Un modelo de red para identificar colocaciones de especialidad y sus propuestas de traducción

María de los Ángeles CRIADO ALONSO [1]
Universidad Rey Juan Carlos (ESPAÑA)
angeles.criado@urjc.es || https://orcid.org/0000-0001-8608-180

Recibido: 22/1/2024 || Aceptado: 20/2/2024

Resumen: En este trabajo se presenta un enfoque innovador que se fundamenta en modelos matemáticos como el *Enriched Line Graph* para la detección y traducción de unidades fraseológicas en textos técnicos o redactados en diversos lenguajes de especialidad. Se ejemplifica la aplicación de este enfoque a través del lenguaje jurídico especializado presente en la legislación del Estado de Virginia (EE.UU.). El método utilizado implica el empleo de un modelo de red compleja multicapa pesada y dirigida, el cual, al ser utilizado sobre un corpus lingüístico debidamente etiquetado, posibilita el análisis e identificación de las relaciones sintagmáticas entre las distintas secuencias de términos. Esta estrategia permite identificar palabras clave, unidades de conocimiento especializado y verbos de especialidad, facilitando así la traducción precisa de textos especializados teniendo en cuenta la complejidad que supone la presencia de estructuras léxicas con relaciones sintagmáticas entre términos para los traductores si no son especialistas en la materia o lenguaje de especialidad en la que se encuentra redactado el texto que están traduciendo. Esta metodología resalta la identificación de secuencias de palabras y unidades fraseológicas características de cualquier corpus textual, lo que contribuye a mejorar la comprensión y traducción de textos de lenguaje común o de cualquier lenguaje especialidad.
Palabras clave: Fraseología. Unidad fraseológica. Lenguaje de especialidad. Traducción. Español. Inglés.

Titre : « Un modèle de réseau complexe pour traduire collocations de spécialité ».
Résumé : Cet article présente une approche innovante basée sur un modèle mathématique appelé *Enriched Line Graph* pour la détection et la traduction d'unités phraséologiques dans des textes techniques ou des textes rédigés en différentes langues spécialisées. L'application de cette approche est illustrée par le langage juridique spécialisé présent dans la législation de l'État de Virginie (États-Unis). La méthode utilisée fait appel à un modèle de réseau complexe multicouche lourd et dirigé qui, utilisé sur un corpus linguistique correctement étiqueté, permet d'analyser et d'identifier des mots-clés, des unités de connaissances spécialisées et des verbes spécialisés, facilitant ainsi la traduction précise de textes spécialisés, compte tenu de la complexité que la présence de structures lexicales, avec des relations syntagmatiques entre les termes, implique pour les traducteurs qui ne sont pas spécialistes de la matière ou de la langue spécialisée dans laquelle le texte qu'ils traduisent est rédigé. Cette méthodologie met en évidence l'identification de séquences de mots et d'unités phraséologiques caractéristiques de tout corpus textuel, ce qui contribue à améliorer la compréhension et la traduction de textes de langue courante ou de toute spécialisée, puisqu'elle est exportable à n'importe quel corpus.
Mots-clé : Phraséologie. Unité phraséologique. Langue spécialisée. Traduction. Espagnol. Anglais.

Title: "A network model to identify idioms in specialty language and find translation proposals."
Abstract: This paper presents an innovative approach based on mathematical models such as the *Enriched Line Graph* for the detection and translation of phraseological units in technical texts or texts written in different specialized languages. The application of this approach is illustrated through the specialized legal language found in the current legislation of the State of Virginia (USA). The methodology involves the use of a heavy and directed multilayer complex network model, which, enables the analysis and identification of syntagmatic relationships between the different sequences of terms and

[1] Grupo de investigación Liynmedia (Universidad Rey Juan Carlos).

words in a corpus. This strategy enables the accurate translation of terms of specialty language and other linguistic structures such as colocations and Phraseological Units for translators who are not specialists in the subject matter or specialty language in which the text they are translating is written. This methodology highlights the identification of keyword sequences, which contributes to improving the comprehension and translation of texts since it is exportable to any corpus.
Keywords: Phraseology. Phraseological Unit. Specialized language. Translation.

INTRODUCCIÓN

Este artículo aborda un novedoso enfoque de traducción orientado a los lenguajes de especialidad basado en la aplicación de modelos de red multicapa provenientes de la teoría de las redes complejas a cualquier corpus de especialidad con la intención de facilitar la tarea al traductor tal como se hizo en trabajos previos de reciente publicación (véase: Criado-Alonso, 2022a). Normalmente, los traductores tienen problemas a la hora de identificar qué secuencia de palabras constituye una unidad fraseológica (UF) o cuándo estamos ante una unidad de conocimiento específico (que puede estar compuesta por uno o varios términos o palabras) en los textos de origen cuando estos pertenecen a una disciplina (especialidad) o área del saber concreta. Por tanto, se pretende arrojar algo de luz a las dificultades que plantea la traducción técnica, también llamada a veces traducción de lenguajes de especialidad.

