



TESIS DOCTORAL

*Metodología para el análisis y traducción del
lenguaje de especialidad matemático basada
en modelos de redes complejas*

Autora:

M^a Ángeles Criado Alonso

Directores:

Elena Battaner Moro

Miguel Romance del Río

Programa de Doctorado en Humanidades: Lenguaje y Cultura

2022

A mis padres

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas aquellas personas que con su apoyo me han ayudado a realizar este trabajo. En primer lugar, mi máximo agradecimiento a la directora de esta tesis doctoral, doña Elena Battaner Moro, por su generosa y valiosa ayuda en la elaboración de la misma. Sin su sabia dirección, su paciencia y su apoyo, este trabajo no habría llegado nunca a su fin. También quiero expresar mi más sincero y profundo agradecimiento a don Miguel Romance del Río, cuya guía y dirección de la parte práctica de la tesis ha sido fundamental para el desarrollo de la misma.

No puedo olvidarme de todos los compañeros, colaboradores y amigos que han estado a mi lado dándome ánimos para que este proyecto llegase a buen puerto. A todos aquellos que me han acompañado y apoyado en todos los momentos y en particular en los más difíciles. A todos vosotros, muchísimas gracias de todo corazón.

Muy especialmente quiero agradecer a mi familia haber permanecido a mi lado durante todo este proceso, sobre todo a mi madre, que ha cuidado siempre de todos nosotros y, de manera muy especial, a mi padre, pues sin su ejemplo, sus enseñanzas y su apoyo jamás podría haber llegado a ser la persona que soy hoy. Muchas gracias a todos.

Índice

Introducción	11
PRIMERA PARTE TEÓRICA	21
Capítulo 1: La traducción y el lenguaje de especialidad	21
1.1 El papel del traductor como terminólogo en el lenguaje de especialidad.....	21
1.2 Lenguaje de especialidad y terminología específica	23
1.3 Las unidades de conocimiento especializado y la traducción. Importancia de la semiótica.....	29
1.4 El lenguaje científico como lenguaje de especialidad	32
1.5 El lenguaje matemático de especialidad	34
1.6 Lenguaje matemático de especialidad del área de redes complejas.....	37
1.7 La teoría de redes complejas.....	40
1.7.1 Introducción	40
1.7.2. Conceptos básicos sobre redes y grafos.....	42
1.8. Recapitulación	48
Capítulo 2. Lingüística de corpus y lingüística computacional. Descripción del corpus. Análisis con <i>Wconcord</i> , <i>Flair</i> y <i>Voyant</i>	51
2.1 Introducción a la lingüística de Corpus. Herramienta para analizar los lenguajes de especialidad	52
2.2 Lingüística de corpus	59
2.3 Descripción del corpus utilizado.....	69
2.4 Estudio tradicional del corpus mediante las herramientas <i>WConcord</i> , <i>Flair</i> y <i>Voyant</i> . Primeros resultados del estudio	82
2.4.1 <i>Wconcord</i> , <i>Flair</i> y <i>Voyant</i>	83

2.5 Palabras clave según las herramientas tradicionales.....	89
2.5.1 <i>Voyant</i> y <i>Wconcord</i> para obtener las palabras clave	91
2.6 Conclusiones. Resumen y rasgos de la morfosintaxis de este tipo de textos. Presentación del estilo que se debe seguir en la redacción de textos	105
2.6.1 Propuestas de traducción basadas en los doce términos identificados como importantes y en las directrices de redacción del apartado anterior.....	111
2.6.2 Recapitulaciones generales sobre el lenguaje de especialidad matemático de la teoría de redes complejas	118
SEGUNDA PARTE PRÁCTICA	121
Capítulo 3: Estudio del corpus mediante Redes complejas.....	121
3.1 Presentación del enfoque para la caracterización del lenguaje matemático....	122
3.2 Proceso de creación de la herramienta: convertir el corpus en una red compleja multicapa	123
3.3 Asociaciones entre capas y palabras: colocaciones	125
3.4 Descripción formal del corpus usando Python.....	127
3.5 Uso de <i>Line Graph</i> y Centralidad del <i>PageRank</i> en nuestra Red.....	130
3.6 Nuestro modelo multicapa en detalle. Resultados.	132
3.7 Resultados computacionales sobre el lenguaje de especialidad presente en el corpus	135
3.8 Las colocaciones y el lenguaje de especialidad. Presentación del Kendall (τ) para este caso	145
3.9 Centralidades <i>PageRank</i> de los términos (unidades) del corpus. Mejora del uso del <i>PageRank</i> sobre la frecuencia absoluta para detectar palabras importantes...	146
3.10 Resultados computacionales: sobre la estructura de este lenguaje matemático de especialidad y sus colocaciones principales	150
3.11 Recapitulaciones sobre el estudio del corpus mediante redes complejas	154

Capítulo 4: Conclusiones	157
Conclusiones parciales.....	157
Conclusiones sobre el estudio del lenguaje de especialidad matemático a través de las redes complejas. Trabajo futuro.	158
Bibliografía	169
Anexos.....	175

Introducción

Dentro del mundo académico, nos encontramos con toda clase de normas que tienen como fin tratar de homogeneizar el estilo lingüístico o de redacción de las distintas publicaciones científicas. Partiendo de la base de que las publicaciones científicas, dado su carácter académico, están redactadas usando lenguaje de especialidad, este trabajo surge de la necesidad de una norma para verbalizar (y redactar) el lenguaje simbólico-numérico matemático ya que, hasta la fecha, no consta norma alguna generalizada sobre cuál es la mejor forma de verbalizar este tipo de lenguaje. Además, es necesario tener adquirido un conocimiento matemático previo, tanto para enfrentar la verbalización del lenguaje simbólico numérico como para traducir de un idioma a otro este tipo de textos. Posiblemente debido a esta circunstancia (entre otras), la mayoría de las traducciones de textos matemáticos no son realizadas por licenciados o graduados en estudios de traducción, sino por especialistas en la materia capaces de expresarse en dos idiomas.

