

Título: «Diseño de herramienta en línea de apoyo al estudiante para la generación automática de fraseología y de marcadores del discurso en el macro-género de trabajos académico-científicos en Inglés»

María Araceli Losey León*, Juan Antonio Guitarte Fernández+

*Departamento de Filología Francesa e Inglesa, Escuela de Ingenierías Marina, Náutica y Radioelectrónica +Facultad de Ciencias

araceli.losey@uca.es

RESUMEN: El presente trabajo muestra los resultados obtenidos conducentes al diseño de una herramienta en línea dirigida al alumnado tanto de grados como de postgrados que sirva de orientación y de apoyo lingüístico en la redacción de textos en inglés. Su utilidad se concreta en posibilitar la elección de la fraseología apropiada en distintos apartados de la elaboración de un trabajo de índole académica y científica en dominios especializados. Se trata de un recurso semiautomático en el entorno *Human Language Technology tool* (HLT) y cuyos constructos fundamentales son un corpus lingüístico especializado compilado ad-hoc, el aprendizaje automatizado o *machine learning*. A pesar de que los resultados obtenidos tras el uso por el alumnado ofrecen evidencias del beneficio de la herramienta en el aprendizaje de la fraseología académica en inglés, estos han de tomarse con cautela ya que tendría que aumentarse el tiempo de uso de la práctica y extenderlo a asignaturas de contenido y al uso de investigadores de los campos científico-técnicos implicados para evaluar el alcance e impacto real del programa.

PALABRAS CLAVE: proyecto, innovación, mejora, docente, *machine learning*, *English for Academic Purposes (EAP)*, explotación de corpus para fines de enseñanza de lenguas, English for Science and Technology, Maritime English, tecnología educativa, minería de textos

INTRODUCCIÓN

Una correcta redacción requiere, además de un manejo preciso de la gramática y de disponer del conocimiento y de la perspectiva necesaria para correctamente cohesionar las oraciones y distribuir las secuencialmente en el texto, plasmar adecuadamente la Terminología del campo específico en el que se halle inscrita. El fin con el que se origina el texto, el medio en el que se genera y el contexto que lo envuelve determinará otros rasgos a los que todo escritor o escritora deberá adaptarse, tales como el tipo de género textual, el registro, el estilo y el enfoque. Si a todo ello añadimos que el producto ha de ser escrito en lengua inglesa, en un contexto especializado como el campo científico-técnico y en el género de los trabajos académicos, la dificultad inicial se acrecienta, especialmente para los estudiantes de 4º año de Grado y para los estudiantes de Postgrado e investigadores que hayan optado por presentar el trabajo/proyecto de fin de grado/máster total o parcialmente en inglés. Desde la experiencia como docente de cursos de inglés específico en el contexto científico-técnico a través de las asignaturas "Inglés marítimo", "Control de tráfico marítimo", "Inglés radioelectrónico", "Inglés técnico para arquitectura naval", "Inglés técnico para ingeniería marítima" e "Inglés técnico" del Máster en transporte marítimo, se han podido constatar dos carencias en los estudiantes a la hora de redactar los trabajos académicos. Una de ellas se observa en la elección inadecuada o en el desconocimiento de la fraseología propia comúnmente asociada a las secciones de resumen, introducción, objetivos, metodología, análisis, resultados y conclusiones. La segunda carencia se localiza en la organización del texto. Hay que tener en cuenta que el estudiante, además de la consulta de textos lingüísticamente más complejos de los que habitualmente ha consultado anteriormente a lo largo de sus estudios, en esta etapa se enfrenta a textos estructuralmente menos habituales y con un

patrón definido, como es el caso de los artículos científicos o de los manuales técnicos de gran extensión. Por este motivo, en los últimos dos cursos he incluido en la programación de ciertas asignaturas de Inglés específico en el contexto marítimo y naval, tanto en el grado como en los estudios de máster, una unidad temática dedicada al inglés académico adaptadas al entorno marítimo. Sin embargo, la incorporación de esta unidad no prevista en los contenidos curriculares de la materia, supone recomponer la temporización de la programación.

Con el fin de dotar al alumnado con mayor tiempo de dedicación y práctica al inglés académico, consideramos interesante la opción de utilizar recursos informáticos que ayuden en la redacción académica. Revisamos cuáles eran las opciones que el mercado de la tecnología educativa en idiomas ofrecía y encontramos varias alternativas orientadas al inglés académico, de entre las que cabe mencionar a Ackermann & Chen 2013 (1), Gardner & Davis 2013 (2); de ellas, algunas se ubicaban en el contexto científico-técnico pero ninguna de ellas abordaba su uso en el contexto marítimo. Por este motivo, consideramos interesante la idea de desarrollar un recurso de aprendizaje y entrenamiento propio. La investigación en fraseología basada en corpus ha abierto nuevos campos desde los que explorar su aprendizaje en la docencia presencial y en la docencia virtual.

