cada ruta que elige al generar un texto.

La pérdida de rasgos semánticos no es el único detrimento que resulta de la pronominalización. En (18), por ejemplo, se observa el uso de *esto* sin un referente claro.

- (18) Esta Conferencia sobre Población y Desarrollo intenta dar soluciones efectivas para la mejora del nivel de vida de los hombres de todo el mundo. Pero, ¿por qué atenta contra ella? Pienso que lo que debería promover dicha Conferencia es una formación profesional en los países con menos recursos y una ayuda efectiva a la familia.

¿Por qué entonces ese miedo a que la población crezca? Creo que se pueden tomar otras medidas que

permitan salvaguardar los derechos humanos. Si **esto** cambia de rumbo aún estamos a tiempo de evitar un gran desastre de dimensiones internacionales. (RAE, s. f., corpus CREA)

Si bien el mensaje es comprensible en su conjunto, no termina de quedar resuelto a qué elemento precedente hace referencia la expresión *esto* en la penúltima línea de la cita. En este caso, hay por lo menos tres candidatos que podrían satisfacer las cualidades del referente:

- *ese miedo a que la población crezca*
- *Creo que se pueden tomar otras medidas que permitan salvaguardar los derechos humanos*
- *se pueden tomar otras medidas que permitan salvaguardar los derechos humanos*

Finalmente, el lector experimentado se percatará de que la unidad léxica en tal instancia está posiblemente vacía, desprovista de referente, y la omitirá a fin de comprender el segmento. Por desgracia, el proceso de filtrar la información tiene un costo cognitivo, y a medida que los microesfuerzos se acumulan, terminan por convertirse en un desgaste mental considerable.

2.6. La forma pronominal *esto*

El vocablo español *esto* proviene del demostrativo neutro latino *iste*, y opera como un pronombre demostrativo neutro. En virtud de su estatus como pronombre, se encuentra desprovisto de contenido semántico en sí; más bien, constituye “un espacio ‘vacío’ que adquiere significado de modo ocasional, dependiendo de las circunstancias del discurso” (Fernández, 1999). Esta clase de palabra tiende a ser unifuncional y no mostrar sincretismos (Parodi y Burdiles, 2016).

En términos de su morfología, la forma pronominal *esto* exhibe una concordancia con la tercera persona del singular; sin embargo, no remite a la noción de persona como tal, porque nunca hace referencia a ninguno de los participantes de una situación comunicativa determinada. Se refiere, en cambio, a algo distinto de la primera o segunda persona gramatical, a la no-persona (Benveniste, 1980). Por otro lado, a causa de su neutralidad de género, designa algo que no es ni masculino ni femenino, pero en español no se reconocen sustantivos con género específicamente neutro. Asimismo, son ausentes flexiones particulares de los adjetivos para concordar en género neutro con los pronombres; ambos hechos son factores que, según la RAE (2016), conllevan a concluir que el neutro no designa un tercer género del español, por lo menos no en el mismo sentido que el masculino y el femenino. Así, la condición neutral de *esto* lo convierte en miembro de una clase gramatical que designa entidades abstractas (RAE, 2016).

En línea con la recuperación de ideas precedentes por medio de los demostrativos, el pronombre personal *esto* es capaz de reproducir conjuntos de “dos o más sustantivos que signifiquen cosas (no personas)” (Bello, 1911). Así, el pronombre *esto* admite como antecedentes oraciones, pronombres o grupos nominales neutros y conjuntos de sintagmas nominales. Además de oraciones, recibe como antecedentes nombres abstractos, por lo general deverbales, interpretados como eventos, situaciones o diversos *statu quo* que se suelen representar mediante oraciones (RAE, 2016). En varios de sus usos, se asemeja al también demostrativo neutro *eso*; por tal razón muestra un comportamiento de alternancia con él, sobre todo por fuera de la literatura, en la oralidad actual; no obstante, al igual que en cualquier otra alternancia, no puede decirse que ambas formas sean totalmente intercambiables.

3. Antecedentes

3.1. Índices cuantitativos

Conviene repasar los conceptos de *legibilidad* y *lecturabilidad* como punto de partida. Sigaud Sellos (2010) resalta la distinción entre ambos términos; pero, aunque presentan similitudes en un sentido general, el autor los discrimina del siguiente modo: señala que la legibilidad alude al aspecto visual de la presentación de un texto (cuestiones como el tamaño, disposición, forma y color de los caracteres); por su parte, la lecturabilidad se refiere a la complejidad inherente de un mensaje escrito, contenida en aspectos como la sintaxis y la semántica. Durante el siglo XIX, tanto en Rusia como en Inglaterra surgió un interés por los factores involucrados en la lecturabilidad; desde entonces, se han desarrollado fórmulas para determinarla cuantitativamente, tales como los índices de Flesch, Dale Chall, SMDG, Fleisch Kincaid, RES y Gunnig Fog (Campos, 2014); independientemente del índice específico, todos comparten el fundamento de que hay ciertas propiedades verbales de los textos que son medibles. Sobre esta base cuantificable es posible obtener un valor objetivo que da cuenta de la complejidad del material escrito.

Por su parte, Hyland (2015) destaca la importancia y utilidad de los análisis de corpus. En su ensayo afirma que dichos enfoques “permiten dotar las investigaciones de una dimensión empírica que sustente las intuiciones, fortalezca las interpretaciones y permita discutir los géneros académicos con mayor confianza”. Ya otras investigaciones se han centrado en el análisis de corpus de manuscritos con distintas finalidades, por ejemplo, la de Daniela Kolbe-Hanna y Benedikt Szmrecsanyi (2015), quienes estudiaron la variación lingüística del complementador *that* en el compendio del idioma inglés *Freiburg English Dialect Corpus*, compuesto por transcripciones de entrevistas orales sostenidas con hablantes de Inglaterra, Gales y Escocia. Ellos emplearon muestras de cláusulas relativas para examinar el comportamiento de dicha partícula: en algunas instancias, *that* queda elidido de las cláusulas relativas, mientras que en otras sí aparece; con base en su conjunto de datos, los autores se sirvieron de herramientas estadísticas para construir un modelo que predice la elisión de *that* con una precisión de 92 %.

El estudio mencionado de C. H. Björnsson se titula *Lecturabilidad de periódicos en 11 idiomas*. En la década de 1960, el autor estableció un índice rudimentario de lecturabilidad que resulta de la siguiente fórmula:

$$\text{longitud de oración} + \text{longitud de palabra} = l_{ix}$$

Donde

longitud de oración: cantidad de palabras por oración

longitud de palabra: porcentaje de palabras que exceden seis letras

l_{ix}: abreviatura de índice de lecturabilidad (*läsbarhetsindex*)

Los criterios comprenden 100 pasajes variados de textos, cuya dificultad fue evaluada por entre 20 y 30 personas. Los elementos lingüísticos se compararon por medio de análisis correlacionales. Por último, se definieron normas para la interpretación del índice de lecturabilidad.

Por otro lado, existen herramientas informáticas que permiten analizar cuantitativamente los textos. En esta categoría se encuentra TAACO (Crossley *et al.*, 2015), un programa de análisis de cohesión dedicado a los textos en inglés. El software cuenta con una licencia de código abierto, es relativamente fácil de operar (en palabras de su creador) y compatible con la mayoría de los sistemas operativos. En su investigación de 2016 en *Behavioral Research*, los diseñadores del programa lo sometieron a prueba para evaluar su capacidad predictiva y concluyeron que la cohesión es una función de las características globales del texto, más que de las locales y generales.

Tabla 1. Características de cohesión, categorizaciones, descripciones y ejemplos en TAACO.

Fuente: Crossley *et al.* (2016).

Característica	Tipo de cohesión	Descripción	Ejemplo
Conectores	Local	Varias listas teóricas y retóricas de conectores.	<i>First, she was rich and happy.</i>
Givenness	Texto general	Relación entre pronombres y sustantivos; incidencia de los demostrativos; artículos definidos	<i>The man was happy he had that.</i>
Type Token Ratio (TTR)	Texto general	Repetición de palabras en un texto	<i>The big dog saw the big cat.</i>
Traslape léxico	Local y global	Solapamiento entre sustantivos, argumentos, tallos, contenido y etiquetas POS (tanto para frases como para	<i>The sun was bright. The day was sunny.</i>

		(párrafos)	
Traslape de sinonimia	de Local global y	Superposición de sinónimos en frases y párrafos	<i>The animal was small. It was a cat.</i>

Otra herramienta computacional es Coh-Metrix (McNamara, D. S., Graesser, A. C., McCarthy, P., y Cai, Z., 2014), un programa que utiliza el procesamiento del lenguaje natural para analizar el discurso. Por medio del cálculo de una serie de índices lingüísticos relacionados con diversos aspectos del lenguaje, puede determinar la calidad, lecturabilidad u otras propiedades específicas de un texto escrito u oral. El sistema analiza múltiples niveles y factores de los textos para ofrecer una perspectiva multidimensional de los mismos. Por ejemplo, mide índices simples como la frecuencia de palabras y la longitud de las frases, y aspectos más complejos, como la cohesión y la complejidad sintáctica. Coh-Metrix se ha aplicado a textos de múltiples sistemas de tutoría educativa y con diversos fines.

El análisis de cohesión de textos a través de plataformas informáticas tiene las ventajas de ser rápido, pues las herramientas de software pueden procesar cantidades ingentes de información en cuestión de segundos. Otro beneficio asociado al procesamiento informático es la consistencia analítica, puesto que el ordenador trabaja con base en algoritmos fijos diseñados específicamente para detectar patrones. Como última ventaja podría listarse la facilidad de acceso que implica una interfaz de usuario; de ese modo, el operador necesita únicamente los conocimientos operativos de la aplicación. No obstantes estas utilidades, hay aspectos positivos asociados al análisis manual, tales como la sensibilidad y flexibilidad analítica o el reconocimiento de características extralingüísticas. Otras ventajas son la comprensión holística del discurso y la adaptabilidad ante las anomalías o divergencias. Por estos motivos, aunados a la ausencia en nuestro idioma de programas de análisis como los mencionados, la presente investigación adopta un acercamiento de análisis manual cualitativo.

3.2. Enfoques de la lingüística de corpus (LC)

La lingüística de corpus es una metodología de investigación basada en el análisis de compendios de textos orales y escritos. Su ventaja principal es que permite observar las manifestaciones del

lenguaje en contextos comunicativos auténticos; es decir, realiza una aproximación empírica a la lengua como la emplean sus hablantes en realidad. A tal efecto, el material que constituye un corpus debe ser representativo, abundante, amplio y variado (Parodi, 2008).

Biber (2015) reconoce dos enfoques principales en los estudios de corpus lingüísticos: el denominado *Corpus-Based* y su contraparte, el *Corpus-Driven* (basado en datos y guiado por datos, respectivamente). El primero consiste en preseleccionar una lista de ítems “potencialmente productivos” y emplearla para analizar las frecuencias y los modos en que se comportan según el contexto donde aparecen; el segundo jerarquiza los datos y, de modo más orgánico, deja que sea el corpus el que manifieste los fenómenos de interés.

3.3. Características de un texto académico

En lo que respecta a las características estilísticas, Alexander Sánchez Upegui recomienda, en su *Manual de redacción académica e investigativa*:

La claridad debe reflejarse en las palabras, frases y párrafos. Ésta se refiere al uso de palabras y frases cortas, lenguaje concreto, ilustrar con ejemplos, recurrir a estructuras que favorezcan la anticipación como títulos enunciativos, entretítulos, resúmenes de información y recuadros explicativos. (2011: 34)

En cuanto a la estructura, las investigaciones incluyen, por lo regular, las secciones correspondientes a *introducción*, *metodología*, *experimentación (resultados)*, *discusión* y *conclusión* (Chandrasekar, 2002). Asimismo, dadas las naturalezas diversas de cada apartado del texto, los autores adoptan métodos distintos para exponer sus ideas y crear la “narrativa científica”; de ese modo, puede observarse un estilo más rebuscado, por ejemplo, en la sección de conclusiones y uno más concreto en los resultados. A continuación, se presentan dos ejemplos extraídos del corpus de estudio:

(19) Conclusiones

Esta contribución ha demostrado, por un lado, la poca relación entre el aparato institucional de cambio climático instalado en el país y la aplicación de políticas y programas en la materia a nivel local; por el otro, la relevancia de considerar las percepciones locales en investigaciones sobre vulnerabilidad

al cambio climático para el diseño de políticas públicas que sean adecuadas a los contextos y expectativas locales. (Corpus propio, 2024)

(20) Resultados

Este apartado presenta los conceptos que tienen las personas sobre el cambio climático y los programas gubernamentales para atenderlo, así como sus percepciones sobre las manifestaciones del fenómeno y sus impactos en las actividades productivas. Aunado a ello, se brindan los datos climáticos referentes al SPI a 12 meses en el periodo 1960-2008 para la localidad de Emiliano Zapata (latitud norte 17°44'31", longitud oeste 91°45'47") con énfasis en los periodos relativamente largos de sequía o precipitaciones fuertes, porque las señales de largo plazo son las que quedan regularmente en la memoria de los usuarios. (Corpus propio, 2024)

3.4. Estudios previos relacionados

Parodi y Burdiles (2016) centraron su atención en un fenómeno similar al abordado en esta tesis, con algunas diferencias específicas: su artículo introduce una mirada profunda al uso de *ellos* como mecanismo de encapsulación anafórica en el discurso de la economía. Construyeron un corpus de 152 textos, compuesto por artículos de investigación, guías de ejercicios, informes de políticas monetarias y manuales. Empero, su aproximación se basa en la categorización del tipo de antecedentes encapsulados en el discurso (cláusulas, complejos clausulares o sintagmas nominales), y de las funciones desempeñadas por el demostrativo *ellos* en cada contexto de aparición (adición, adversatividad, causalidad, finalidad o temporalidad). Tras el análisis, se comprobó que *ellos* suele encapsular cláusulas y complejos clausulares con mucha mayor frecuencia que sintagmas nominales; además, se observó una predominancia marcada de los complejos clausulares sobre las cláusulas en los géneros de artículo de investigación e informes de políticas monetarias. En lo respectivo al comportamiento funcional de *ellos*, se evidenció una tendencia muy fuerte hacia el establecimiento de relaciones de causalidad.