Al no existir pautas de redacción del lenguaje matemático de especialidad, tampoco existen pautas de traducción. Por comparación, es fácil pensar que las pautas de redacción que deben seguirse ante textos de especialidad matemática han de ser similares –por no decir iguales– a las de otros lenguajes de especialidad plagados de neologismos. Sin embargo, determinar la estrategia que debe seguirse para traducir textos matemáticos es complicado, ya que los distintos lenguajes de especialidad suelen haber sido objeto de estudios lingüísticos que han dado lugar a la identificación de distintos problemas de redacción y traducción (el lenguaje matemático, por su parte, ha sido muy poco estudiado). Por ejemplo, podemos ver cómo, en González (2016), la autora nos habla de distintos tipos de préstamos, neologismos y calcos en el lenguaje de especialidad económico y describe distintas pautas dependiendo no sólo del tipo de neologismo identificado (formal, semántico o de préstamo), sino dependiendo también del grado de especialización textual y de factores como el área económica en la que se pueda englobar el texto en cuestión. Este tipo de conclusiones también son necesarias para redactar correctamente textos matemáticos, de ahí la necesidad de analizar este lenguaje cuando es verbalizado y considerarlo un lenguaje de especialidad como tal. El

lenguaje matemático cumple con todas las características para ser considerado como lenguaje de especialidad y, además, añade una: los símbolos específicos utilizados en el lenguaje simbólico-numérico matemático son, a su vez, un lenguaje en sí mismo, que puede ser comprendido sin necesidad de utilizar el lenguaje común en diversas ocasiones (Σ , ∞ y la expresión $2+2=4$ son solo un ejemplo de expresiones matemáticas que no necesitan del lenguaje verbalizado para ser comprendidas).

El objetivo principal de este trabajo es alcanzar una mayor comprensión de los lenguajes de especialidad y, más concretamente, del lenguaje de especialidad matemático utilizado por la comunidad científica que publica sobre la teoría de redes complejas, con vistas a ofrecer pautas de traducción y redacción de este lenguaje de especialidad. Para ello se presenta una nueva perspectiva de análisis utilizando la teoría de redes complejas y un nuevo modelo matemático diseñado para estudiar y analizar las relaciones entre palabras de especialidad y cuándo estas forman una unidad de significado específico. Este modelo se denomina *Enriched Line Graph* y ha sido presentado por primera vez en Criado-Alonso et al. (2021a)¹.

Entendemos por lenguaje de especialidad aquel que tiene un vocabulario distinto tal y como establece Pinchuk (1977). Este es el caso del lenguaje matemático verbal, que goza de una terminología exclusiva. Sin embargo, no existe ningún tipo de guía sobre cómo verbalizar el lenguaje matemático, muy poco estudiado en términos lingüísticos pese a ser un lenguaje transversal a muchísimas disciplinas. Incluso en el lenguaje común (o vehicular) se utilizan términos matemáticos que alteran su significado fuera del ámbito de las matemáticas.

Esta tesis doctoral arroja nueva luz sobre enfoques anteriores descritos por autores como Parodi (2008), cuya orientación va enfocada a desentrañar los sistemas del lenguaje a través de conjuntos de textos paralelos (o corpus lingüísticos especialidad, que ayudan a identificar las palabras clave de manera tradicional). Esto, en el caso de

¹ Este artículo (Criado-Alonso et al., 2021) es uno de los trabajos publicados sobre la investigación desarrollada en esta Tesis. La Escuela Internacional de Doctorado de la URJC establece como requisito para la defensa de la Tesis doctoral la publicación de un artículo de impacto que, en nuestro caso, es Criado-Alonso et al. (2020). Una relación precisa de los trabajos publicados en torno a esta Tesis doctoral se puede encontrar en Referencias bibliográficas y se corresponden con los siguientes trabajos: Criado-Alonso (2021) y Criado-Alonso et al. (2020, 2021a y 2021b).

corpus de textos de lenguaje matemático no es efectivo ya que hay muchas unidades de conocimiento especializado compuestas por conjuntos de términos y, sobre todo, por pares de términos.

El significado de las palabras o términos no viene dado por un único elemento o palabra, sino por un conjunto de términos, y, este hecho, llevado al lenguaje de especialidad, supone una dificultad para un lector no experto en la materia o ciencia a la que pertenece un texto escrito en un determinado lenguaje de especialidad. Suponiendo, por ejemplo, que se desee comprender un texto escrito en lenguaje técnico o especializado, el lector tendrá que saber cuáles son estas asociaciones de términos para entender correctamente los conocimientos y el mensaje que pretende transmitir el autor de dicho texto. Complicando un poco más la tarea de comprensión, podemos pensar que el lector está ante un texto que no se encuentra escrito en su lengua materna y que desea traducirlo (como sucede en muchos casos con las publicaciones científicas del área de la matemática aplicada).