FIN, OBJETO DE ANÁLISIS Y ALCANCE

El objetivo general del proyecto de innovación para la mejora docente es el diseño de una herramienta en línea dirigida a estudiantes y a investigadores que sirva de guía y apoyo para la redacción del género de textos académicos en Inglés. El recurso aborda las áreas científico-técnicas presentes en las enseñanzas impartidas en los campus de Puerto Real y Bahía de Algeciras de la Universidad de Cádiz que están

especialmente vinculadas al Campus de Excelencia Internacional del Mar.

Los objetivos específicos son:

1. Crear un recurso de consulta y de generación de textos académicos en inglés propios y en el entorno de sus enseñanzas.
2. Utilizar un corpus lingüístico especializado como base para dotar al programa informático de muestras de uso de la fraseología académica prototípica en la redacción de textos en Inglés especializado en el género académico de los trabajos, proyectos y artículos de investigación de revistas especializadas.
3. Explotar las muestras reales de los corpus lingüísticos especializados como herramienta docente.

El usuario final al que primariamente está dirigida la herramienta es el alumnado del Grado en Náutica y Transporte Marítimo, del Grado en Marina, del Grado en Ingeniería Radioelectrónica y del Máster en transporte marítimo de la Escuela de Ingenierías Marina, Náutica y Radioelectrónica. Su uso es extensible al alumnado de las dos menciones del Grado en arquitectura naval e ingeniería marítima. La configuración de la herramienta y los resultados obtenidos podrían ser objeto de estudio en algunas asignaturas del Máster en comunicación internacional así como de las asignaturas "Análisis del discurso" y "Nuevas tecnologías aplicadas a la Lengua Inglesa" del Grado en Estudios Ingleses y de la asignatura "Industrias de la lengua" del Grado en Lingüística y Lenguas Aplicadas.

Una vez implementada la herramienta el análisis se centrará en los resultados obtenidos a partir del uso de la herramienta por los estudiantes.

La novedad de la herramienta reside en tres aspectos:

1. El recurso aborda los campos de la navegación marítima, las máquinas navales y la ingeniería radioelectrónica al haberse detectado esta ausencia en las herramientas de escritura académica existentes en la actualidad.
2. Se ha desarrollado una herramienta basada en "machine learning" o aprendizaje automatizado.
3. La base de datos está constituida por un corpus lingüístico especializado *ad hoc*, que ha sido especialmente compilado para este proyecto, el corpus STAR-AB.

Conviene aclarar que si bien, en un primer momento, se pretendía abordar el artículo de investigación, la limitación de disponibilidad de un curso académico para llevar a cabo el proyecto nos condujo a centrarnos en la sección del resumen únicamente y en ella nos hemos basado a modo de prueba de concepto. El resumen condensa la información que se despliega en el artículo y también está compuesto de secciones con fraseología prototípica y coincidente con las de las secciones del artículo. Hemos re-adaptado el número de secciones ya que el análisis no se suele plantear en el resumen con el mismo detalle que en el artículo de investigación pero, en líneas generales, sí se ha podido contar con abundante fraseología académica que acompaña y es característica de las secciones del artículo de investigación.

DESCRIPCIÓN DE LA BASE DE DATOS

Como hemos indicado, la herramienta se nutre del corpus STAR-AB. Está compuesto de textos escritos extraídos de medios electrónicos y concretamente de resúmenes de artículos de investigación procedentes de revistas incluidas en el listado JCR 2016. Acudimos a este listado porque los criterios

de calidad de las revistas aquí indexadas buscan asegurar la calidad del inglés escrito.

Tal y como se describe en Losey (3), con carácter previo a la compilación del corpus, se concretaron los campos clasificados como científico-técnico consultando distintas páginas web de instituciones y organismos educativos. Finalmente, se adoptó la clasificación de los comités técnicos de ANECA complementados con ciertos campos de conocimiento del catálogo de la UNESCO. Actualmente, el corpus está constituido por tres subcorpus procedentes de tres divisiones académicas:

- Subcorpus STAR-ABENG: Dedicado a la Ingeniería y Tecnologías de la Producción y de la Construcción; dotado de 394 resúmenes.
- Subcorpus STAR-ABMN: Centrado en la Navegación Marítima; dotado de 141 resúmenes.
- Subcorpus STAR-ABAS: Especializado en las Ciencias Aplicadas; dotado de 83 resúmenes.

