

Conocer la lengua a través de los corpus: la herramienta METOOL, retos para el análisis de los marcadores discursivos

MARÍA LUISA CARRIÓ-PASTOR

Catedrática de Universidad
Universitat Politècnica de València
Departamento de Lingüística Aplicada
Camino de Vera, 14
46022 Valencia
E-mail: lcarrio@upv.es
Tel. 963877530

CONOCER LA LENGUA A TRAVÉS DE LOS CORPUS: LA HERRAMIENTA METOOL, RETOS PARA EL ANÁLISIS DE LOS MARCADORES DISCURSIVOS

RESUMEN: La identificación y etiquetado de los marcadores discursivos que se utilizan en la lengua en el registro de artículos científicos representan un reto en estos momentos, dada su naturaleza versátil. Es por ello que hemos creado una herramienta que identifica los marcadores discursivos y orienta a los investigadores en el uso del metadiscursivo. Para llevar a cabo este proyecto hemos seleccionado tres disciplinas distintas, la ingeniería, la medicina y la lingüística. La finalidad del proyecto es doble: primero, caracterizar las estrategias retóricas que identifican el discurso académico y, después, identificar los patrones retóricos entre disciplinas. Para ello se ha diseñado la herramienta 'METOOL' con la Universidad de Wolverhampton. A través del análisis de corpus y de la medición estadística de la capacidad de involucrar al lector y de convencerlo de los argumentos que se esgrimen, se pueden identificar, por ejemplo, estrategias de persuasión o mitigación y proponer alternativas.

PALABRAS CLAVES: corpus; metadiscursivo; marcadores discursivos; herramienta; *METOOL*.

SUMARIO: 1. Introducción. 2. El análisis. 3. Resultados. 4. Conclusiones.

UNDERSTANDING LANGUAGE THROUGH CORPORA: METOOL, CHALLENGES FOR THE ANALYSIS OF DISCURSIVE MARKERS

ABSTRACT: The identification and labelling of the discursive markers used in scientific papers represent a challenge, given their nature. That is why we have created a tool that identifies discursive markers and guides researchers in the use of metadiscourse. To carry out this project we have selected three different disciplines, engineering, medicine and linguistics. The purpose of the project is twofold: first, to characterise the rhetorical strategies that identify academic discourse and then, to identify the rhetorical patterns among disciplines. For this, 'METOOL' was designed in collaboration with the University of Wolverhampton. Through corpus analysis and the identification of statistical frequencies, this tool analyses the ability to involve readers and convince them with, for example, persuasion or mitigation strategies.

KEY WORDS: corpus; metadiscourse; discourse markers; tool, *METOOL*.

SUMMARY: 1. Introduction. 2. The analysis. 3. Results. 4. Conclusions.

CONNAITRE LA LANGUE A TRAVERS LE CORPUS : L'OUTIL METOOL, ENJEUX POUR L'ANALYSE DES MARQUEURS DISCURSIFS

RÉSUMÉ : L'identification et l'étiquetage des marqueurs discursifs utilisés dans la langue dans le registre des articles scientifiques représente un défi à l'heure actuelle, étant donné sa nature polyvalente. C'est pourquoi nous avons créé un outil qui identifie les marqueurs discursifs et guide les chercheurs dans l'utilisation du metadiscours. Pour mener à bien ce projet, nous avons sélectionné trois disciplines différentes, l'ingénierie, la médecine et la linguistique. Le but du projet est double : premièrement, caractériser les stratégies rhétoriques qui identifient le discours académique et, plus tard, identifier les schémas rhétoriques entre les disciplines. Pour cela, l'outil « *METOOL* » a été conçu avec l'Université de Wolverhampton. Grâce à l'analyse de corpus et à la mesure statistique de la capacité à engager le lecteur et à le convaincre des arguments avancés, il est possible d'identifier, par exemple, des stratégies de persuasion ou d'atténuation et de proposer des alternatives.

MOTS CLÉS : corpus ; métadiscours ; marqueurs discursifs ; outil ; *METOOL*.

SOMMAIRE : 1. Introduction. 2. L'analyse. 3. Résultats. 4. Conclusions.

Fecha de Recepción 10/07/2020
Fecha de Revisión 24/07/2020
Fecha de Aceptación 01/09/2020
Fecha de Publicación 01/12/2020

DOI: <http://dx.doi.org/10.25267/Pragmalinguistica.2020.i28.13>

Conocer la lengua a través de los corpus: la herramienta METOOL, retos para el análisis de los marcadores discursivos

MARIA LUISA CARRIÓ-PASTOR

1. INTRODUCCIÓN

La Lingüística Aplicada se ha ocupado, tradicionalmente, del análisis de la lengua para buscar patrones que nos ayuden a entender la manera en la que organizamos los signos lingüísticos. Por ello, muchos investigadores adoptan una metodología basada en análisis de corpus para poder identificar esos patrones con más exactitud, apoyándose en datos objetivos. Este tipo de método es bastante común en los análisis de las colocaciones (Carrió-Pastor, 2017), en los estudios que se centran en la identificación de patrones (Carrió-Pastor, 2015, 2016b; Carrió-Pastor y Muñiz, 2015) o en trabajos que identifican funciones que realizan ciertos signos lingüísticos en un contexto determinado (Carrió-Pastor, 2014; Alonso-Almeida y Carrió-Pastor, 2015).

Desde un punto de vista pragmático, los signos lingüísticos se pueden interpretar de formas distintas según el contexto, como bien apuntaba Halliday (1998), ya que el significado de las palabras no es monosémico, si no que muchas palabras se manifiestan de forma polisémica. Este hecho fue observado por la Escuela de Birmingham en los años 70 del siglo pasado, siendo Sinclair (1987) uno de los pioneros en este campo, ya que comprendió la importancia de los datos objetivos y de poder demostrar mediante frecuencias, los cambios que surgen en las distintas lenguas. Por lo tanto, la lingüística de corpus fue adoptada en numerosos estudios para demostrar hipótesis sobre la evolución de la lengua. Varios estudiosos adoptaron este método para realizar análisis sobre la variación diacrónica del lenguaje, pero en la actualidad se utiliza con frecuencia para identificar patrones que van cambiando, realizando con ello estudios de corpus sobre la variación sincrónica. Ello se debe, fundamentalmente, al uso global de la lengua, ya que Internet ha puesto al abasto un gran abanico de posibilidades para que podamos comunicarnos de forma efectiva y rápida, haciendo que la lengua evolucione más rápidamente y, con ello, que existan una mayor variedad de patrones lingüísticos que cambian según el contexto.

Por esos motivos, podemos afirmar que la lingüística de corpus nos aporta datos valiosos que nos permiten identificar nuevos significados de las palabras, sobre todo en un entorno tan cambiante como el actual. Hoy en día, la comunicación ya no se realiza de forma unilateral, si no que estamos continuamente con la lengua en una multitud de medios. Por un lado, en la comunicación escrita en formato papel (periódicos, libros, informes...), por otra, en la comunicación escrita virtual (redes sociales, foros de debate, WhatsApp, mensajes de texto...), así como la comunicación oral (entrevistas, programas televisivos, noticias, discursos políticos...). Esta multitud de medios de comunicación está propiciando que la lengua varíe y progrese de

forma mucho más rápido de lo que ha hecho en los últimos siglos y, por ello, existe la necesidad de identificar los patrones lingüísticos que son fruto de esta rápida evolución del lenguaje.