La traducción técnica ha estado siempre muy ligada al mundo de los lenguajes de especialidad y, a su vez, al de la terminología. Estudios como los de Cabré (2004) dan a entender que, de algún modo, los traductores ejercen como selectores de términos porque incluyen nuevos términos en el léxico (ya sea voluntaria o involuntariamente) cuando se enfrentan a problemas de traducción que tienen que ver con realidades que se encuentran en los textos de origen y que no existen en la cultura meta. Estudios como el de Martín Camacho (2004) o el de Alcalá y García Antuña (2011) explican cómo se va creando el léxico de especialidad con la inclusión de términos exclusivos de determinadas áreas del saber y palabras «terminologizadas»[2] que alteran su significado cuando se dicen en un contexto determinado (entre otros mecanismos de creación léxica). Por ejemplo, el término «integral» tiene un significado muy diferente en un texto matemático que el que tiene, en la colocación «pan integral». Llegados a este punto, podemos resaltar que estudios como el de Solano (2012) conciben las unidades fraseológicas como un instrumento útil para enseñar y aprender un lenguaje de especialidad, ya que pueden proporcionar ejemplos concretos de cómo se utilizan las colocaciones en un contexto específico. Por otra parte, artículos como el de Moreno Fernández (1999) o el de Alcaraz (2004) y obras como la de Cabré y Gómez de Enterría (2006) abordan la relación entre la enseñanza de las lenguas de especialidad (o lenguas para fines específicos) y la variación lingüística. Moreno Fernández reflexiona sobre la importancia de tener en cuenta la variación lingüística en la enseñanza de las lenguas de especialidad, ya que los estudiantes necesitan aprender a utilizar el lenguaje técnico de manera adecuada en diferentes contextos y situaciones. También destaca la importancia de la investigación en este campo para poder responder preguntas como qué exponentes lingüísticos[3] son más adecuados

[2] Cabré (1993) denomina «terminologización» al proceso por el cual las palabras de la lengua general pasan a los distintos ámbitos de especialización con significados precisos en cada caso, y por lo tanto, diferenciados entre sí. Es destacable el hecho de que puede darse también un tercer transvase de unidades de una lengua de especialidad a otra, a veces con el consiguiente cambio de significado o bien el referente conceptual en ambas disciplinas.

[3] Para Moreno Fernández (1999), el exponente lingüístico se refiere a cualquier elemento lingüístico que expresa una determinada categoría gramatical, función sintáctica o significado en un idioma. En el contexto de las lenguas de especialidad y la variación lingüística, los exponentes lingüísticos pueden

para cada situación y cómo se pueden enseñar de manera efectiva. En este contexto, un exponente lingüístico se refiere a una palabra, una UF o construcción gramatical que se utiliza en un campo específico de la lengua para fines específicos. Estos exponentes lingüísticos son propios de cada campo profesional o académico y se utilizan para comunicarse de manera efectiva en situaciones especializadas. Por ejemplo, en el español jurídico, los exponentes lingüísticos pueden incluir términos legales y expresiones formales que no se utilizan en otros contextos del español. Estos exponentes lingüísticos aparecen en forma de secuencias de palabras ordenadas, por lo que consideramos una UF como una secuencia de palabras que se repiten de manera ordenada, formando una estructura que alude a una realidad concreta dentro del lenguaje de especialidad. En este punto, y para evitar ambigüedades en las nomenclaturas utilizadas en este artículo, cabe destacar que cualquier exponente lingüístico tiene, en este sentido, en común con las colocaciones[4] el hecho de que son secuencias de palabras (lo mismo sucede con la traducción, en algunos casos, de paremias y refranes) que para el traductor han de contemplarse como un concepto que debe traducir. Es notable que lo que en el texto origen se expresa con una o varias palabras (UF) en el texto meta quizá necesite de una sola palabra o término de especialidad para ser verbalizado

El conocimiento y uso adecuado de estos exponentes lingüísticos, UF y colocaciones es fundamental para el desarrollo de la competencia comunicativa en situaciones especializadas. Es precisamente en este punto donde la traducción automática y las herramientas computacionales y de traducción asistida empiezan a ser insuficientes para el traductor no especializado que se enfrenta a un texto redactado en un lenguaje de especialidad concreto. De hecho, existen diversos estudios sobre la identificación de colocaciones a través de corpus lingüísticos más allá del recuento de la frecuencia de aparición de las palabras como los de Blanco San Martín *et al.* (2019) y los de Domínguez (2014), que ponen de manifiesto la importancia de la identificación y del uso de las colocaciones para los hablantes de un idioma. En este artículo profundizaremos en el análisis de un modelo de red multicapa pesada y dirigida publicado por primera vez en el año 2020 en la publicación de Criado-Alonso *et al.* (2020), que permite procesar un corpus de especialidad para extraer las unidades de significado compuestas por dos palabras o términos de especialidad y así facilitarle la tarea al traductor. Este modelo de red compleja es exportable a cualquier corpus de especialidad y es capaz de contemplar las asociaciones entre distintos términos identificando así las colocaciones e, incluso, las UF (concebidas como sucesiones de palabras) de los distintos lenguajes de especialidad.