En este punto es fundamental estar familiarizado con las colocaciones o combinaciones de palabras tal y como las conciben autores como Koike (2008) en forma de locuciones o conjuntos de palabras de especialidad que, asociadas, forman unidades de significado –en el caso de los lenguajes de especialidad, unidades de conocimiento específico– donde queda reflejado cómo las palabras que aparecen de forma consecutiva adquieren un significado diferente. Con vistas a traducir un texto de manera adecuada, es necesario entender los conceptos de especialidad que, a veces, no vienen expresados exclusivamente por un término específico, sino por dinámicas entre dos o más términos. A veces, un término puede ser considerado como específico –es decir, puede estar terminologizado– debido a las palabras con las que dicho término aparece combinado, sin formar parte de este contexto de especialidad en caso de no aparecer en lo que Santamaría-Pérez (2009) llama “unidades de conocimiento especializado”.

Para entender en profundidad las “unidades de conocimiento especializado”, es importante contextualizar dónde aparecen que, en este caso, es en el lenguaje de especialidad. Por ejemplo, en el caso del término “primo”, si aparece en el ámbito de las matemáticas acompañado de la palabra número, formando la colocación “número primo”, se obtiene un significado diferente del que un hablante piensa si verbaliza la palabra primo fuera del contexto de especialidad.

El primer objetivo de esta tesis es demostrar que el matemático es, efectivamente, un lenguaje que entra dentro del espectro catalogable como “lenguajes de especialidad” y, tras un análisis del mismo, mostraremos que cumple con las características descritas en Pérez Hernández (2002) para definir o caracterizar un lenguaje de especialidad (especificidad del tema; restricción léxica, semántica y sintáctica; reglas gramaticales específicas y poco frecuentes; frecuencia de aparición de estructuras lingüísticas particulares; macro-estructuras de texto homogéneas; y, por último, el uso de símbolos específicos). El hecho de que exista más de un tipo de lenguaje matemático (o distintos sub-lenguajes matemáticos) dificulta aún más la tarea del traductor que desea especializarse en este tipo de lenguaje. Serrano Gómez (2005) ya sienta el precedente de que se necesita hablar de algún tipo de subcategorización añadida para poder analizar y extraer conclusiones referentes a este lenguaje de especialidad; esto motiva a que esta tesis doctoral se acote al lenguaje de especialidad que tiene que ver con la teoría de redes complejas.

Distintos expertos como Baker ya hablaban en fechas tempranas de 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 (Baker, 1993). Para poder utilizar la lingüística de corpus como herramienta de análisis del lenguaje de especialidad, es necesario contar con un corpus lingüístico representativo de esa área de especialidad tal y como apuntan diversos estudios [véanse por ejemplo los criterios de representatividad expuestos en Atkins et al., (1991)]

El segundo objetivo es realizar un análisis lingüístico sobre un corpus lingüístico lo suficientemente representativo. Para este caso, se han utilizado los *extended abstracts* de la revista IJCSS (*International Journal of Complex Systems in Sciences*) incluyendo los volúmenes 1 a 6 publicados entre abril de 2011 y noviembre de 2016 (<http://www.ij-css.org>), cuyas áreas de especialidad matemática están centradas en la matemática aplicada en general y muy en particular en la teoría de grafos o redes complejas. Se trata de un corpus redactado en inglés.

Dicho en otras palabras: se utiliza la teoría de redes complejas para determinar el lenguaje de especialidad que tiene que ver con la teoría de redes compleja. Así, cada una de las palabras del corpus se convierte en un nodo (o vértice) y, siempre que

aparezca una palabra a continuación de la otra, aparecerá una arista dirigida entre dichos nodos (considerando, en principio, que las dos palabras están relacionadas) estableciendo qué término va delante y qué término va detrás.

No es la primera vez que se crea una red compleja para analizar una lengua. De hecho, el modelo de red lingüística considerado en este trabajo surge de la naturaleza variable del texto, que está siempre orientado hacia adelante, y es por eso por lo que se utilizan enlaces dirigidos y ponderados para representar las relaciones entre las unidades lingüísticas como en Cong & Liu (2014), Martincic S et al. (2016) o Masucci A & Rodgers (2006). De este modo, la relación de co-ocurrencia se establece entre dos palabras o unidades lingüísticas adyacentes en una frase, en la que la dirección del enlace refleja la secuencia de las palabras, y el peso en el enlace refleja la frecuencia de aparición de esa secuencia de dos palabras enlazadas (se trata de una red compleja dirigida pesada). La aplicación de la teoría y las herramientas computacionales de las redes complejas al estudio de las lenguas no es una tendencia científica nueva [(Borge-Holthoefer & Arenas, 2010), Cárdenas JP. et al (2014), Criado-Alonso et al.(2020)].