Mediante el análisis de corpus se aportan datos objetivos que apoyan nuestra investigación, asemejando los estudios lingüísticos a las ramas del saber que tradicionalmente se han basado en la observación y en el análisis de casos como, por ejemplo, la medicina, la ingeniería, etc. Los investigadores en lingüística, de esta forma, no explican opiniones sobre muestras reducidas, sino que, con el análisis de corpus, los lingüistas se basan en muestras objetivas de lenguaje real, con ello, los resultados son demostrables y, por lo tanto, verificables al basarse en datos objetivos.

Basándome en estos principios, este capítulo se centra en una aplicación de la lingüística de corpus para el estudio de los marcadores discursivos que son la base de los patrones lingüísticos que organizan la comunicación desde un punto de vista textual y de la interacción con el lector (Thompson, 2001; Hyland, 2005). Me baso, en este caso, en una interpretación funcional del lenguaje, clasificando los marcadores según su función en la comunicación y dentro de los análisis dedicados al metadiscurso. Existen varias definiciones de metadiscurso, entre ellas destacamos las que menciona Bekke (2005, p. 8):

Según Connor (1996: 94), "el metadiscurso le permite al escritor mostrar al lector cómo las partes del texto se relacionan entre sí y expresar su propia evaluación del contenido y su actitud". Para Fairclough (1992: 122), el metadiscurso se define como el lenguaje que se utiliza en un texto "cuando el hablante/escritor está situado por encima y fuera de su propio discurso y está en una posición de controlarlo y manipularlo".

Así, pues, podemos aglutinar todos los elementos lingüísticos que organizan el lenguaje y ayudan a conectar el emisor con el receptor en dos categorías, es decir, la categoría textual y la categoría interpersonal (Thompson, 2001; Hyland, 2005). Éstas se diferencian en que la primera está compuesta por todas aquellas palabras que utilizamos para otorgarle cohesión y coherencia al discurso, mientras que la segunda está compuesta por palabras o sintagmas que nos ayudan a expresar al oyente o lector lo que deseamos transmitir y, con ello, convencerlo de nuestras ideas o creencias. Varios son los estudios de los marcadores discursivos, principalmente centrado en el análisis de la lengua inglesa (Hyland, 2005; Mur Dueñas, 2011; Carrió-Pastor, 2016a, 2016b, 2016c, 2018, 2019) y la española (Briz, 2001, 2007; Albelda, 2008; Albelda y Briz, 2010; Albelda y Cesteros, 2011; Salas Valdebenito, 2015; Moya y Carrió-Pastor, 2018a, 2018b, 2018c; Flores, 2020). Así mismo, muchos de ellos se centran en el análisis del lenguaje académico (Hyland, 2005; Mur-Dueñas, 2011; Carrió-Pastor, 2014, 2016a, 2016b, 2016c, 2017; Alonso-Almeida, 2015; Alonso-Almeida y Carrió-Pastor, 2017), pero también hay un interés cada vez más creciente en otro tipo de géneros, aplicándolo a la comunicación mediada por ordenador, al lenguaje coloquial, al lenguaje político, entre otros registros (cf. Albelda y Briz, 2010; Carrió-

Pastor y Muñiz, 2015a, 2015b; Moya y Carrió-Pastor, 2018a, 2018b, 2018c; Alonso-Almeida y Carrió-Pastor, 2019).

Las taxonomías que se han propuesto para clasificar los tipos de marcadores discursivos han sido diversas. Las más conocidas son las de Thompson (2001) y Hyland (2005). Sin embargo, estos autores, aunque describen las categorías y subcategorías en detalle, no incluyen los marcadores que se pueden incluir en cada categoría o subcategoría. Este hecho dificulta que otros estudiosos puedan utilizar dichas taxonomías, ya que los marcadores se han de interpretar en el contexto. Mur Dueñas (2011) y Salas Valdebenito (2015) sí que incluyen un listado de los marcadores que han hallado en su investigación, lo cual resulta muy útil como punto de partida en otros estudios. En este proyecto nos hemos basado en las propuestas de estas dos autoras para poder, a partir de los marcadores que han encontrado en sus corpus, verificar su función en nuestro corpus e incorporar otros marcadores que se desprendan de nuestros análisis. Las dos categorías en que comúnmente se dividen los marcadores discursivos, la textual y la interpersonal, se dividen en varias subcategorías que identifican la naturaleza de la organización discursiva y de la interacción que se crea con el lector u oyente. Por un lado, la categoría textual se divide en las siguientes subcategorías que podemos identificar como:

- a) **Marcadores lógicos.** Son aquellos elementos del discurso que relacionan unos elementos con otros, enlazando los enunciados y guiando el procesamiento de la información. Esta subcategoría se subdivide en tres: en primer lugar, los marcadores aditivos, que añaden información como, por ejemplo, *además, por otro lado*, etc., en segundo lugar, los marcadores adversativos o de contraste que contraponen un sentido a otro como, por ejemplo, *sin embargo, pero*, etc. y, por último, los marcadores continuativos o consecutivos, que tiene un valor de cohesión interna, por ejemplo, *pues, por ese motivo*, etc.
- b) **Códigos de glosa.** Se refieren a las expresiones que utilizamos para explicar el significado de la proposición. Esta subcategoría se divide en marcadores de reformulación, que son los elementos que reformulan una idea como son: *es decir, en concreto, esto es*, etc. y marcadores de ejemplificación que muestran ejemplos al receptor como, *por ejemplo, como, tales como*, etc.
- c) **Secuenciadores.** Son los marcadores del discurso que nos indican la secuenciación que el emisor desea realizar en el discurso, con lo cual marca el ritmo de la argumentación. Ejemplos de *secuenciadores* son: *primero, segundo, a continuación, finalmente*, etc.
- d) **Topicalizadores.** Estos elementos del discurso nos marcan el tema del mensaje, focalizando la atención sobre el punto principal del mensaje; algunos ejemplos son: *en cuanto a, respecto a*, etc.
- e) **Marcadores del discurso.** Incluye marcadores que guían al lector no desde el punto de vista de la coherencia de los párrafos, sino de todo el texto; como, por ejemplo, *hasta ahora, en resumen*, etc.

- f) **Enunciadores.** Esta categoría indica que el autor va a realizar una acción, como es el caso de los marcadores *vamos a, seguimos*, etc.
- g) **Marcadores endofóricos.** Estos marcadores tienen la función de apuntar o señalar, en forma anafórica o catafórica, a otras partes del discurso. Se dividen en marcadores anafóricos, los que se refieren a partes anteriores del discurso (por ejemplo, *anterior, previo, al principio, antes*, etc.) y catafóricos, los que se refieren a partes del discurso posteriores (por ejemplo, *siguiente, seguidamente*, a continuación, etc.).
- h) **Marcadores evidenciales.** Esta es una de las subcategorías más controvertidas. Si nos ceñimos a la concepción de Hyland (2005), se refiere a aquellos elementos del lenguaje que nos indican que existen evidencias de lo indicado en una proposición. Se dividen en personales, es decir que la evidencia que se aporta es una persona, por ejemplo, *como señala X* y en impersonales, cuando se refieren a una tabla, un gráfico o figura como evidencia, como en *como se indica en la tabla*, etc.