1. LOS LÍMITES DE LA TRADUCCIÓN AUTOMÁTICA EN LOS LENGUAJES DE ESPECIALIDAD.

La traducción de especialidad, la lingüística de corpus y el uso de textos paralelos están estrechamente relacionados en el ámbito de la traducción especializada. La lingüística de corpus es una herramienta muy útil para traductores y lingüistas ya que se puede definir como una metodología capaz de sistematizar y analizar conjuntos extensos de datos orales y escritos ordenados con criterios lingüísticos, literarios, culturales y sociales con el propósito de dar cuenta de la lengua en uso desde diversas concepciones valiéndose de herramientas computacionales, tal y como se expone en el libro de Bernal e Hincapié (2018). El uso de la lingüística de corpus como herramienta para la traducción de textos es algo indiscutible (el texto

variar dependiendo del contexto comunicativo y las características específicas de la lengua utilizada en ese ámbito particular. Estos exponentes pueden manifestarse a través de rasgos léxicos, gramaticales, fonéticos, entre otros, que son distintivos de las lenguas de especialidad.

[4] Una colocación es, según la obra de Koike (2001), una asociación de dos palabras que, cuando aparecen juntas una seguida de la otra adquieren un significado concreto.

paralelo es siempre un buen recurso), donde más espacio existe para la mejora es en las herramientas computacionales o programas de análisis de textos más convenientes. Criado-Alonso (2022a) plantea en su tesis doctoral cómo las herramientas de análisis textual tradicionales son eficaces a la hora de contemplar las relaciones paradigmáticas entre los términos de un corpus, pero no para contemplar las relaciones sintagmáticas existentes entre los mismos.

Estudios como el de González Boluda (2010) ponen de manifiesto que las herramientas de traducción automática son útiles para el traductor, aunque siguen necesitando la supervisión del ojo humano (o traductor experto) para no cometer errores. La ineficacia de las herramientas de traducción automática viene dada por la propuesta de traducciones literales y por la ausencia de la identificación de las UF presentes en los textos de origen. Está claro que la traducción literal o la traducción automática se quedan cortas ya que, como hemos visto en el apartado anterior, a veces, las distintas disciplinas presentan conceptos con los que el traductor debe estar familiarizado o incluso conceptos que existen únicamente en la cultura de origen y que son intraducibles (se entiende por intraducibilidad los casos en los que en un texto presenta una «laguna léxica). En estos casos, un traductor automático va a dar al usuario una propuesta que le llevará al falso sentido. Por eso, normalmente, los traductores de especialidad recurren a herramientas como la lingüística de corpus, que es una herramienta muy útil puesto que, como se ha señalado anteriormente, permite recopilar y analizar, con la ayuda de sistemas de computación, grandes cantidades de textos en diferentes idiomas y campos especializados con el objetivo de identificar patrones lingüísticos y terminológicos. Este tipo de herramientas, junto con la utilización de textos paralelos, es decir, textos con contenidos relacionados (o similares) configuran una metodología adecuada para comparar y contrastar las diferencias lingüísticas y culturales entre las lenguas (de nuevo, permitiendo al usuario comparar conceptos). Estos textos paralelos o corpus lingüísticos son especialmente útiles para la traducción de especialidad, ya que permiten identificar las diferencias entre los términos y expresiones utilizados en diferentes idiomas y campos especializados y hallar propuestas de traducción para las UF presentes en los textos de origen.

2. USO DE CORPUS PARA LA TRADUCCIÓN DE LENGUAJES DE ESPECIALIDAD.

La creación de corpus con textos paralelos bilingües es una práctica común en la traducción de especialidad, ya que permite comparar y contrastar los textos en diferentes idiomas y campos especializados Estos corpus se utilizan para identificar patrones lingüísticos y terminológicos en diferentes idiomas y campos especializados, lo que facilita la traducción de textos técnicos y científicos. El problema aparece cuando tratamos, por ejemplo, con distintos sistemas de derecho y ordenamientos jurídicos. Así, si consideramos un testamento redactado en cualquier país donde se haga uso de la *common law*, encontraremos probablemente un documento que es el *probate of a will*. En castellano, no hay un equivalente directo y, si usamos la traducción automática, llegaremos a una expresión muy lejana de la realidad como por ejemplo «legalización de un testamento», siendo esta la traducción facilitada por el traductor automático de uso muy extendido *DeepL*. El *probate of a will* es un documento parecido a un poder notarial que le da capacidad jurídica a una persona designada por el testador para satisfacer todas las deudas que el testador dejara y liquidar todos los activos que pudieran generar deudas y repartir la herencia que quede entre los herederos legítimos. En este caso concreto, una alternativa adecuada para obtener una traducción ajustada a la realidad consistiría en analizar y buscar la secuencia de las palabras *probate+of+a+will* para obtener la explicación que el traductor necesita saber para dar una propuesta de traducción adecuada para esta agrupación de términos, unidad de conocimiento específico o UF (en este caso «autorización para la gestión de la herencia», o algo similar, podría ser una traducción adecuada, en cualquier caso, siempre