Las redes multicapa constituyen un marco unificado que permite modelar 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. Para este análisis en concreto nos interesa considerar una red multicapa, tal y como quedó establecido en Criado-Alonso et al, (2020), compuesta por cuatro capas:

1. Una capa en la que hemos incluido las palabras específicas (principalmente adjetivos y sustantivos) de la lengua de especialidad, que llamamos capa léxica (*lexical*). 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".

2. La capa de verbos (*verbs*), en la que hemos incluido todos los verbos independientemente de su forma conjugada.

3. Una capa, a la que denominamos capa de enlaces (*linking*) 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. Y, por último, una capa, a la que llamamos capa-resto (*remaining*), en la que colocamos el resto de 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.

En este punto, es importante resaltar que se ha realizado un proceso exhaustivo de limpieza sobre el corpus objeto de análisis de tal manera que no se ha tenido en cuenta el lenguaje simbólico-numérico matemático. Dicho de otro modo, no se han considerado como palabras los elementos del lenguaje artificial contemplados en las distintas fórmulas. También es notable el hecho de que el lenguaje matemático de redes complejas está compuesto por distintos términos categorizables en las distintas capas de las que se compone la red creada (dejando a un lado el lenguaje simbólico-numérico).

La principal aportación de esta nueva perspectiva es que las palabras importantes ya no son palabras importantes porque se repitan muchas veces, sino que son palabras importantes porque están asociadas a otras palabras que también lo son (es decir, se conectan unos nodos de alto grado con otros nodos considerados de alto grado) y esto nos permite identificar matemáticamente las unidades de conocimiento especializado (incluso las colocaciones de especialidad, que son las unidades de conocimiento especializado compuestas por dos términos o palabras) que aluden a un concepto concreto dentro de este campo de especialidad. Dicho de otro modo, se pasa del estudio de las relaciones paradigmáticas de los términos del corpus a poder contemplar el funcionamiento de las relaciones sintagmáticas. El uso del concepto de *Line Graph* (caracterizado por el hecho de que sus nodos se corresponden con las aristas del grafo original) junto con la red compleja multicapa pesada y dirigida es lo suficientemente sensible como para determinar cuáles son los nodos que formarían estas colocaciones de especialidad.

Cualquier secuencia de dos unidades lingüísticas o palabras colocadas consecutivamente en un texto puede representarse mediante el enlace entre esas dos

palabras, que pasarían a ser un nodo en el *Line Graph* asociado. Por lo tanto, una colocación puede identificarse con el enlace que une las palabras que la componen y, en consecuencia, con el nodo correspondiente del *Line Graph* de la red considerada. Es importante mencionar aquí que existe una conexión, y en algunos casos una relación directa entre las propiedades de un grafo y las de su *Line Graph* asociado (Criado et al., 2018). Está claro, intuitivamente hablando, que el *PageRank* de una arista de un grafo dirigido y ponderado $G = (X, E)$, es decir, la frecuencia con la que el caminante aleatorio pasa por la arista $(i, j) \in E$ está relacionada con el *PageRank* del propio nodo i , ya que cada vez que un caminante aleatorio pasa por la arista $i \rightarrow j$ este pasa también por el nodo i . En el estudio citado, se presenta un método para calcular el *PageRank* de las aristas de un grafo ponderado y dirigido a partir de su *Line Graph* asociado (Criado et al., 2018). Así, siguiendo a Criado et al. (2018), podemos obtener el *PageRank* de la arista $i \rightarrow j$ mediante la expresión $LPR(i \rightarrow j) = PR(i) * pij$, en la que pij es la probabilidad de que el paseante aleatorio siga su camino desde el vértice i hacia el vértice j en lugar de ir hacia otro nodo directamente conectado con él.

Para llevar a cabo un estudio en profundidad del lenguaje de especialidad matemático relacionado con las redes complejas, se ha dividido esta tesis doctoral en dos partes. La primera parte está dedicada fundamentalmente a la traducción y los lenguajes de especialidad, y pone el énfasis en la dificultad de traducir los textos que pertenecen a un área concreta del saber (es decir, a un área de especialidad). La segunda parte de esta tesis presenta una nueva perspectiva de análisis de textos basada en la teoría de las redes complejas con un enfoque totalmente novedoso a través de una red compleja dirigida pesada multicapa, así como los resultados obtenidos sobre el corpus que ha sido objeto de análisis. En los anexos que aparecen adjuntos también se puede encontrar una lista de palabras y colocaciones (o unidades de conocimiento específico) que son “clave” para entender el lenguaje de especialidad matemático (además, se desarrollan propuestas de traducción de algunos de estos términos clave a lo largo de los distintos capítulos de este trabajo). Estas palabras y colocaciones clave se pueden obtener a partir de la observación y modelación de la red compleja multicapa anteriormente descrita de manera casi automática, lo que supone un gran avance ya que, previamente, se identificaban todos estos elementos en base a consideraciones

tradicionales tales como su frecuencia de aparición. Además, gracias a esta técnica se pueden identificar las unidades de significado compuestas por más de un término.