Aunque los autores no lo mencionan explícitamente, cabe suponer que el corpus analizado se compuso de textos con algún nivel de revisión editorial. Entonces, hay que plantearse las implicaciones que tales características imponen en los resultados que obtuvieron, dado que los procesos editoriales suelen desembocar en textos más pulidos en términos de recursos cohesivos, nivel de coherencia y lecturabilidad. Por otra parte, la revisión editorial a menudo adopta una mirada estricta de la lengua escrita, de manera que se eliminan las formas pronominales en favor

de marcadores discursivos y conectores ilativos estandarizados. En definitiva, los textos originales difieren en gran medida de material que termina publicado.

Si bien Parodi y Burdiles ofrecen un panorama valioso sobre el comportamiento de *ello* como anáfora encapsuladora en textos altamente elaborados y revisados, la presente investigación se ciñe a un enfoque distinto, al trabajar con materiales en diferentes niveles de revisión editorial. Así, es posible observar el comportamiento de la anáfora cohesiva con *esto* en condiciones cercanas a la producción académica en progreso, que podría considerarse como un contexto lingüístico más natural que artificial.

4. Metodología

4.1. Tipo de investigación

La presente es una investigación de métodos mixtos; es decir, se emplea una amalgama de elementos cualitativos y cuantitativos. En cuanto a la porción cualitativa del análisis, se llevó a cabo una inspección manual de cada uso de *esto* como anáfora encapsuladora con el fin de determinar los parámetros cuantitativos; en este aspecto se encuentran las frecuencias de aparición del pronominal *esto* y los niveles de anidación de las cláusulas donde figura, así como las longitudes oracionales de cada ejemplo.

El análisis preliminar de las unidades de muestra se llevó a cabo mediante un procedimiento guiado por datos, en el sentido descrito en la sección 3.2. Enfoques de la lingüística de corpus (LC) (Biber, 2015). Los estudios de corpus guiados por datos involucran el análisis de grandes cantidades de datos por medio de herramientas computacionales. Se preparan elementos textuales para examinarlos a conciencia en busca de tendencias e irregularidades. En el contexto del discurso académico, el enfoque permite dilucidar los modos por los cuales los autores construyen significados, transmiten mensajes y argumentan sus ideas; dicho sea de paso, también se evidencian las posturas disciplinarias. Esta clase de estudio presta atención al uso del metadiscurso, los patrones léxicos y la estructura gramatical de los escritos.

Desde tal perspectiva, se realizó una indagación previa dentro del corpus de estudio. Como se mencionó anteriormente, existe una extensa variedad de fenómenos lingüísticos dignos de un análisis a profundidad en la escritura académica. Se muestran a continuación algunos candidatos que se consideraron preliminarmente:

- Anáforas formadas con la estructura *el/la/lo cual*. En estas instancias se observa cómo la oración se subordina y extiende hasta 83 palabras; pero, regularmente, el antecedente se encuentra justo previo a la construcción, lo que dificulta la generación de ambigüedad:
- (21) El departamento de Nariño según datos reportados por el Ministerio de Agricultura y desarrollo Rural-MADR ocupa el sexto puesto en producción de cacao a nivel nacional y 76% de la producción de cacao se concentra en el municipio de Tumaco, **el cual** está ubicado en el sur occidente de Colombia sobre la Costa Pacífica, **la cual** se caracteriza por tener un clima tropical húmedo y cuenta

con aproximadamente 17 809 ha de cacao sembrado y el rendimiento promedio no supera las 0.466 kg ha⁻¹. (Corpus propio, 2024)

- Anáfora formada con la estructura *misma(s)/mismo(s) que*. Al igual que la opción anterior, la ambigüedad se resuelve rápidamente en el contexto; empero, cabe resaltar la longitud oracional del ejemplo, de 71 palabras.
- (22) El análisis de nuestro sistema ambiental, y sus elementos, deriva de la información existente, tanto en la literatura científica, como en las fuentes oficiales de información socioeconómica y de gobernanza, por ello, se han seleccionado a la revisión sistemática, el meta-análisis y el análisis de contenido como las metodologías para la obtención y el análisis de la información, **mismas que** se consideran oportunas por sus procesos exhaustivos, rigurosos, y validables.
- Introducción de gerundios como predicados adicionales. En estos casos, los gerundios, al ser formas verbales desprovistas de concordancia, pueden generar ambigüedades en cuanto a los agentes que realizan las acciones. El problema se agrava si además la puntuación es inadecuada, como puede observarse en el ejemplo (23):
- (23) El aprovechamiento de los traspasios puede contribuir en las tres esferas de la sustentabilidad, pues combate la degradación ambiental, se reduce el consumo de recursos, mejora la salud y el estado nutricional de las personas **promoviendo** la producción de manera sustentable de alimentos suficientes, sanos e inocuos para una población creciente y con tendencia a lograr la seguridad alimentaria. (Corpus propio, 2024)
- Usos concatenados del pronombre relativizador *que*. Este fenómeno puede contribuir a la anidación de las cláusulas y perjudicar la comprensión lectora, pero se descartó en virtud de que no mostraba gravedad ni prevalencia:
- (24) La agricultura periurbana, es una actividad adaptable, móvil y en constante renovación, **que** se ubica en una zona de transición **que** aún no ha sido completamente asimilada por la mancha urbana, ubicada en un espacio híbrido, **que** posee tanto rasgos rurales como urbanos. (Corpus propio, 2024)
- Ambigüedad en el uso del relativizador *quien*. En (25), no queda resuelta la ambigüedad respecto de quién es maestro en la Universidad, pues el relativizador *quien* puede remitir tanto al señor Chabelo como a su hijo. En ese mismo ejemplo puede avistarse otra imprecisión en la frase preposicional *como la Dra. Guzmán*. De forma estricta, el sintagma podría complementar la oración de, al menos, dos maneras distintas: *Él, como la Dra.*

Guzmán, ha invitado a investigadores... y Él ha invitado a investigadores, como la Dra. Guzmán... Independientemente de cuál es la interpretación principal y cuál la marginal, estos casos fueron infrecuentes y por ese motivo se descartaron.

- (25) Un caso en particular es el hijo del señor Chabelo, quien es maestro en la Universidad Autónoma del Estado de Morelos en la facultad de Ciencias Agropecuarias. Él ha invitado a investigadores a conocer los viveros como la Dra. Guzmán. (Corpus propio, 2024)

Se seleccionó de entre las alternativas la anáfora cohesiva formada con el pronominal *esto*, pues se manifiesta con frecuencia en el corpus de estudio. Además, la riqueza de contextos de aparición abre el campo para una interpretación lingüística detallada. En (26) se muestra una oración de 118 palabras donde se ejemplifica este fenómeno.

- (26) En México, también podemos ver un proceso de envejecimiento de la población rural, en donde, según un reporte de la SAGARPA, la edad promedio de las personas dedicadas al campo era de 55.6 años, con una gran proporción de responsables de más de 60 años, **esto**, aunado a una falta de relevo generacional, impulsado por la inflexibilidad en el mercado de tierras, baja rentabilidad de la actividad agropecuaria, carencias educativas para reforzar el capital humano, falta de incentivos para mantener y acceder a este sector productivo, que también tienen alto nivel de marginación y pobreza, y un bajo nivel educativo, situaciones que contribuyen a la migración rural de jóvenes y adultos, datos preocupantes para el campo mexicano. (Corpus propio, 2024)

En lo respectivo al alcance de la investigación, dadas las condiciones de tiempo y el poder de procesamiento disponible, esta tesis se limita al examen descriptivo de la anáfora con *esto* en textos académicos. En esa línea, un acercamiento descriptivo implica únicamente la observación e ilustración detallada de los usos de *esto* como mecanismo de cohesión textual; no es la intención de este trabajo emitir juicios normativos acerca de su adecuación, sino poner de relieve el impacto sobre la lecturabilidad en el ámbito de la investigación científica. Se presentan los resultados de las observaciones en un conjunto de ejemplos elegidos específicamente por su prominencia, con base en su frecuencia, complejidad estructural o porque representan casos especialmente ilustrativos del patrón lingüístico bajo estudio.

Finalmente, ya que el alcance de esta tesis es acotado, se espera que sus hallazgos sirvan como punto de partida para investigaciones futuras, que profundicen en los efectos pragmáticos y retóricos de la anáfora encapsuladora en distintos géneros discursivos.

4.2. Población o unidad de análisis

Los datos, como ya se ha mencionado anteriormente, fueron extraídos de la plataforma Open Journal System; se trata de manuscritos de investigación en el área de agronomía, que sus autores han sometido al proceso de revisión por pares de la revista *Perspectivas de la Ciencia y la Tecnología* (PCT); se incluyen también tesis de maestría recabadas del repositorio institucional de la Benemérita Universidad Autónoma de Puebla (BUAP); por último, se eligieron artículos publicados en la Revista Mexicana de Ciencias Agrícolas (RMA). El propósito de controlar las tres categorías fue realizar una comparación entre los textos revisados (los publicados en la RMA) y los inéditos (tesis y envíos a revisión), y confirmar o rechazar la hipótesis de que la anáfora con *esto* es característica de los textos menos pulidos. En cuanto al enfoque del análisis, se adoptó un procedimiento concurrente cuantitativo-cualitativo.

La selección de la temática en ciencias agrícolas obedece a un criterio de conveniencia, puesto que los manuscritos inéditos en la plataforma OJS son un recurso escaso; por ese motivo constituyeron el eje para la selección del resto de textos. Adicionalmente, se consideran todos los documentos como equiparables dentro de cada categoría (inéditos, publicados y tesis); es decir, no se hizo distinción entre investigaciones originales, revisiones de literatura o metanálisis. En seguida, se muestran las descripciones de ambas revistas (PCT y RMA) como se muestran en sus portales en línea:

Perspectivas de la Ciencia y la Tecnología es una revista semestral publicada por la Universidad Autónoma de Querétaro, a través de la División de Investigación y Posgrado de la Facultad de Ingeniería. Los contenidos incluyen artículos originales, reseñas y estudios que han resultado de proyectos que siguen un riguroso proceso de investigación o de una aplicación innovadora en las áreas de estudio de la Ingeniería. (Revistas UAQ, 2023)

La *Revista Mexicana de Ciencias Agrícolas*, tiene como objetivo publicar escritos originales e inéditos, relacionados con las ciencias agrícolas y a fines, en español e inglés, que pueden presentarse en forma de artículos científicos, nota de investigación, ensayo y descripción de cultivares. (Revista Mexicana de Ciencias Agrícolas, 2023)

Por su parte, la BUAP ofrece diversos programas de posgrado relacionados con la agronomía, la economía agrícola y el manejo sustentable del terreno. Por ese motivo se eligió como la fuente de las tesis analizadas en la presente investigación.

El corpus recopilado se compone de sesenta documentos, veinte por cada categoría y, en total, se contabilizaron 573 842 palabras (o *tokens*). Si bien se trata de una cifra limitada, sobre todo para un análisis a gran escala de la escritura académica, las dimensiones del corpus permitieron observar patrones y tendencias de los textos que vale la pena describir. Finalmente, se muestraron esas 3 categorías porque se asocian al nivel de la investigación científica. En cierta medida, pueden considerarse muestras representativas de los esfuerzos de investigadores mexicanos para comunicar efectivamente sus descubrimientos.

4.3. Implicaciones éticas

La presente investigación se dedica al análisis de lecturabilidad de tres tipos de textos académicos: tesis de maestría, artículos publicados y manuscritos de investigación. Respecto de las primeras dos clases, se trata de documentos extraídos de fuentes de acceso abierto. En cuanto a la tercera, el acceso se realizó por medio de la plataforma OJS, específicamente durante los procesos de revisión por pares a doble ciego, de modo que los recursos están anonimizados. No se recopilan datos personales ni profesionales de autores, más allá de la corroboración de su grado académico.

Por otro lado, cada uno de los géneros elegidos atravesía procesos de validación distintos. Los artículos académicos suelen pasar por un riguroso proceso de dictamen por pares, lo que con frecuencia implica intervenciones editoriales y estilísticas destinadas a mejorar la claridad y la cohesión del texto. En contraste, las tesis de maestría son productos de formación, elaborados por estudiantes en vías de consolidarse como académicos, mientras que los manuscritos corresponden a trabajos aún pendientes de revisión formal. Es preciso reconocer tales diferencias y manifestar explícitamente que el presente trabajo no pretende descalificar a ninguno de los géneros.