identificando el país de donde procede el documento como fuente de derecho).

La lingüística de corpus tiene sus defensores y sus detractores, pero no cabe duda de que, como afirman Baker *et al.* (1993), son muchas las posibilidades que ofrece la lingüística de corpus al mundo de la traducción en un espectro que abarca desde las comparaciones entre distintos registros hasta las comparaciones entre distintos idiomas o lenguas. La diferencia entre la lingüística de corpus y la lingüística computacional consiste en que la lingüística computacional basa sus estudios en la mejora del procesamiento automático de los corpus lingüísticos, es decir, es la parte informática del análisis del corpus. La lingüística computacional es la que se encarga de los programas que analizan los corpus textuales.

En concreto, una propuesta muy novedosa que planteamos en este trabajo consiste en utilizar un modelo de red compleja para analizar el lenguaje y así identificar las colocaciones de especialidad tal y como se ha hecho en los trabajos previos de Criado-Alonso *et al.* (2020, 2021, 2022a y 2022b). La idea se basa en dividir los términos de un corpus en diferentes capas según su naturaleza, de manera que el corpus queda configurado como una red compleja multicapa pesada y dirigida en la que se basan los resultados obtenidos a partir de esta nueva estructura y representación del corpus. Hasta el momento presente, esta técnica se ha utilizado sobre un corpus basado en el lenguaje matemático de especialidad, pero es evidente que se trata de un modelo exportable a cualquier disciplina siempre y cuando se procesen los datos utilizando la metodología establecida con el apoyo computacional adecuado.

2.1. Utilización del nuevo modelo para el estudio del lenguaje de especialidad. Casos previos

Está claro que el lenguaje matemático cumple con los criterios para ser considerado un lenguaje de especialidad. No hay que confundir el lenguaje matemático con el lenguaje artificial, ya que el lenguaje simbólico numérico lleva a su alrededor expresiones verbales. Además, tal y como afirma Wood (2001), el lenguaje científico está plagado de términos matemáticos. En este punto cabe señalar que los símbolos matemáticos constituyen un lenguaje y que no necesitan de expresión verbal para ser comprendidos, por lo que resulta difícil hacer lecturas opuestas de este lenguaje. Sin embargo, es posible malinterpretar la verbalización que hay alrededor de uno o varios símbolos concretos (sobre todo si no se es experto en la materia) no verbalizando de manera adecuada el lenguaje de especialidad matemático.

El lenguaje matemático está presente en prácticamente todas las áreas del saber, pero se han escogido textos (documentos) pertenecientes al ámbito de las redes complejas para poder elaborar un corpus específico en el corpus elaborado en los estudios previos de Criado-Alonso *et al.* (2020, 2021, 2022a y 2022b). Sin esta elección habría sido casi imposible la selección de documentos. Es importante señalar que el lenguaje de especialidad matemático y, en particular, el referido al campo de las redes complejas, está dividido a su vez en cuatro sublenguajes: el simbólico numérico matemático, el vehicular, el de especialidad de redes y aquel perteneciente al de la especialidad sobre el que se esté aplicando el estudio de redes complejas (por ejemplo, un artículo sobre «redes complejas neuronales» tendría vocabulario específico del ámbito de las redes complejas y de biosanidad).

Los textos analizados que conforman el corpus en Criado-Alonso (2020) son un total de 89 documentos del que hemos obtenido un corpus de lenguaje matemático de especialidad de redes complejas compuesto por 147 637 palabras y 25 210 frases (entendiendo por frase la secuencia de palabras que aparecen entre punto y punto). Concretamente, los documentos son *extended abstracts* de la revista *International Journal of Complex Systems in Science* entre los años 2011 y 2016 cuyas publicaciones entran dentro de esta área de especialidad. Es importante señalar que acotar y definir el corpus no han sido tareas fáciles en absoluto. Para ilustrar la mejora que supone asumir la representación de este lenguaje como una red compleja mencionaremos que el corpus se ha analizado, en un primer momento, utilizando herramientas de análisis ya existentes,

