



Análisis automático de la complejidad sintáctica presente en tres géneros de la disciplina médica: Entradas de Wikipedia Médica, Artículos de Investigación Científica Médica y Casos Clínicos

Alumna: Bárbara Jerez Cuevas

Profesor Guía: Dr. Walter Adrián Koza

Tesis para optar al grado de Licenciado en Lengua y Literatura Hispánica

Becaria Proyecto Fondecyt 11130469

5 de Enero de 2016

Índice

I.- INTRODUCCIÓN	3
II.- MARCO TEÓRICO	6
1.Complejidad Sintáctica	6
1.1. Estructuras Sintácticamente Complejas	11
1.1.1 Nominalización	12
1.1.2 Voz Pasiva	13
1.1.3 Subordinación	14
1.1.3.1 Subordinada sustantiva	14
1.1.3.2. Subordinada adjetiva	16
2.Géneros Profesionales de la Medicina	17
2.1 Géneros Discursivos	17
2.2 Discurso Académico	18
2.3 Entradas Wikipedia	19
2.4 Artículos de Investigación Científica	20
2.5 Casos Clínicos	21
3. Lingüística Computacional	22
3.1 Software Nooj	24
III.METODOLOGÍA	25
3.1 Pregunta de investigación	25
3.2 Objetivos	25
3.2.1 Objetivo general	25
3.2.2 Objetivos específicos	25
3.3 Tipo y alcance de la Investigación	25
3.4 Tareas desarrolladas	26
3.4.1 Recolección Corpus	26
3.4.2 Trabajo Computacional	28
3.4.2.1 Gramática Nominalización	28
3.4.2.2 Gramática Voz Pasiva	29
3.4.2.3 Subordinada sustantiva	30
3.4.2.4 Subordinada adjetiva	30
3.5 Etapa de análisis y resultados obtenidos	31
3.5.1 Identificación complejidad sintáctica	31
3.5.2 Clasificación complejidad sintáctica	31
IV.-RESULTADOS OBTENIDOS	32
4-1.1 Resultados Entrada Wikipedia	32
4.1.2 Artículos de Investigación Científica	33
4.1.3 Resultados Casos Clínicos	33
4.2 Descripción y Comparación	34
4.3 Análisis	35
V.-CONCLUSIÓN	39
VI.-BIBLIOGRAFÍA	42

I. Introducción

Actualmente, los trabajos de investigación relacionados con la complejidad sintáctica poco a poco han ido incrementando. Este fenómeno remite a la noción de oraciones complejas en su estructura. La producción de dichas oraciones resulta de la selección y combinación de ciertos elementos “sintácticamente complejos” y su uso estaría justificado en la medida en que permitirían comunicar de manera más efectiva las ideas que el hablante desea transmitir (Hunt, 1970). El uso de estos elementos está supeditado a diversos factores tanto externos como internos al sujeto como por ejemplo, la educación, el desarrollo cognitivo o la naturaleza del texto.

Uno de los motivos de interés que provoca el estudio de este fenómeno es que a partir de su análisis se puede extraer variadas respuestas respecto al desarrollo del lenguaje de los seres humanos (Hunt, 1970) y de sus habilidades comunicativas. Asimismo, también es posible caracterizar textos escritos según el grado de complejidad, lo que tiene un papel preponderante en las investigaciones sobre lecturabilidad y rendimiento lector (Campos, Contreras, Rifo, Véliz & Reyes, 2014; Véliz, 1999). El estudio progresivo de este fenómeno ha permitido que autores como Hunt (1970) haya descubierto variables que inciden directamente en el aumento o escaso desarrollo de la complejidad sintáctica. Elementos como la edad, la escolaridad y el tipo de discurso influirían en que el sujeto elabore un número mayor o menor de oraciones sintácticamente complejas. Por otro lado, Nefedova (2005) considera que la complejidad sintáctica favorece el despertar de la reflexión y la formación de significados. Lo anterior desde una perspectiva del lector, quien interpreta los recursos sintácticos estilísticos y de este modo descubre los significados implicados en ellos, reconstruyendo en su mente los significados intencionales que la sintaxis compleja vehicula formalmente.

Ahora bien, una de las formas de conocer de qué modo se manifiesta la complejidad sintáctica es analizando producciones orales y escritas para identificar aquellos elementos sintácticamente complejos que se emplean con mayor y menor frecuencia en los textos. En relación con esto último, a partir de la literatura que se ha consultado para la elaboración del presente trabajo (Nefedova, 2005; Di Tullio, 2005; Gómez Manzano, 2002; Gutierrez-Clellen, 1994; Hunt, 1970), se consideran elementos sintácticamente complejos los

fenómenos tales como la nominalización, la voz pasiva, la coordinación, la subordinación, los incisos y la elipsis, entre otros.

Si bien, las investigaciones en torno a la complejidad sintáctica son cada vez más, no existen estudios que analicen la complejidad sintáctica específicamente en ámbitos profesionales como la medicina, ya que la preferencia es estudiarla en campos educativos (Veliz, 1999; Rodríguez, 2002; Nefedova, 2005; Meneses, Ow & Benítez, 2012; Checa-García, 2013; López, 2013; Rodríguez, 2016). Por otro lado, son escasos los estudios que dan cuenta de este fenómeno desde la perspectiva de la lingüística computacional, de hecho, los trabajos que se han consultado refieren al fenómeno de manera accesoria y en relación con otros objetivos, como ser el desarrollo de herramientas de traducción automática (Salvador Mata, 1985) de evaluación de nivel de conocimiento lingüístico en estudiantes de español como lengua extranjera (Ferrero, 2011), o la simplificación automática de textos (Saggion, Marimon & Ferrés, 2015)

Es por esta razón que surge el objetivo de conocer y describir de qué manera se manifiesta la complejidad sintáctica en el dominio médico a partir del análisis automático de tres géneros: Entradas de Wikipedia (EWM), Artículos de Investigación Científica (AICM) y Casos Clínicos (CCM). La elección de esta área se justifica en, por un lado, la importancia que tiene la medicina en la sociedad, es decir, preservar la integridad física y psíquica de los seres humanos, y, por otro, la enorme cantidad de textos médicos que circulan y se incrementan diariamente, lo que necesario contar con herramientas que permitan procesarlos de manera automática.

El análisis se realizó a partir de la observación del grado de complejidad sintáctica que presentaban según la aparición de tres variables: la nominalización, la voz pasiva y la oración subordinada. Para dicho análisis se utilizó el software Nooj que permite utilizar herramientas para crear y mantener fuentes lexicales, así como gramáticas, sintácticas y morfológicas. Asimismo, otra de sus características es el desarrollo de operaciones morfológicas sobre los textos mientras se lleva a cabo un análisis sintáctico, a su vez, puede procesar varios textos en varios idiomas, siendo una herramienta multilingüe. Cabe señalar que la cobertura y precisión del software como de las gramáticas es bastante alto, entregando veracidad a los resultados obtenidos (Bonino, 2012; Koza, 2016).

La elección de las tres estructuras sintácticas radica por un lado, en que presentan una estructura sintáctica diferente, lo cual permite una descripción amplia y detallada de la sintaxis de los textos médicos. Asimismo, la elección del corpus surge de la importancia de la medicina para el conocimiento humano y este estudio aporta con evidencia empírica respecto a la conformación sintáctica de textos científicos. A su vez el presente estudio como una aproximación de la conformación sintáctica de tres textos del área médica pretende sentar bases a futuros estudios en los que se profundice más sobre el fenómeno de la complejidad sintáctica en el área médica.

La presente tesis se estructura en los siguientes apartados: En la sección II, se presenta el Marco Teórico, en donde se explica el significado de la complejidad sintáctica, así como la descripción de los elementos sintácticamente complejos, la caracterización de los géneros profesionales y una breve señalización a la lingüística computacional. En la sección III, se menciona la Metodología, en la cual se explica las etapas y fases que se desarrollaron para la selección del corpus, su descripción y posterior comparación. En la sección IV, se indica los Resultados y análisis, apartado que reporta los datos obtenidos y su análisis a partir del número de apariciones de los tres elementos sintácticamente complejos estudiados. Finalmente, en la sección V, se presenta la Conclusión, en donde se realiza una reflexión respecto a los resultados obtenidos, ligados a la complejidad sintáctica presente en textos del área médica.

II. Marco Teórico

En este capítulo se presenta el Marco Teórico que sustenta la presente investigación. En la cual se realiza un recorrido por las diferentes concepciones de fenómeno estudiado, así como las investigaciones en torno a la complejidad. Posteriormente se revisan y describen los elementos sintácticamente complejos que guían este trabajo. Para finalizar con una breve caracterización de la Lingüística Computacional y del software Nooj.

1) Complejidad Sintáctica

Tradicionalmente la complejidad sintáctica ha sido estudiada por muchos autores, quienes se interesaron en comprender y caracterizar el fenómeno en detalle. Lo anterior, dio origen a diferentes definiciones, las cuales estaban guiadas por las corrientes teóricas que en dicho momento se encontraban en apogeo: la Teoría Formal y la Teoría Funcional. La primera de estas teorías, está fuertemente influenciada por el trabajo del lingüista Noam Chomsky (1965), quien en la década de los 50 señaló que la gramática de una lengua puede ser representada como una lista de reglas abstractas, como las usadas por matemáticos. El autor sostiene que esas reglas sirven para explorar los límites del lenguaje y que dichas limitaciones tenían origen neurológico. Además, agrega que todos los seres humanos nacen con una facultad innata para el lenguaje y por ende, la tarea de la lingüística formal es descubrir cómo funciona y en qué consiste dicha facultad (Martin & Rothery, 1993).

Con base en lo anterior, la complejidad sintáctica de tipo formal, va asociada a una cierta complejidad cognitiva. En teoría al poner en relación una mayor cantidad de elementos implicaría una dificultad mayor, por consiguiente, operaciones mentales más elaboradas. (Martin & Rothery, 1993) De este modo, la complejidad sintáctica desde esta teoría se aborda con una perspectiva mental, a partir de las operaciones cognitivas que el individuo elabora para comunicar sus ideas. Ahora bien, los lingüistas de la Teoría Funcional se han preocupado por un rango de intereses variados. Han sido influidos por los trabajos del profesor Michael Halliday (1979) esta lingüística se interesa por el estudio de los rasgos prácticos y reales del lenguaje, pues tienen una orientación más sociológica, es decir, más preocupados por las relaciones y uso del lenguaje en sociedad, tratando de entender cómo las formas en que el lenguaje es usado ha conformado su estructura. Esta visión social del lenguaje, ha llevado a los teóricos funcionales a crear gramáticas semánticamente orientadas

que muestren cómo las personas emplean el lenguaje para construir significado con la finalidad de llevar adelante su vida y sus relaciones sociales (Martin & Rothery, 1993).