Por otro lado, la categoría interpersonal, que es la que incluye marcadores que intentan transmitirle al receptor la idea que desea el emisor, se divide en las subcategorías:

- **Atenuadores o mitigadores.** En esta subcategoría se incluyen los marcadores que se utilizan para mitigar la opinión o aseveración del emisor, con el fin, muchas veces de persuadir al receptor. Los marcadores que se incluyen en esta subcategoría son: *poder, considerar, intentar, apuntar, bastante, casi*, etc.
- **Intensificadores o enfatizadores.** Se trata de marcadores utilizados por el emisor para enfatizar la opinión o la proposición que se realiza. Algunos de estos marcadores son: *mostrar, determinar, constatar, cierto, evidente*, etc.
- **Marcadores de actitudinales.** Son aquellas partículas que expresan la evaluación o apreciación que el hablante hace de los hechos explicados en una proposición. Dentro de esta subcategoría podemos identificar a: *principal, importante, relevante, gran, necesario, bueno, adecuado*, etc.
- **Marcadores relacionales.** En esta subcategoría se incluyen los marcadores que involucran al receptor en el discurso y se dirigen a él/ella, que hace que el receptor se sienta parte del mensaje y hace que se involucre en la comunicación. Se dividen en directivos personales, en los que se engloba al receptor en el mensaje (por ejemplo, *comencemos, miremos, y otros verbos imperativos de primera persona plural, vosotros, usted, nos*, etc.) y en directivos (por ejemplo, *deber, haber que*, etc.).
- **Automenciones.** Son los marcadores que se refieren a los narradores de la comunicación; se dividen en referencias personales como, por ejemplo, *comenzamos, nuestro*, etc. refiriéndose a los autores de la proposición y en autocitas, cuando en el emisor se cita a sí mismo en el discurso.

Todas estas subcategorías están compuestas por elementos que podríamos denominar como parte de la comunicación *no necesaria*, puesto que podemos prescindir de ellos para comunicarnos, pero que, a la vez, forman

parte intrínseca de lo que deseamos transmitir, ya que los hablantes utilizamos estos recursos retóricos que nos permiten organizar el pensamiento y, de la misma forma, convencer o persuadir a los receptores del mensaje de nuestras intenciones.

Este capítulo se centra en describir y aportar ejemplos de la investigación basada en análisis de corpus que estamos llevando a cabo varios miembros del Grupo de Investigación de Lenguas de Especialidad (GALE) de la Universitat Politècnica de València como parte del proyecto concedido por el Ministerio de Economía y Competitividad en 2016 (referencia FFI2016-77941-P), cuyo propósito es identificar los elementos que componen las categorías textuales e interpersonales en corpus de artículos científicos de distintas ramas del saber. Gracias a la financiación del Ministerio se ha podido llevar a cabo una investigación pionera en su campo ya que se basa en el análisis comparativo de los patrones lingüísticos de la lengua inglesa y española tanto desde el punto de vista textual como interpersonal. La hipótesis de partida que hemos planteado en este proyecto es que los marcadores discursivos se usan de manera distinta en diferentes disciplinas. Por ello, la finalidad de proyecto es doble. En primer lugar, caracterizar el discurso científico y las estrategias retóricas que identifican cada discurso con sus características peculiares para, a continuación, identificar los patrones de la variación retórica entre disciplinas.

Los objetivos que me planteo en este artículo son tres. En primer lugar, describir la metodología seguida en el proyecto desde una perspectiva de la lingüística de corpus. En segundo lugar, voy a identificar los retos que han surgido en el proyecto para poder realizar una herramienta que identifique los marcadores discursivos polisémicos, con lo cual se ha tenido en cuenta el contexto específico dentro del género del lenguaje académico. Finalmente, voy a describir los proyectos futuros que nos planteamos a partir de este proyecto.

Este artículo se ha estructurado de la siguiente forma, primero, se han descrito los estudios previos que se han centrado en el análisis de marcadores discursivos en la lengua española e inglesa y que se han tenido en cuenta para escribir este capítulo. Posteriormente, se ha detallado el análisis llevado a cabo, distinguiendo entre la selección del corpus y la descripción del método que se ha llevado a cabo. A continuación, se han descrito los resultados alcanzados hasta ahora y los retos y, finalmente, en las conclusiones se identifican los hitos alcanzados y los retos que tenemos que afrontar en el futuro.

2. EL ANÁLISIS

2.1. CORPUS

La primera tarea llevada a cabo ha sido identificar los corpus existentes en estos momentos, tanto en lengua inglesa como española, para observar

de qué forma de habían recopilado y si podíamos utilizar uno de estos corpus para los análisis de los marcadores discursivos. Por una parte, los corpus de español más conocidos son: *Corpus de Referencia del Español Actual* (CREA), *Corpus de conversación coloquial del Grupo Val.Es.Co, Proyecto para el Estudio Sociolingüístico del Español del España y de las Américas* (PRE-SEEA), *Corpus del español actual* (CEA), *Corpus de aprendices de español* (CAES) y los corpus que podemos encontrar en los recursos IULA. Por otra parte, los corpus más conocidos de la lengua inglesa son: *British National Corpus* (BNC), *Lancaster-Oslo/Bergen Corpus- LOB Corpus*, *Corpus of Contemporary American English* (COCA), *University Centre for Computer Corpus Research on Language* (UCREL- Lancaster) y los corpus incluidos en el *Centre for English Corpus Linguistics* de la Universidad de Lovaina. Una vez identificados los corpus disponibles, se decidió recopilar y procesar nuestro propio corpus con el fin de poder identificar los marcadores discursivos de las tres especialidades de artículos científicos que se tuvieron en cuenta para este proyecto: la medicina, la ingeniería y la lingüística.

Así pues, el corpus del estudio está compuesto por dos corpus del lenguaje académico, uno de la lengua inglesa y otro de lengua española y cada uno de estos corpus se subdivide en las especialidades indicadas en el párrafo anterior. La intención de este corpus es que sirva de referencia, una vez etiquetados los marcadores discursivos, y que, a partir de este corpus, se puedan identificar patrones de marcadores discursivos en los tres campos específicos de investigación seleccionados. Se seleccionaron estos tres macro-corpus sabiendo que cada uno se puede dividir en decenas de sub-corpus según su especificidad como, por ejemplo, en medicina, la pediatría, la cirugía, etc.; en lingüística, la enseñanza de lenguas, la lexicografía, el análisis del discurso, etc.; y en ingeniería, la mecánica, la electrónica, la topografía, la geomática, la química, etc. Esta subdivisión de los corpus se realizará en una segunda fase del proyecto.

Los dos corpus de inglés y español están formados por tres sub-corpus compuestos por artículos científicos de diferentes campos del saber. Los tres sub-corpus están compuestos de sesenta artículos científicos de la disciplina de medicina, sesenta de la disciplina de ingeniería y sesenta de la disciplina de lingüística. El corpus en español se compone de 220 artículos científicos y el corpus en inglés de 230 artículos. Todos los artículos recopilados fueron publicados en revistas internacionales incluidas en *Journal Citation Report* o en *Scopus* desde 2016 a 2019. El corpus de español está compuesto por 1.668.209 palabras (592.321 en lingüística, 421.674 en medicina y 654.214 en ingeniería) y el corpus en inglés por un total de 1.692.911 palabras (636.620 en lingüística, 371.548 en medicina y 684.743 en ingeniería). El número de palabras recopiladas en los dos corpus representa un total de 3.361.120 palabras.

Se estableció, para los artículos de multiautoría, que un 60% de los autores fuera hablantes nativos de la lengua del corpus. En el caso del español, se procuró que fueran hablantes procedentes de España. En el caso del inglés, se recopilaron artículos de autores estadounidenses y británicos, cuya

nacionalidad y lengua materna se discernió tras identificar su filiación y, en caso de duda, se le mandó un correo a los autores para que indicaran su nacionalidad.