en concreto, *Voyant, Wconcord*[5] *y Flair*. En este punto es importante señalar que los resultados obtenidos a través de estos programas no permitían extraer las relaciones sintagmáticas de este lenguaje. Las tres herramientas mencionadas de análisis textual de corpus a través de las cuales extraen conclusiones sobre el lenguaje matemático de especialidad se basan en valores tales como la frecuencia absoluta y relativa de aparición de las palabras o términos del corpus. Estas tres herramientas de análisis ofrecen resultados complementarios entre sí que permiten llegar a conclusiones relevantes sobre el lenguaje de los textos que se están analizando como es la identificación de las palabras clave (que no de las colocaciones clave). Sin embargo, aunque identificar las palabras clave de un texto es algo básico para el traductor a la hora de entender el significado y sentido correcto de un documento, una de las principales carencias de estas herramientas consiste en la ausencia de métodos correctos para identificar qué elementos constituyen una unidad de conocimiento específico (o, lo que en los apartados anteriores llamábamos colocación y/o UF que, en este caso, pueden pertenecer al ámbito de los lenguajes de especialidad). Por otro lado, la herramienta de análisis *Flair* ofrece información relativa a la sintaxis y la dificultad de los textos según el MCERL, mientras que *Voyant*, mucho más compleja, alude a palabras clave, frecuencia y contexto (ausencia o presencia de lenguaje académico) entre otros aspectos.

Es la representación y utilización del corpus como una red compleja multicapa la que ha permitido extraer las relaciones sintagmáticas que tienen que ver con las colocaciones o unidades de significado de especialidad de este lenguaje en concreto. En lo que se refiere a los términos del lenguaje de especialidad analizado, encontraremos términos del lenguaje de especialidad de redes complejas, así como términos del área de investigación específica a la que el documento concreto pertenece y en la que se presentan aplicaciones de este tipo de estructuras matemáticas. Debido a esto último, estudiar las palabras clave del lenguaje de especialidad matemático de las redes complejas conlleva una dificultad añadida. En concreto, se trata de distinguir entre los términos de esta área matemática concreta y los términos de especialidad que quedan fuera de este campo. El proceso de limpieza y etiquetado del corpus del presente estudio ha sido muy largo ya que se han tenido que tomar múltiples decisiones sobre qué constituye una palabra y qué no. Como aparece reflejado en la tesis de Criado-Alonso (2022a), los resultados obtenidos a través de estos programas de análisis de textos tradicionales (*Wconcord, Flair* y *Voyant*) no permitían extraer las relaciones sintagmáticas de este lenguaje, si bien se puede afirmar que estas herramientas de análisis ofrecen algunos resultados complementarios entre sí. La novedad y la diferencia en el uso de esta novedosa metodología se aprecia tras concebir el corpus como una red pesada y dirigida de cuatro capas y es, precisamente, el poder identificar las colocaciones de especialidad entre los términos. Usando esta metodología, una palabra (convertida en un nodo) ya no es importante porque se repita muchas veces, sino porque se relaciona con otras palabras que también son importantes. Ha sido la representación y utilización del corpus según el modelo de red pesada y dirigida de cuatro capas la que ha permitido extraer las relaciones sintagmáticas que tienen que ver con las colocaciones o unidades de significado de especialidad de este lenguaje específico. En este punto conviene señalar que la mayoría de las publicaciones del corpus analizado en el caso previo descrito, están redactadas en inglés por hablantes no nativos.

En cualquier caso, asimismo es destacable que el lenguaje matemático de especialidad no es el único que se entremezcla con otra disciplina pudiendo así generar textos que contienen dos especialidades. Lo mismo sucede con el lenguaje jurídico, que puede mezclarse (y de hecho se

[5] *Wconcord* es una herramienta de análisis textual que se puede descargar gratuitamente desde https://www.linglit.tu-darmstadt.de/index.php?id=linguistics y fue desarrollada por el Technische Universität Darmstat dentro del Institut für Sprach und Literaturwissenschaft (Instituto de estudios Lingüísticos y Literarios de la Universidad Técnica Darmstat).

mezcla muy a menudo) con el lenguaje financiero, el lenguaje laboral y el lenguaje fiscal. En definitiva, el traductor de especialidad, muchas veces se enfrenta a textos donde aparecen términos pertenecientes a más de una disciplina.

3. METODOLOGÍA. USO DE UN MODELO DE RED COMPLEJA PESADA Y DIRIGIDA

Cabe señalar que los programas *Flair Wconcord y Voyant* muestran que existe una gran variación de los términos o palabras clave en función del texto (o el corpus) que se escoja, ofreciendo en muchas ocasiones diferentes resultados para un mismo texto (véase Criado-Alonso, 2022a). La utilización complementaria de *Wconcord* y *Voyant* permite identificar un total de doce términos clave exclusivos del lenguaje de especialidad matemático relativo a la teoría de las redes complejas: *matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering* y *clustered*. El resto de téminos clave que identifican la combinación de *Voyant* y *Wconcord* no son, en realidad, términos clave del lenguaje de la teoría redes complejas sino términos del lenguaje de la especialidad científica específica sobre la que se está aplicando esta teoría. Estos términos clave varían de texto a texto del corpus al aplicarle *Wconcord*. Por ejemplo, «*optical communications*» es una palabra clave de uno de los textos del corpus, que no es una colocación clave del lenguaje de redes.