Cabe señalar que con el tiempo varios han sido los autores que han abarcado el estudio del lenguaje desde esta perspectiva, aportando con relevante información al respecto. De este modo, surge el enfoque Funcional-discursivo que con los trabajos de Berman (1994), quien en conjunto con su equipo de trabajo afirman que la adquisición del lenguaje se desarrolla como un proceso natural y cultural, en el que confluyen elementos como la experiencia, el ambiente social y las estructuras cognitivas de los sujetos. A partir de esto, señalan que el discurso se convierte en una guía para la sintaxis y que la complejidad sintáctica no puede ser analizada sin tomar en cuenta elementos como la secuencia (argumentativa, narrativa, descriptiva, expositiva, etc.) y a la modalidad de los textos, es decir, si son orales o escritos. (Crespo, Alfaro & Góngora, 2011)

A modo general, este enfoque propone que la complejidad depende del tipo de relaciones entre cláusulas y entre paquetes clausulares al interior del texto, lo que provoca un uso más retórico de estas unidades. A partir de esa definición se logra percibir que la unidad de análisis que instaura este enfoque es el “Paquete clausular” que es un conjunto de cláusulas (unidad nuclear definida por criterios sintácticos y semánticos) vinculados por criterios sintácticos, temáticos y discursivos que funcionan como una unidad al interior del texto mayor (Crespo, Alfaro & Góngora, 2011). Tomando en cuentas esta unidad de análisis el método que emplearon los investigadores para analizar la complejidad sintáctica en la producción de discurso fue el cualitativo, pues el interés se encontraba en caracterizar la sintaxis compleja ya no según el número de apariciones de ciertos recursos sintácticos, sino que el foco estuvo en analizar la manera en que el paquete clausular variaba según elementos discursivos como el tema, el tipo de texto y otros recursos sintácticos. Finalmente, con el tiempo descubren que con la edad aparecen relaciones más complejas y un manejo más consciente de los objetivos retóricos. En otras palabras este enfoque permitió poner atención a elementos que otros enfoques no habían considerado como el tema y el tipo de discurso (Crespo, Alfaro & Góngora, 2011).

Como se logra apreciar la complejidad sintáctica ha sido estudiada desde diferentes perspectivas y con variados objetivos. Por ello, a continuación se describirán una selección de tres estudios relacionados con este fenómeno que permitirán esclarecer la información y los diferentes descubrimientos que han brotado como fruto de dichas investigaciones. El primero se relaciona con los aportes de Hunt (1970) respecto a las diferencias y relación entre complejidad sintáctica y madurez sintáctica. El segundo estudio es realizado por Veliz (1999) quien se enfocó en averiguar la influencia de la organización del discurso en la complejidad sintáctica. Finalmente, Nefedova (2005) se centra en estudiar la complejidad sintáctica a partir de seis elementos sintácticamente complejos en textos literarios.

De los estudios relacionados con la complejidad sintáctica cabe señalar los aportes realizados por Hunt (1965,1970) quien se centró en variables que influían en el desarrollo de una sintaxis compleja. De este modo, factores como la edad, la escolaridad y el tipo de discurso se convierten en elementos a considerar para analizar la complejidad sintáctica de una producción textual. El autor formula con claridad una diferencia entre Complejidad sintáctica y Madurez sintáctica. La primera, se relaciona con la elaboración de oraciones complejas en su estructura, es decir, el uso de transformaciones como la subordinación o el empleo de conectores. Mientras que la segunda, se refiere a la habilidad para producir oraciones con mayor complejidad estructural, adecuando el empleo de las transformaciones necesarias para escribir una secuencia oracional. Por ende, un escritor inmaduro sintácticamente no sabría utilizar las transformaciones adecuadas, escribiendo enunciados complejos muy densos que terminan no siendo efectivos para la comunicación.

Del mismo modo, y desde la perspectiva de análisis el discurso, Mónica Veliz (1999) quiso determinar si el modo de organización del discurso afecta la complejidad sintáctica de textos escritos por adultos lingüísticamente competentes. La autora con este trabajo reveló que sí incide el modo discursivo en la complejidad sintáctica de textos escritos. De esta manera, textos argumentativos presentan una mayor presencia de complejidad sintáctica, mientras que el descriptivo exhibe una sintaxis compleja intermedia por sobre el narrativo que manifiesta una complejidad sintáctica considerablemente menor a los otros tipos de textos. De esto se puede inferir que quizá esto se deba a que a temprana edad el tipo de texto más

cercano y enseñado es el narrativo, lo que facilitaría su apropiación y por ende la producción de textos menos complejos sintácticamente, al ya poseer estructuras mentales asociadas al texto narrativo.

Por otro lado, Nefedova (2005) analiza la complejidad sintáctica en textos literarios como un elemento que propicia la formación de los significados desde cuatro idiomas diferentes: inglés, español, alemán y ruso. Este estudio arrojó que los ocho recursos de la complejidad sintáctica que se utilizan con más frecuencia en las obras literarias son las repeticiones, el paralelismo, el asíndeton, el polisíndeton, la combinación del asíndeton y el polisíndeton, la inversión, la frase intrusiva, la parcelación, la elipsis. Estos elementos según la investigadora son utilizados por el escritor de manera consciente con el fin de captar la atención del lector y formar significados particulares, pues: “al enriquecer el texto con el uso de los recursos sintácticos estilísticos, el emisor ofrece al receptor una posibilidad de comunicación con sus personajes y consigo mismo” (Nefedova, 2005).

Asimismo, la autora agrega que la complejidad sintáctica se considera como una categoría sintáctica que favorece el despertar de la reflexión y la formación de significados. La *reflexión*, según la autora se entiende como estado de actividad mental relacionada con la producción de los significados y con los procesos de la realización de estos significados en forma de conocimientos en la actividad práctica. “Al interpretar los recursos sintácticos estilísticos, el receptor descubre los significados implicados en ellos y forma en su conciencia unas imágenes, es decir, reflexiona o reconstruye en su mente los significados intencionales que la sintaxis compleja vehicula formalmente” (Nefedova, 2005). Respecto a esto la autora indica que la importancia de los recursos de la complejidad sintáctica como estímulo para la reflexión está explicada por el hecho de que la observación de las estructuras sintácticas es una condición necesaria para la comprensión de textos literarios.

Finalmente, la investigadora explica que la sintaxis compleja o la complejidad sintáctica es una característica esencial para un texto que posea una intensa organización significativa y que se utiliza intencionalmente por el emisor para formar los significados. Asimismo, agrega que la sintaxis compleja es típica sobre todo para el texto literario con criterios estéticos. Los

criterios estéticos forman la base del lenguaje poético y se reflejan en el uso de los recursos estilísticos que actúan como la forma de la obra literaria. (Nefedova, 2005).

Cabe destacar que la relevancia de este estudio para la presente investigación es fundamental, ya que este trabajo al igual que el de Nefedova (2005) pretende describir la complejidad sintáctica de un corpus determinado. Ahora bien, la diferencia radica en que la actual investigación analizó textos del área médica, mientras que la autora se centró en analizar la complejidad en textos literarios. Además las variables de análisis que la investigadora aborda son seis, y este trabajo se enfoca en analizar tres elementos sintácticamente complejos. Si bien las diferencias no son tantas, es relevante este estudio para la presente investigación ya que se considera como una base teórica influyente, respecto a los procedimientos metodológicos como en su percepción respecto al fenómeno estudiado.

A modo general, como se logra apreciar los estudios sobre complejidad sintáctica son variados, y sin duda han demostrado aportar con información relevante para distintas áreas del conocimiento humano, destacándose sobre todo el ámbito pedagógico, ya que entender cómo funciona este fenómeno lingüístico permite orientar al docente a potenciar el desarrollo de la competencia lingüística en estudiantes, favoreciendo la madurez sintáctica y la producción de textos cada vez más completos a nivel sintáctico y semántico.

Ahora bien, una vez que se han señalado las diferentes perspectivas de estudio de la complejidad sintáctica y los aportes en torno a este fenómeno es necesario señalar cuál de los enfoques guiará este trabajo y cómo se comprenderá la complejidad sintáctica. En primer lugar, la investigación realizada se inserta en el Enfoque Funcional, principalmente porque se considera como un factor relevante en los resultados el tipo de texto que se está estudiando, así como el propósito comunicativo que cada texto posee. Puesto que fin comunicativo como el tipo de textos serán factores importantes a considerar a la hora de elaborar las conclusiones respecto a la complejidad sintáctica en los tres géneros abordados.

En segundo lugar, para señalar la definición de complejidad sintáctica que este trabajo desarrollará es preciso definir oración Kernel, la cual es “a simple declarative construction

with only one verb. A kernel sentence is always active and affirmative. Also known as a *basic sentence* or a *kernel*” (1).

El concepto fue creado por Harris (1958) y aparece por primera vez en los trabajos de Noam Chomsky (1965). Un ejemplo de oración Kernel es: “El hombre abrió la puerta”, como se logra apreciar la oración posee un verbo conjugado en tiempo indicativo con voz activa. De este modo, el presente trabajo entenderá complejidad sintáctica como toda aquella oración que no coincida con las características de una oración núcleo, se consideró esta noción como base de esta investigación, porque al ser una oración simple permite identificar con mayor precisión la manifestación de elementos sintácticamente complejos en el texto.

1.1 Estructuras sintácticamente complejas

Los elementos que se consideran sintácticamente complejos han variado a lo largo del tiempo, según cada autor y de acuerdo con la perspectiva de estudio a la que se adhiera (Nefedova, 2005; Di Tullio, 2005; Gómez Manzano, 2002; Gutierrez-Clellen, 1994; Hunt, 1970). De este modo a continuación se indicará una selección de los elementos sintácticamente complejos, considerados como los más representativos. Posteriormente se profundizará en tres estructuras complejas que se analizarán en el presente trabajo. Las estructuras sintácticamente complejas son : las oraciones yuxtapuestas, las oraciones coordinadas, la elipsis, el asinetón, el polisíndeton, hipérbaton, la nominalización, la voz pasiva y la oración subordinada. El presente trabajo abordará las últimas tres estructuras nombradas. La razón de su selección radica en que son elementos bastante interesantes de analizar, no tan solo porque su uso es frecuente en los textos, sino también porque existen escasos estudios que centren su objetivo en analizar específicamente estas tres estructuras por ende, se consideró como un aporte de información para los estudios relacionados con la complejidad sintáctica. A continuación se describen las tres estructuras que se identificaron en el corpus.