Asimismo, se ha procurado mantener el enfoque en las características lingüísticas y discursivas de los textos, evitando cualquier juicio de valor sobre la competencia profesional o intelectual de los autores. Las observaciones y resultados aquí planteados se orientan a visibilizar fenómenos y patrones lingüísticos; no existe intención alguna de evaluar la experticia profesional de quienes compusieron los documentos.

Finalmente, es cierto que señalar las limitaciones de cohesión en determinados textos podría interpretarse como una crítica negativa. No obstante, el propósito de la presente tesis trasciende la descripción de los problemas identificados, pues busca contribuir a la reflexión sobre la enseñanza de la escritura académica y la importancia de atender la cohesión como un factor fundamental para la transmisión clara y efectiva del conocimiento.

4.4. Criterios de inclusión

Los materiales de estudio se recopilaron según los siguientes criterios:

- Se consideraron únicamente documentos redactados en español de México, para descartar diferencias debidas a la nacionalidad de los autores.
- Se seleccionaron textos directamente relacionados con el ámbito de la agricultura, por los motivos señalados anteriormente.
- Se limitó la inclusión a trabajos escritos por académicos con grado de maestría o superior, pues se presupone que son quienes muestran más pericia en términos de redacción.

Por otra parte, la selección de las unidades de análisis se efectuó en función de un criterio específico:

- Se emplea la unidad lingüística *esto* como método de encapsulación. Es decir, se eligieron los segmentos donde, en lugar de un encapsulador nominal, aparece la forma pronominal *esto*.
- En ese sentido, se descartaron las formas *este/a + sustantivo, esta, estos y estas*.
- El criterio formal para la búsqueda fue que *esto* apareciera como inicio de oración (después de un punto) o de cláusula relativa (después de coma).

4.5. Técnicas e instrumentos

Puesto que el procedimiento requirió de la creación de un corpus a medida, se recurrió a las herramientas informáticas TagAnt (Anthony, 2022) y AntConc (Anthony, 2023), ambas desarrolladas por el equipo de ingeniería lingüística de AntLab. El instrumento TagAnt es un programa de código abierto y licencia libre de etiquetado basado en categorías gramaticales (POS, por sus siglas en inglés), diseñado para el análisis de textos en lingüística de corpus. Su función

principal es reconocer los *tokens* (muestras) dentro de un texto y clasificarlos de manera explícita para su posterior análisis. Los *tokens* son elementos básicos utilizados en análisis de corpus y procesamiento del lenguaje natural (PLN); en términos sencillos, podría decirse que cada palabra dentro del texto constituye un *token*. AntConc, por su parte, pertenece a la misma familia de software, pero su enfoque se centra en el análisis de corpus que ya han sido debidamente etiquetados; proporciona herramientas como visualización de líneas de concordancia, aparición de palabras clave en contexto y nubes de palabras; además, permite al investigador realizar búsquedas de palabras específicas, frases e incluso estructuras como colocaciones o coligaciones. Ambas aplicaciones informáticas gozan de aceptación en estudios de lingüística de corpus gracias a que han mostrado efectividad y precisión para el procesamiento de grandes volúmenes de texto.

Se encontró necesario incorporar la aplicación de edición de texto Notepad++ como parte del proceso para realizar modificaciones en el corpus a fin de depurarlo. En este contexto, las ediciones se dirigieron específicamente a la limpieza de los textos como una medida para asegurar la coherencia y uniformidad en el conjunto de datos, y evitar cualquier interferencia durante las búsquedas. En el apartado 4.6.1. Pretratamiento de los datos, se proporciona una descripción completa de la metodología empleada.

Finalmente, se utilizó también la aplicación de hojas de cálculo de Microsoft Excel como método para el conteo de los datos. Se consignó un ejemplo por línea y se aprovecharon las columnas para disponer los contextos previo y posterior, así como el nivel de anidación y las observaciones cualitativas conforme iban apareciendo.

4.6. Procedimientos

4.6.1. Pretratamiento de los datos

Los manuscritos de investigación se ciñen, por lo regular, a una estructura establecida con la finalidad de facilitar la organización del texto y, por ende, el modo de presentar la información al público objetivo; lo usual es que se compongan de apartados como *resumen*, *introducción*, *metodologías*, etc. Por ese motivo, fue necesario someter el material a un proceso manual de limpieza de datos por medio de la herramienta de búsqueda y sustitución de Notepad++: resultó indispensable descartar aquellas secciones que entorpecieran el análisis debido a que no constituyen manifestaciones de uso del lenguaje escrito, sino que conforman una especie de

“plantilla” para la exhibición del texto. En dicho proceso de limpieza, se tomó la decisión de descartar las secciones de *Abstract*, puesto que se redactan en un idioma distinto al español, y no aportan información pertinente al análisis. En el caso de los encabezados de sección, su eliminación responde al hecho de que solamente fungen como marcadores estructurales, más no como elementos del flujo narrativo del contenido. Asimismo, se eliminaron los apartados de *Referencias/Bibliografía*, dado que no representan usos reales del lenguaje, sino especificaciones de fuentes que, lejos de contribuir al objetivo del estudio, generan interferencias. Del mismo modo, se removieron de manera manual elementos como figuras, tablas, ecuaciones y demás unidades cuya naturaleza gráfica o simbólica las hacía incompatibles con la metodología de la lingüística de corpus.

Otros elementos depurados fueron las llamadas a referencias bibliográficas dentro del cuerpo del texto, correspondientes a los aparatos críticos del Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) y la Asociación Estadounidense de Psicología (APA), ya que podrían devolver resultados falsos en las búsquedas, al introducir sintagmas nominales ajenos a la estructura expositiva o argumentativa de los escritos. También es necesario resaltar que este paso se llevó a cabo de manera semiautomática: la detección de las cadenas de texto contenidas entre paréntesis o corchetes se realizó mediante la búsqueda automática por software; empero, dicha medida se sustenta sobre el supuesto de que estos signos de puntuación se reservan a la atribución de créditos a las fuentes de investigación en los formatos APA e IEEE, la especificación de siglas y otras funciones complementarias dentro de los documentos. A tal efecto, no se realizó una revisión exhaustiva de la totalidad del corpus para corroborar que los paréntesis y corchetes se empleasen únicamente con los propósitos mencionados; en ese sentido, se reconoce esta parte del análisis como una limitación, resultante de las restricciones de tiempo inherentes al desarrollo de esta investigación, pero razonable en términos generales tras una revisión superficial del material.

4.6.2. Etiquetado del corpus

El etiquetado morfosintáctico o morfológico se refiere a la designación explícita de cada uno de los *tokens* de un texto en términos de *partes de la oración*, sin tener en cuenta las relaciones sintácticas. A su vez, las partes de la oración denotan la categoría gramatical de cada palabra y pueden incorporar información morfológica.

Los archivos originales en formato de procesador de texto se sometieron al etiquetado por medio del software TagAnt una vez que tuvo lugar la limpieza de datos descrita en el apartado anterior. Aquí cabe especificar que el programa ofrece algunas alternativas aplicables a los parámetros de salida. Para el caso de este análisis se optó por el etiquetado de la forma *palabra+lema+categoría gramatical*; es decir, la estructura de las etiquetas muestra en primer lugar la palabra, seguida de su lema correspondiente y posteriormente de la categoría grammatical. Cada uno de estos elementos va separado por un guion bajo. Así, algunas etiquetas de ejemplo son:

```
comunidad_comunidad_NOUN
científica_científico_ADJ
encuentran_encontrar_VERB
ellos_el_PRON
```

Figura 1. Las palabras *comunidad*, *científica*, *encuentran* y *ellos* etiquetadas con los parámetros descritos.

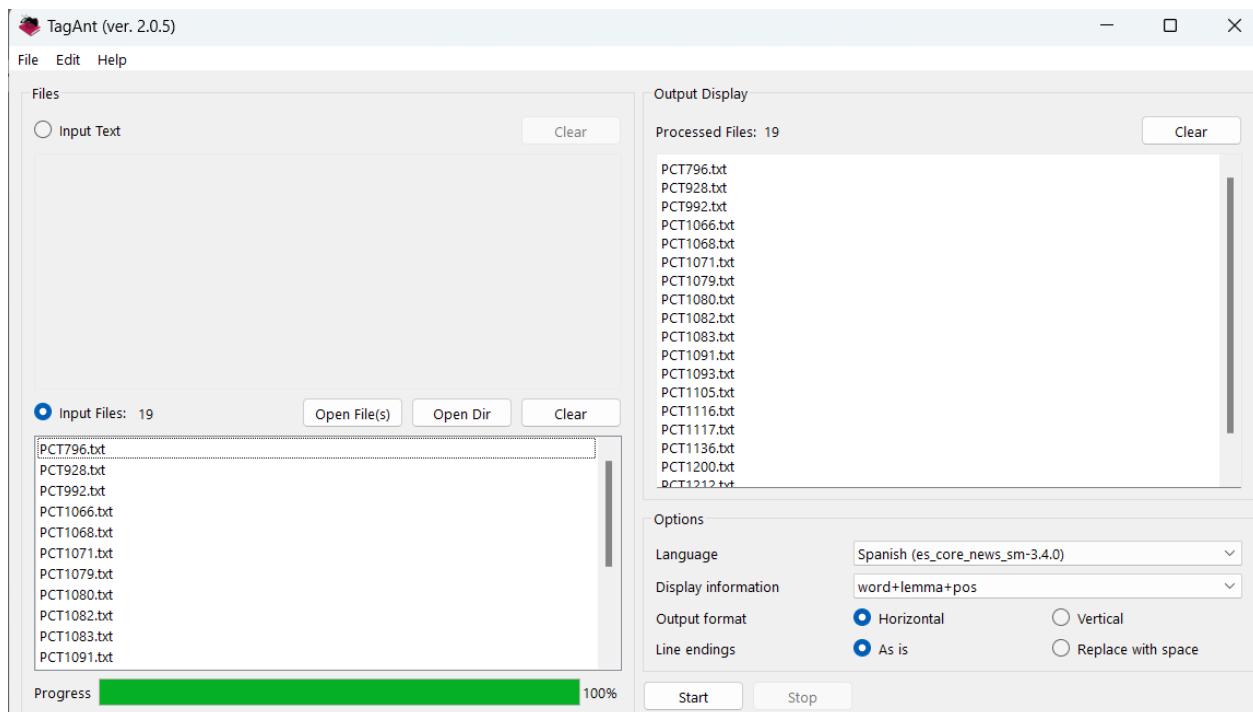


Figura 2. Área de trabajo de TagAnt.

En la Figura 1 se ilustra el área de trabajo de TagAnt, donde se realizó el etiquetado del corpus. Puede observarse en el recuadro inferior izquierdo la lista de archivos fuente en formato de texto llano (.txt). Del lado derecho se despliega la configuración correspondiente al tipo de etiquetado:

- En la opción *Language* debe especificarse el idioma del texto original. Quizás parezca trivial, pero si no se determina correctamente el lenguaje, el proceso da como resultado un corpus inutilizable. Se hace esta aclaración porque es un aspecto fácil de pasar por alto.
- El menú desplegable *Display information* puede establecerse el orden de etiquetado de cada token; como ya se mencionó anteriormente, se optó por la opción word+lemma+pos.
- *Output format* indica la disposición visual del archivo de salida. En esta opción se aplicó la organización horizontal a fin de que el texto fuese legible “de corrido” en la pantalla del gestor de corpus AntConc.
- *Line endings* configura si deben eliminarse los saltos de línea en el archivo original. Se optó por no aplicar dicho arreglo, a fin de contabilizar los párrafos en cada documento.

Durante la asignación de etiquetas, se identificó un inconveniente relacionado con el reconocimiento incorrecto de ciertos elementos por parte del software. Este problema tuvo como principal causa la presencia de letras mayúsculas en algunos ítems, situación que causaba el reconocimiento de ciertos ítems como nombres propios en lugar de sustantivos simples (por ejemplo, al inicio de cada oración). Dado que el uso de mayúsculas y minúsculas no es un aspecto al que se preste especial atención durante la redacción de los originales, el problema se convirtió en una interferencia importante durante la realización del análisis. Para enfrentar dicha situación, fue pertinente realizar ajustes específicos en el corpus para garantizar una asignación precisa de etiquetas y mejorar la exactitud del reconocimiento de los elementos en cuestión.

ESTRATEGIAS_ESTRATEGIAS_PROPN

RESUMEN_RESUMEN_PROPN

Figura 3. Etiquetas inadecuadas para los tokens “estrategias” y “resumen”.

Se reconocen como nombres propios.