Para describir el proceso de análisis y extracción de conclusiones, se puede hacer un recorrido a través de la descripción de los apartados contemplados en el índice de esta tesis doctoral y de cada una de sus dos partes. La primera parte se encuentra dividida en dos capítulos. En el primer capítulo, se habla sobre los distintos lenguajes de especialidad y del papel que ejerce la figura del traductor como terminólogo (cuando se traducen textos de algún área específica del saber, se adaptan los términos específicos a la lengua meta. En esta parte de la tesis, se hablará sobre esos mecanismos de adaptación). También en el primer capítulo se hace un repaso sobre los tipos de términos que aparecen en las lenguas de especialidad (terminología específica), y se presentan tanto las unidades de conocimiento específico o unidades de conocimiento especializado como las técnicas de traducción que se han de seguir para lograr un texto meta fiel al original (para ello, en la sección 1.4 se habla de la importancia de la semiótica en la traducción y en las unidades de conocimiento especializado).

Una vez presentadas las peculiaridades y características del lenguaje de especialidad y las dificultades que este tipo de lenguaje plantea a la figura del traductor, los apartados 1.5 y 1.6 justifican que el lenguaje científico –y más concretamente el lenguaje matemático– han de ser considerados como un lenguaje de especialidad ya que las características expuestas en los apartados anteriores se ven sobradamente satisfechas (entre otras, el uso de palabras y términos específicos, la presencia de colocaciones y fraseologías consideradas de especialidad, el uso del lenguaje artificial y la aparición de palabras que se muestran terminologizadas).

La siguiente sección (1.7) del capítulo 1 habla sobre el motivo por el que el lenguaje de especialidad matemático no solo es un lenguaje de especialidad en sí mismo, sino que a su vez está compuesto por diversos lenguajes de especialidad. En este apartado se habla del lenguaje matemático de especialidad del área de las redes complejas puesto que es precisamente esta área matemática la que se utilizará para desarrollar la herramienta de análisis textual presentada en la segunda parte de esta tesis doctoral.

El apartado 1.8 del capítulo 1 desarrolla los conceptos necesarios para entender el funcionamiento de la herramienta de análisis textual planteada en la parte dos de la

tesis (entre otras las siguientes: qué es una red compleja, qué es un grafo, qué es una red asortativa, o una red multicapa, a qué nos referimos cuando hablamos de *PageRank* o en qué consiste el concepto de *Line Graph*).

En el capítulo 2 de la primera parte teórica, se presenta la metodología de análisis que tradicionalmente se ha utilizado en el ámbito de los lenguajes de especialidad: la lingüística de Corpus. A lo largo de los primeros apartados de este capítulo se habla sobre el motivo por el cual el uso de un corpus es el mejor mecanismo de análisis para este caso y se hace un análisis del estado actual de la situación en lo que a lingüística de corpus se refiere. En este capítulo se presenta el corpus que ha sido objeto de análisis (basado en textos que tratan sobre la teoría de redes complejas) y se hace una presentación sobre el lenguaje matemático de especialidad de redes complejas a través de varias herramientas de análisis textual ya existentes (*Wconcord*, *Flair* y *Voyant*) de tal forma que se pueda apreciar la mejora que supone el uso de la herramienta de análisis que se propone en la segunda parte de la tesis. El capítulo dos propone las primeras conclusiones sobre el lenguaje matemático de especialidad de redes complejas en lo que se refiere a la identificación de palabras clave y otras características generales de estilo imitables por la figura del traductor y de todo aquel que tenga que redactar documentos en este lenguaje de especialidad en concreto. Además, se ofrece una propuesta de traducción sobre algunas unidades que se han identificado como recurrentes dentro del corpus objeto de análisis.

La segunda parte de esta tesis doctoral está formada por el capítulo 3 y las conclusiones. En el capítulo tres se describe de manera exhaustiva la metodología empleada y se describe cómo se ha creado la red compleja dirigida multicapa a través de la cual se han podido extraer conclusiones sobre el lenguaje de especialidad. Se explica cómo cada una de las palabras del corpus se ha convertido en un nodo, el criterio para establecer una conexión o arista entre los nodos y se presentan las capas consideradas y el proceso que han seguido las palabras o nodos para ser incluidas en cada una de estas capas. También se pone de manifiesto el hecho de que las asociaciones entre las distintas capas de esta red forman las colocaciones de especialidad que resultaron ser, precisamente, las unidades de conocimiento específico que caracterizan este lenguaje de especialidad (esto se corrobora tanto en el apartado 3.3 como en el hecho de que las colocaciones sean el elemento más característico de

este tipo de lenguaje, según se destaca en el apartado 3.10). En el capítulo 3 se hace una nueva descripción formal del corpus, pero, en esta ocasión, en lugar de utilizar las herramientas tradicionales (*Wconcord*, *Flair* y *Voyant*), se utiliza el lenguaje de programación Python 3.7 y, en el apartado 3.7 se presenta al lector cómo el uso del concepto de *Line Graph* y la centralidad de *PageRank* sobre nuestra red han sido determinantes para extraer los resultados que se presentan.