1.1.1 Nominalización

Noam Chomsky con su ensayo “Remarks on Nominalizations” (1970) aborda la nominalización como el germen de lo que será su propuesta denominada Hipótesis Lexicalista. En esta propuesta Chomsky señala que nombres como destrucción, rechazo o crítica no derivan de transformaciones de nominalización, sino que se introducen como nombre directamente en el Lexicón, entendido como : “el componente de la gramática en que se refleja las unidades significativas de la lengua, o bien solo sus unidades exclusivamente léxicas” (Rello, 2006). De esta manera, Chomsky indica que no existirían dos entradas diferentes en el Lexicón, puesto que este poseería las raíces con unos rasgos de subcategorización estricta y de selección establecidos, pero sin los rasgos denominados categoriales (N, V, Adj). De esta forma la elección de uno de los términos, ya sea nominal, verbal o adjetivo se realizará según el nudo de categorías que lo denomine en la derivación sintáctica. Posteriormente señala que reglas fonológicas particulares establecerán la forma fonológica correspondiente.

La nominalización es un procedimiento de construcción de sustantivos a partir de un verbo, es decir, dichos elementos son transformados en sustantivos, convirtiéndose en un elemento nominal o sustantivado, por ejemplo el verbo destruir se nominaliza al decir la destrucción. Esta operación guarda relación con la producción del individuo, por ello, si un escritor la emplea en su texto es un indicio de madurez intelectual o un indicador de un apropiado desarrollo cognitivo. M.A.K. Halliday (2009) señala que la nominalización facilita condensar gran cantidad de información en solo una cláusula nominal con varios modificadores, además Halliday (2009) agrega que la nominalización tiene como finalidad la transición de la proposición del concepto, o también conocida como “abstracción sustantiva

Por otra parte, en el ámbito escolar, cuando la nominalización está presente en los textos académicos, leer aquellos que poseen excesivos uso del recurso resulta bastante complejo para lectores no expertos, pues se les dificulta comprender el proceso de metaforización implícito (Cinto, 2009). Para guiar la comprensión es sumamente importante la mediación del docente con sus alumnos y seleccionar material que gradualmente aumente en dificultad para poco a poco desarrollar la comprensión a niveles más complejos sintácticamente.

1.1.2 Voz pasiva

La voz pasiva es una categoría gramatical del verbo que indica que el sujeto sufre o recibe la acción realizada por otro. (Navarro, Hernández & Rodríguez, 1994). Se diferencia de la voz activa en que este tipo de oración expresa el sujeto que realiza la acción; un ejemplo de oración en voz activa citado por Navarro, Hernández & Rodríguez (1994) es: “el médico recetó un antibiótico” esta misma oración se puede transformar en pasiva y queda: “un antibiótico fue recetado por el médico”, respecto a esta transformación los autores señalan que la relación existente entre el sujeto y el complemento no sufre grandes cambios ya que la queda claro que en ambas oraciones la relación de sentido entre el antibiótico y el médico es la misma.

A su vez, los autores clasifican la voz pasiva en dos tipos: voz pasiva perifrástica y voz pasiva refleja. La primera de ellas se conforma con el participio del verbo que expresa la acción y el verbo ser como auxiliar (perífrasis verbal ser + participio) un ejemplo de ello es: “las crisis epilépticas fueron descritas ya por Hipócrates”, como se logra apreciar en este tipo de oración el participio coincide en género y número con el sujeto (Navarro, Hernández & Rodríguez, 1994). Además agregan que el empleo de la pasiva con ser ha disminuido mucho en español ya que la gran mayoría tiende a utilizar la forma activa, la razón radica en que: “La influencia humanística y culta lo mantuvo, pero el incremento de las construcciones con sentido pasivo conseguidas con el signo se, creciente hasta hoy, ha restringido más y más su uso.”

Por otro lado, la pasiva refleja es otra manera de expresar una acción sin otorgarle sujeto gramatical, en este tipo de oración, el sustantivo que acompaña el verbo es un sujeto gramatical, por ende el verbo debe ir en singular o plural de acuerdo al sustantivo, un ejemplo es el siguiente: “se realizó un estudio sobre la bilirrubina”, “se observaron varios casos de pandemia”. En el ámbito médico, los autores acusan que en castellano la preferencia se inclina hacia el uso de la voz activa más que la pasiva, mientras que en alemán, francés y sobre todo en inglés se produce una preferencia por la voz pasiva. De esta manera, expresan que textos médicos producidos en inglés presentan un uso excesivo de voz pasiva, sobre todo cuando al traducirlos al español. Esto provoca que la lectura del texto sea mucho más

lenta y poco amigable, por esta razón los autores entregan varios consejos para evitar este problema en los textos médicos, algunos son:

1. Evitar la traducción de muchas pasivas inglesas por pasivas castellanas
2. Cuando el agente del verbo es desconocido, carece de importancia o no interesa declararlo, es insustituible la utilización de una voz pasiva, ya sea perifrástica o pronominal
3. No utilicemos una construcción pasiva si es posible decir lo mismo con una forma activa.

El presente trabajo selecciona la voz pasiva por ser un elemento que se manifiesta con frecuencia, lo cual nos permite observar si en función de otras estructuras sintácticamente complejas es mayor o menos su uso, de esta forma lograr realizar una breve comparación de lo señalado por Navarro, Hernández y Rodríguez (1994) de la presencia de voz pasiva en textos médicos escritos en español.

1.1.3 Subordinación

La oración subordinada o secundaria se encuentra incorporada a la principal o subordinante, y establece, con ella, la misma relación que guardan con el verbo los elementos sintácticos de la oración simple. Por ende, la oración subordinada desempeña, dentro de la oración principal, la misma función sintáctica que corresponde a un sustantivo, a un adjetivo, o a un adverbio, y será equivalente de alguna de estas tres clases de palabras. En consecuencia, las oraciones subordinadas se clasifican en tres tipos, oración subordinada sustantiva, adjetiva o adverbial. Para efectos de la de la presente investigación se definirán la subordinada sustantiva y adjetiva, pues ambas adquieren relevancia en el desarrollo de este trabajo como variable de análisis.

1.1.3.1 Subordinada sustantiva

Las oraciones subordinadas sustantivas se caracterizan porque sintácticamente les corresponden todas las funciones que cumple un sustantivo en relación con la proposición principal de la que forman parte como elementos constituyentes o modificantes. Las subordinadas completivas son anteceditas por conjunciones (que, si); por formas interrogativas (qué, quién, quiénes, cuáles, cuál, dónde, cómo, cuándo, cuánto) (Rodríguez, 2012)

A continuación se presentan algunos de los tipos de oraciones subordinada existentes:

Tipos de oración subordinada	Característica	Ejemplo
Sujeto	La oración subordinada cumple la función de un sujeto.	“Que Juan sea una buena persona no significa que lo puedan tomar por tonto”
Complemento Directo	Para identificarla se puede sustituir la oración subordinada por “lo”	“Me comunicó que vendría”
Atributo	La oración principal debe tener un verbo copulativo delante del nexo	“Él parecía que estaba cansado”
Término de preposición	La subordinada es exigida por la preposición	“Las hermanas de la casa tienen ganas de que elimines la ropa quemada”

Selección basada en apuntes de estudios teóricos y prácticos del lenguaje.

1.1 3.2 Subordinada Adjetiva

La oración subordinada adjetiva es aquella que cumple la función propia de un adjetivo, por tanto actúan como modificador de un elemento sustantivo. Los pronombres relativos son: “que”, “cual”, “cuales”, “cuyo”, “cuyos”. Esta oración puede ser dividida en dos tipos: Adjetiva explicativa (no restrictiva) y adjetiva especificativa (restrictiva). A continuación se presenta una descripción de cada una:

La subordinada adjetiva explicativa es aquella que se limitan a explicar una característica del antecedente. Se caracteriza porque se escribe entre comas, y su función es detallar la información que se entrega.

1. Los jugadores, **que habían sido seleccionados**, se reunieron ayer

En el ejemplo anterior, la subordinada adjetiva se refiere a que todos los jugadores fueron seleccionados.

En cambio, la subordinada adjetiva especificativa es aquella que delimita y restringe el significado de la palabra de la oración principal a la que se refiere. Además estas oraciones pueden reemplazarse por un adjetivo. De la siguiente manera:

2. Un gato **que esta solo** produce lástima → Un gato **solitario** produce lástima

Otro ejemplo de subordinada adjetiva especificativa es:

3. El pantalón **que me compré ayer** se rompió

En el ejemplo anterior, la subordinada especificativa se refiere únicamente al pantalón que se compró ayer.

2) Géneros profesionales de la medicina

2.1 Géneros Discursivos

El estudio de los géneros discursivos con sus consiguientes variaciones según la disciplina a la que pertenecen, ha sido un foco de estudio reciente en lo que se refiere al ámbito lingüístico, dado el temprano interés en el análisis de corpus de textos completos. Esto ha llevado a la aceptación del desarrollo de una teoría del género, basada en una descripción profunda de textos académicos especializados. En este sentido, se manifiesta la importancia de un análisis contrastativo que permita dar cuenta de las variables existentes entre los diversos ámbitos disciplinares.

La caracterización de los géneros discursivos y la variación de estos según cada disciplina ha cobrado importancia en el ámbito lingüístico en los últimos años. Si bien, la búsqueda de taxonomías más o menos rigurosas desde una mirada textualista cuenta con un repertorio mayor, la aceptación consensuada de la teoría del género basado en el análisis de corpus totales ha cobrado relevancia recientemente. Por este motivo, diversos autores han trabajado en la descripción detallada de los distintos géneros que circulan en el diario vivir.