Una vez probada la relación y la existencia de estos cuatro sublenguajes matemáticos dentro de la teoría de redes complejas (el simbólico-gráfico-numérico matemático, el lenguaje de especialidad matemático verbalizado, el lenguaje natural en el que se exprese el texto y el lenguaje del área de especialidad al que se esté aplicando la teoría de redes complejas) es cuando podemos empezar a aplicar la visión proporcionada por la teoría de redes complejas sobre su propio lenguaje de especialidad para extraer conclusiones lingüísticas. Esta visión permite determinar qué pares de palabras han de aparecer juntas y determinan una unidad de conocimiento específico, así como analizar el funcionamiento de este lenguaje de especialidad a la hora de, por ejemplo, crear verbos de especialidad.

Para nuestro análisis nos interesa considerar una red multicapa, tal y como quedó establecido en Criado-Alonso *et al*. (2020), compuesta por cuatro capas, colocando los nodos (i.e., las palabras) en cada una de dichas capas según su naturaleza:

1. Una capa en la que hemos incluido las palabras específicas (principalmente adjetivos y sustantivos) de la lengua de especialidad, que llamamos capa léxica. En esta capa, hemos incluido los términos exclusivos de la especialidad lenguaje matemático junto con los términos cuyos referentes se ven alterados cuando no aparecen en un contexto matemático. Consideramos que todos estos términos son «unidades léxicas» o «palabras específicas relevantes» por cumplir con los criterios terminológicos de especialidad de Cabré Castellví (2004).
2. La capa de verbos, donde hemos incluido todos los verbos independientemente de su conjugación.
3. Una capa, a la que denominamos capa de enlaces (*linking layer)* en la que colocamos, como su propio nombre indica, las palabras de enlace. En esta capa hemos incluido las palabras con relevancia gramatical para la construcción de la frase. Preposiciones, conectores, adverbios de frecuencia y determinantes responden a esta descripción.
4. Por último, una capa, a la que llamamos capa-resto, en la que colocamos el resto de las palabras (sin caracterizar). Hemos incluido las palabras que no forman parte del lenguaje de especialidad matemático bien porque no sufren ninguna alteración cuando aparecen en él o bien porque son palabras del lenguaje cotidiano o de la disciplina a la que se está aplicando (en este caso, la teoría de redes complejas).

3.1. Exportación de la metodología a otros lenguajes de especialidad

Esta metodología es exportable a cualquier corpus de especialidad que se quiera crear. Ya hemos visto cómo el lenguaje jurídico, a menudo, se puede mezclar con el financiero o con el laboral, por lo que una red multicapa dividida en cuatro capas también podría ser efectiva para un análisis similar al ya realizado en este campo, aunque para analizar corpus que incluyan textos con términos y expresiones de varios lenguajes de especialidad, podría resultar adecuado considerar un modelo de red de más de cuatro capas (las cuatro especificadas en el modelo descrito, junto con una capa léxica para cada lenguaje de especialidad incluido en el corpus). A modo de ejemplo, en la frase original que veremos a continuación, y considerando nuestro modelo de cuatro capas, vamos a utilizar distintos números entre paréntesis para clasificar en qué capa iría cada uno de los términos e ilustrar la utilidad que tendría hacer un análisis del lenguaje jurídico-financiero presente en la legislación vigente del Estado de Virginia[6].

Al distribuir los términos principales de la siguiente frase en diferentes capas descritas, de manera que el número asociado a cada término indique la capa a la que pertenece nos encontramos con términos que pertenecerían a la capa léxica del lenguaje jurídico y términos que pertenecerían a la capa léxica del lenguaje financiero, por lo que, en el modelo de cuatro capas, dependiendo del lenguaje de especialidad analizado, unos u otros irían a la capa «resto» (por eso, en este caso, si se quiere estudiar simultáneamente ambos lenguajes de especialidad, sería necesario crear una red de cinco capas en lugar de cuatro). Incluyendo en el modelo una capa léxica «extra», es posible estudiar simultáneamente textos que mezclan dichos lenguajes de especialidad. Así pues, analicemos el original: «*The* (4) *trustee* (1) *shall* (3) *invest* (3) *and* (4) *manage* (3) *trust* (1) *assets* (2) *as* (4) *a* (4) *prudent* (5) *investor* (2) *would* (3) *by* (4) *considering* (3) *the* (4) *purposes* (1), *terms* (1), *distribution* (2) *requirements* (2), *and* (4) *other* (4) *circumstances* (1) *of* (4) *the* (4) *trust* (1)».