2.2. MÉTODO

Una vez recopilado el corpus, se procedió a etiquetarlo y para ello se necesitaba una herramienta que fuera capaz de contextualizar el etiquetado y que pudiera dar ejemplos de los marcadores metadiscursivos en las diferentes disciplinas escogidas y que conformaban el corpus. Se propuso el diseño de la herramienta *METOOL* al doctor Constantin Orasan (Universidad de Wolverhampton, Gran Bretaña), que accedió a colaborar en el proyecto mediante la colaboración del profesor Ruslan Mitkov, director del *Research Institute for Information and Language Processing* en la Universidad de Wolverhampton. Para el diseño de la herramienta, se siguieron las indicaciones de María Luisa Carrió-Pastor que se pautaron según las necesidades del corpus a recopilar y analizar. Tras varias reuniones, el Dr. Orasan presentó la herramienta en la que se fueron implementando las capacidades a medida que avanzaba el proyecto. Una vez diseñada *METOOL*, se tuvo en cuenta que pudiera transformarse, partiendo de un etiquetador, en un analizador de textos académicos y que pudiera convertirse en una herramienta *inteligente* para ofrecer en el futuro una evaluación de los marcadores discursivos de artículos académicos según su temática y su categoría. Se prefirió realizar una duplicidad de la herramienta para ubicar el corpus en español, ya que ello evitaba interferencias en el etiquetado de los diferentes marcadores discursivos en español e inglés.

Una vez realizada la herramienta y recopilado el corpus que serviría de referencia, se procedió a insertarlo en la herramienta y a etiquetar, en las diferentes categorías y subcategorías que se han detallado en el apartado de Introducción, los marcadores discursivos. Anteriormente ya se habían insertado todos los marcadores por categorías y subcategorías, definiéndolos y adjudicándoles diferentes colores para poder distinguirlos. Se incluyó la posibilidad de marcar un caso como conflictivo y la posibilidad de marcar la polaridad negativa, neutra o positiva del marcador textual.

El grupo de trabajo que se encargó del etiquetado y selección del corpus está compuesto por diez investigadores que estaban inicialmente en el proyecto o que se han ido incorporando paulatinamente tras indicar su interés en el proyecto o por estar realizando su tesis doctoral sobre marcadores discursivos. Los investigadores son investigadores afiliados a la Universitat Politècnica de València, la Universitat Jaume I, la Universidad de Las Palmas de Gran Canaria y la Universidad de Wolverhampton.

La intención de este proyecto es recopilar corpus académicos paralelos que nos aporten patrones lingüísticos de los marcadores discursivos de dos o más lenguas, como lo son, en este caso, el inglés y el español. Nuestra intención futura es que se amplíe a otras lenguas para que se puedan iden-

tificar los diferentes patrones retóricos de la lengua. Con ello, nuestro propósito es analizar los elementos lingüísticos que se utilizan para unir las distintas partes del discurso o para convencer a los lectores académicos de los hallazgos de los investigadores.

Este proyecto es novedoso en tanto que, tras el etiquetado del corpus base de *METOOOL*, los escritores de artículos académicos podrán saber el porcentaje de marcadores discursivos que utilizan en su artículo, así como la categoría y, con ello, determinar si ese es el porcentaje más frecuente, por subcategoría y por campo de conocimiento. Esta herramienta propondrá la incorporación o eliminación de marcadores según los valores que se aportarán desde el corpus de referencia.

Durante el proceso de etiquetado fueron necesarias varias reuniones del grupo de trabajo con los investigadores. Para ello, nos reunimos dos veces durante congresos realizados dentro del seno del proyecto, se realizaron sesiones por Skype, se abrió un grupo en *Google Groups* para intercambiar mensajes y se realizó un entrenamiento para poder etiquetar aquellas partículas más complejas, como fueron los atenuadores.

Finalmente, tras el etiquetado de todo el corpus en inglés, se procedió a etiquetar el corpus en español después de introducir, como se había hecho en la herramienta en inglés, todos los marcadores y sus categorías. Una vez etiquetados los dos corpus, que son los que van a servir de base para poder desarrollar la herramienta para convertirla en un analizador de los marcadores discursivos que se utilizan en textos académicos, ya se pudieron realizar análisis de los patrones que se siguen en las tres disciplinas. Los análisis cuantitativos se realizan por casos y se pueden extraer ejemplos de los textos por disciplinas, verificando la anotación de cada uno de los investigadores que ha etiquetado el corpus.

3. RESULTADOS

Una vez diseñada la herramienta que nos facilitaría el etiquetado del corpus y tras varios meses de trabajo en colaboración con la Universidad de Wolverhampton, se pudo presentar públicamente. La interfaz es simple, como se observa en la figura 1, e indica los cuatro aspectos importantes para este proyecto, la configuración de los marcadores, la inserción del corpus, la anotación de corpus y el análisis de los datos. El doctor Orasan y la doctora Carrió-Pastor son los administradores de la herramienta y los encargados de su configuración, de introducir los marcadores discursivos y de insertar los textos del corpus.

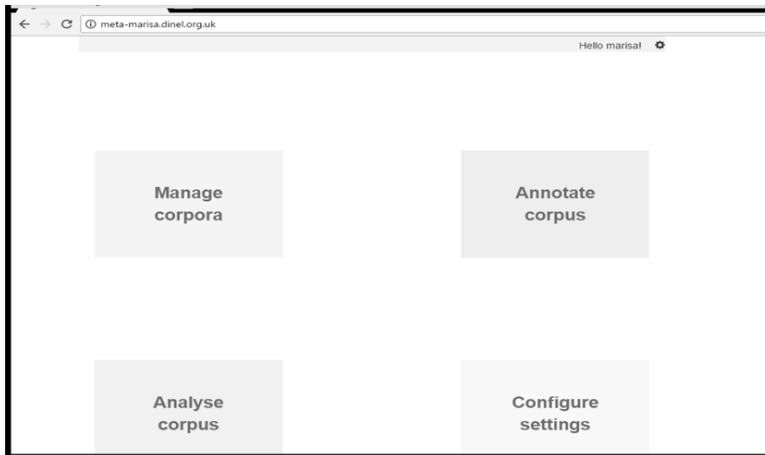


Figura 1: Interfaz de METOOL

La configuración de la herramienta se ha realizado en las distintas categorías y los marcadores se han distribuido de forma alfabética para que sea fácil saber si ya se ha incluido, como se observa en la figura 2:

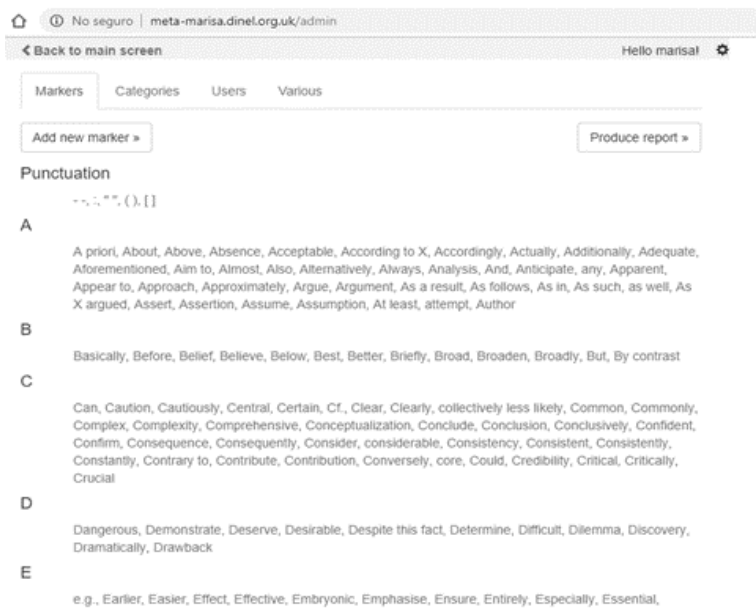


Figura 2: Marcadores discursivos en inglés en orden alfabético

El corpus se insertó en tres sub-corpus según su área de conocimiento en un corpus de ingeniería, otro de medicina y otro de lingüística, como se observa en la figura 3:

The screenshot shows the METOOL interface with a browser address bar at the top displaying 'meta-marisa.dinel.org.uk/select-corpus'. Below this, there are three sub-corpus boxes, each with an 'Annotate' button and a trash icon. The first box is for 'Academic discourse. Engineering papers' (65 texts, 23484 types, 684743 tokens, filters: 2016, 21st century, 2017, Male). The second box is for 'Academic discourse. Medicine papers' (87 texts, 15678 types, 371548 tokens, filters: 21st century, Academic English Medicine). The third box is for 'Academic discourse. Linguistics papers' (80 texts, 19357 types, 636620 tokens). To the right, a 'Filter conditions' panel allows filtering by 'Gender of author' (Male, Female), 'Genre' (Academic English - Engineering, Academic English Medicine, Academic English Linguistics), 'Specific field' (Cognitive linguistics, Semantics, Morphology, Syntax, Prosody), and 'Year' (2016, 21st century, 2017, 2015, 2014). At the bottom of the filter panel are buttons for 'Filter on all', 'Filter on any', and 'Reset filters'.

Figura 3: Sub-corpus de *METOOL* por área de conocimiento

Una vez insertados los textos en *METOOL*, se dividieron para que cada uno de los investigadores procediera a etiquetarlos. Cada texto y anotación ha sido revisado por otro de los miembros del grupo para evitar errores. A continuación, en la figura 4 se puede observar un texto antes de ser etiquetado y después de ser etiquetado. La herramienta identifica los marcadores en color amarillo para que sean fácilmente vistos en el corpus y una vez se le asigna una categoría y subcategoría, aparecen en el color de cada categoría. Para evitar la influencia en la anotación de los distintos anotadores, no se podían consultar los otros textos, ya que aparecían en color amarillo, como si estuvieran sin anotar.



Figura 4: Texto sin etiquetar (arriba) y texto etiquetado (abajo)

Después de etiquetar los textos, se puede proceder a su análisis y obtención de resultados, que puede hacerse por expresiones, por corpus o por categorías, como se observa en la figura 5:

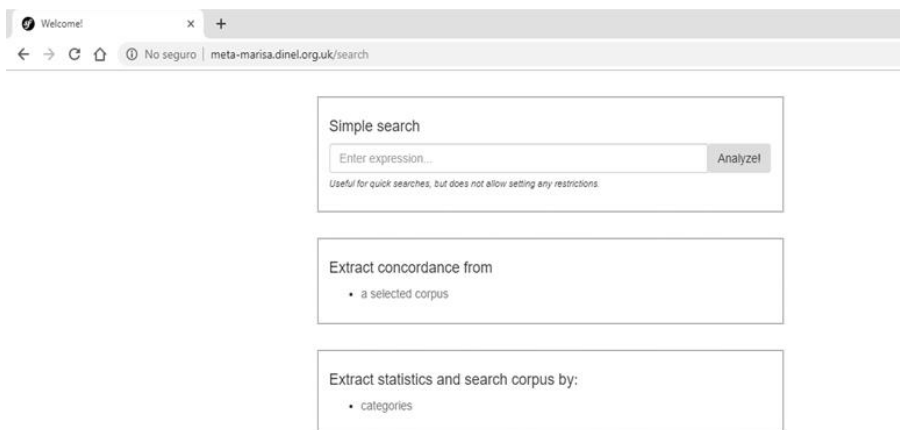


Figura 5: Análisis de los marcadores discursivos en *METOO*

Al buscar un marcador metadiscursivo se puede observar quién lo ha etiquetado (izquierda), ver las categorías y subcategorías y se puede consultar el texto para comprobar que ha sido etiquetado de forma adecuada según su contexto (derecha), como se aprecia en la figura 6:

checkbox	Author	Term	Description	Category	Count
<input checked="" type="checkbox"/>	marisa	can	It is used to express possibility, (it can be substituted by may)	Hedges- Modal verbs	69
<input checked="" type="checkbox"/>	marisa	can	used to express possibility in the present	Hedges- Modal verbs	21
<input checked="" type="checkbox"/>	marisa	can	used to express possibility in the present	Booster- verb	1
<input checked="" type="checkbox"/>	marisa	can	Not a marker		3
<input checked="" type="checkbox"/>	francisco.vizcaino	can	It is used to express possibility, (it can be substituted by may)	Hedges- Modal verbs	161
<input checked="" type="checkbox"/>	francisco.vizcaino	can	used to express possibility in the present	Hedges- Modal verbs	4
<input checked="" type="checkbox"/>	Fran	can	It is used to express possibility, (it can be substituted by may)	Hedges- Modal verbs	207
<input checked="" type="checkbox"/>	Fran	can	It is used to express possibility, (it can be substituted by may)	Epistemic stance	1
<input checked="" type="checkbox"/>	Fran	can	It is used to express possibility, (it can be substituted by may)	Epistemic	3
<input checked="" type="checkbox"/>	Fran	can	It is used to express possibility, (it can be substituted by may)	Dynamic	9
<input checked="" type="checkbox"/>	Fran	can	It is used to express possibility, (it can be substituted by may)	Hedges- Verbs	2
<input checked="" type="checkbox"/>	Fran	can	It is used to express possibility, (it can be substituted by may)	Hedges_ adjectives	2

All None

The annotation

1.	formally specified experi - ment designs can be used to enable jurisdictions to work together	+
Annotator: marisa Sense selected: It is used to express possibility. (it can be substituted by may) Category: Hedges/Hedges- Modal verbs Polarity: 2 Notes: Source: Go to annotation		
2.	one of which is how interpersonal relations can best be studied from a pragmatic perspective	+
Annotator: marisa Sense selected: It is used to express possibility. (it can be substituted by may) Category: Hedges/Hedges- Modal verbs Polarity: 1 Notes: Source: Go to annotation		
3.	The spatial relation expressed in (1) can be expressed in the Bantu language Mushunguli	+
Annotator: francisco.vizcaino Sense selected: It is used to express possibility. (it can be substituted by may) Category: Hedges/Hedges- Modal verbs Polarity: 1 Notes: Source: Go to annotation		
4.	The spatial relation expressed in (1) can be expressed in the Bantu language Mushunguli	+
5.	1 . In contrast , (1) is acceptable and can be true regardless of the distance between	+
6.	1 . In contrast , (1) is acceptable and can be true regardless of the distance between	+
7.	unacceptability is illustrated in (6) . (6) can be made acceptable in the context given by	+
8.	unacceptability is illustrated in (6) . (6) can be made acceptable in the context given by	+
9.	near the Ground , a simple LCC with ha - can be used to describe the location of a Figure	+
10.	near the Ground , a simple LCC with ha - can be used to describe the location of a Figure	+
11.	speaker is sufficient . A simple LCC with ha - can also be used acceptably even when the Ground	+

Figura 6: Ejemplo de muestra de etiquetado y de contextualización de los marcadores

Finalmente, en la figura 7 se puede ver el análisis cuantitativo por cada categoría y subcategoría del marcador discursivo:

Statistics per categories	
Corpus: Academic discourse. Medicine papers	
Boosters: 1085	booster- noun: 174
	Booster- verb: 328
	Booster- adjectives: 131
	Booster- adverbs: 440
	Booster- phrases: 12
Logical markers: 3165	Additive: 2573
	Contrastive: 406
	Consecutive: 181
Code Glosses: 366	Reformulation: 132
	Exemplification: 234
Frame markers: 244	Sequencers: 99
	Topicalisers: 56
	Discourse labels: 27
	Announcers: 62

Figura 7: Análisis cuantitativo de los marcadores en METOOL por subcategorías

Estos datos nos permiten identificar los patrones por el tipo de corpus que estudiamos a la vez que nos aportan información sobre el tipo de marcadores que se utilizan con más frecuencia. Así mismo, nos permite ver el

contexto de los marcadores con los ejemplos de los textos, las colocaciones más frecuentes y su significado contextualizado. El etiquetado del corpus de referencia que servirá para utilizar la herramienta como revisor de los elementos retóricos ha sido un proceso largo, pero los frutos que hasta ahora hemos podido extraer consideramos que nos permiten identificar los patrones retóricos de las diferentes ramas del saber dentro del género académico.