En relación con los géneros discursivos es que el término “comunidad discursiva” ha sido acuñado por diversos autores, sin embargo, aún no se llega a un consenso en torno a su definición. Con respecto a lo anterior, Jara (2009), menciona que el origen del término se remonta a diversas disciplinas que se señalaban como “colegio invisible” a los distintos académicos que, si bien no se agrupaban de manera física, manejaban los mismos conocimientos disciplinares y objetivos. Cada comunidad discursiva posee géneros particulares que son característicos para alcanzar un objetivo determinado. Bajtin (1990) señala que cada esfera social está estrechamente relacionada con el uso de la lengua, por lo mismo, existen grandes cantidades de enunciados que son elaborados de manera estable en cada esfera lo que da origen a los denominados géneros discursivos. Posteriormente, Swales (1991) menciona el concepto de propósito comunicativo, el cual marcará la diferencia entre un género y otro. De esta manera el propósito comunicativo corresponde al objetivo que tiene el texto informar, explicar, describir, narrar; estas funciones serían cruciales en la diferencia entre géneros.

Otra definición a considerar es la propuesta por Parodi (2008) quien afirma que, desde una perspectiva psicosociodiscursiva del lenguaje, los géneros discursivos se articulan a partir de tres dimensiones: social, cognitiva y lingüística, complementándose entre sí, siendo la dimensión lingüística la encargada de establecer un nexo entre las dos restantes satisfaciendo la necesidad anclada a un propósito comunicativo, según Parodi, los géneros, en su manifestación concreta son:

Variedades de una lengua que operan a través de conjuntos de rasgos lingüístico-textuales co-ocurrentes sistemáticamente a través de las tramas de un texto, y que se circunscriben lingüísticamente en virtud de propósitos comunicativos, participantes implicados (escritores y comprendedores), contextos de producción, ámbitos de uso, modos de organización discursiva, soportes y medios, etc.

De lo que se puede desprender que los géneros variarán en cuanto varíe la disciplina a la cual están adscribiendo.

2.2 Discurso académico

Siguiendo a Ibáñez, Moncada y Santana (2014) el discurso académico se entiende como la manifestación del lenguaje (predominantemente escrito) en reductos o de circulación académica, por medio del cual se transmiten los conocimientos o información específicos de una disciplina en relaciones de interacción de los miembros de la comunidad discursiva. Cada género discursivo posee un propósito comunicativo particular que, según Bhatia (1993), es la característica primordial de los géneros. Asimismo, Jara (2009) señala que el propósito comunicativo da forma y estructura interna al texto, lo que posibilita el estudio del género, pues al cumplir con un propósito comunicativo particular, se constituye como un criterio para hacer análisis y la consiguiente distinción entre diversos géneros.

En base a los diferentes géneros discursivos y reconociendo la importancia de su análisis para los estudios de lingüística, la presente investigación enfoca su análisis en el ámbito de la medicina, principalmente porque el discurso especializado de la medicina ha incrementado con el tiempo, acumulando géneros de distinta índole, algunos más complejos que otros, es por ello que esta tesis se centrará en trabajar con en tres textos relacionados con el área médica: Entradas de Wikipedia médicas, Artículos de Investigación científica médicos y Casos Clínicos médicos.

A continuación se describirán brevemente cada género:

2.3 Entradas Wikipedia

Wikipedia, “enciclopedia libre” conocida porque “todos” pueden acceder para editar la información que es presentada al público. Con el tiempo ha logrado ser una de las páginas que recibe a diario miles de visitas, convirtiéndose en el buscador de información más relevante de la web, principalmente por presentar la información de forma amigable, precisa y comprensible a todo público. A lo largo del tiempo Wikipedia ha generado un gran número de información o literatura sobre variados temas sobre todo de carácter científico, por esta razón uno de los aspectos más preocupantes es si esta información puede considerarse verídica y de calidad, en cierto sentido una fuente confiable de información. Aibar, et al (2016) señalan que la gran parte de los editores acude a fuentes confiables de información para extraer los datos en torno al tema. A su vez, dichas fuentes son primarias, la mayoría provenientes de connotadas revistas científicas, las cuales aparecen citadas al final del artículo. Además, los editores muestran un alto conocimiento sobre el tema, redactando y expresándose de forma clara, utilizando campos semánticos relacionados al tema.

En la misma línea, Aibar. Et.al (2016) en su investigación relacionada con Wikipedia Española destaca dos características relevantes de esta enciclopedia virtual. Por un lado, se relaciona con que la plataforma es la principal fuente de información científica para todas las personas, y que además permite motivar a las personas a participar en la plataforma como editores. La segunda, según este estudio es que se muestra como una herramienta que puede emplearse como enseñanza de la ciencia en todos los niveles educativos. Como fuente primaria de información permite a los estudiantes acceder a datos verídicos, de esta forma es bastante útil como información introductoria sobre el tema en cuestión, incluso profesores acuden con frecuencia a la enciclopedia libre para acceder a temas que no manejan con tanta experticia.

En conclusión, se ha considerado en el corpus las Entradas Wikipedia médicas por ser entregar información que puede considerarse científica y confiable, además de ser literatura a la cual gran parte de las personas accede y accederá por mucho tiempo más, de esta forma surge la motivación de conocer este tipo de textos en su construcción sintáctica, específicamente en la complejidad que estas producciones poseen con el fin de generar datos

que permitan continuar con estudios sobre las entrada Wikipedia desde la complejidad o con otro objeto de estudio.

2.4 Artículos de investigación científica

El artículo de investigación científica destaca por ser un medio de publicación académica-profesional más utilizada, con los años se ha convertido en una fuente de información científica de alta calidad y cada vez son más los científicos que tienen preferencia en publicar por este género. Una de las razones de su alta demanda se origina en la obtención de beneficios que implica publicar, tales como becas, obtención de proyectos y promoción de un tema específico, entre otros (Almunia, 2007).

El propósito de los Artículos de investigación radica en dos aspectos, por un lado, publicar un artículo en importantes revistas del área académica implica credibilidad y otorga reconocimiento al estudio. Por otro lado, el propósito de este connotado género radica en difundir los resultados de una investigación en foros profesionales, logrando el reconocimiento del estudio (Cortés & Cruz, 2001).

Este género se caracteriza por estar escrito de forma objetiva e imparcial, a su vez los Artículos de investigación, específicamente sus resultados tiene como una de sus labores integrar los conocimientos a la práctica profesional, en otras palabras los Artículos se basan en hallazgos que deben ser veraces, útiles, reales y que aporten con importante información a la disciplina a la que se adscribe. Cabe destacar que la parte más importante de los Artículos de Investigación son los resultados descubiertos en el estudio, ya que estos son la base del estudio, deben ser originales y no puede repetirse un mismo estudio (Henríquez & Zepeda, 2004)

Finalmente, es preciso señalar otra característica de los Artículos es la labor de acumular literatura disciplinar para que esta quede a disposición de quien estime conveniente informarse respecto a un problema científico. Para lograr lo anterior, el artículo debe ser original, es decir, mostrando las evidencias obtenidas por primera vez, esto permite la constante renovación de información para el área de estudio, de esta forma, el artículo tiene la posibilidad de participar en eventos de la disciplina o ser divulgado por revistas científicas, conferencias y otros medios de presentación.

2.5 Casos clínicos

Dentro de las publicaciones profesionales, el género Caso Clínico ha logrado posicionarse como uno de gran importancia dentro de la comunidad discursiva donde circula. El objetivo del Caso Clínico es el reporte médico de un caso o de varios casos, por ejemplo reporte de enfermedades, la evolución de esta, o bien casos inusuales (Burdiles, 2012). El caso Clínico se caracteriza por ser un género que facilita el traspaso de información, en el ámbito profesional, entre interlocutores en simetría de conocimiento, es decir, de experto a experto. Asimismo, se caracteriza por ser dentro de los géneros académicos de la medicina el más breve y simple, ya que se refiere específicamente a un solo tema de forma profunda.

Además, un rasgo que se destaca del Caso Clínico es su función pedagógica, ya que suele ser un género bastante utilizado para la enseñanza de la disciplina en el proceso de formación profesional. A su vez, cumple una función de ser un medio de acumulación de información científica y de alta calidad profesional para el área médica, por lo mismo se mantiene vigente en su preferencia de publicación, ya que adquiere relevancia el uso de casos clínicos en el área investigativa como fuente de información primaria y confiable. De este modo, Vega (2015) señala que una de las formas más importantes de aprendizaje se realiza mediante la escritura de Casos Clínicos, el individuo realiza la aplicación de sus conocimientos disciplinares ya además recibe retroalimentación respecto a sus decisiones del caso, mejorando su proceder en situaciones a futuro.

A su vez, el autor menciona una función importante de la escritura de caso clínicos para estudiantes o profesionales del área disciplinar, ya que dejando de lado la experiencia en el currículum, el autor menciona que estudiar, escribir y resolver un caso clínico “es la única manera de difundir las enseñanzas que han dejado, haciéndolas perdurables en el tiempo, de modo que pueden servir en el futuro a los médicos que se encuentren en situaciones similares” (Vega, 2015). De esta manera, el Caso Clínico posee una función importante como reporte y transmisión a la comunidad médica y científica los resultados de la historia de una enfermedad, además de caracterizarse por ser breve no sobrepasando las 1.500 palabras, por ende tiende a ser bastante preciso, específico y ordenado cronológicamente según los pasos llevados a cabo en el tratamiento o procedimiento médico.

3. Lingüística computacional

La lingüística computacional o lingüística informática compone un campo de la ciencia de carácter interdisciplinar, relacionado con la informática y la lingüística. Su objetivo es crear modelos computacionales que reproduzcan diferentes aspectos del lenguaje humano y que faciliten el estudio o tratamiento informatizado de las lenguas, en otras palabras, permite sistematizar el lenguaje humano para comprender su funcionamiento a través de un método objetivo y específico. El surgimiento de la lingüística computacional se remonta aproximadamente a finales de la segunda guerra mundial, pues en dicho contexto eran varios los científicos-técnicos de Estados Unidos y la Unión Soviética que comenzaron a estudiar la forma de elaborar programas de traducción entre el inglés y el ruso. Existía un gran interés en el desarrollo de estos proyectos de parte de las fuerzas armadas y las fuerzas de inteligencia, principalmente, porque sus resultados al ser positivos cooperarían con la superioridad intelectual de cada país, por ello fueron los principales gestores en apoyar económicamente las investigaciones.