Según se puede observar, los términos que aparecen señalados con el 1 en la frase anterior y resaltados en azul en la imagen representan términos pertenecientes al lenguaje jurídico, los términos señalados con el 2 son términos pertenecientes al lenguaje económico o financiero, los términos resaltados con un 3 irían a la capa verbal, los términos señalados con un 4 formarían parte de la *linking layer* y los términos señalados con un 5 formarían parte de la capa resto. Si quisiéramos utilizar sólo cuatro capas, habría que determinar si los términos que van a la capa resto son los del lenguaje financiero o los del lenguaje jurídico. Normalmente estudiaríamos las colocaciones desde el lenguaje jurídico, así que *assets, investor* y *distribution* formarían parte de la capa resto si hacemos una red de cuatro capas. Sin embargo, este modelo planteado de cinco capas, siguiendo pautas similares a las presentadas en Criado-Alonso *et. al* (2021), permite identificar y estudiar las colocaciones jurídico-financieras tales como *trust assets* (expresión que refiere a los activos que existen en un fideicomiso), con qué frecuencia aparecen, si son características o no de un corpus que se puede diseñar en base a este lenguaje además de otras particularidades de este lenguaje de especialidad. Esta metodología de estudio de lenguajes de especialidad a través de redes complejas es exportable a cualquier corpus. Al ser exportable a cualquier corpus, el uso de este método permite el análisis del lenguaje más allá del contexto del lenguaje de especialidad y abordar el estudio de, por ejemplo, las paremias definidas por Sevilla y Crida (2013). Una paremia se concibe como «un enunciado breve y sentencioso, que corresponde a una oración simple o compuesta, que se ha fijado en el habla y que forma parte del acervo sociocultural de una comunidad hablante» (Sevilla y Crida, 2013: 206). Al poder identificar las paremias y las secuencias de palabras presentes en un corpus, se le facilitará al traductor y al estudiante de una lengua (sea esta de especialidad o no) la traducción, el aprendizaje y el estudio de dicha lengua.

[6] https://law.lis.virginia.gov/vacodefull/title64.2/chapter7/article9/

CONCLUSIONES

Las herramientas tradicionales de traducción automática son insuficientes al tratar con lenguajes de especialidad ya que muy a menudo los textos de origen presentan realidades que no existen en los textos meta, y los traductores automáticos ofrecen propuestas erróneas de traducción. El uso de textos paralelos y corpus textuales ayudan a identificar unidades de conocimiento específico en los distintos originales que permiten dar una descripción cercana al mensaje original sin crear un concepto inexistente en la cultura meta. El modelo y la visión aportada por la utilización de las redes multicapa pesadas y dirigidas en el ámbito de los lenguajes de especialidad facilitan la identificación de las unidades de conocimiento específico difíciles de traducir utilizando herramientas de traducción automática convencionales. La red multicapa de cinco capas propuesta en este capítulo permitiría analizar el lenguaje de especialidad jurídico-financiero y sus características basándose en el precedente de análisis que ya se ha hecho sobre el lenguaje matemático de especialidad.

Los modelos de redes complejas presentados en Criado-Alonso *et al.* (2021) permiten realizar un estudio sintagmático del léxico y determinar qué elementos constituyen una unidad de conocimiento específico a diferencia de otros programas de análisis textual tradicionalmente empleados en el mundo de la traducción, constituyendo además un marco unificado que ha permitido analizar las propiedades estructurales de los lenguajes de especialidad explorando la interacción entre las unidades lingüísticas y la formación de asociaciones de estas unidades, constituyendo nuevas estructuras que formarán parte de la terminología del lenguaje de especialidad, es decir, determinando qué elementos constituyen una colocación de especialidad.

La metodología asociada a este tipo de representaciones (redes complejas) permite no solo localizar las colocaciones propias de los lenguajes de especialidad, sino también identificar las palabras más importantes de un corpus siguiendo un nuevo sistema más allá del recuento ITF basado en frecuencia absoluta e inversa utilizado por herramientas clásicas de análisis textual (eg. *Voyant*) así como la identificación de las unidades de conocimiento especializado formadas por más de un término e incluso la identificación de verbos de especialidad (eg *cluster together* – agrupar), lo que resulta esencial para la correcta traducción de los lenguajes de especialidad.