La solución se encontró en el software de procesamiento de textos Notepad++. Dado que TagAnt da por sentado que los sustantivos constituyen nombres propios cuando inician por o se componen solamente de mayúsculas, fue necesario encontrar y sustituir todas las letras mayúsculas por minúsculas. Este procedimiento se realizó por medio de la búsqueda de expresiones regulares. Una

expresión regular, a menudo abreviada como *regex* o *regexp*, es una cadena de caracteres que define un patrón de búsqueda. Se utiliza para tareas como buscar, emparejar o manipular secuencias de caracteres en herramientas de programación y procesamiento de texto:

- Coincidencia de patrones específicos
- Validación de entradas, como números de teléfono o contraseñas
- Sustitución o modificación de texto basadas en patrones
- Por ejemplo, la *regex* `[\d+]` representa “uno o más dígitos”, por lo que encontraría la cadena “123” en “abc123xyz”

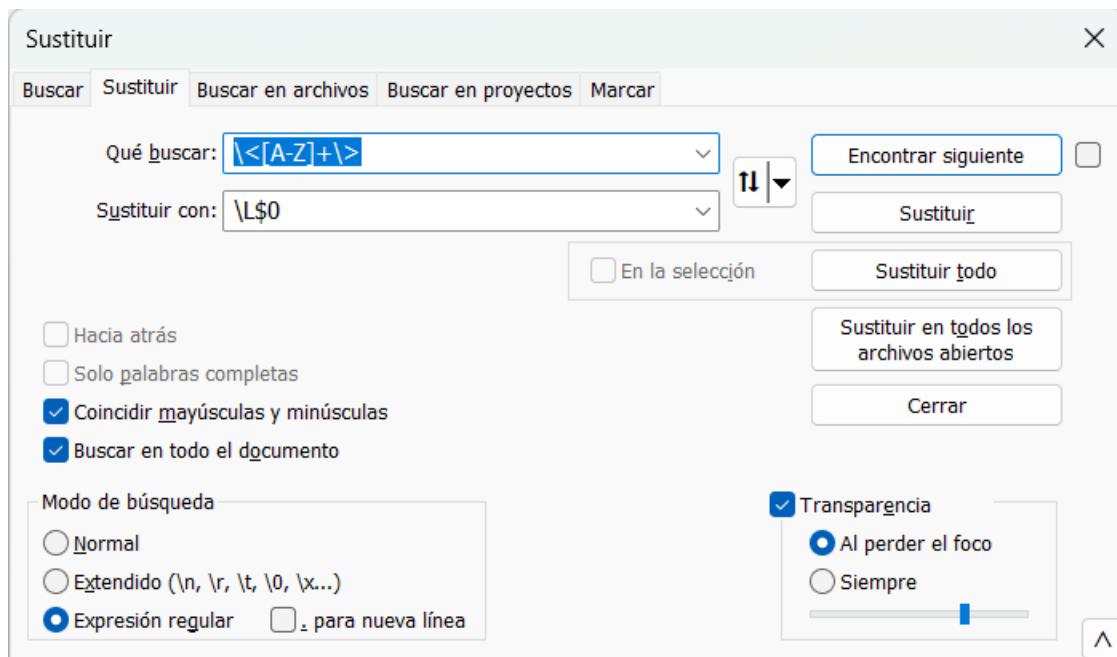


Figura 4. Ventana de sustitución en Notepad++.

Se emplearon expresiones regulares para realizar los cambios necesarios para la depuración del material.

En el caso descrito, la expresión regular `\<[A-Z]+\>` representa, en lenguaje natural, “cadenas de texto conformadas dos o más letras mayúsculas”. Una vez implementada esta modificación en todos los archivos del corpus, el etiquetado se efectuó sin mayores complicaciones.

4.6.3. Compilación del corpus

En esta sección se detallan los procedimientos que se llevaron a cabo para construir el corpus de estudio. Se señalan las características del gestor de corpus que AntConc pone a disposición y los

parámetros que se establecieron para crear un material funcional para los propósitos de esta investigación.

En la Figura 4 se muestra la ventana inicial de AntConc, con las diversas opciones de procesamiento disponibles en el programa. Para el presente estudio fueron relevantes únicamente las secciones *KWIC* (palabras clave en contexto) y *File View* (vista de archivo), las cuales permiten visualizar las líneas de concordancia para un ítem de búsqueda y su contexto amplio de aparición, respectivamente.

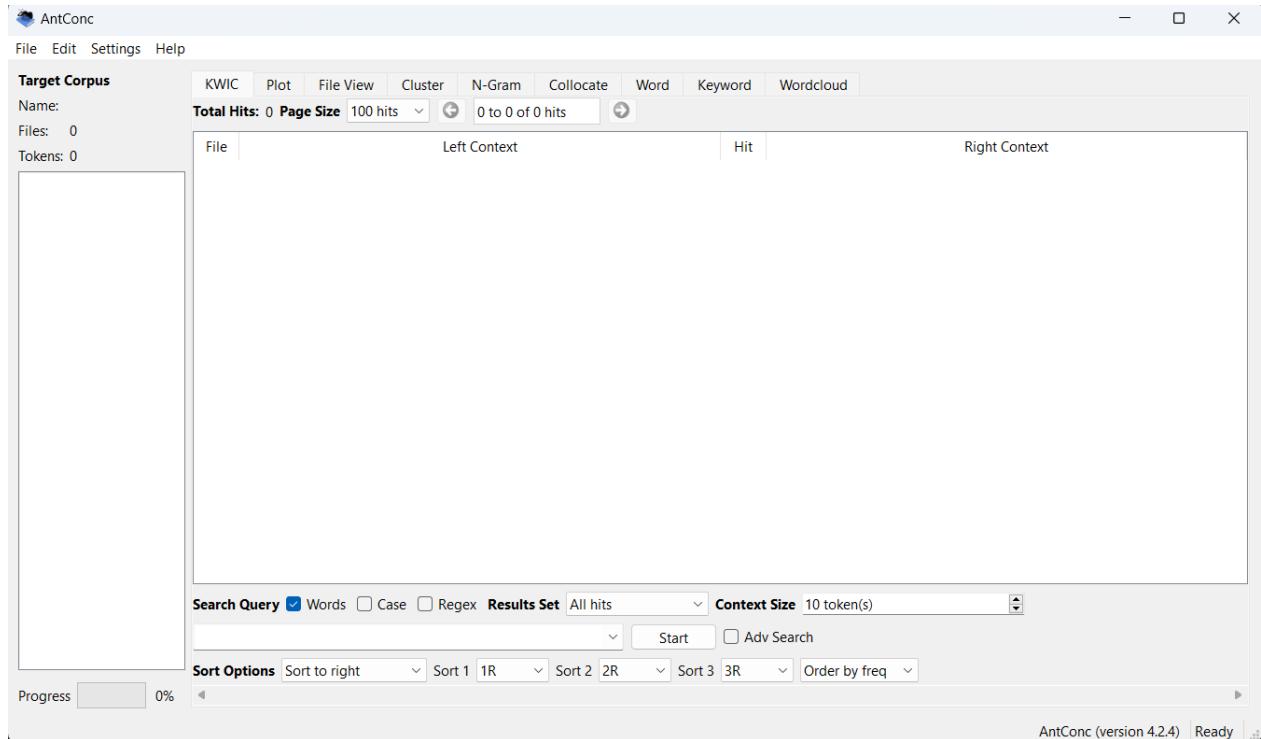


Figura 5. Ventana inicial de AntConc.

En la barra de herramientas de la ventana principal del programa se encuentra el menú *File*; al desplegarlo y seleccionar la opción *Open corpus manager*, el software muestra una lista de los corpus disponibles. Dicha ventana secundaria también permite crear un nuevo corpus a partir de archivos individuales en formato de texto.

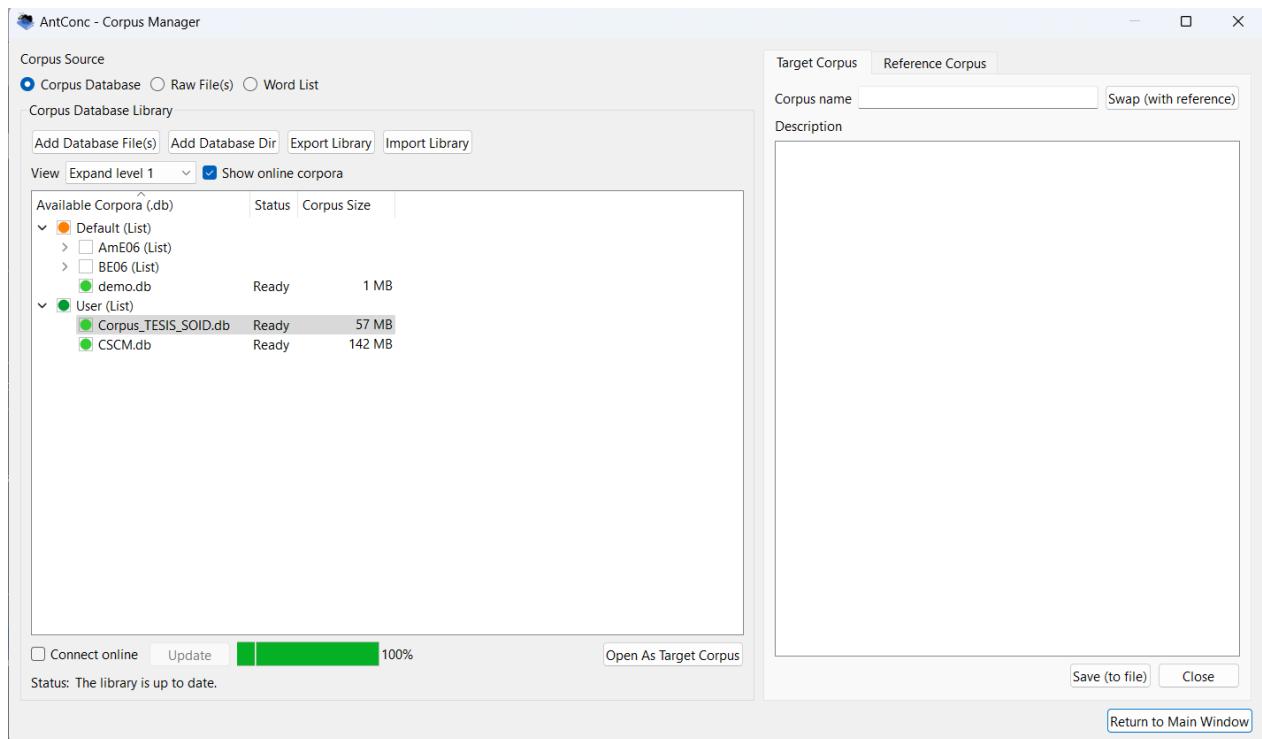


Figura 6. Gestor de corpus de AntConc.

Al seleccionar el apartado *Raw File(s)* en la parte superior de la pantalla, se abre la opción para seleccionar los elementos textuales que conformarán el nuevo corpus. En este cuadro de diálogo se cargaron todos los archivos correspondientes a los manuscritos de investigación, tesis de maestría y artículos científicos publicados en formato de texto plano (.txt).

En cuanto a los parámetros técnicos del nuevo corpus, se destaca el Formato de Transformación Unicode de 8 bits (UTF-8), que corresponde a un tipo de codificación estándar de amplia compatibilidad. Por otra parte, la opción *simple_word_pos_headword_indexer* indica al programa que el material textual cuenta con un etiquetado previo. El resto de opciones se dejó intacto, pues su configuración por defecto era favorable para el análisis.

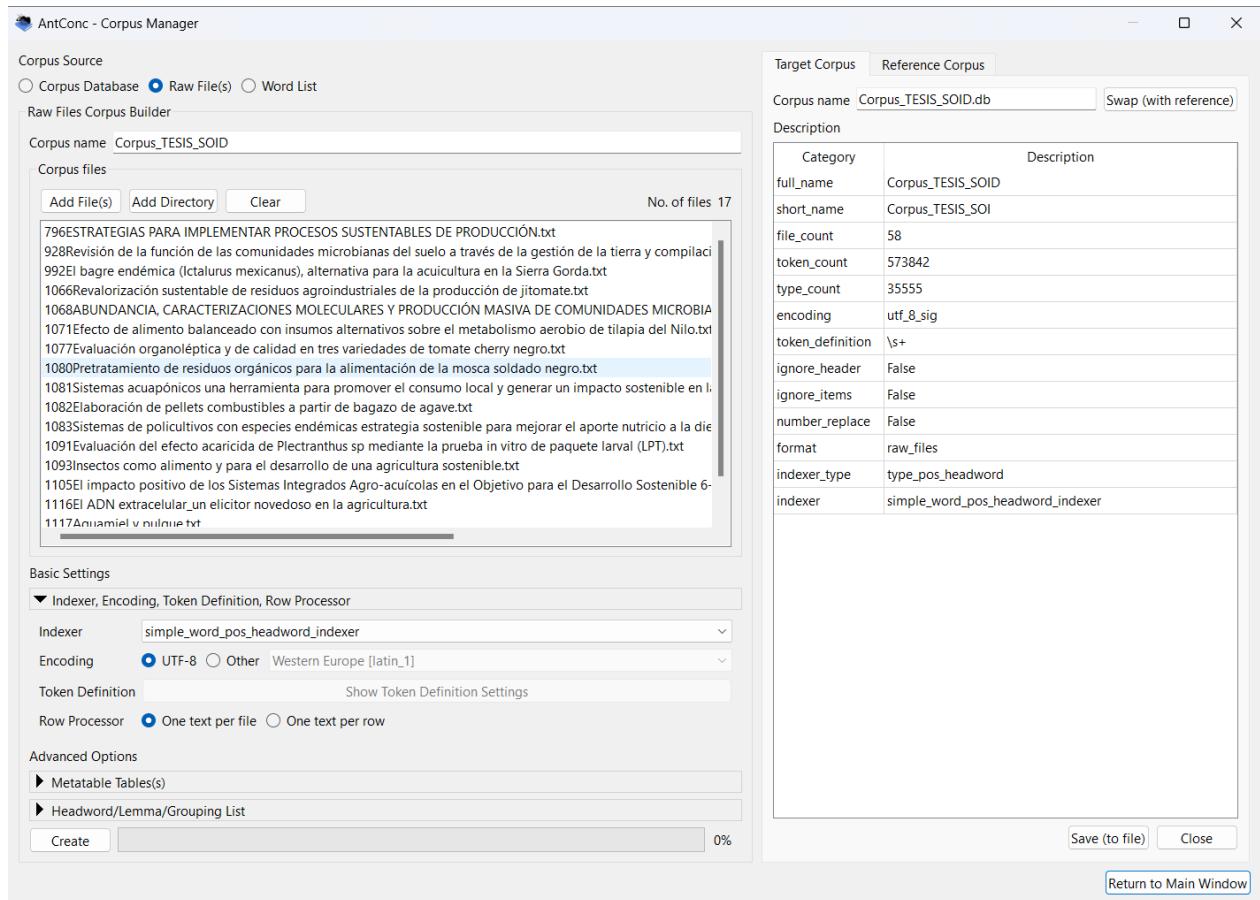


Figura 7. Creación del corpus en AntConc.