Durante los años cuarenta y cincuenta se produjeron variados avances en dos áreas centrales para el desarrollo de tecnologías de procesamiento del lenguaje: la teoría de autómatas, área que se originó con los trabajos de Turing (1937) y los modelos de la teoría de la información, representada por la investigación de Shannon (1948). Finalizando los años cincuenta los estudios se enfocaron en dos campos principales: en primer lugar, el simbólico, el cual se divide en dos corrientes de estudio, por un lado aquella que se interesó por el análisis sintáctico, liderada por Chomsky (1965); y por otro, la orientada hacia la inteligencia artificial, representada por Minsky (1968) junto con Shannon (1948). En segundo lugar, el estocástico en el que destacan los informáticos, trabajando mediante estadísticas y probabilidades, cuyas investigaciones dieron origen al método Bayes del reconocimiento óptico, además de muchas otras aplicaciones. Aunque los orígenes de la lingüística computacional se enmarcan en un contexto principalmente lingüístico, actualmente se ha visto redefinida sus aplicaciones, debido a las oportunidades de utilizar la explotación de los textos digitales, incluso de la propia web.

Cabe destacar que si bien, el desarrollo de esta lingüística es relativamente reciente, a esta le preceden dos líneas relevantes de investigación: la Lingüística computacional teórica y la

lingüística computacional aplicada, además, Gómez Guinovart (2000) considera relevante añadir una tercera línea que denominó la informática aplicada a la lingüística. En relación con la primera de estas líneas de investigación, la lingüística computacional teórica, se enfoca en construir modelos lingüísticos en términos formales, que dichos modelos sean implementables a los distintos niveles de descripción lingüística y finalmente la comprobación en el ámbito computacional de la correspondencia del modelo y sus predicciones. Respecto a la segunda, la lingüística computacional aplicada, se destaca por ser una línea mayoritariamente tecnológica, que se centra en diseñar sistemas informáticos que sean capaces de gestionar, comprender, producir y traducir enunciados orales y escritos en lenguaje natural. Finalmente, respecto a la aplicación de la informática en la lingüística, es denominado por Gómez Guinovart (2000) como toda investigación lingüística que emplee programas computacionales o software para su desarrollo, destacando dentro de estas la lingüística de corpus.

La lingüística de corpus según Guinovart (2000) se entiende como “el estudio empírico de la lengua a partir de los datos que proporciona el análisis de ejemplos reales de producciones lingüísticas (orales o escritas) almacenadas en un ordenador.” De esta manera, mientras mayor sea el conocimiento que se posea de la estructura de la lengua a nivel morfológico, sintáctico, semántico y pragmático y de cómo el ser humano emplea la lengua, mejor serán los resultados para manejar, almacenar o extraer información e interactuar con los computadores. También la lingüística de corpus se caracteriza por trabajar principalmente con compilaciones de textos reales, por ende sus resultados son de gran importancia a nivel de conocimiento de la estructura de la lengua, y han sido aplicados en áreas como la lexicografía, la construcción de gramáticas y la traducción automática.

A modo general, respecto a la aplicación de la informática en la lingüística es preciso señalar que el etiquetado tiene una gran importancia, pues por medio de este se codifica y organiza la información que se analiza, el etiquetado morfológico agrupa cada palabra según su categoría gramatical correspondiente. Los corpus etiquetados morfológicamente permiten trabajar con palabras no ambiguas respecto a su categoría gramatical y no presenta grandes dificultades en su utilización, pues reconoce sin ningún problema por ejemplo, la voz pasiva, los pronombres preverbiales y el “lo” seguido de adjetivo. No obstante, cuando se refiere a

un análisis de un corpus a nivel sintáctico, existen menores investigaciones que la trabajen, debido a la complejidad de la tarea y a la dificultad que presenta la automatización de la sintaxis de manera precisa en los programas. En consecuencia, los estudios de lingüística computacional que analicen la complejidad sintáctica son muy escasos por lo que los pocos trabajos que existen son muy valiosos y útiles para el campo de la lingüística y la informática.

3.1 Software Nooj

Nooj es un software creado por Max Silberztein en el año 2002, su función es realizar análisis morfológico, sintáctico y semántico en lenguas naturales. Una de las características más relevantes del programa es que es multilingüe, lo cual permite el análisis en 46 lenguas distintas, entre ellas se encuentra: inglés, español, ruso, italiano, alemán, etc. Nooj puede ser aplicado para diferentes objetivos lingüísticos: análisis de textos literarios, Investigación y extracción desde diarios o corpus técnicos, formalización de fenómenos lingüísticos y aplicaciones computacionales (análisis automático de textos).

Asimismo, incluye herramientas para crear y mantener fuentes lexicales así como gramáticas sintácticas y morfológicas. Este programa permite ejecutar el análisis morfológico mientras a la vez se desarrolla el sintáctico, lo que es bastante eficaz y rápido con los resultados. Por ejemplo, puede transformar oraciones en frases pasivas, o también permite manejar la concordancia morfosintáctica de texto analizado. Tiene la capacidad de procesar cientos de archivos de textos, además de contar con herramientas que sirven para limpiar, revisar, conservar, modificar y compartir archivos.

Referente a las tareas de reconocimiento el usuario tiene la posibilidad de crear extractores para reconocer unidades semánticas en textos que sean demasiados extensos, tales como nombre de personas, expresiones técnicas, financieras o de alguna pareja específica, fechas, ubicaciones, entre otros. Además existen dos tipos de recursos lingüísticos, el primero de ellos es el diccionario que asocia palabras o expresiones con una o más propiedades sintácticas, paradigmas de inflexiones o derivaciones (por ejemplo, cómo conjugar o nominalizar verbos), etc. Y el segundo es las gramáticas que son empleadas para representar una gran cantidad de fenómenos lingüísticos, desde niveles ortográficos y morfológicos a niveles sintagmáticos.

III. Metodología

En esta sección, se describe la metodología que sustenta este trabajo de investigación. Para ello, se presentan (i) la pregunta de investigación; (ii) Objetivos; (iii) Tipo de investigación; (iv) tareas desarrolladas, y (v) Etapas de Análisis y Resultados Obtenidos.

3.1 Pregunta de investigación

La pregunta de investigación que guiará este trabajo es ¿Cómo se presenta la complejidad sintáctica en tres géneros del ámbito de la medicina: Entradas de Wikipedia Médicas, artículos de investigación científica Médica y casos clínicos?

3.2 Objetivos

3.2.1) Objetivo general:

Describir la complejidad sintáctica que se manifiesta en los siguientes tres géneros: Entradas de Wikipedia de Medicina (EWM), Artículos de Investigación Científica Médica (AICM) y Casos Clínicos Médicos (CCM).

3.2.2 Objetivos específicos:

- Identificar la complejidad sintáctica en presentes en EWM, AICM y CCM
- Determinar el grado de complejidad que presenta cada género
- Comparar la complejidad sintáctica manifestada en cada género

3.3 Tipo y alcance de la Investigación

El enfoque de la investigación es el denominado cuantitativo, pues el propósito de ésta es describir en detalle el fenómeno estudiado. Para dicho fin se estableció una categoría de análisis que orientó los datos obtenidos y la posterior conclusión de los resultados. Por otro lado, el alcance de este estudio es exploratorio-descriptivo, ya que se plantea como objetivo principal la descripción del grado de complejidad sintáctica en textos del ámbito de la medicina, lo anterior se realiza sin plantear teorías o ideas previas, en otras palabras, sin tener antecedentes respecto a cómo se manifiesta la complejidad sintáctica en estos géneros del área médica (Hernández, Fernández y Baptista, 2006).

3.4 Tareas desarrolladas

3.4.1 Recolección del Corpus

Se trabajó con parte de los corpus compilados en el marco del Proyecto Fondecyt 11130469 (Koza, 2015; Koza, Muñoz & Mánquez, 2015): EWM y AICM, y con parte del CCM-2009, compilado por Burdiles (2012). Para efectos de la presente investigación se han extraído de cada uno 57.000 mil palabras aproximadamente, lo que da un total de 171.000 palabras analizadas. A continuación se describen brevemente cada uno de ellos.

Las entradas EWM fueron extraídas directamente del sitio web, se partió de la entrada “Medicina” y se continuó con los enlaces que se encontraban de especialidades médicas, lo que suma un total de 60 entradas, las cuales son las siguientes:

Alergología	Cirugía general y del aparato digestivo	Epidemiología	Hepatología	Medicina física y rehabilitación	Neonatología
Análisis clínicos	Cirugía oral y maxilofacial	Estomatología y odontología	Hidrología médica	Medicina forense	Neurocirugía
Anatomía patológica	Cirugía ortopédica y traumatología	Farmacología clínica	Infectología	Medicina intensiva	Neumología
Anestesiología y reanimación	Cirugía pediátrica	Gastroenterología	Inmunología	Medicina interna	Neurofisiología clínica
Angiología y cirugía vascular	Cirugía plástica	Genética	Medicina de emergencia	Medicina nuclear	Neurología
Bioquímica clínica	Cirugía torácica	Geriatria	Medicina del trabajo	Medicina preventiva	Obstetricia
Cardiología	Dermatología	Ginecología	Medicina deportiva	Microbiología y parasitología	Oftalmología
Cirugía cardiovascular	Endocrinología y nutrición	Hematología	Medicina familiar y comunitaria	Nefrología	Oncología médica

Oncología radioterápica	Otorrinolaringología	Pediatría	Proctología	Psiquiatría	Radiología o radiodiagnóstico
Reumatología	Salud pública	Traumatología	Toxicología	Urología	Cardiología

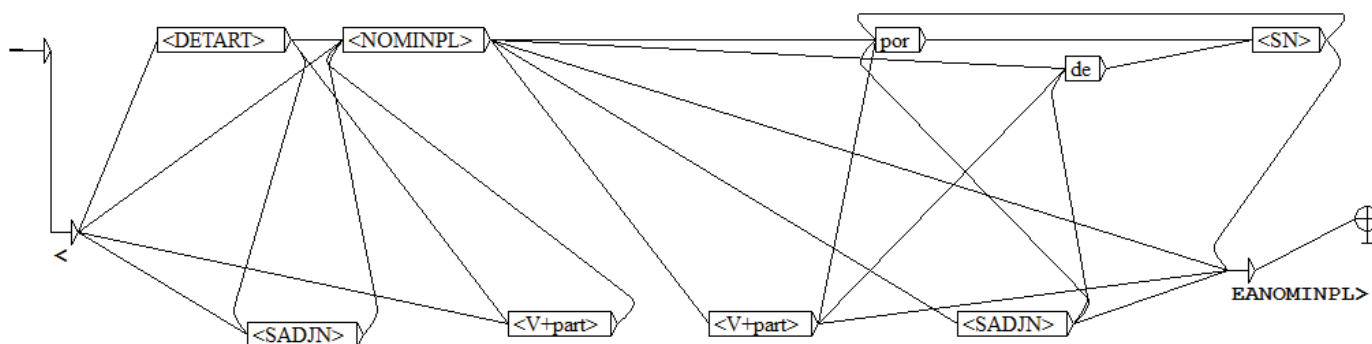
El CCM-2009 contiene casos clínicos aparecidos entre los años 1999 y 2008, suma un total de 969 textos y 801.224 palabras. Los textos reunidos pertenecen a las especialidades de Parasitología; Neuropsiquiatría; Enfermedades Respiratorias; Otorrinonaringología y Cirugía; Infectología; Pediatría; Obstetricia; Cirugía y Ciencias Biomédicas, y Medicina Interna y Especialidades derivadas. El corpus de artículos de información científica médica (AICM-2014) fue construido en el seno de este proyecto, extraído de revistas médicas chilenas indexadas en la plataforma Scielo. Incluye artículos de investigación aparecidos entre los años 2006 y 2013, suma un total de 305 textos y 1.333.281 palabras. Incluye las especialidades de Cardiología, Farmacología, Neurología y Oncología. La diferencia con la cantidad de textos del CCM-2009 se debe a que los casos clínicos presentan una extensión mucho más breve que los artículos. Cabe destacar que el corpus es bastante amplio en cantidad y especialidades, por ende para efectos de la presente investigación, al tratarse de una primera aproximación al fenómeno de la complejidad sintáctica en textos del área médica, se optó por tomar 57.000 palabras de cada género.