En el capítulo 3 se ofrecen los resultados del estudio sobre la red multicapa, así como los resultados computacionales relativos al lenguaje de especialidad utilizado en los textos de los que está compuesto el corpus. Como se decía en el párrafo anterior, en el apartado 3.8 del capítulo 3, a través del índice τ de Kendall, se puede observar el comportamiento de las colocaciones de especialidad de dos términos como unidades de conocimiento específico características de este lenguaje de especialidad frente a las unidades de conocimiento específico compuestas por un único término. En el apartado 3.9 se muestra al lector la mejor que implica utilizar el algoritmo *PageRank* sobre las medidas estadísticas tradicionales como la frecuencia absoluta para detectar las palabras más importantes dentro de un corpus (utilizando el criterio seguido por el algoritmo *PageRank*, una palabra es importante no sólo porque se repite muchas veces, sino porque aparece junto a otra palabra que también es importante). Todo este estudio permite hablar en el apartado 3.10 de resultados computacionales sobre la estructura del lenguaje matemático de especialidad que tiene que ver con la teoría de redes complejas y sus colocaciones ofreciendo una serie de palabras clave y colocaciones clave que ayudan a caracterizarlo.

Finalmente, en el capítulo dedicado a la discusión y conclusiones, se presentan las conclusiones derivadas del estudio realizado sobre este lenguaje de especialidad en el que hemos utilizado la teoría de redes complejas para su análisis, así como los motivos de por qué es mejor o más apropiado emplear esta técnica y no la “tradicional” que se sustenta sobre la frecuencia absoluta de aparición de los términos y su comparación con textos paralelos. Por último, se reflexiona acerca de las ventajas e inconvenientes que implica el uso de esta nueva perspectiva y se hace una presentación del trabajo futuro que se podrá abordar mediante el empleo de esta técnica.

PRIMERA PARTE TEÓRICA

Capítulo 1: La traducción y el lenguaje de especialidad

1.1 El papel del traductor como terminólogo en el lenguaje de especialidad

El mundo de la traducción (o escribir un texto en otro idioma) está muy unido al mundo de los lenguajes de especialidad y, a su vez, al de la terminología específica. Para estudiar el lenguaje de especialidad desde el punto de vista de la traducción es necesario empezar hablando de la traducción como tal y de los orígenes de la enseñanza de idiomas para fines específicos.

Durante los años sesenta, la lengua inglesa dominaba la comunicación especializada internacional, tal y como se resalta en estudios como los de Alcalá & Antuña (2011). Esta comunicación especializada en inglés hizo que los académicos no anglófonos quisieran aprender a expresarse en esta lengua para que sus trabajos fueran más conocidos, formándose en lo que en enseñanza se conoce como “inglés con propósitos específicos” (*English for Specific Purposes, ESP*). No solo se puso de moda la enseñanza del inglés con propósitos específicos, sino que también se suscitó la curiosidad por investigar el funcionamiento de lo que se llamaron LFE (Lenguas para Fines Específicos). No es sorprendente que las lenguas estén llenas de términos específicos si la finalidad de un texto es la de comunicar una idea en un área muy concreta del saber.

Haciendo un repaso a la historia se puede entender por qué el inglés fue el vehículo de comunicación principal de las lenguas para fines específicos. Tras la Segunda Guerra Mundial, se produjo un incremento considerable de la actividad económica, científica y tecnológica que requirió de la utilización de una sola lengua para la comunicación internacional: el inglés. Gracias al potencial económico de Estados Unidos, esta nación se encontraba ya en una posición ventajosa para convertirse en la ostentadora de la lengua internacional por lo prolífico de sus publicaciones especializadas. Los términos o palabras específicas usados en este tipo de comunicación hicieron que la terminología como disciplina entrase en juego en el panorama de las

ciencias, ya que busca explicar cómo el conocimiento especializado se estructura en unidades conceptuales que denominan realidades en un sistema de expresión muy concreto de modo que, sin los términos (o terminología específica), la comunicación especializada no sería posible. Así, el traductor se valdrá de la terminología para expresar el conocimiento especializado con adecuación.

En todo caso, es destacable el hecho de que el traductor necesita de la terminología, pero no ha de ser considerado exclusivamente como un terminólogo, ya que la acepción de terminología variará dependiendo del enfoque que utilizan distintos autores para estudiar las palabras. Se puede entender la terminología como una actividad encaminada a la recopilación de términos (y el uso de los términos o palabras) según las distintas situaciones comunicativas.

Estudios como los de Cabré Castellví (2004, págs. 89-125) dan a entender que, de algún modo, los traductores ejercen como terminólogos 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. Estos problemas de traducción están muy ligados a los lenguajes de especialidad; en un primer momento, diremos que los lenguajes de especialidad son aquellos que permiten a los especialistas en una materia transferir su conocimiento.

La terminología que se emplea en el lenguaje de especialidad (y su funcionamiento) es la base para determinar de qué trata un texto que nos habla sobre un área concreta del saber. Cabré Castellví (2004, págs. 89-125) establece una comparación entre la figura del traductor y un selector de términos. La ordenación del pensamiento y la conceptualización representan la dimensión cognitiva de la terminología y, por tanto, el traductor de especialidad es una figura mediadora que escoge qué términos o expresiones utiliza en el idioma hacia el que traduce para transmitir la totalidad del mensaje del texto origen.

Pese a que no se trata de la misma disciplina, la traducción y la terminología están muy unidas, sobre todo si tenemos en cuenta que se sirven la una de la otra para transmitir conocimientos. Cabré Castellví (2004) resalta el hecho de que no se puede saber exactamente cuál fue el punto en el que se utilizó por primera vez un término para expresar conocimiento especializado (es decir, el origen de la terminología). Tampoco

se sabe con exactitud el momento en que nació el conocimiento especializado en contraste con el conocimiento general, aunque se intuya que el momento más favorable fuera aquel en el que el colectivo humano distribuyó las funciones sociales.