4.6.4. Identificación de los usos de *esto* como anáfora encapsuladora

Una vez construido el corpus, se procedió a identificar las apariciones de la forma *esto*. Puede apreciarse en la Figura 8 cómo luce la pantalla de resultados. La palabra clave introducida en el campo inferior de búsqueda se muestra en color azul al centro del recuadro principal; se encuentran además a la derecha e izquierda el resto del contexto de la línea de concordancia. Asimismo, el apartado de vista de archivo (Figura 9) despliega el contexto del archivo original, de modo que el usuario pueda consultar el ejemplo y observar sin limitación qué elementos preceden o suceden cada instancia.

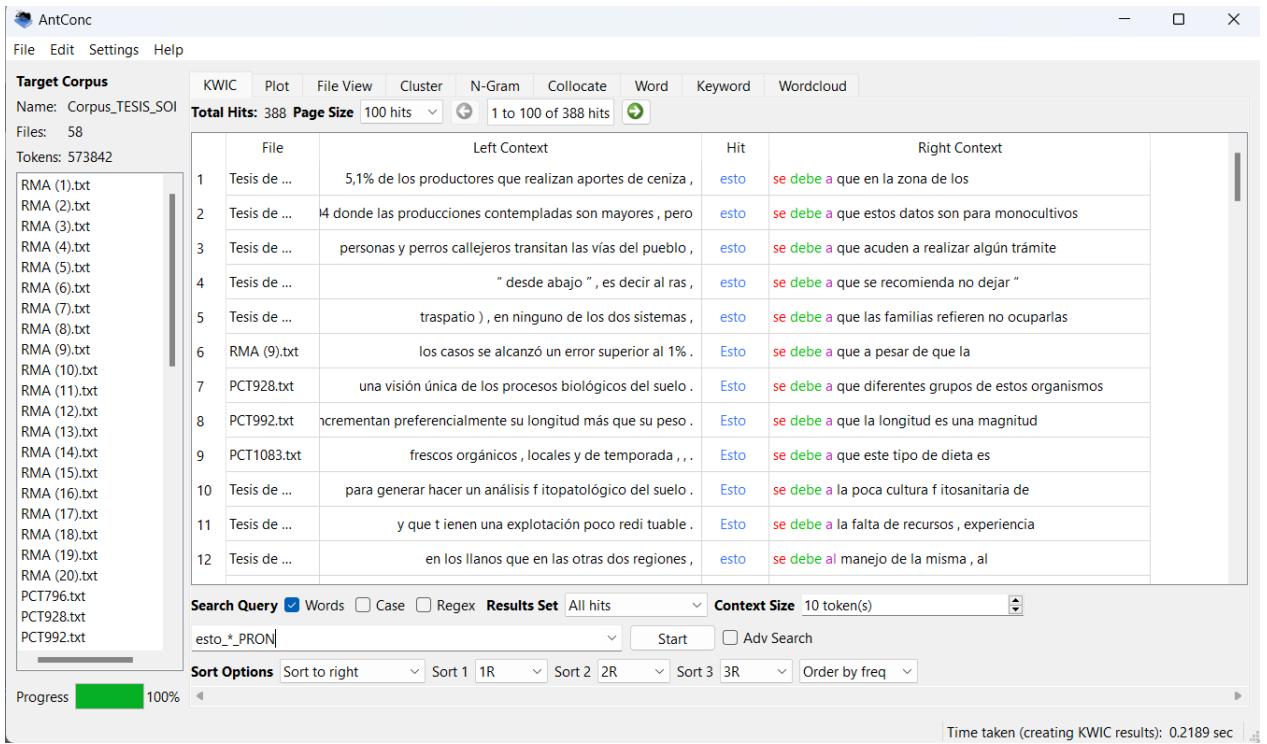


Figura 8. Ventana de trabajo de AntConc: vista de líneas de concordancia.

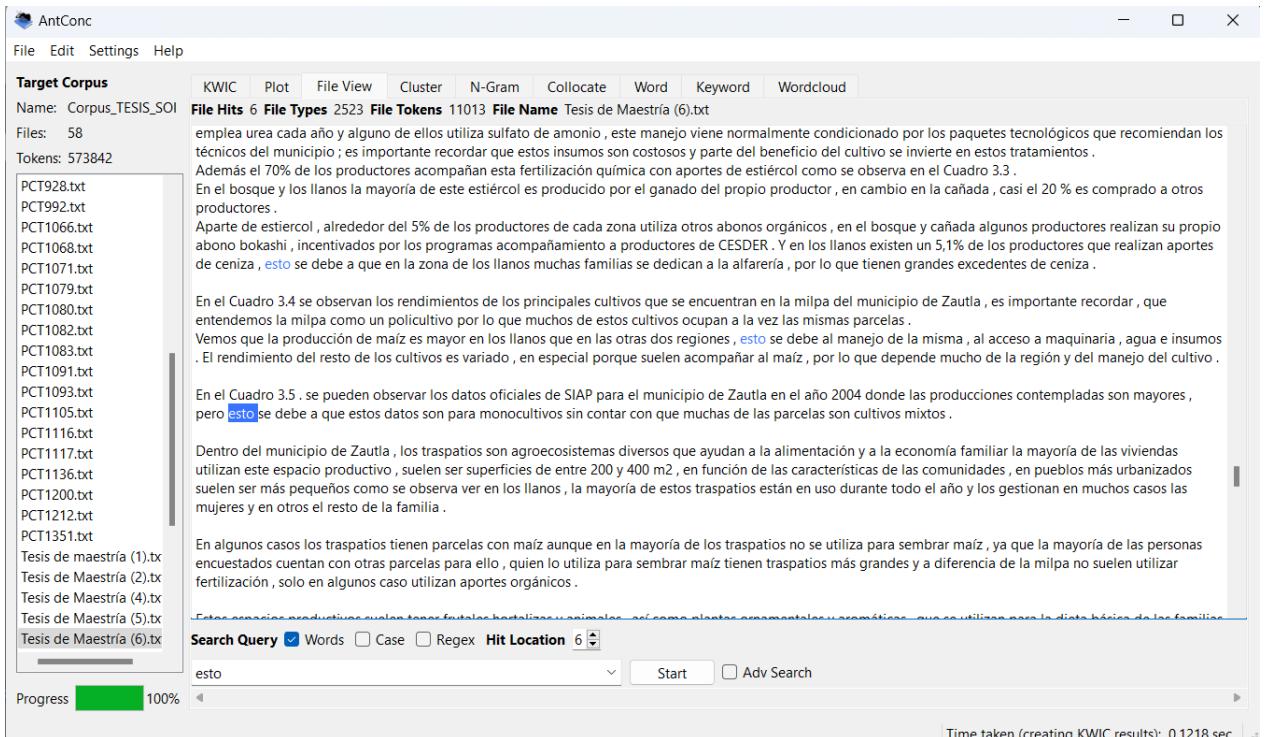


Figura 9. Ventana de trabajo de AntConc: vista del contexto amplio.

En seguida, se extrajeron los resultados y se enlistaron en una hoja de cálculo en Microsoft Excel, donde se categorizaron en función del tipo de relación cohesiva que se entabla con base en el pronominal *esto*; durante este paso, se adoptó la clasificación de adición, adversatividad, causalidad, finalidad y temporalidad propuesta por Parodi y Burdiles (2016), además de una categoría adicional para algunos casos marginales: localización. Para ilustrar el método de categorización, se presentan a continuación los ejemplos (27)-(32).

La adición permite incluir información extra, complementar ideas o agregar detalles. Ejemplos de conectores son *además*, *también* y *por otra parte*:

- (27) En el ensayo con 16 materiales establecido en Julio, la incidencia de la enfermedad fue menor, **esto** se comprobó porque las variedades testigo fueron las mismas en los dos ensayos, JANASA, Azufrasin y Azufrado Higuera, las que mostraron menor incidencia en el segundo ensayo. (Corpus propio, 2024)

La adversatividad sirve para introducir contrastes u oposiciones entre ideas. Se utilizan términos como *pero*, *sin embargo* y *aunque*:

- (28) **Sin embargo**, **esto** no quiere decir que sea así, ya que como indica Carreño y colaboradores, cuando se aplican grandes cantidades de agroquímicos se tiene como consecuencia la contaminación del suelo, del aire y de acuíferos subterráneos, así como también graves problemas de salud. (Corpus propio, 2024)

La causalidad relaciona causas con efectos en el discurso. Conectores comunes son *porque*, *ya que*, *debido a*:

- (29) Y en los llanos existen un 5.1% de los productores que realizan aportes de ceniza, **esto se debe a que** en la zona de los llanos muchas familias se dedican a la alfarería, por lo que tienen grandes excedentes de ceniza. (Corpus propio, 2024)

La finalidad indica el propósito o la intención de una acción. Algunos ejemplos son *para que*, *con el fin de*, *a fin de*:

- (30) El aislamiento de colonias bacterianas se realizó mediante el procesamiento de siete ejemplares vivos de *A. mellifera*, las cuales se colocaron durante un minuto a una temperatura de 4 °C, **esto con el fin de** disminuir sus funciones y procesarlos vivos. (Corpus propio, 2024)

La temporalidad establece relaciones de tiempo entre eventos o acciones. Conectores como *antes*, *después*, *mientras*, *al mismo tiempo* son frecuentes:

(31) Asimismo, las grandes obras industriales han afectado la vida y tranquilidad de los pueblos. Tal es el caso del Gasoducto del Proyecto de Integración Morelos que provocó una fuerte movilización en defensa de la tierra, un conflicto que ha durado muchos años, **esto comenzó en octubre de 2011**, cuando la CFE otorga la concesión del Gasoducto Morelos a una empresa española para realizar el proyecto en las comunidades de San Lucas Tulcingo, y San Juan Amecac, municipio de Atzizihuacan, poblaciones que se encuentran dentro de la zona naranja de alto riesgo del volcán Popocatépetl.

La localización ayuda a situar un evento en un contexto espacial específico. Ejemplos incluyen *en*, *junto a*, *donde*:

(32) **Esto es en el valle de San Nicola Buenos Aires** pero antes en los años de 1980 a 1990, se cultivaban grandes extensiones de papa roja en las faldas del Citlalpetel, en su parte Poniente en el estado de Puebla y en su parte Oriente en el estado de Veracruz.

También se registraron los contextos amplios para habilitar el análisis lingüístico cualitativo por medio de la observación y registrar el nivel de anidación de las oraciones. En este caso, se entiende por nivel de anidación la distancia, medida en cláusulas subordinadas, entre *esto* y la cláusula principal.

4.6.5. Conteo de palabras por oración en los documentos

Se aludió en la 2.1. Descripción del problema que la memoria de trabajo es una capacidad mental limitada. Tanto es así, que los manuales de estilo recomiendan mesura en la cantidad de palabras contenidas en una oración. A partir de dicha premisa, se efectuó un conteo del total de oraciones en cada texto; esta parte del procedimiento también se llevó a cabo en Notepad++: se contabilizaron las instancias de puntos sucedidos por una letra mayúscula o un salto de línea. Cada una de estas apariciones se contabilizó como una oración. Cabe señalar que la depuración de los datos garantizó que no hubiera interferencias de otros elementos, tales como cifras con decimales o abreviaturas. Consecutivamente, se calculó la proporción general de palabras por oración en todos los documentos. Para realizar el cómputo, se consultó el menú *Propiedades* que el sistema operativo Windows despliega en el explorador de archivos. Dicha opción exhibe la cantidad de palabras en los documentos de Microsoft Word.

Contenido	
Estado de contenido	
Tipo de contenido	application/vnd.openxmlformats-offi...
Páginas	56
Contar palabras	23719
Contar caracteres	130457
Número de líneas	1087
Número de párrafos	307
Plantilla	Normal.dotm
Escala	No
Vínculos obsoletos	No
Idioma	
...	