El modelo de red multicapa descrito es adaptable al tratamiento de un corpus que incorpore términos y unidades de conocimiento específico pertenecientes a más de un lenguaje de especialidad añadiendo una capa adicional que incluya los términos léxicos de cada uno de estos lenguajes. Por ello, el modelo original de red compuesta con cuatro capas se puede adaptar añadiendo nuevas capas tal y como sucede con el lenguaje jurídico-financiero, para el que es más adecuado utilizar un modelo de cinco capas. El estudio de este lenguaje de especialidad a través una red compleja de cinco capas utilizando una técnica similar a la descrita en Criado-Alonso *et al*. (2021) permitiría crear un diccionario de términos, colocaciones y/o unidades fraseológicas aceptables de traducción para estos casos. Además, futuros estudios lingüísticos, desarrollados a partir de la exportación de esta metodología a otros corpus permitirían el estudio y análisis de otro tipo de estructuras lingüísticas tales como las paremias o cualquier tipo de UF o exponente lingüístico que esté compuesto por uno o varios términos y que pueda ser de interés desde la perspectiva del terminólogo, del lingüista o del traductor.

REFERENCIAS BIBLIOGRÁFICAS

Alcalá, A.; García Antuña, M. (2011). Lenguas de especialidad y lenguas para fines específicos: precisiones terminológicas y conceptuales e implicaciones didácticas. *XX Congreso Internacional de la Asociación para la Enseñanza*. Cádiz: Asociación para la Enseñanza del Español como Lengua Extranjera-ASELE, 907-932.

Alcaraz Varó, E. (2004). El lenguaje profesional y académico. *Nuevas formas de aprendizaje en lenguas extranjeras*. Ministerio de Educación, Cultura y Deporte, 91-114.

Baker, M.; Francis, G.; Tognini-Bonelli, E. (1993). *Text and technology: in honour of John Sinclair.* John Benjamins Publishing.

Bernal Chávez, J.; Hincapié Moreno, D. (2018). *Lingüística de Corpus.* Bogotá: Instituto Caro y Cuervo.

Blanco San Martín, L.; Ferreira Cabrera, A.; Blanco San Martín, E. (2019). Colocaciones léxicas verbo-nominales en un corpus de aprendices de Español como Lengua Extranjera de nivel A2 y B1. *Letras de Hoje*, 54(3), 407-416.

Cabré, M. (1993). *La terminología: teoría, metodología y aplicaciones.* Barcelona: Empúries.

Cabré, M. T. (2004). La terminología en la traducción especializada. G. García, y García Yebra, *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco-Libros, 89-125.

Criado-Alonso, M.; Battaner Moro, E.; Aleja, D.; Romance del Río, M.; Criado Herrero, R. (2020). Using complex networks to identify patterns in specialty mathematical language. A new approach. *Social Networks and Data Mining*, 10(1), 1-10.

Criado-Alonso, M.; Battaner Moro, E.; Aleja, D.; Romance, M.; Criado, R. (2021). Enriched line graph: A new structure for searching language collocations. *Chaos, Solitons and Fractals*, 142, 110509.

Criado-Alonso, M. (2022a). *Metodología para el análisis y traducción del lenguaje de especialidad matemático basada en modelos de redes compleja*. Tesis Doctoral. Universidad Rey Juan Carlos.

Criado-Alonso, M.; Aleja, D., Romance, M.; Criado, R. (2022b). Derivative of a hypergraph as a tool for linguistic pattern analysis. *Chaos, Solitons and Fractals*, 163, 112604.

DeepL. (s.f.).https://www.deepl.com/es/translator

Domínguez, U. (2014). Las colocaciones en un corpus de aprendices valón y flamenco. *La enseñanza de ELE centrada en el alumno*, 977-986.

Flair. (s.f.). http://sifnos.sfs.uni-tuebingen.de/FLAIR/

González Boluda, M. (2010). Estudio comparativo de traductores automáticos en línea: Systran, Reverso y Google. *Núcleo*, 22(27), 187-216.

Koike, K. (2001). Colocaciones léxicas en el español actual: análisis formal y léxico-semántico.
Universidad de Alcalá.

Martín Camacho, J. (2004). *Los procesos neológicos del léxico científico. Esbozo de clasificación.* (Vol. XXVII). Online: Anuario de Estudios Filológicos. http://hdl.handle.net/10662/846

Moreno Fernández, F. (1999). Lenguas de especialidad y variación lingüística. *Lenguas para fines específicos (IV)*, 3-14.

Moreno Ortiz, A.; Pérez Hernández, C.; Faber, P. (1999). Lexicografía computacional y lexicografía de corpus. *Revista española de lingüística aplicada, monográfico*, 1, 175-214.

Sevilla Muñoz, J.; Crida Álvarez, C. (2013). Las paremias y su clasificación. *Paremia*, 22, 104-114. https://cvc.cervantes.es/lengua/paremia/pdf/022/009_sevilla-crida.pdf

Solano Rodríguez, M. (2012). Las unidades fraseológicas del francés y del español: tipología y clasificación. *Paremia*, 21, 117-128.
https://cvc.cervantes.es/lengua/paremia/pdf/021/011_solano.pdf

Voyant. (s.f.). https://voyant-tools.org

Wood, A. (2001). International scientific English: The language of research scientists around the world. *Research perspectives on English for academic purposes,* 1, 71-83.