Paralelamente, es igualmente difícil determinar cuándo nació la traducción como acto de traslación de una idea expresada originalmente en otra lengua o en otro sistema de expresión (aunque se tenga la certeza de que el origen fue oral). Cabré Castellví (2004, págs. 89-125) destaca que sí que constan testimonios históricos de actividades de traducción en crónicas de viajes y conquistas, ya que se conservan muestras de documentos originales que son prueba de que la actividad traductora ya existía en la práctica.

En esta línea conceptual sobre el acto de traducir, si entendemos el concepto de traducción como “el traslado del discurso de una lengua a otra” (Schleiermacher & Yebra, 2000, pág. 224), podemos toparnos con muchas dificultades a la hora de realizar esa tarea o traslado, especialmente si el traductor en cuestión se encuentra ante un texto escrito en lenguaje de especialidad. Para poder acercarnos a las dificultades señaladas, pasaremos a describir o definir qué se entiende por lenguaje de especialidad.

1.2 Lenguaje de especialidad y terminología específica

El lenguaje es la materia en torno a la que giran tanto la terminología como la traducción ya que se trata del “sistema expresivo que refleja la concepción que los hablantes tienen de la realidad y que permite a los individuos relacionarse y expresar sus ideas y pensamientos” (Cabré Castellví, 2004, pág. 94). Cabe resaltar que la terminología y la traducción tienen un origen práctico ya que tratan de resolver problemas de comprensión en un texto origen². La terminología tiene un origen que se vincula a la necesidad del individuo de expresar la realidad de modo preciso y exacto (se usan términos específicos de manera inconsciente). De este modo, la terminología específica se desarrolla en respuesta al interés de los científicos por llegar a acuerdos para fijar conceptos y denominaciones en sus respectivos ámbitos de manera que la comunicación entre especialistas sea posible y unívoca.

² En traducción se habla de texto origen y de texto meta, siendo el texto meta la traducción del primero.

El auge del mundo científico incrementó el interés por el léxico especializado, ya que fue entonces cuando eran más necesarios los estudios terminológicos tal como resalta Santamaría-Pérez (2006). Una vez se ha definido el concepto de traducción en los apartados anteriores, también hay que profundizar en el concepto de terminología como tal. Santamaría Pérez afirma que la terminología se encuentra definida sobre la base de tres acepciones diferentes: “disciplina, práctica y producto generado por esa práctica” (Santamaría Pérez, 2006, pág. 3). Entendemos de este modo la terminología como la disciplina que se ocupa de los términos; por otro lado, encontramos la práctica terminológica donde se estudian y analizan los principios que rigen la recopilación de los distintos términos. Y en último lugar, y siguiendo lo propuesto por Santamaría-Pérez, nos topamos con la tercera manera de concebir la terminología como el conjunto de términos de una materia especializada.

Las dos primeras definiciones propuestas nos ofrecen una concepción de la terminología que no es tan reciente como la tercera. Podemos pensar en la elaboración, desde el siglo XVIII, de glosarios terminológicos, vocabularios y otras herramientas relacionadas, por poner solo un ejemplo, con el léxico de la mano de Lavoisier o Berthold en química o Linneo en las áreas de la botánica y la zoología. Cuando estamos dentro de un área de especialidad (o que no todo el mundo domina), dentro del ámbito científico, es necesario desarrollar una denominación de los nuevos conceptos.

El léxico especializado utilizado en el mundo científico se caracteriza por adoptar palabras procedentes de otros ámbitos léxicos³ –pudiendo adoptarse de otro idioma– y por la creación de términos por medio de mecanismos *lexicogénicos*⁴. Martín Camacho (2004, pág. 159) afirma que las estrategias empleadas en la conformación del léxico científico se pueden agrupar en dos grandes bloques: adopción de léxico científico y creación del léxico científico.

³ La habilitación de términos a partir de palabras del vocabulario común de la misma lengua a un ámbito del saber específico es lo que entendemos por terminologización. Según Martín Camacho (2004, pág. 160), mediante la terminologización se atribuye a una palabra del léxico común un nuevo significado, más preciso y específico que el originario.

⁴ Un mecanismo lexicogénico es un mecanismo a través del cual se genera un término o palabra.

Dentro del primer bloque –cuando el léxico científico es adoptado–, Martín Camacho (2004, págs. 159-163) establece una clasificación de términos en el léxico científico en función del procedimiento o proceso de adopción que el término especializado ha sufrido. Se observan hasta cuatro fenómenos distintos: la habilitación de términos a partir de palabras del vocabulario común de la misma lengua (terminologización); el paso de términos de unas ciencias a otras (trasvases); la adopción de términos de otras lenguas coetáneas (extranjerismos) y la conversión en términos de elementos léxicos procedentes de las lenguas clásicas (cultismos). El fenómeno de la terminologización puede ilustrarse con los ejemplos señalados por el propio autor Martín Camacho (2004, pág. 160): ácido (química), acoplar(se) (física), bastón (biología), clase (botánica, zoología, lingüística), diferenciación (biología, matemáticas), dinámica (física), dispersión (física, química, matemáticas), familia (biología), narcótico (medicina), tálamo (anatomía, botánica) o trabajo (física).