3.4.2 Trabajo computacional

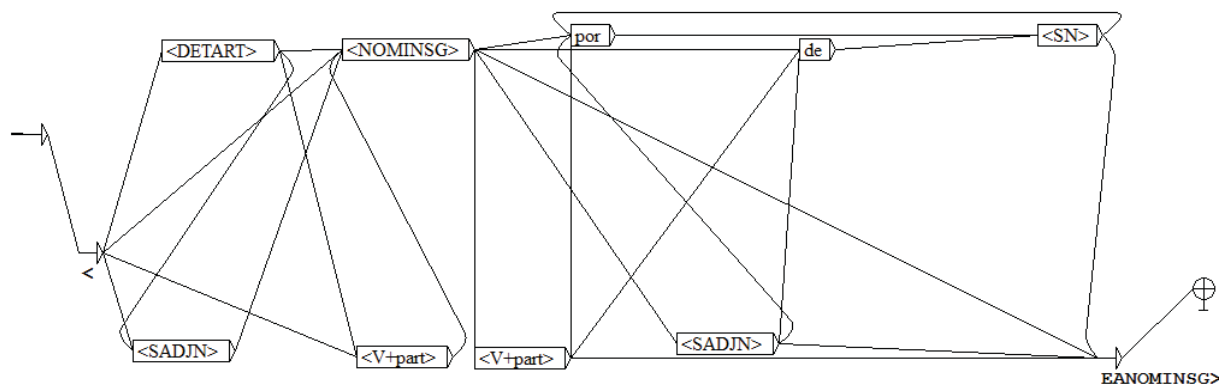
Para realizar el análisis automático del corpus se utilizó el software Nooj con el cual se crearon gramáticas adecuadas para detectar los elementos sintácticamente complejos seleccionados. Por ello fueron creadas 4 gramáticas. Una gramática para detectar la nominalización, otra para la detección de la voz pasiva, una tercera para la subordinada completiva o sustantiva y la cuarta gramática para la subordinada relativa o adjetiva. Las cuatro gramáticas son ilustradas mediante las siguientes imágenes.

3.4.2.1 Gramática para la detección de la Nominalización plural y singular

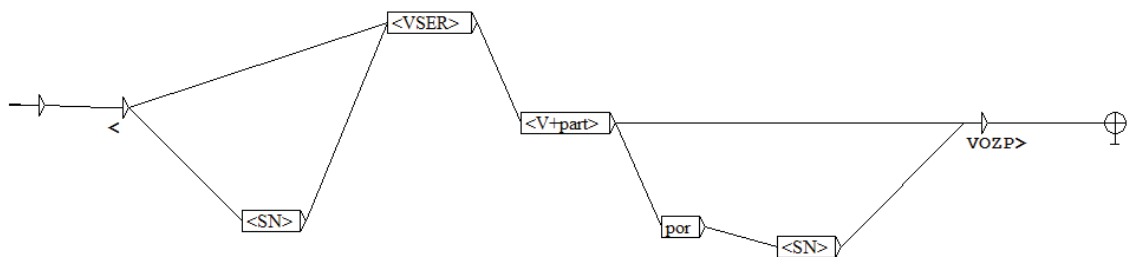
Nominalización plural



Nominalización singular

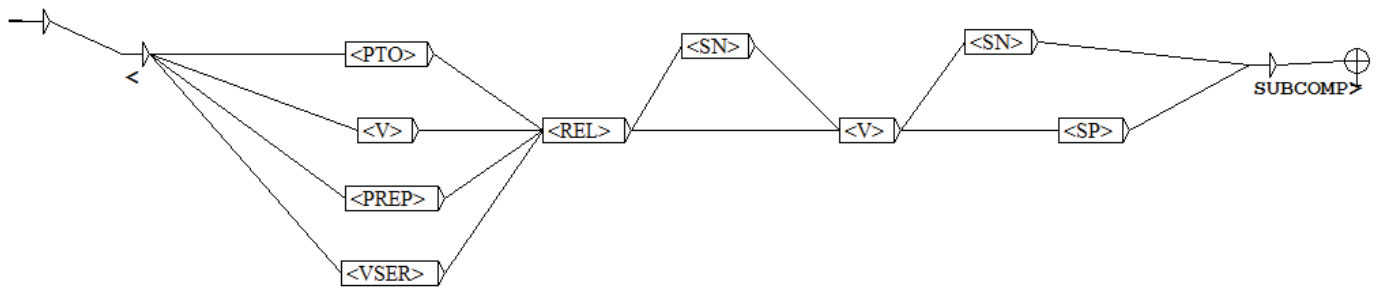


3.4.2.2 Gramática para la detección de la voz pasiva

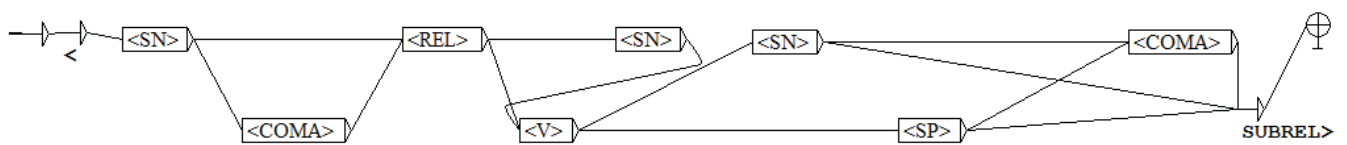


Para efectos de este trabajo e no se considera la pasiva con sé, la cual quedará para trabajos futuros

3.4.2.3 Gramática subordinada completiva o sustantiva



3.4.2.4 Gramática subordinada relativa o adjetiva



3.5 Etapas de análisis y resultados obtenidos

3.5.1 identificación de la complejidad sintáctica

Para evaluar la confiabilidad del output de Nooj, se lo probó en un fragmento del corpus EWM (12.000 palabras) y se obtuvieron los resultados lo suficientemente adecuados con un 89,40% de precisión, un 92,20 de cobertura y 90,82% de medida F. Los resultados muestran que el algoritmo de Nooj es bastante confiable y de alta precisión en la detección de los elementos analizados. Posteriormente a la evaluación manual del software, se prosiguió con el análisis computacional para lo cual fueron utilizadas las gramáticas antes ilustradas.

3.5.2 Clasificación de la complejidad sintáctica

La clasificación de la complejidad sintáctica se estableció a partir del número de apariciones de los tres elementos analizados: nominalización, voz pasiva y subordinación. Esto, a fin de determinar qué elemento es utilizado con mayor o menor frecuencia según el género del área médica y por ende verificar si los géneros analizados difieren en cuanto a la complejidad sintáctica que presentan.

A tales efectos, se describió el número de apariciones de cada elemento, posteriormente se extrajo el total de apariciones de todos los elementos, generando como resultado la complejidad sintáctica de cada texto a partir del número de apariciones. Finalmente, se llevó a cabo una comparación de los tres géneros estudiados, con el fin de establecer una correlación entre un género y otro.

IV. Resultados Obtenidos

El presente capítulo reporta los resultados del análisis del corpus, que tuvo los resultados que se detallan a continuación :

4.1 Resultados

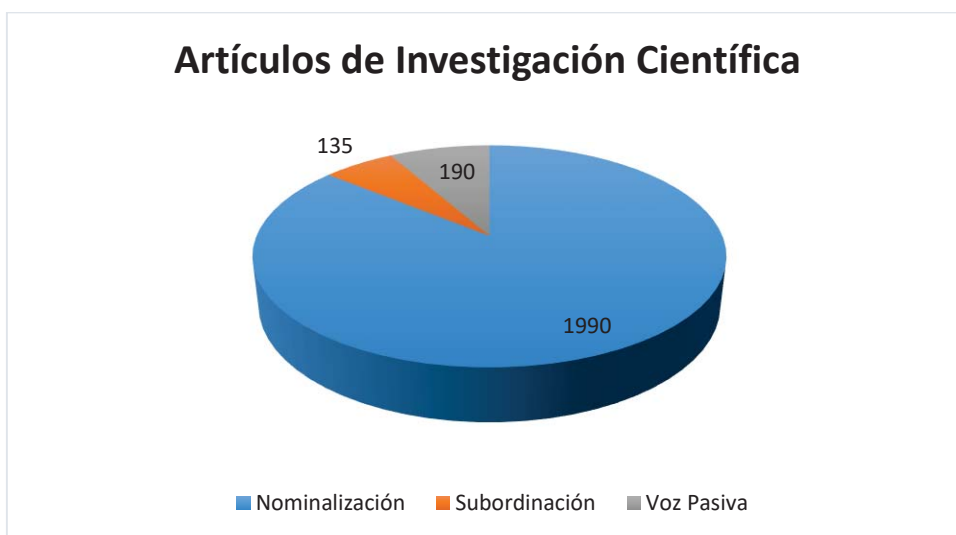
4.1.1 Resultados entrada Wikipedia

El corpus Entrada Wikipedia presentó 1.743 casos de nominalización, 220 casos de subordinación y finalmente 141 apariciones de voz pasiva. Cabe destacar que de un total de 57.000 mil palabras analizadas, este género manifestó un total de 2.104 casos de elementos sintácticamente complejos.