4. CONCLUSIONES

En este artículo se ha detallado el proceso seguido para la recopilación de un corpus que se ha dividido en tres campos específicos, el de la ingeniería, la medicina y la lingüística, con el propósito de identificar los distintos patrones retóricos que se utilizan dentro del mismo género, la comunicación académica, específicamente el artículo de investigación. Se ha explicado la forma en que se ha recopilado, así como el etiquetado mediante una herramienta para realizar un corpus de entrenamiento que sirva como base para analizar textos similares. Nuestro propósito es que la herramienta pueda identificar con precisión las estrategias retóricas y que logre realizar cálculos estadísticos que nos permitan conocer las frecuencias de cada uno de los marcadores y así compararlos de forma objetiva, aunque el número de palabras no sea el mismo.

Con todo ello, nuestro propósito es que nos permita distinguir las distintas categorías metadiscursivas y realizar observaciones en el caso en el que un marcador tenga varios significados, es decir, se utiliza con dos sentidos, como, por ejemplo, el caso de *nosotros*, el cual puede ser identificado como una automención o como un marcador relacional que incluye al lector. Somos conscientes que es un proceso largo y nada fácil, pero creemos que es importante que *METOOOL* pueda ser *inteligente* e identifique marcadores polisémicos para que el escritor de artículos académicos pueda conocer cómo utilizarlos y si los usa de manera adecuada. Saber cómo utilizar los marcadores discursivos es una competencia que se ha de desarrollar en los escritores académicos, ya que ellos han de saber cómo explicar sus hallazgos y convencer al lector que lo que están investigando es esencial para el avance de la ciencia y la sociedad. Para lograr esto, han de tener un dominio avanzado de la retórica para poder utilizar los marcadores de forma eficaz.

Por ello, el propósito de la herramienta *METOOOL* es el de ayudar a científicos a mejorar su forma de comunicación, identificando los marcadores discursivos que utilizan y comparándolos con los que utilizan otros académicos de su misma especialidad y género. Con ello, se aporta un dato objetivo para que el escritor tenga un referente de la forma en la que se utilizan los marcadores metadiscursivos, ya que también mostrará ejemplos en otros textos por especialidades.

Somos conscientes que existen limitaciones en este proyecto, puesto que la coordinación de un equipo de trabajo no es tarea fácil y la colaboración a distancia es compleja. Así mismo, la herramienta también tiene sus limitaciones y vamos avanzando en la implementación de mejoras poco a poco,

como lo hemos hecho en el cambio de servidor, en la forma de coordinar el etiquetado, etc. Así pues, los planes futuros del proyecto son, por un lado, desarrollar esta herramienta para que pueda utilizarse en el análisis de las variaciones de una lengua para determinar sus variantes sincrónicas. Ello puede incluso ayudar a identificar los patrones de la evolución de esa lengua, puesto que actualmente se ha acelerado ese proceso debido a la velocidad de la comunicación y las consiguientes influencias de otras lenguas.

Por otro lado, otra aplicación que se puede llevar a cabo es identificar los elementos del metadiscurso que hacen que un discurso sea fácil de entender, atractivo y persuasivo. Con ello, se pueden proponer maneras efectivas de comunicación para los profesionales que lo necesitan como, por ejemplo, los expertos en anuncios, los políticos y, en general, aquellos que necesitan comunicarse diariamente en su entorno laboral.

Adicionalmente, podemos identificar los marcadores retóricos según los niveles de conocimiento de una lengua extranjera, por ejemplo, siguiendo las directrices del Marco común europeo de referencia para las lenguas, que divide la adquisición de una lengua extranjera en los niveles A1, A2, B1, B2, C1 y C2. Esta información se podría utilizar para incorporarlos en la enseñanza de una lengua extranjera progresivamente; después de analizar textos e identificar los marcadores que se utilizan de forma más frecuente en cada nivel de aprendizaje se podrían incluir en los manuales de aprendizaje con ejemplos.

Aunque este proyecto se ha centrado en el análisis de textos académicos para el etiquetado de los marcadores discursivos y establecer patrones, *METOOL* también puede identificar los marcadores discursivos que se usan en otros géneros o tipos de comunicación, como el lenguaje oral, el digital, etc. Inicialmente nos centramos en el lenguaje académico ya que, al tratarse de textos escritos y de un género que ha sido ampliamente analizado, se dispone de información que nos ha servido para identificar los marcadores y poderlos incluir en la herramienta de forma sencilla y ágil. No obstante, se pueden incorporar nuevos marcadores en la herramienta que pertenezcan a otro tipo de géneros o manifestaciones del lenguaje, con lo cual, *METOOL* puede analizar e identificar patrones de cualquier tipo de comunicación, así como se pueden incluir diversas lenguas. En resumen, esas son las futuras líneas de investigación que trazamos en este proyecto y que deseamos implementar cuando *METOOL* ya sea capaz de aportar retroalimentación a los escritores académicos.

Finalmente, esperamos que este trabajo, a pesar de no poder presentar una versión definitiva de *METOOL*, pueda hacer ver que los lingüistas hemos de poder detectar y analizar las diferentes formas del lenguaje de forma objetiva y la tecnología nos puede ayudar a ello. Una colaboración más estrecha entre los investigadores que se dedican al procesamiento del lenguaje natural y los investigadores en lingüística aplicada es cada vez más necesaria para poder avanzar en la identificación de patrones y en la comprensión de los procesos del lenguaje.