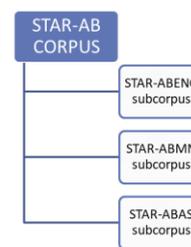


Figura 1. Composición del corpus STAR-AB (Scientific-Technical Research Articles-Abstract Corpus)

Las funcionalidades de la herramienta son:

- a) Disponibilidad de ejemplos de textos reales específicos de los resúmenes de artículos de investigación de los dominios y subdominios implicados.
- b) Proporcionar fraseología distintiva y particular en función de la sección del resumen.
- c) Ofrecer distintas alternativas y formas de expresión extraídas de los textos del corpus.

Además de contribuir a la mejora del aprendizaje de fraseología concreta, se estima que aumentaría la seguridad y la satisfacción del estudiante durante la etapa de elaboración de los trabajos.

METODOLOGÍA DE TRABAJO

La Lingüística computacional es el soporte de este trabajo, es decir, se utilizará el marco informático del procesamiento del lenguaje natural para el desarrollo de una herramienta semiautomática generadora de fraseología característica de las secciones de un resumen de un artículo científico a partir de la explotación de los datos extraídos de un corpus electrónico elaborado al efecto. La metodología en la investigación lingüística ha sido fundamental y se despliega en varios apartados que no se presentan al no ser objeto del presente trabajo.

Una vez compilado el corpus STAR-AB se ha procedido mediante técnicas de minería de textos a la extracción de la

fraseología académica por cada una de las secciones del resumen. Para ello hemos empleado el programa de concordancias lingüísticas Antconc de Anthony (4). A modo de ejemplo, se muestra en la figura 1 los 47 concordancias obtenidas para “study” tras la búsqueda realizada en el subcorpus de resúmenes de Ingeniería aeroespacial que contiene un total de 147 resúmenes.

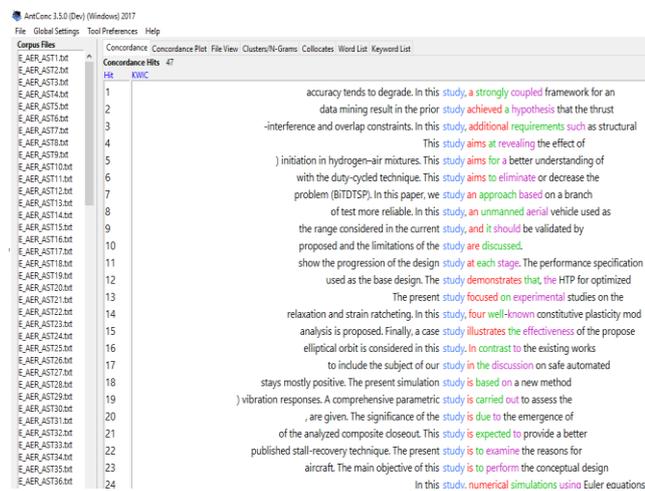


Figura 2. Captura de pantalla del total de 47 concordancias de “study” obtenidas con el programa de concordancias AntConc

A partir del corpus el estudiante podrá acceder a través de la herramienta que hemos diseñado a visualizar, reconocer, identificar y asociar la fraseología académica prototípica en su rama de estudio y adecuada para los siguientes fines:

- a) estructurar y secuenciar el resumen,
- b) introducir el estado de la cuestión,
- c) transmitir el objetivo que el autor se ha fijado,
- d) designar y describir la metodología de trabajo,
- e) aludir al proceso de análisis,
- f) describir unos resultados,
- g) incidir en aspectos que aún no han sido estudiados,
- h) reivindicar un enfoque,
- i) mostrar los resultados,
- j) presentar un corolario,
- k) o para sugerir recomendaciones futuras.

PRUEBA PILOTO DE LA HERRAMIENTA

A principios del primer y del segundo semestre se llevó a cabo una pequeña encuesta en la que se les pedía al alumnado que evaluaran el interés que podría tener una herramienta con las funcionalidades expuestas. Todos los alumnos de las asignaturas implicadas constituyeron el grupo experimental siendo un total de 45 alumnos. De ellos, 32 percibían que habría mucha dificultad en adquirir las competencias de redacción académica en inglés y en un contexto técnico, mientras que 11 manifestaron bastante dificultad y solo 2 indicaron que la dificultad sería media.

La prueba se desarrolló una vez terminada la impartición de la asignatura ya que el inicio de la programación de la herramienta se retrasó hasta abril de 2018 debido, fundamentalmente, a que el número de textos y los campos previstos inicialmente a abordar en el corpus se multiplicaron.