4.1.2 Resultados Artículos de Investigación Científica (AICM)

En cuanto a las estructuras sintácticamente complejas que se manifiestan en los artículos de investigación científica analizados, se aprecia 1.990 casos de nominalización, 135 casos de subordinación para finalizar con 190 apariciones de voz pasiva. Lo anterior muestra que los AICM presentan 2.315 casos de estructuras sintácticamente complejas en su redacción.



4.1.3 Resultados Casos Clínicos

Por último, en los Casos Clínicos se distingue 200 casos de nominalización, mientras que la subordinación presenta casi la mitad con 112 casos, y la voz pasiva aparece en 100 oportunidades. Esto entrega un total de 412 elementos sintácticamente complejos que constituyen un caso clínico de 57.000 palabras.



4.2 Descripción y comparación de la complejidad sintáctica

Ahora bien, para una mayor comprensión se muestra un cuadro que permite visualizar con mayor claridad la comparación de los resultados entre los tres géneros.

Corpus	Nominalización	Subordinación	Voz Pasiva	Total
EWM	1743	220	141	2104
AICM	1990	135	190	2315
CCM	200	112	100	412

Como se aprecia en el cuadro anterior, respecto a la nominalización el género que presenta más casos son los artículos de investigación científica con 1990 apariciones, le siguen la Entrada de Wikipedia con 1743 casos y en último lugar los casos clínicos que solo manifiesta 200 apariciones, un número considerablemente menor en relación con los dos género antes nombrados. En cuanto a la Subordinación los resultados son distintos, ya que esta vez la Entrada de Wikipedia manifiesta 220 casos de Subordinación a diferencia de las 135 apariciones de los Artículos de Investigación, al igual que la vez anterior los Casos Clínicos nuevamente presentan un número menor con 112 casos de Subordinación. En cuanto a la Voz Pasiva nuevamente los Artículos manifiestan un mayor número con 190 casos, mientras que las Entradas Wikipedia poseen 141 casos, lo cual no varía demasiado de la Entradas

Wikipedia. Finalmente, los casos clínicos muestran 100 casos de voz pasiva siendo nuevamente el menor número de apariciones de un elemento sintáctico.

Cabe destacar que frente a la Voz Pasiva los tres géneros manifestaron resultados similares. Un caso similar ocurre con la nominalización entre los artículos de investigación y las entradas Wikipedia, ya que no muestran grandes diferencias en el número de apariciones. Ahora bien, para la subordinación se produce una leve diferencia, ya que las entradas Wikipedia presentan una cantidad mayor con 100 apariciones más. De esta manera, es necesario señalar cuáles son las implicancias de estos resultados en cuanto a complejidad, destacando las semejanzas y diferencias. A continuación se describe un análisis a nivel de los resultados obtenidos.

4.3 Análisis

Como se logra apreciar a partir del número de apariciones de cada elemento analizado, el género con una elevada complejidad sintáctica en su redacción correspondería a los Artículos de Investigación Científica, pues manifiestan un número de elementos sintácticamente complejos mayor que el resto. No obstante, es preciso señalar que si bien en cantidad los AICM presentan una mayor complejidad, la diferencia es mínima con las Entradas Wikipedia, ya que ambos géneros están conformados sintácticamente con un número de estructuras similares. Esto quizá se podría relacionar con el propósito que persigue ambos géneros, ya que el fin del AIC es dar a conocer resultados de investigaciones en el área científica para aspirar al reconocimiento.

A su vez, las EW tienen como propósito informar a la comunidad de manera rápida, comprensible y precisa sobre temas científicos. De este modo, un punto en común de ambos géneros es el público al que va dirigido los textos, varían desde lectores expertos a no expertos por ende la explicación y profundidad en el contenido deberá ser mayor para que sea comprensible a todo quien lo lea, en consecuencia al emplear más oraciones y necesitar mayores recursos sintácticos la complejidad sintáctica aumentaría.

Por otro lado, en un nivel medio de complejidad se encuentran las Entradas de Wikipedia, ya que de los tres géneros, este posee un total de 2.315 casos en total de estructuras sintácticas complejas. Se destaca que la diferencia con los Artículos de Investigación Científica es muy

baja como ya se señaló con anterioridad, la similitud en el objetivo de los dos géneros influye en la sintaxis de los textos. Finalmente, el género que manifestaría una complejidad sintáctica baja es el Caso Clínico, debido a que en su estructura solo se manifestaron en total 412 casos de complejidad sintáctica, cifra baja en comparación a los otros dos géneros trabajos.

Esta minoría se debe a que el propósito del Caso Clínico es informar específicamente sobre la evolución de alguna enfermedad o los procedimientos que se llevaron a cabo en un procedimiento médico, por ello el público al cual va dirigido este género es restringido a expertos en medicina, ya sea médicos o estudiantes que se van integrando al área. De este modo, la causa principal sería la extensión de los textos. Por otro lado, la modalidad del género: descriptivo y narrativo, concentra gran cantidad de información nueva, puesto que se trata de la descripción de una enfermedad poco usual o, directamente, desconocida.

A partir de lo señalado anteriormente cabe mencionar un alcance, el primer análisis que se llevó a cabo con los resultados abordó el número de apariciones por estructura sintáctica. Ahora bien, para fines del presente trabajo se decidió otorgar un grado de dificultad diferente a cada estructura sintáctica de acuerdo con el número de transformaciones que presenta. De acuerdo con esto y según la oración Kernel: “Juan destruyó la ciudad” las estructuras sintácticas quedarían con el siguiente grado de complejidad.

Estructuras sintácticas	Número de transformaciones	Ejemplo
Nominalización	4 : T1: complemento del nombre T2: Complemento agente T3 y T4: cambio del verbo a sustantivo y cambio de categoría	La destrucción de la ciudad por Juan
Voz pasiva	3: T1: Complemento agente T2: cambio del sujeto	La ciudad fue destruida por Juan

	T3:cambio en la voz del verbo	
Subordinación	1: T1:Desplazamiento de un elemento sintáctico	La ciudad que Juan destruyó

De esta manera, al considerar que cada elemento sintáctico es distinto uno de otro se pueden diferenciar según su grado de complejidad. Como resultado, la nominalización debido a que presenta 4 transformaciones en su estructura tiene un grado de complejidad mayor. Asimismo, la voz pasiva en su conformación presenta tres transformaciones, que la cataloga con un grado de complejidad intermedio y finalmente la subordinación que solo presenta una transformación es evaluada con el grado de complejidad más bajo

Tomando en cuenta esto, los resultados según el grado de complejidad serían los siguientes.

El género que presentó un alto grado de complejidad es el AICM, el cual presentó 1990 casos de nominalización. El resultado en complejidad se mantuvo, ya que en el análisis anterior, según el número de apariciones de cada estructura nuevamente el AICM presenta un mayor uso de estructuras sintácticas, en este caso un mayor grado de complejidad al utilizar una estructura sintáctica compuesta por cuatro transformaciones. Por otro lado, las EWM manifestaron un total de 1743 nominalizaciones, monto que si bien es menor, la diferencia entre ambas (AICM y EWM) es mínima. Situación distinta sucede con el Caso clínico ya que posee menos nominalizaciones (112) apariciones, diferencia considerable, en comparación a los otros dos géneros.

En la misma línea, el grado de complejidad medio es para la voz pasiva, según esta estructura el texto que más la emplea es el Artículo de Investigación con 190 casos, a diferencia de los 141 de las entradas Wikipedia y los 100 de los Casos Clínicos. De esta forma, en un grado de complejidad media-alta se encontraría los AICM ya que en su producción emplea mayor número de nominalizaciones y también mayor número de voz pasiva, consideradas las dos estructuras con un número de transformaciones superior a dos. Respecto al género que es

catalogado con el grado de complejidad menor es la EWM, ya que presenta un mayor uso de subordinación, estructura sintáctica que exige únicamente una transformación.

A modo de síntesis, se logra distinguir que en cuanto al grado de complejidad, los resultados son los siguientes. El género que emplea tiene un alto grado de complejidad es el AICM, ya que presenta un mayor uso de la nominalizaciones. Asimismo, en un grado de complejidad media-alta nuevamente se presente el AICM ya que también utiliza en gran cantidad la voz pasiva, que posee un grado de complejidad intermedio. Por último el género con menor grado de complejidad es la EWM, ya que utiliza en gran mayoría la subordinación, estructura que solo requiere una transformación, en comparación de las otras estructuras sintáctica que exigen más de dos.

Para finalizar el análisis de los datos obtenidos, cabe señalar respecto a la complejidad sintáctica en tres textos del área médica, no existe una gran diferencia en el número de estructuras sintácticas empleadas. Entradas Wikipedia y Artículos de investigación Científica poseen un número bastante similar de elementos sintácticos. La diferencia significativa se produce con el Caso Clínico, el cual en las tres variables analizadas manifestó un número bajo de apariciones. Como ya se dijo anteriormente, se debe al propósito que persigue cada género y al lector ideal de cada texto.

La complejidad sintáctica en textos médicos se muestra como un elemento que claramente varía según el género, siendo un factor importante el propósito comunicativo de cada texto, tal como señala Hunt (1970) la naturaleza del texto será un factor relevante en el desarrollo de la complejidad sintáctica. Esto se vio reflejado en la comparación de la complejidad en las Entradas Wikipedia, los Artículos de Investigación Científica y los Casos Clínicos, en donde los dos primeros manifestaron una complejidad sintáctica mayor, con leve grado de diferencia, mientras que la complejidad en los Casos Clínicos fue considerablemente menor. Otro postulado que es preciso retomar son los de Navarro, Hernández & Rodríguez (1994) quienes señalan el abuso de la voz pasiva en textos médicos de origen inglés; en el caso del español no sucede lo mismo, ya que de las tres estructuras analizadas, la voz pasiva es el

elemento menos empleado en los géneros, lo que coincide con lo expuesto por los autores al señalar que en español latino la preferencia es por el uso de la voz activa.

V. Conclusiones

En la presente investigación, se describió la complejidad sintáctica en tres géneros del ámbito de la medicina, con el propósito de caracterizar y analizar el fenómeno, a partir del número de apariciones de tres estructuras sintácticas: la nominalización, la subordinación y la voz pasiva. Para ello se utilizó un corpus de textos del área médica compuesto por Entradas Wikipedia Medicas, Artículos de Investigación Científica Médicos y Casos Clínicos Médicos, recolectadas dentro del marco del proyecto Fondecyt 11130469 (Koza, 2015). En esta ocasión se utilizó el software Nooj (Silberztein, 2016) como herramienta de detección de las estructuras sintácticas, esto se justifica por el gran alcance y precisión de la herramienta, entregando resultados confiables.