Figura 10. Atributos de archivo. Se aprecia el conteo de palabras de una de las tesis de maestría incluidas en el análisis.

5. Resultados

En este apartado se presentan los datos recopilados con base en la metodología descrita, así como un análisis pormenorizado del fenómeno de estudio.

5.1. Apariciones de *esto*

Dentro del corpus analizado, se registraron 218 instancias de la palabra *esto* empleada como recurso de cohesión en una posición donde se esperaría un encapsulador nominal. Es imprescindible considerar que los textos correspondientes a tesis de maestría son notablemente más extensos que sus contrapartes. Por esa razón es quizás esperable que se encuentren más casos contenidos en ellos. Para corroborar tal suposición, fue necesario efectuar una normalización de los datos, a fin de reconocer hasta qué punto el desbalance se debe al tamaño de las muestras. En ese sentido, la Tabla 2 presenta tanto el conteo absoluto de los casos como la ratio de aparición por cada mil palabras.

Tabla 2. Repeticiones de *esto* por tipo de texto.

Tipo de texto	Absoluto	Relativo (por millar de palabras)
Manuscrito	28	4.54
Publicado	23	3.64
Tesis	167	4.07
Total	218	

Adicionalmente, la Figura 11 grafica los conteos generales para el tipo de relación cohesiva que se manifiestan en la totalidad del corpus tras la categorización de los 218 ejemplos identificados. La gráfica de barras permite apreciar una fuerte tendencia hacia dos clasificaciones funcionales en todos los tipos de texto: la adición y la causalidad. El resto de funciones goza de escasa prominencia, lo cual es un fenómeno que coincide con los resultados de estudios anteriores en cuanto al modo en que se construyen estos textos (Parodi y Burdiles, 2016).

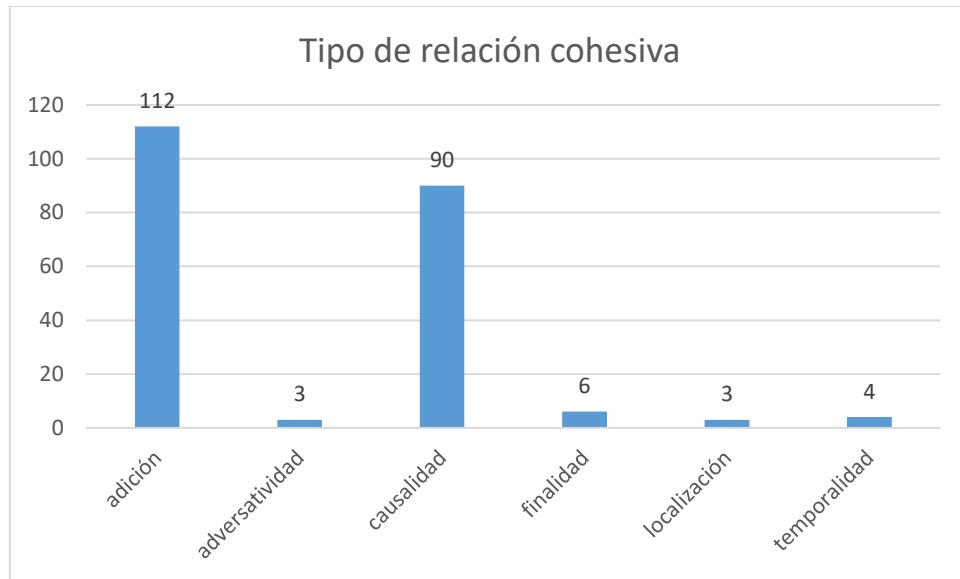


Figura 11. Conteo de las relaciones de cohesión en el corpus general.

En lo sucesivo se presentan los resultados clasificados por tipo de escrito a fin de ilustrar con mayor claridad las diferencias entre cada género. Resalta que la variedad funcional, tanto en artículos publicados como inéditos (Figuras 12 y 13), es extremadamente limitada, con una sola aparición en ambas clasificaciones que difiere de la causalidad o la adición. Adicionalmente, la distribución entre las funciones de adición y causalidad son inversas: los manuscritos tienden más hacia el empleo de *esto* en construcciones que implican causalidad; los publicados, se apegan más a la adición de ideas por medio de dicho pronominal.

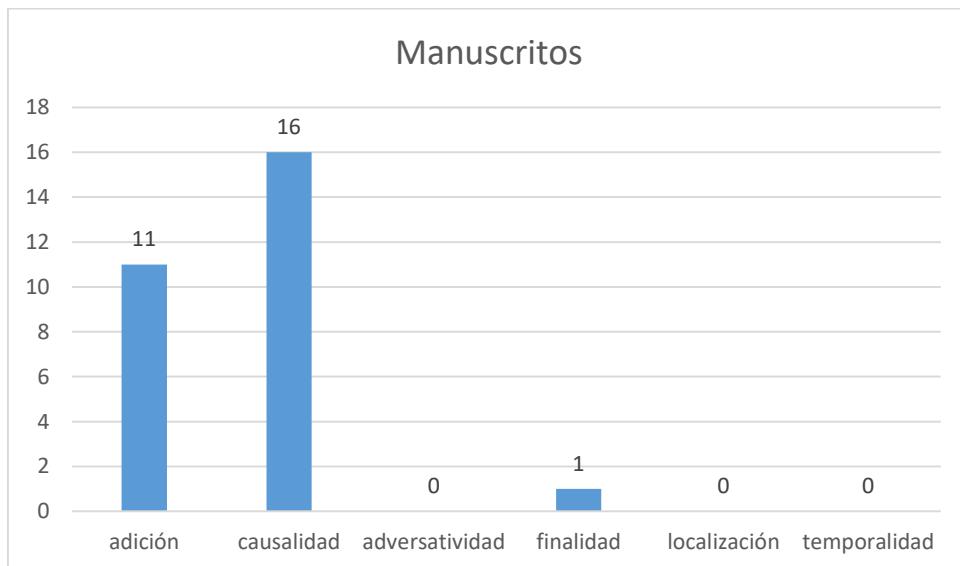


Figura 12. Conteo de las relaciones de cohesión en los manuscritos de investigación inéditos.

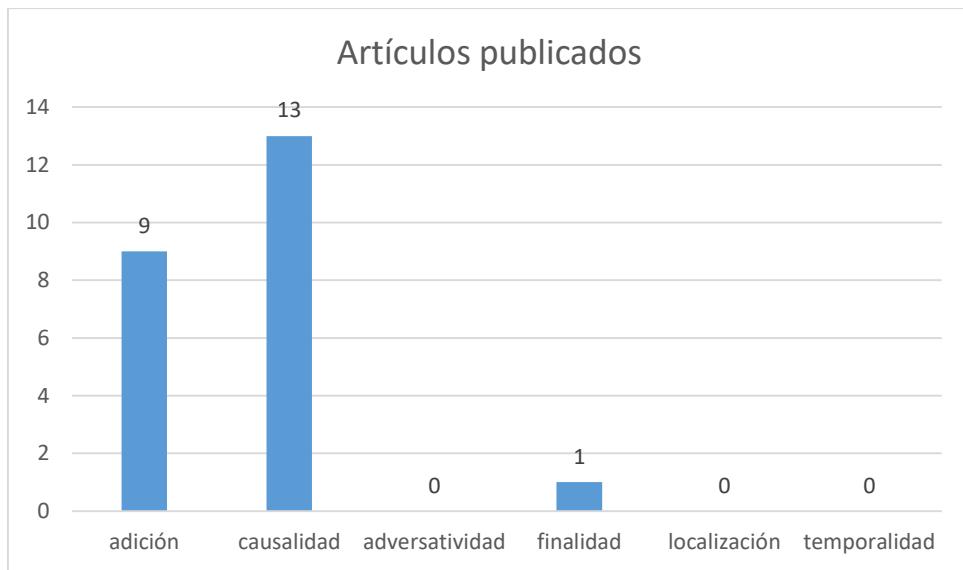


Figura 13. Conteo de las relaciones de cohesión en los artículos publicados.

Finalmente, en la Figura 14 se ilustra cómo entre los documentos correspondientes a tesis de maestría aparecen más el resto de clasificaciones funcionales. No obstante, la predominancia de la causalidad y la adición por sobre el resto de funciones se sostiene. Aquí cabe recordar que las tesis de maestría son documentos con mucha mayor extensión, por lo que se justifica este comportamiento.

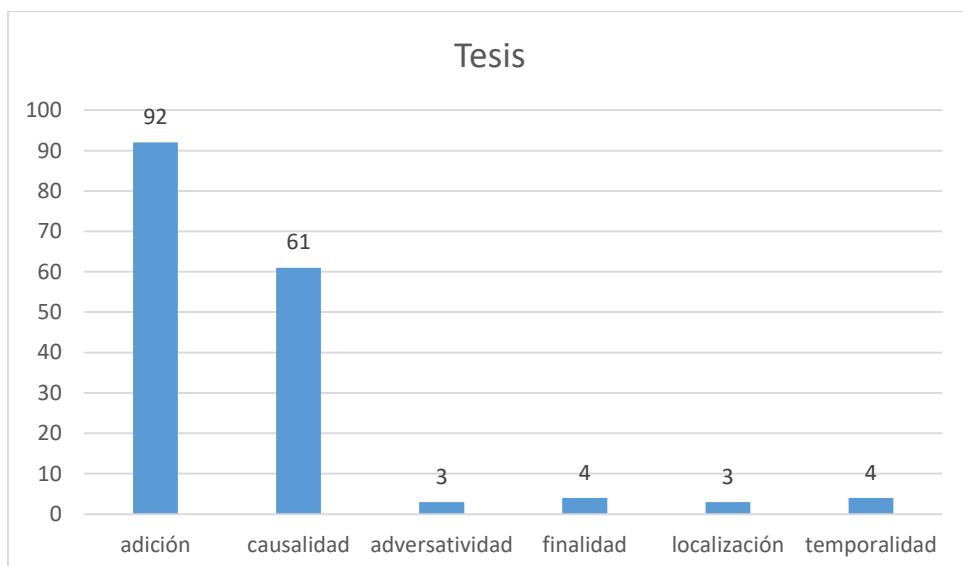


Figura 14. Conteo de las relaciones de cohesión en las tesis de maestría.

5.2. Longitud oracional

Se detallan primero las características generales de los textos en términos de longitudes oracionales, clasificados según el tipo de obra. Los artículos publicados alcanzan un promedio por debajo del límite de 35 palabras recomendado como la cantidad máxima para favorecer la comprensión lectora. Por otro lado, tanto las tesis de maestría como los manuscritos de investigación superan esa frontera. Así, las cifras parecen apuntar a una moderación en este rubro por parte de los publicados. No obstante, un ANOVA revela que no existen diferencias significativas entre las medias de las longitudes oracionales para los tres tipos de texto, con un valor *p* muy superior a la tolerancia de 0.05 (Tabla 4). La ausencia de distinción en este aspecto solo es indicativa de que una métrica basada en cantidades de palabras por oración es insuficiente para distinguir entre un tipo de texto y otro.

Tabla 3. Estadísticos descriptivos entre los grupos.

Tipo de texto	Promedio	Varianza	DE
Manuscrito	36.547	62.275	7.891
Publicado	33.35	57.344	7.573
Tesis	34.95	34.052	5.835

Tabla 4. ANOVA de las longitudes oracionales por tipo de texto.

Casos	Suma de Cuadrados	gl	Cuadrado Medio	F	p
Tipo	99.640	2	49.820	0.976	0.383
Residuales	2857.467	56	51.026		
<i>Nota.</i> Suma de Cuadrados Tipo III					

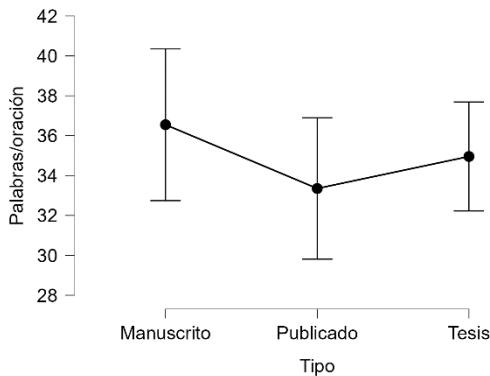


Figura 15. Gráfico descriptivo de las longitudes oracionales por tipo de texto.

En cuanto al análisis de la extensión promedio de las oraciones en el corpus, los resultados reflejan una tendencia general hacia la conformidad con el límite establecido de 35 palabras por oración, indicador de accesibilidad y claridad textual. Sin embargo, un escrutinio detallado del material revela que, más allá de este promedio general, existen numerosas excepciones notables donde la longitud de las oraciones se excede de manera significativa. Por su impacto en la comprensión y fluidez del texto, estos casos particulares merecen un examen específico que permita ilustrar cómo las oraciones extensas pueden romper con la accesibilidad general del texto.

A continuación, se presenta una disección pormenorizada de los casos más representativos, evidenciando la presencia de oraciones extraordinariamente largas y explorando las posibles razones detrás de su estructura.