REFERENCIAS

- ALBELDA, M. & BRIZ, A. (2010): "Cortesía y atenuantes verbales en las dos orillas a través de muestras orales", Aleza, M. y Enguita, J. (coords.): *La lengua española en América: normas y usos actuales*, Valencia: Universitat de València, pp. 237-260.
- ALBELDA, M. & CESTERO, A. (2011): "De nuevo sobre los procedimientos de atenuación", *Español Actual*, 96, pp. 121-155.
- ALBELDA, M. (2008): "Atenuantes en Chile y España: distancia o acercamiento", *Actas del Tercer Coloquio del Programa EDICE*. Estocolmo: Universidad de Estocolmo, pp. 89-113.
- ALONSO-ALMEIDA, F. (2015): "On the mitigating function of modality and evidentiality. Evidence from English and Spanish medical research papers", *Intercultural Pragmatics*, 12(1), pp. 33-57.
- ALONSO-ALMEIDA, F. & CARRIÓ-PASTOR, M. L. (2019): "Constructing legitimation in Scottish newspapers: The case of the independence referendum", *Discourse Studies*, 21(6), pp. 621-635.
- ALONSO-ALMEIDA, F. & CARRIÓ-PASTOR, M. L. (2017): "Variation and Function of Modals in Linguistics and Engineering Research Papers in English", Marín-Arrese, J. I., Lavid-López, J., Carretero, M., Martín de la Romero, E. D., Rosa, M. V. y Pérez Blanco, M. (coords.), *Evidentiality and Modality in European Languages. Discourse-pragmatic perspectives*, pp. 277-311. Berna: Peter Lang.
- ALONSO-ALMEIDA, F. & CARRIÓ-PASTOR, M. L. (2015): "Sobre la categorización de *seem* en inglés y su traducción en español. Análisis de un corpus paralelo", *Revista Signos*, 48, pp. 154-173.
- BEKE, R. (2005): "El metadiscurso interpersonal en artículos de investigación", *Revista Signos*, 38(57), pp. 7-18.
- BRIZ, A. (2001): *El español coloquial en la conversación: esbozo de pragmatología*, Barcelona: Ariel.
- BRIZ, A. (2007): "Para un análisis semántico, pragmático y sociopragmático de la cortesía atenuadora en España y América", *Linguística Española Actual*, 29(1), pp. 5-40.
- CARRIÓ PASTOR, M. L. (2014): "Cross-cultural variation in the use of modal verbs in academic English", *Sky, Journal of Linguistics*, 27, pp. 153-166.
- CARRIÓ PASTOR, M. L. (2015): "Identification of Rhetorical Moves in Business E-mails Written by Indian Speakers of English", Darics, E. (coord.), *Digital Business Discourse*, Londres: Palgrave MacMillan, pp. 226-242.
- CARRIÓ-PASTOR, M. L. (2016a): "A contrastive study of interactive metadiscourse in academic papers written in English and in Spanish", Alonso Almeida, F., Cruz García, L. y González Ruiz, V. (coords.), *Corpus-based studies on language varieties*, Berna: Peter Lang, pp. 80-102.
- CARRIÓ-PASTOR, M. L. (2016b): "A contrastive study of the hedges used by English, Spanish and Chinese researchers in academic papers", Alonso Almeida, F. et al. (coords.), *Input a word, analyze the world: Selected approaches to Corpus Linguistics*, Newcastle upon Tyne: Cambridge Scholars, pp. 477-492.
- CARRIÓ-PASTOR, M. L. (2016c): "Mitigation of claims in medical research papers: A comparative study of English and Spanish writers", *Communication & Medicine*, 13, pp. 1-25.
- CARRIÓ-PASTOR, M. L. (2017): "Verbal phraseology: An Analysis of Cognitive verbs in Linguistics, Engineering and Medicine Academic Papers" y Mitkov, R. (coord.), *Computational and Corpus-based Phrase-*

- ology. *Lecture Notes in Artificial Intelligence*, Berlin: Springer, pp. 325–336.
- CARRIÓ PASTOR, M. L. (2019): “Authorial engagement in business emails: a cross-cultural analysis of attitude and engagement markers”, Sancho Guinda y C. (coord.), *Engagement in Professional Genres*, Amsterdam: John Benjamins, pp. 47-65.
- CARRIÓ-PASTOR, M. L. & MUÑIZ CALDERÓN, R. (2015a): “A contrastive analysis of metadiscourse features in business e-mails written by non-native speakers of English”, *Procedia, Social and Behavioral Sciences*, 173, pp. 214-221.
- CARRIÓ PASTOR, M. L. & MUÑIZ CALDERÓN, R. (2015b): “Identification and causes of lexical variation in Chinese Business English”, *English Today*, 31(1), pp. 10-15.
- FLORES, N. (2020): *Linguistic mitigation in English and Spanish: How speakers attenuate expressions*, Londres: Routledge.
- HALLIDAY, M. A. K. (1998): *El lenguaje como semiótica social. La interpretación social del lenguaje y su significado*, México: S. L. Fondo de cultura económica de España.
- HYLAND, K. (2005): *Metadiscourse: Exploring Interaction in Writing*, Londres: Continuum.
- MOYA, P. & CARRIÓ-PASTOR, M. L. (2018a): “Análisis comparativo de los marcadores de compromiso en los comentarios sobre noticias digitales en España y Chile”, *Onomázein. Revista de lingüística, filología y traducción*, número especial: 4, pp. 26-48.
- MOYA, P. & CARRIÓ-PASTOR, M. L. (2018b): “Estrategias de intensificación en los comentarios digitales sobre noticias” *Spanish in Context*, 15(3), pp. 369-391.
- MOYA, P. & CARRIÓ-PASTOR, M. L. (2018c): “La atenuación en los comentarios sobre las noticias digitales en periódicos de España y Chile”, *Onomázein, Revista de lingüística, filología y traducción*, 40, pp. 56-76.
- MUR DUEÑAS, P. (2011): “An intercultural analysis of metadiscourse features in research articles written in English and in Spanish”, *Journal of Pragmatics*, 43, pp. 3068-3079.
- SALAS VALDEBENITO, M. (2015): “Una propuesta de taxonomía de marcadores metadiscursivos para el discurso académico-científico escrito en español”, *Revista Signos*, 48(87), pp. 95-120.
- SINCLAIR, J. (1987): *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*, Oxford: Collins.
- THOMPSON, G. (2001): “Interaction in Academic Writing: Learning to Argue with the Reader”, *Applied Linguistics*, 22(1), pp. 58-78.

ANEJO 1: Marcadores discursivos en inglés utilizados en METOOL (basado en Mur Dueñas, 2011).

1. CATEGORÍA TEXTUAL:

- A) Marcadores lógicos. Aditivos: *in addition, furthermore, moreover, similarly, also, further*; adversativos: *however, in contrast, but, yet, rather, nevertheless, instead, still* y consecutivos: *thus, therefore, as a result, consequently, then, as such*.
- B) Códigos de glosa. Marcadores de reformulación: *in other words, that is, i.e., paréntesis, punto y coma y guiones* y marcadores de ejemplificación: *for example, for instance*.
- C) Secuenciadores: *first, second, finally, to begin with, next, on the one hand*.
- D) Topicalizadores: *with regard to, concerning, turning to, in terms of*.
- E) Marcadores del discurso: *thus far, in sum, in brief, briefly*.

- F) Enunciadores: *aim to, will, seek to*.
 G) Marcadores endofóricos. Anafóricos: *as noted earlier, in this paper, above, previously, before* y catafóricos: *see Table 1, the next section, in Fig. 1, following, below, next*.
 H) Evidenciales. Personal: *according to X, as X argued, in X's study* e impersonal: *previous, past research, previously*.