Se identificó la complejidad sintáctica y se clasificó según el número de apariciones de los tres elementos analizados. Posteriormente, se llevó a cabo un análisis en dos niveles: uno relacionado con el número de apariciones de las tres estructuras, el otro respecto al grado de complejidad de cada estructura sintáctica. Entregado de esta manera resultados que variaron según cada análisis.

De esta manera, y cumpliendo con los objetivos específicos de esta investigación, se puede observar que en el primer análisis, relacionado con el número de apariciones de cada estructura, se observó que el género que posee una mayor complejidad sintáctica era la Entrada de Wikipedia, si a esto se le suma que el género Artículo de Investigación Científica es el segundo con una complejidad intermedia, se destaca que la diferencia entre ambos resultados fue mínima. Por lo tanto, la complejidad sintáctica en estos dos textos del ámbito médico es de naturaleza similar. Situación que no sucede con el Caso Clínico, ya que de los tres fue el género que presentó una disminuida complejidad, al estructurarse con menor cantidad de elementos sintácticamente complejos.

Por otra parte, en el segundo análisis, esta vez del grado de complejidad de cada estructura los resultados fueron diferentes. Se descubrió que el género con un grado de complejidad mayor fue el Artículo de Investigación Científica, ya que en su escritura presentaba un

número elevado de nominalizaciones, consideradas complejas en su uso, pues requieren cuatro transformaciones en su estructura. Por el contrario, el texto que presentó menor cantidad de nominalizaciones fue el Caso Clínico por ende su estructuración no es tan compleja como en el caso de los AICM. Referido a la voz pasiva, clasificado con un nivel de dificultad en su aplicación intermedia, el género que presentó más usos de este elemento fue el Artículo de Investigación Científica, del mismo modo sucede con la nominalización, catalogada con un grado de complejidad mayor, nuevamente el Artículo de Investigación presentó un uso mayor, por ende se considera a los AICM con un grado de complejidad en su uso medio alto ya que manifiesta en la mayoría de su redacción dos de los elementos catalogado como alto e intermedio en el uso. Finalmente, respecto al Caso Clínico, manifestó un número elevado solo en la nominalización, pero que en comparación a los otros dos géneros sigue siendo muy baja, lo mismo sucede con las otras dos estructuras, por ende el grado de complejidad en el Caso Clínico vuelve a ser nuevamente el menor.

Además, se observó que un factor relevante en la complejidad sintáctica lo conforma la naturaleza del texto, ya que los dos géneros que manifestaron una alta complejidad sintáctica según el número de apariciones, fue las Entradas de Wikipedia y Artículos de Investigación, justamente géneros que tienen como objetivo informar sobre variados temas del área científica, el primero se caracteriza por entregar información de forma clara y comprensible, el segundo por entregar los resultados de una investigación buscando el reconocimiento en la disciplina. Mientras que el Caso Clínico que manifestó una baja complejidad sintáctica tiene como fin informar sobre un tema específico ya sea una enfermedad u otro padecimiento. Lo anterior, también se relaciona con el público al que va dirigido el texto, ya que en el caso de las Entradas Wikipedia y los Artículos de Investigación Científica, van dirigido tanto a público especializado como a público no experto, pero sobre todo a público no experto, por lo tanto la creación de los textos debe ser lo suficientemente comprensibles para toda persona, empleando mayores oraciones para explicar, aumentando el número de estructuras sintácticas en dicha labor. En cambio, el público de los Casos Clínicos tiende a ser del tipo experto, por ende se asume que el público conoce y maneja varios de los términos empleados, economizando oraciones, por lo tanto disminuyen la extensión del texto, disminuyendo también las posibilidades de utilizar más estructuras sintácticas complejas.

Finalmente, es necesario señalar que la presente investigación como estudio de aproximación al fenómeno pretendió sentar algunas bases respecto al fenómeno estudiado y no cabe duda que aún queda mucho por ahondar en esta temática. Un mirada bastante interesante de analizar a futuro, sería abordar la complejidad sintáctica analizando el número de transformaciones que se presentan por oración, esto brindaría información bastante innovadora, (Hunt, 1970) respecto a cuantas transformaciones se manifiestan por oración, detallando mucho más la complejidad de los textos médicos. También, otro aporte se podría observar agregando más estructuras sintácticas al análisis de textos médicos, esto entregaría información más completa respecto al comportamiento sintáctico del ámbito de la medicina. Finalmente, una última proyección de estudio, se apreciaría en abordar un análisis de otros ámbitos del géneros médicos como recetas, fichas de pacientes , realizando una comparación entre géneros escritos únicamente por médicos y otros escritos por editores no necesariamente del área médica, con esto se lograría realizar una comparación entre madurez sintáctica entre escritores expertos y legos (Hunt,1970).

VI. Bibliografía

- Bonino, R. (2015). Una propuesta para el tratamiento de los enclíticos en NooJ. *Nfosur Revista*, 7, 31-40.
- Berman, R., & SloBin, D. (1994) Relating events in narrative: a crosslinguistic development study, mahwah, nJ: Lawrence erlbaum.
- Burdiles, G. (2012). *Descripción de la organización retórica del género caso clínico de la medicina a partir del corpus ccm-2009*. Tesis Doctoral. Chile: Pontificia Universidad Católica de Valparaíso.
- Carlino, P. (2013). Alfabetización académica diez años después. *Revista mexicana de investigación educativa*, 18(57), 355-381.
- Carrasco, F. (1973). Sobre el formante de “voz pasiva” en español. *Revista Española de Lingüística*, 3, 333-342.
- Crespo, N., Alfaro, P., & Góngora, B. (2011). La medición de la sintaxis: evolución de un concepto. *Onomazéin*, 24, 155-172.
- Córtés, M., & Cruz, M. (2001). El análisis del Género Artículo de Investigación. *Revista de Lenguas para Fines Específicos*, 7, 32-50.
- Crespo, N., Benítez, R., & Ramos, C. (2005). Una propuesta de medición de las inferencias en la comprensión del discurso oral”, en mauricio pilieux (ed): Los Contextos del Lenguaje, Valdivia: Universidad Austral de Chile, 142-151.
- Duseel, I. (2010). Cultura participativa y producción de los saberes: reflexiones sobre los usos pedagógicos de Wikipedia. en: Guías para usar la Wikipedia en el aula. Buenos Aires, Educ.ar, disponible en: <http://wikipediaenelaula.educ.ar/cultura-participativa.html>.

- Ferrero, P. (2011). Definición y análisis de parámetros lingüísticos cuantitativos para herramientas automáticas de evaluación aplicables al Español como Lengua Extranjera. Tesis Doctoral. Madrid: Universidad Autónoma de Madrid.
- Gómez Guinovart, J., & Palomar, M. (coords.) (1998): Lengua y Tecnologías de la Información, número monográfico de Novática, 133.
- Gutiérrez-Clellen., & Hofstetter, R. (1994). "Syntactic complexity in spanish narratives", *Journal of Speech and Hearing Research* 37, 645-654.
- Herrero, J. (1992). Algunas consideraciones en torno al complemento agente. *Revista Española de Lingüística*, 22, 339-359.
- Halliday, M. (1979). El lenguaje como semiótica social, México DF: *Fondo de Cultura económica*.
- Hunt, K. (1965). Grammatical structures written at three grade levels. National Council of Teachers of English Research Report N° 3, Champaign, ill.: national Council of Teachers of english. —, 1970: "Syntactic maturity in Schoolchildren and Adults", *Monographs of The Society for Research* 35 (1), 1-67.
- Koza, W. (2012). *Los signos de puntuación en el análisis automático de textos. El caso de la coma*. Tesis doctoral. Rosario: Universidad Nacional de Rosario
- Koza, Walter. (2015). Propuesta de extracción automática de candidatos a término del dominio medico procesando información lingüística. Descripción y evaluación de resultados. *Alfa. Revista de Lingüística*, 59(1) 113-127.

- Koza, W. (2014). Funciones gramaticales de la coma: Clasificación e implantación computacional. *Revista signos*, 47(86), 412-434.
- Merino, A. (2009). Cómo escribir documentos científicos (parte 2). Caso Clínico. *Salud en Tabasco*, 5(1), 852-853.
- Moreno, A. (1998). *Lingüística Computacional*. Madrid Síntesis
- Meneses, A., Ow, M., & Benítez, R. (2012). Complejidad Sintáctica: ¿modalidad comunicativa o tipo textual? Estudio de casos de producciones textuales de estudiantes de 5° básico. *Onomázein*, 1, 65-93.
- Navarro, F., Hernández, F., & Rodríguez, L. (1993). Uso y abuso de la pasiva en el lenguaje médico escrito. *Medicina Clínica*, 103, 461-464.
- Nefedova, N. (2005). La complejidad Sintáctica como Recurso del despertar de la reflexión (Refleksia). Tesis doctoral
- Parodi, G. (Ed.) (2008). Géneros académicos y géneros profesionales: Accesos discursivos para saber y hacer. Valparaíso: *Ediciones Universitarias de Valparaíso*.
- Rodríguez, A. (2015). Análisis de la complejidad sintáctica de textos producidos por alumnos con baja visión, con ceguera y con visión normal. *Revista Iberoamericana de Educación*, 1. 1681-5653.
- Saggion, H., Marimon, M., & Férres, D. (2015). Simplificación automática de textos para la accesibilidad de colectivos con discapacidad: experiencias para el español y el inglés. *IX Jornadas Científicas Internacionales de Investigación sobre Personas con Discapacidad*. Madrid: *Universidad de Salamanca*.
- Salvador, F. (1985). Los índices de complejidad sintáctica, instrumentos de evaluación de la expresión escrita: estudio experimental en el ciclo medio de E.G.B. Enseñanza & Teaching. *Revista interuniversitaria de didáctica*, 3, 59-82.

Silberztein, M. (2005). NooJ: A Linguistic Annotation System For Corpus

Processing, en línea, disponible en: <http://www.aclweb.org/anthology/H05-2>

Véliz, M. (1988). Evaluación de la madurez sintáctica en el discurso escrito, *Revista de Lingüística Teórica y Aplicada* 26, 105-141