- (33) Lo anterior es posible mediante la integración de más de un genotipo para los cultivos, **esto debido a** que muchas especies están interconectadas y dependen unas de otras para sobrevivir por lo que confieren diferentes servicios ecosistémicos para sustentar la vida en la Tierra y las sociedades y economías humanas. (Corpus propio, 2024, tipo de texto: manuscrito inédito)

El ejemplo exhibido en (33) consta de 51 palabras, un excedente de 45.7 % (casi la mitad) sobre el rango de 35. Pero la longitud oracional no es la única limitante para la comprensión lectora en este caso; simultáneamente, el texto ofrece cláusulas subordinadas en *esto debido a, y por lo que*. Como se expresó anteriormente, la anidación recursiva de la cláusula puede llegar a complicar la lecturabilidad. En el contexto de la recursividad, cabe añadir que la forma pronominal *esto* (empleada aquí en su caracterización como recurso de causalidad) remite al lector a la proposición de la cláusula anterior: *Lo anterior es posible mediante la integración de más de un genotipo para*

los cultivos. Dicho constituyente, a su vez, inicia en la construcción defictiva *lo anterior*, la cual debe rescatarse, de la misma manera, del contexto precedente. Ahora bien, es cierto que el discurso se construye progresivamente sobre la información dada (Givon, 1981); no obstante, es crucial reconocer que el alto nivel de dependencia entre cláusulas y referencias pronominales puede sobrecargar al lector, especialmente en textos con objetivos explicativos o científicos. La interacción constante entre los elementos anafóricos, como *lo anterior* y *esto*, y la estructura subordinada no solo exige mayor esfuerzo cognitivo, sino que puede generar interrupciones en el flujo de comprensión.

- (34) En un estudio describieron el espectro electromagnético al que responde el ojo compuesto de la MSN, **esto** es fundamental para la fabricación de fuentes de iluminación artificial, ya que permite prescindir del espectro que no genera una respuesta por parte de los fotorreceptores de la MSN, promoviendo el ahorro de recursos energéticos por parte de los productores y una maximización de la postura. (Corpus propio, 2024, tipo de texto: manuscrito inédito)

De nueva cuenta, en (34) se desarrolla una oración cuya longitud resulta desmedida; en este caso, con 64 palabras, se alcanza un excedente de 82.8 %. Por otra parte, en esta instancia se muestra un uso de *esto* como recurso de adición de información, que parece remitir a la predicción anterior (*esto = en un estudio describieron el espectro electromagnético al que responde el ojo...*), o al sintagma nominal encabezado por *el espectro...*, tal que no queda claro qué elemento debe rescatarse para asignarle la posición de sujeto en el predicado *es fundamental...*

En lo respectivo a la anidación de las cláusulas, aparecen las construcciones: *ya que permiten...*, *que no genera...* y *promoviendo el ahorro...*; de esta manera, la recursividad clausular demanda un esfuerzo adicional por parte del lector para permanecer en el hilo argumentativo del texto. Este tipo de estructura, con múltiples cláusulas encadenadas, no solo incrementa la complejidad sintáctica, sino que también dificulta la identificación de las ideas principales y secundarias, prolongando el procesamiento cognitivo necesario para comprender el mensaje.

- (35) Este beneficio se relaciona al posible efecto nutraceutíco que los productos pudieran obtener durante su desarrollo; un ejemplo de **esto** es la cantidad de fibra soluble e insoluble, cantidad de fenoles y flavonoides, aumento de aminoácidos y otros compuestos bioactivos que se encuentran en los frutos y/o flores del cultivo. **Esto** puede aportarle al consumidor una mayor cantidad y calidad de nutrientes, ayudando así a prevenir o mejorar su estado de salud (Corpus propio, 2024, tipo de texto: manuscrito inédito).

De forma contrastante, (35) se muestra más en línea con la limitante de longitud, si se considera que el punto y coma marca una frontera entre oraciones, al igual que el punto. En ese sentido, la extensión de los segmentos no supone una carga tan pesada; no obstante, es visible cómo la forma *esto* prevalece como recurso de cohesión entre ideas. En la primera aparición, parece referenciar al *posible efecto nutracéutico...*, aunque compiten por el puesto de referente los sintagmas nominales *su desarrollo* y *este beneficio*; a pesar de la ambigüedad anafórica en este sentido, los posibles referentes se encuentran suficientemente relacionados para favorecer la comprensión del mensaje. La segunda instancia de *esto*, por otra parte, remite aparentemente al conjunto de proposiciones anteriores, o a la abstracción creada por ellas, sin necesariamente generar ambigüedad; o quizás el contexto es suficientemente difuso para que cualquier ambigüedad sea despreciable. Ambas apariciones corresponden a la función de adición informativa.

En cuanto a la recursividad clausular, no se observa un nivel alto de anidación que pueda comprometer la lecturabilidad del párrafo. Con todo, el ejemplo (35) es una muestra de que la medida en la longitud favorece la lecturabilidad al restringir la cantidad de referentes que puedan generar confusiones.

- (36) Sin embargo, coincide con los resultados de Toyes-Vargas *et al.*, quienes explican que **esto**, probablemente sea un mecanismo de adaptación de las plantas a las zonas con temperaturas altas, debido a que una reducción de la fluidez de las membranas en la planta ayudaría a disminuir la evapotranspiración, **esto** pudiera lograrse al incorporan (sic) en la membrana vegetal una mayor concentración de un AG saturado como el C16:0. (Corpus propio, 2024, tipo de texto: artículo publicado)

Vale la pena recalcar en (36) el uso de *esto* sin un referente claro; de hecho, el sujeto implícito en la cláusula anterior dificulta aún más la identificación de un antecedente que satisfaga la proposición general. No es evidente en el texto cómo debe reconstruirse el significado, de modo que el segmento es confuso, y parece ser que *esto* actúa únicamente como un recurso para añadir información. Desde un punto de vista práctico, ambigüedades de este tipo podrían retrasar el proceso de revisión y edición, ya que los revisores tendrían que dedicar tiempo adicional a entender y corregir estas estructuras. En entornos científicos y académicos, donde la publicación o distribución de los textos está sometida a criterios rigurosos de calidad y claridad, estos problemas representan un desafío significativo. La segunda aparición de *esto* en el segmento podría referir a los sintagmas *una reducción de la fluidez*, *disminuir la evapotranspiración* o a la proposición

compuesta por ambos. En cualquier caso, no surge una ambigüedad, puesto que cualquiera de los referentes resultaría en una interpretación similar de la oración.

Respecto de la extensión, 65 palabras representan un excedente de 85.7 % a la longitud límite, y en cuanto a la anidación, hay una cláusula de relativo en *quienes explican que*, una subordinada en *debido a que* y una más en *esto pudiera lograrse*; el complejo clausal, que ocasiona la patente longitud, aunado al referente incierto del pronominal *esto*, termina perjudicando la lecturabilidad del párrafo.

- (37) Por otra parte, la alimentación es el factor que más influye en la concentración de ALC en leche de rumiantes y el pastoreo es la estrategia nutricional más eficaz y económica para incrementar el contenido de ALC. **Esto** se debe, a que los forrajes frescos tienen una concentración alta de ácido linoleico y α -linolenico, precursores de ALC y ácido vaccénico en rumen. Lo cual causa una tasa mayor de escape de ALC y AV hacia la glándula mamaria y en consecuencia existe disponibilidad mayor de sustrato para que actúe la enzima $\Delta 9$ -desaturasa, que es la responsable de sintetizar alrededor de 90% de ALC total en leche de rumiantes. (Corpus propio, 2024, artículo publicado)

En el caso de (37), la palabra *esto* funge como medio para añadir la información siguiente; por otro lado, refiere a *el pastoreo es la estrategia nutricional más eficaz y económica...* Empero, podría argumentarse que dicho antecedente compite con el complejo clausal formado por los predicados *es el factor que más influye...* y *es la estrategia nutricional...*, ya que se presentan coordinados al mismo nivel por la conjunción *y*. Es decir, algunos lectores podrían necesitar un esfuerzo adicional para separar el complejo, aislar el segundo predicado e interpretarlo como sujeto de la oración siguiente. Finalmente, la longitud no representa un problema en las dos primeras oraciones del párrafo, al contrario de lo que ocurre en la tercera: con 46 palabras, se ve un excedente de 31.4 %, sostenida en una anidación causada por *para que actúe...y que es la responsable...*

- (38) Respecto a la variación entre los resultados de las simulaciones numéricas en las rápidas de los aforadores, para los escenarios analizados; se encontró que una reducción del área hidráulica a la entrada del aforador no tiene un efecto relevante en el comportamiento del flujo desarrollado en la rápida, pues como se puede ver en el Cuadro 7, en ninguno de los casos se alcanzó un error superior al 1%. **Esto** se debe a que a pesar de que la estructura se modifica en el fondo por la acumulación de sedimentos, aún mantiene el diseño de un aforador con constricciones laterales, los cuales según Bos y Wijbenga ya se han utilizado con anterioridad para la medición de caudales en canales con transporte de sedimentos, además recientemente en el trabajo de Aali y Vatankhah se reportó un error

promedio cercano a 2.2% en la medición de caudales utilizando canales simples con una contracción trapezoidal. (Corpus propio, 2024, tipo de texto: artículo publicado)

Expresar una relación cohesiva de causalidad con el segmento anterior es la función de *esto* en el ejemplo (38); asimismo, los referentes que podrían entrar en competencia son *en ninguno de los casos se alcanzó un error superior al 1% y se encontró que una reducción del área hidráulica a la entrada del aforador no tiene un efecto relevante...* No obstante, dado que ambos se hallan, a su vez, en una relación de causalidad, la anáfora se construye con éxito en cualquiera de sus interpretaciones.

En este caso, la longitud de la oración iniciada en *Esto se debe a...* se extiende hasta 83 palabras, lo cual supone un exceso de 165.7 %. Dicha cifra representa una sobrecarga formidable para la memoria de trabajo, que bien podría distribuirse en segmentos más mesurados. Por último, en cuanto a la recursividad, destaca la construcción inicial *Esto se debe a que a pesar de que*, la cual encadena dos relativizadores en notoria proximidad. Pasando por alto su naturaleza cacofónica, la frase subordina una cláusula en cuanto inicia la oración, y posteriormente aparece una nueva subordinada en *los cuales*. En síntesis, el segmento sufre de una longitud exacerbada y una anidación múltiple que perjudican la lecturabilidad.

- (39) Al comparar la información previa con datos registrados del INEGI, se observa que dicha información no concuerda, ya que estos registros no mencionan la contaminación de suelo, agua y aire que los colaboradores de este estudio si perciben, **esto** posiblemente se debe a que las encuestas del INEGI son basadas en respuestas “si” y “no”, cerrando así la apertura de los habitantes para hablar de sus necesidades o problemáticas en su contexto. (Corpus propio, 2024, tipo de texto: tesis de maestría)

Con 77 palabras, (39) exhibe un excedente de 120 % al umbral de 35 palabras. Pero lo que llama la atención es el uso de *esto*, ubicado cerca del centro de la oración, porque es justo en esa posición que duplica el conteo de palabras. De tal modo, *esto* actúa como un ancla para la adición de una nueva idea, pero podría sustituirse por una encapsulación nominal (*tal omisión o tal discrepancia*) como inicio de una nueva idea, a fin de seccionar la predicación y permitir al lector procesar el primer bloque de información antes de continuar al siguiente.

- (40) En general el paisaje que se observa presenta una parcela dispar, no uniforme, ya que hay pequeñas partes donde la siembra fue precaria en tanto que hay otros espacios donde se dio en abundancia, **ello** se expresa principalmente por las características de las plantas especialmente por su tamaño, las hay pequeñas, otras largas y con un tallo grueso de color verde y rojo, **esto** expresa la biodiversidad del

amaranto, ya que la multiplicidad en cuanto a tonalidades se aprecia no sólo en la base de la planta, también en las hojitas verdes, y las flores o panojas.

Ahora bien, en (40) pueden apreciarse las formas pronominales *ello* y *esto* desempeñando una función similar de adición. La extensión de este ejemplo de 98 palabras, 180 % por encima del umbral, evidencia cómo también se compromete el hilo argumentativo del autor conforme crece la longitud. El problema es notorio en la frase *ello se expresa principalmente por...*, donde *ello* remite por proximidad a la predicación *la siembra fue precaria*, pero por sentido a *el paisaje presenta una parcela dispar*. Más adelante, en la frase *esto se expresa en...*, se perfilan las características variadas de las plantas presentes en la parcela, de forma que es hasta entonces que puede resolverse la ambigüedad planteada en *ello*.