2. CATEGORÍA INTERPERSONAL:

- A) Mitigadores o atenuadores: *may, would, can, might, could, appear to, seem, suggest, indicate, expect, predict, propose, consider, tend to, believe, see, view, interpret, imply, argue, hypothesize, think, posit, point to, feel, anticipate, regard, theorize, assume, speculate, want, relatively, typically, likely, potentially, often, perhaps, not necessarily, about, somewhat, usually, commonly, quite, nearly, partially, just, almost, theoretically, probably, marginally, roughly, approximately, basically, maybe, normally, ostensibly, partly, presumably, virtually, tentatively, sometimes, slightly, cautiously, likely, potential, possible, hypothesized, common, unlikely, indicative, typical, apparent, feasible, presumed, probable, proposed, unclear, prone to, hypothesis, likelihood, argument, possibility, view, idea, probability, attempt, assumption, prediction, notion, conceptualization, perspective, tendency, implication, proposition, belief, expectation, feasibility, in general, in part, at least, to our knowledge, a priori, in theory, in broad terms, in our judgement, to some extent*.
 B) Enfatizadores: *determine, show, demonstrate, reveal, highlight, confirm, emphasize, conclude, hold, underscore, establish, assert, prove, know, clearly, significantly, generally, largely, particularly, indeed, widely, highly, primarily, consistently, strongly, actually, mostly, especially, extensively, entirely, essentially, dramatically, substantially, always, fully, considerable, clear, vast, evident, substantial, evidence, fact, majority, assertion, conclusion, in fact, for the most part, of course, to a large extent, in effect*.
 C) Marcadores actitudinales: *important, consistent, significant, critical, interesting, difficult, key, well + adjetivo, necessary, robust, unique, valuable, complex, major, best, better, comprehensive, effective, main, useful, easier, problematic, relevant, surprising, fine-grained, (the) first, good, reasonable, contrary to, broad, confident, desirable, generalizable, great, limited, primary, serious, short of, suggestive, tremendous, true, worthwhile, acceptable, central, core, hard, inconsistent, in-depth, influential, intriguing, meaningful, new, notable, promising, underdeveloped, valid, worth, adequate, crucial, dangerous, embryonic, essential, exorbitant, fundamental, hopeful, inconclusive, marginal, missing, narrow, noteworthy, obvious, opaque, paradoxical, poor, ripe, satisfactory, skewed, stricter, sufficient, underspecified, unexplored, unfortunate, vexing, wise, support, extend, contribute, fail, deserve, ensure, expand, ignore, neglect, overemphasize, broaden, go (beyond), lack, merit (attention), overlook, respond, support, importance, limitation, insight, contribution, complexity, value, paradox, caution, consistency, hurdle, significance, validity, absence, credibility, dilemma, discovery, failure, key, shortcomings, strengths, wealth, only, importantly, interestingly, usefully, unfortunately, surprisingly, reliably, paradoxically, critically, conclusively, broadly*.
 D) Marcadores relacionales o de compromiso. Directivos personales: *we incluso, us, our, one, you*, preguntas, imperativos: *see, note that, cf.*, y directivos: *should, need to, must*.
 E) Automenciones. Referencias personales: *we exclusivo, us, our; autocitas*.

ANEJO 2. Marcadores del discurso en español utilizados en *METOOOL* (basado en Mur Dueñas, 2011)

1. CATEGORÍA TEXTUAL:

- A) Marcadores lógicos. Aditivos: *además, por otro lado, por otra parte, por su parte, igualmente, asimismo, y, adicionalmente, a su vez, de igual forma, de igual modo, análogamente, incluso*; adversativos: *sin embargo, no obstante, por el contrario, pero, en cambio, ahora bien, a pesar de ello, con todo, aun así*; y consecutivos: *por tanto, así, por ello, así pues, por lo tanto, pues, por este motivo, por consiguiente, por esta razón, entonces, en consecuencia*.

- B) Códigos de glosa. Marcadores de reformulación: *es decir, en concreto, esto es, concretamente, esto significa, se trata de, en otras palabras, nos referimos a, en todo caso, en cualquier caso, finalmente, por último, en definitiva, para finalizar, en conclusión, resumiendo, en resumen, en síntesis, paréntesis, punto y coma y guiones* y marcadores de ejemplificación: *como, por ejemplo, tales como, ej.*
- C) Secuenciadores: *primero/a, en primer lugar, en segundo/a, en segundo lugar, por último, finalmente, el/la primer/a, el/la segundo, por un lado, por otra parte, a continuación.*
- D) Topicalizadores: *en cuanto a, (con) respecto a, por lo que se refiere a, en relación a/ con, relacionado con, por lo que respecta a.*
- E) Marcadores del discurso: *en resumen, hasta ahora, en breve, brevemente.*
- F) Enunciadores: *vamos a, seguimos, deseamos, concluimos.*
- G) Marcadores endofóricos. Anafóricos: *anterior/es, hipótesis X, anteriormente, previo, previamente, apartado, anterioridad* y catafóricos: *siguiente(s), apartado/ capítulo, delante, como sigue, seguidamente, epígrafe, a continuación, en la tabla, la figura, gráfico, cuadro, continuamos, seguimos.*
- H) Evidenciales. Personal: *según X, como X comenta, el estudio de X,* e impersonal: *estudios anteriores, anteriormente, como en la tabla/ figura/ gráfico se explica.*

2. CATEGORÍA INTERPERSONAL:

- A) Mitigadores o atenuadores: *poder, verbos en forma condicional (-ría), parecer, caber, soler, observar, considerar, plantear, pensar, apreciar, indicar, pretender, entender, tratar de, sugerir, intentar, suponer, tender a, deducir, esperar, creer, implicar, interpretar, querer, estimar, prever, asumir, predecir, apuntar, bastante, casi, aproximadamente, posiblemente, normalmente, quizá (s), probablemente, frecuentemente, habitualmente, tal vez, previsiblemente, ligeramente, aparentemente, apenas, posible, planteado, propuesto, probable, pretendido, aproximado, indicado, hipótesis, probabilidad, posibilidad, tendencia, idea, planteamiento, visión, interpretación, intento, propuesta, observaciones, concepción, percepción, argumentos, en general, en parte, en principio, en cierta medida, en cierta forma, hasta cierto punto, a priori, en términos generales.*
- B) Enfatizadores: *mostrar, determinar, destacar, comprobar, confirmar, corroborar, demostrar, afirmar, poner de manifiesto, constatar, verificar, resaltar, revelar, concluir, evidenciar, enfatizar, subrayar, saber, remarcar, probar, arrojar, fundamentalmente, significativamente, especialmente, principalmente, ampliamente, generalmente, efectivamente, siempre, sustancialmente, claramente, obviamente, realmente, indudablemente, netamente, predominantemente, profusamente, claro, evidente, cierto, demostrada, amplio, considerable, inequívoco, hecho, determinación, evidencia, mayoría, argumentos, muestra, afirmación, verdad, de hecho, en- mayoría, en efecto, sin duda, sin lugar a dudas, en gran parte, en gran medida, en buena medida, de manera sustancial, de manera significativa, de un modo amplio.*
- C) Marcadores actitudinales: *principal, importante, relevante, gran, necesario, bueno, adecuado, válido, escaso, amplio, lógico, conveniente, fundamental, fiable, clave, especial, fuerte, interesante, apropiado, básico, coherente, difícil, fácil, grandes, imprescindible, mejor, pobre, significativo, sólido, útil, aceptable, congruente, curioso, decisivo, esencial, esperanzador, excesiva, exhaustivo, influyente, nuevo, positivo, preciso, satisfactorios, acertada, central, clave, conduyente, confuso, deseable, desfasado, excelente, indispensable, irremediable, nefastas, (el) primer, problemático, prudente, radical, razonable, replicado, riguroso, robusto, sencillo, singular, sorprendente, suficiente, vital, convenir, aportar, garantizar, contribuir, ir (más allá), limitar, ignorar, subsanar, importancia, limitación, problema, fiabilidad, aportación, interés, dificultad, validez, carencia, complejidad, relevancia, falta, problema' tica, utilidad, conveniencia, disparidad, diversidad, esencia, escasez, obstáculo, trascendencia, verdad, sólo, únicamente, excesivamente, mejor, meramente, negativamente, solamente, suficientemente, adecuadamente, debidamente, fielmente, paradójicamente, rotundamente.*
- D) Marcadores relacionales o de compromiso. Directivos personales: *inclusivos mos, nos, nuestro, uno, vosotros(es), usted(es), vuestro, su, suy,* preguntas, imperativos: *veamos,* y directivos: *deber, haber que.*
- E) Automenciones: referencias personales: *-mos, nos, nuestro-exclusivo; autocitas.*