De la estimación inicial de 50 textos se incorporaron al corpus 618 textos y de la estimación inicial de cobertura de 3 dominios en el ámbito de la navegación marítima, se amplió a 13 subdominios del área científico-técnica.

Por consiguiente, la prueba piloto se desarrolló en el mes de junio de 2018. Para ello localizamos al alumnado y solicitamos su participación, siendo finalmente 23 el número de alumnos participantes. Todos habían tenido una preparación previa a través de una unidad didáctica durante el curso. Al final de la prueba con la herramienta los alumnos tuvieron que redactar un breve resumen con un máximo de 150 palabras en las que se les daba un tema, un método y datos numéricos sobre el análisis. La prueba fue individual y todos tenían los mismos datos. Evaluamos cómo resolvieron la redacción teniendo en cuenta los indicadores siguientes: objetivo (a), método (b), análisis (c) y resultado/conclusiones (d). El nivel de éxito no fue el 100% pero el 68% de los alumnos utilizó la fraseología apropiada en el indicador (a), el indicador (b) fue resuelto con éxito por solo un 25% de los alumnos. En la misma dirección y con menor porcentaje se encuentra el indicador (c) que fue resuelto solo por un 8% del total. En cuanto al indicador (d) fue resuelto favorablemente por el 92% del alumnado participante. Realizando un balance general podríamos concluir que los resultados no son muy satisfactorios pero sí son suficientemente satisfactorios.

DISCUSIÓN Y CONCLUSIONES

Se ha desarrollado e implementado una herramienta que no solo proporciona la búsqueda de modelos de resúmenes en inglés sobre un dominio concreto sino que proporciona la práctica, la ejecución de la composición escrita y la corrección.

Nuestra propuesta se ha encaminado a trasladar el aprendizaje automatizado o *machine learning* al servicio de la tecnología educativa, poniendo en valor los corpus lingüísticos especializados como medio de enseñanza basada en datos.

En cuanto a los resultados obtenidos, a pesar de que ofrecen evidencias del beneficio de la herramienta en el aprendizaje de la fraseología académica en inglés, estos han de tomarse con cautela por las siguientes razones (a) la prueba con la herramienta y el test de redacción se realizaron en la misma jornada no existiendo un tiempo de asimilación; (b) el alumnado participante pertenecía a niveles de grado y postgrado y no cruzamos los resultados obtenidos en función de este valor; (c) el número de alumnos participantes ha sido muy bajo. No obstante, si se comparan los resultados se observa que el nivel de fracaso es mayor en la fraseología de los indicadores (b) y (c) que en la fraseología de los indicadores (a) y (d). Posiblemente se deba a que estos últimos suelen desarrollarse con marcadores discursivos y giros más comúnmente utilizados o a que se hallan en cadenas sintácticas más sencillas.

Actualmente estamos aumentando el número de textos en el corpus para enriquecer el programa. Como medida futura, consideramos interesante extender el uso de la herramienta a asignaturas de contenido y al uso de investigadores de los campos científico-técnicos implicados para evaluar el alcance e impacto real del programa.

REFERENCIAS

- (1) Ackermann, K. and Y.-H. Chen. Developing the Academic Collocation List (ACL)- A corpus-driven and expert-judged approach, *Journal of English for Academic Purposes*, 2013, 12: 235-247.
- (2) Gardner, D. & Davies, M. A New Academic Vocabulary List. *Applied Linguistics*, 2013, 35, 305-327.
- (3) Losey-León, M.-A. Researching the Phraseological Units of the Academic Genre through a Domain-specific Corpus of Engineering, Maritime Navigation and Applied Sciences' Research Articles in English. Actas del XXXVI Congreso Internacional de la Asociación Española de Lingüística Aplicada, AESLA, 2018. Universidad de Cádiz, Cádiz, 19-21 de abril de 2018 (en prensa).
- (4) Anthony, L. AntConc (Version 3.5.6) [Computer Software]. Tokyo, Japan: Waseda University, 2018. Available from <http://www.laurenceanthony.net/software>. Último acceso el 28 de septiembre de 2018.

AGRADECIMIENTOS

A la Unidad de Innovación Docente del Vicerrectorado de Recursos Docentes y de la Comunicación de la Universidad de Cádiz por la concesión de la realización de los Proyectos de Innovación y Mejora Docente (Proyecto Sol-201700083912-tra) en la convocatoria 2017-2018 cuyos resultados se presentan en este trabajo, así como la financiación de 1 beca obtenida por el alumno de 3º curso del Grado en Matemáticas, D. Juan Antonio Guitarte Fernández.