- (41) Desde Venezuela, Jesús Núñez, explica el abordaje de los saberes “procesos y productos de la creación humana” vinculándoles con las distintas generaciones: abuelos, jóvenes, adultos o niños, quienes poseen dispares grados de dominio, **esto** es susceptible de entenderse desde la pedagogía de los saberes, **esto** es una iniciativa que busca impulsar una educación acorde a las características y condiciones de los escenarios rurales, **esto** es en la cotidianeidad, **esto** descansa en los señalamientos de Paulo Freire, aspecto que Núñez lo adapta y o da como resultado el reconocimiento de tipos de saberes, a saber aquellos salvaguardados, hibridados, sustituidos y emergentes. (Corpus propio, 2024, tipo de texto: tesis de maestría)

El párrafo mostrado en (41) destaca por su uso encadenado del pronominal *esto* y su longitud de 101 palabras (exceso de 188.6 %). Además, en su primera aparición, la anáfora *esto* remite a un antecedente que se encuentra relativamente lejano: la frase nominal *el abordaje de los saberes*. Seguidamente, el texto presenta de nueva cuenta una anáfora en *esto es una iniciativa*, cuyo referente ya es más difícil de ubicar. Incluso, pueden articularse dos interpretaciones para resolver la referencia anafórica:

- *la pedagogía de los saberes es una iniciativa que busca impulsar una educación acorde a las características y condiciones de los escenarios rurales.*
- *entender esto (el abordaje de los saberes) desde la pedagogía de los saberes es una iniciativa que busca impulsar una educación acorde a las características y condiciones de los escenarios rurales.*

Ante ambas posibilidades, corresponde al lector decantarse por una u otra interpretación en función de su conocimiento o instruyéndose al respecto por sus propios medios. Una revisión rápida en línea puede ayudar a disipar la duda en este caso: la pedagogía de los saberes de Paulo Freire es un método educativo centrado en la importancia de los conocimientos previos y la experiencia de los estudiantes como base para el aprendizaje (Cruz Aguilar, 2020); sin embargo, el alcance de su enfoque no se limita de ningún modo al ámbito de la ruralidad, por lo que la primera interpretación puede ser descartada.

Acto seguido, *Esto es en la cotidianeidad* es una frase breve que podría integrarse a la cláusula anterior como un complemento circunstancial. Inmediatamente después aparece *esto descansa sobre...* La reiteración de *esto* en la misma línea ahora aparenta encapsular la totalidad del párrafo hasta ese punto, de modo que el comportamiento del pronominal se presume inconsistente a lo largo del segmento. Sería quizás mejor para la organización y presentación de las ideas el uso de los encapsuladores nominales. Se ejemplifica en (42) una estructura alternativa del segmento aprovechando dicho recurso y dejando intacto cualquier otro aspecto de la redacción (se resaltan en negritas las ediciones propuestas).

- (42) Desde Venezuela, Jesús Núñez, explica el abordaje de los saberes “procesos y productos de la creación humana” vinculándoles con las distintas generaciones: abuelos, jóvenes, adultos o niños, quienes poseen dispares grados de dominio. **La tarea** es susceptible de entenderse desde la pedagogía de los saberes, y **su enfoque** es una iniciativa que busca impulsar una educación acorde a las características y condiciones de los escenarios rurales **en la cotidianeidad**. **Este modelo educativo** descansa en los señalamientos de Paulo Freire, aspecto que Núñez lo adapta y o da como resultado el reconocimiento de tipos de saberes, a saber aquellos salvaguardados, hibridados, sustituidos y emergentes.

Como ya se afirmó anteriormente, dentro del corpus de estudio se detectaron 218 ejemplos, de los cuales se seleccionaron los ejemplos (33) a (42) para un análisis detallado. Si bien realizar un análisis individual de todos los casos sería una tarea altamente laboriosa y poco práctica, también resulta innecesaria, dado que el resto de iteraciones del fenómeno comienzan a mostrar patrones recurrentes y problemáticas similares. Por tal motivo, se optó por centrarse en los ejemplos más representativos que ilustran con precisión el fenómeno lingüístico investigado.

6. Conclusiones

La presente investigación se basó en un estudio de la anáfora formada con la forma pronominal *esto* como recurso de cohesión en textos académicos, con énfasis en su impacto sobre la lecturabilidad y la generación de ambigüedades. A partir de los resultados, se deduce que el abuso o empleo sin cautela de *esto* como mecanismo anafórico tiene influencia directa en la claridad textual. Específicamente, porque tiende a dificultar la identificación precisa de referentes y, en consecuencia, la comprensión global del texto termina comprometida.

La palabra *esto* constituye un recurso valioso para la economía discursiva y la continuidad topical, pero su neutralidad gramatical —la ausencia de concordancia genérica y numérica— amplifica el potencial de ambigüedad. Más aún cuando varios candidatos referenciales se perfilan en la estructura sintáctica inmediata. Este fenómeno se observó prevalentemente en pasajes donde las formas pronominales remiten a otras que comparten la naturaleza deíctica, lo que produce un efecto de opacidad referencial que juega contra la transparencia del discurso.

Un hallazgo destacable es que la ambigüedad anafórica presenta una distribución heterogénea en el corpus; de hecho, parece intensificarse cuando *esto* redirige a referentes de baja especificidad o cuando el contexto sintáctico no facilita la recuperación anafórica. En contraste, si los referentes posibles son complejos clausulares que exhiben relaciones de causalidad o correferencialidad, la ambigüedad disminuye sensiblemente, aun cuando varios posibles antecedentes se encuentran en competencia.

Contrario a la hipótesis inicial, el abuso de *esto* como anáfora pronominal no es distintivo de los textos inéditos, sino que se presenta de manera recurrente tanto en manuscritos como en publicaciones académicas revisadas. En ese sentido, su proliferación no delata una deficiencia en la revisión editorial, sino que parece responder a dinámicas discursivas propias del género académico. Alternativamente, podría argumentarse que su prevalencia obedece a convenciones de la oralidad filtradas al territorio de lo escrito que a meros descuidos editoriales. Sea cual sea el motivo, lo cierto es que son necesarios procesos de revisión lingüística paralelos a las evaluaciones técnicas para asegurar que el texto cumpla con su función comunicativa.

En cuanto a la extensión oracional, se respeta el límite de 35 palabras por oración sugerido para favorecer la lecturabilidad, pero únicamente en términos generales. Ya en lo particular, se identificó una amplia variabilidad en el corpus, con oraciones que sobrepasan o incluso duplican dicho umbral. Además, se observó que entre más extensa la oración, mayor es la tendencia del autor a

perder el hilo argumentativo, de modo que en ocasiones establece relaciones cohesivas forzadas. Las divagaciones de este tipo pueden resultar, por ejemplo, en la creación de pares causa-efecto entre ideas que no necesariamente construyen estructuras causales legítimas. La inconsistencia resultante compromete la coherencia local y afecta la interpretación global del material.

Otro punto a resaltar es que, con frecuencia, la simple sustitución de la anáfora pronominal cohesiva por un encapsulador nominal puede ser suficiente para favorecer la lecturabilidad del texto. La precisión que permiten los procesos de encapsulación nominal es eficaz para reducir la ambigüedad, acotar los referentes y facilitar la recuperación anafórica en aras de producir un discurso más claro y accesible.

En suma, la anáfora pronominal con *esto* constituye un recurso de cohesión ampliamente utilizado en el discurso académico, capaz de establecer diversas relaciones semántico-discursivas como adición, causalidad, adversatividad, temporalidad, localización y finalidad. De hecho, los resultados coinciden con los de Parodi y Burdiles respecto de la unidad lingüística *ello*, en tanto que los usos se concentran en la expresión de relaciones causales y la adición de información. Sin embargo, su potencial para establecer correferencialidad puede derivar en ambigüedades que afectan la claridad y la lecturabilidad de los textos.

Agradecimientos

Esta tesis no es el fruto de mi esfuerzo individual. Agradezco profundamente a mi familia, mi esposa y amigos de siempre, a mis compañeros de trabajo en el Despacho de Publicaciones de la Facultad de Ingeniería. Sin su paciencia, cariño y, sobre todo, su apoyo emocional e intelectual, no habría sido capaz de culminar este trabajo. Extiendo también mi agradecimiento al Consejo Nacional de Humanidades, Ciencias y Tecnologías por la beca otorgada para la realización de este proyecto.

Referencias

- Anthony, L. (2023). AntConc (v.4.2.4) [Software]. Waseda University. [Index of /software/antconc/releases](https://index.of/software/antconc/releases)
- Anthony, L. (2022). TagAnt (v.2.0.5) [Software]. Waseda University. [Index of /software/tagant/releases](https://index.of/software/tagant/releases)
- Bello, A. (1911). *Gramática de la lengua castellana*, París: Roger y Chernovitz Editores.

- Benveniste, É. (1980). La naturaleza de los pronombres, en *Problemas de Lingüística General. Tomo I*, 172-178, México: Siglo XXI Editores.
- Biber, D. y Reppen, R. (2015). *The Cambridge Handbook of English Corpus Linguistics*. Reino Unido: University Printing House.
- Björnsson, C. H. (1983). Readability of newspapers in 11 languages. *Reading Research Quarterly*, 18(4).
- Campos Saavedra, D. et al. (2014). Complejidad textual, lecturabilidad y rendimiento lector en una prueba de comprensión en escolares adolescentes. *Universitas Psychologica*, 13(3), 15-26.
- Crossley, S. A., Kyle, K. y McNamara D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behav Res*, 48: 1227-1237.
- Cruz Aguilar, E. (2020). La educación transformadora en el pensamiento de Paulo Freire. *Educere*, 24(78), 197-206.
- Etchepareborda, M. C., y Abad-Mas, L. (2005). Memoria de trabajo en los procesos básicos del aprendizaje. *Revista de neurología*, 40(1), 79-83.
- Fernández, O. (1999). El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos, en Ignacio Bosque y Violeta demonte (coords.), *Gramática descriptiva de la lengua española*, Madrid: Espasa Calpe, 1209-1273.
- Fillmore, C. J. (1968). The Case for Case, en E. Bach y R. T. Harms (eds.) *Universals in Linguistic Theory*. Londres: Holt, Rinehart and Winston, pp. 1-25.
- Flores Aguilar, M. D. (2018). La escritura académica en estudios de ingeniería: valoraciones de estudiantes y profesores. *Revista de la Educación Superior*, 47(186), 23-49.
- Fuentes Arderiu, X., Antoja Ribó, F. y Castiñeiras Lacambra, M. J., (2003). *Manual de estilo para la redacción de textos académicos y profesionales*. España: DAP Badalona.
- Givon, T. (1983). *Topic continuity in discourse: A quantitative cross-language study*. John Benjamins Publishing.
- (2001). *Syntax: an introduction*, John Benjamins.
- Halliday, M. A. K. y Hasan, R. (1976). *Cohesion in English*. Routledge.
- Hyland, K. (2015). Corpora and written academic English, en Biber, D. y Reppen, R. (2015). *The Cambridge Handbook of English Corpus Linguistics*. Reino Unido: University Printing House.

- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Kintsch, W. y Keenan, J. (1973). Reading rate and retention as a function of propositions in the base structure of sentences. *Cognitive Psychology*, 5, 257-274.
- Kintsch, W., y Monk, D. (1972). Storage of complex information in memory: Some implications of the speed with which inferences can be made. *Journal of Experimental Psychology*, 94(1), 25-32. <https://doi.org/10.1037/h0032781>
- Kolbe-Hanna, D. y Szmrecsanyi, B. (2015). Gramatical variation, en Biber, D. & Reppen, R. (2015) *The Cambridge Handbook of English Corpus Linguistics*. Reino Unido: University Printing House.
- McNamara, D. S., Graesser, A. C., McCarthy, P., y Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge: Cambridge University Press.
- Nadal Sanchis, L. (2024). El procesamiento de la encapsulación anafórica nominal y pronominal en hablantes de español L2: un estudio experimental con eyetracking. *Revista de Lenguas para Fines Específicos*, 30, 9-36.
- Padró, L. Stanilovsky, E. (2012). FreeLing 3.0: Towards Wider Multilinguality Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA. Istanbul, Turkey. May.
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada*, 46 (1), 93-119.
- Parodi, G. y Burdiles, G. (2016). Encapsulación y tipos de coherencia referencial y relacional: el pronombre ello como mecanismo encapsulador en el discurso escrito de la economía, *Onomázein Revista de lingüística, filología y traducción*, 33, 107-129.
- Real Academia Española (2016). *Nueva gramática de la lengua española. Morfología. Sintaxis. Fonética y fonología*. España: Espasa Libros.
- (s. f.). Banco de datos (CREA Anotado) [en línea]. Corpus de referencia del español actual (CREA). <http://www.rae.es>
- (s. f.). *Glosario de términos gramaticales*, [versión 1.0 en línea]. <https://www.rae.es/gtg/correferencia>
- Romero González, A. N. (2020). La escritura académica de estudiantes universitarios de Humanidades a partir de sus producciones. Estudio transversal. *Revista mexicana de investigación educativa*, 25(85).

- Riffo, B. y Véliz, M. (2011). Modelo de evaluación de la comprensión lectora. (Informe de avance proyecto Fondef D08I1179). Universidad de Concepción, Concepción, Chile.
- Sánchez Upegui, A. A. (2011). *Manual de investigación Académica e investigativa: Cómo escribir, evaluar y publicar artículos*. Colombia: Católica del Norte Fundación Universitaria.
- Schmid, H. G. (2018). Shell nouns in English: a personal roundup, *Caplletra*, 64, 109-128.
- Urizar, C. y Rodríguez, I. (2022). Encapsuladores nominales *hecho, idea, tema y problema* en un corpus de escritura académica en español. *Estudios de Lingüística Aplicada*, 40(74).
- Van Dijk, T. A. (1977). *Text and context: Explorations in the semantics and pragmatics of discourse*. Longman.
- (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Lawrence Erlbaum Associates.
- Véliz, M. (1999). Complejidad sintáctica y modos del discurso. *Estudios Filológicos*, 34, 181-192.