En el segundo bloque, hablamos de términos de nueva creación. La lexicogénesis de los términos específicos procesos se divide en dos bloques: los términos creados que tienen en cuenta los componentes morfológicos de las palabras (morfemáticos) frente a aquellos que no atienden a la estructura morfológica (no morfemáticos).

Martín Camacho (2004, págs. 165-166) pone ejemplos de creación léxica morfemática con afirmaciones como “en el término botánico *labiado*, labio adquiere un significado distinto al del léxico estándar”. Asimismo, habla de términos de creación morfemática o cuyo morfema conserva inalterado el significado de la lengua común como “calor en calorímetro”. En la línea de los términos morfemáticos, también considera los términos científicos cuyos morfemas adoptados del latín y del griego se utilizan únicamente para formar palabras científicas especializadas; son muy frecuentes en el vocabulario de las ciencias las formas endo-, aniso-, -oma, -osis (afijos de palabra); aero, patía, anemo (raíces de palabra).

Los términos no morfemáticos de los que nos habla el autor se crean generalmente a partir de la siglación (uso de siglas como en el caso de SIDA o ADN); a partir de abreviaturas (“el léxico científico se emplea Rh (reducción de *Rhesus*, nomenclatura del género de primates en el que se descubrió este componente de la sangre)”, Martín Camacho (2004, pág. 165); a partir de un proceso de acronimia como en biónica (del gr. bio~ y electrónica); a partir de nombres propios (por ejemplo, los

términos “angstrom”, “atlas”. Llamen la atención especialmente los de este último grupo de nombres propios, ya que Martín Camacho, plantea que estas palabras son “parte de un compuesto sintagmático: “aparato de Golgi, campo de Higgs, círculo de Mohr, cuerpo de Malpigio, número de Avogadro”. Estos compuestos sintagmáticos son, en el ámbito de la traducción, unidades de conocimiento especializado. Una vez más, se observa el papel del traductor como terminólogo cuando escoge cómo traducir o acuñar un término por primera vez.

Una vez se ha contextualizado la traducción y la terminología de especialidad, se llega al punto en el que existe una necesidad de establecer con exactitud qué es una lengua de especialidad (o LESP, tal y como es mencionada por diversos autores). Autores como Alcalá y Antuña (2011) reflexionan sobre la falta de consenso que existe sobre este concepto como tal. De nuevo, nos encontramos con un término que es acotado por diversos autores de manera dispar; incluso la elección de la denominación “lengua de especialidad” es objeto de controversia dentro de la lingüística aplicada, ya que:

Han sido muchos y muy variados los intentos por definir y delimitar las LESP con respecto al lenguaje general. Tales intentos discurren desde los que defienden las LESP como simples variantes léxicas de una lengua determinada (Rey, 1976; Quemada, 1978; Rondeau, 1983), hasta aquellos que han postulado la existencia de los lenguajes de especialidad como códigos lingüísticos diferenciados del lenguaje general. (Alcalá & Antuña, 2011, pág. 915).

Existen muchas denominaciones sobre qué es el lenguaje de especialidad y la terminología específica. A la hora de establecer una disciplina que estudie este fenómeno lingüístico que tiene que ver con el uso de términos de especialidad (y de la que se nutren todo tipo de profesiones relacionadas con el lenguaje), es primordial establecer denominaciones y conceptos. Para ilustrar la pluralidad de enfoques, el estudio de Alcalá y Antuña (2011, pág. 911) resalta que:

La necesidad de una delimitación conceptual y terminológica se justifica no solo por la abundancia de términos, sino –y esto es lo más importante– porque toda disciplina que reclame un estatus científico debe ser cuestionada y debe someter su terminología a revisión crítica (Gómez de Enterría, 2006: 49; 2009: 14). Términos como lenguas de especialidad (Gómez de Enterría, 2009), lenguajes de especialidad (Cabré Castellví, 1993), lenguas especializadas (Lerat, 1997), lenguajes especializados por la temática

(Sager et alii, 1980), lenguas especiales (Rodríguez Díez, 1980), lenguaje de la ciencia y de la técnica (Gili Gaya, 1964), tecnolecto (Haensch, 1987), lengua técnica (Quemada, 1978), microlengua (Balboni, 1982), lengua de minoría (Hernán Ramírez, 1979), lenguajes con fines específicos (Beaugrande, 1987), lenguajes específicos, lenguaje científico (Gutiérrez Rodilla, 2005), lenguaje sectorial (Beccaria, 1973), discurso científico (Grabarczyk, 1988), sottocodice (Berruto, 1997), sublenguaje (Lehrberger, 1982), etc. Han convivido en casi toda la bibliografía relativa al tema como equivalentes terminológicos. (Alcalá & Antuña, 2011, pág. 911)

Esta amplia bibliografía sobre el tema es prueba de que existen posiciones distintas ante la concepción de la terminología como disciplina: en primer lugar, autores como Hoffman (1998), conciben la terminología como una disciplina autónoma conectada históricamente a otras disciplinas pero que tiene sus propias bases o fundamentos y es capaz de establecerse de manera autónoma; en segundo lugar, según otros autores como Rondeau (1983), se puede hablar de la idea de que la terminología no es una disciplina autónoma, sino que es Parte de otras disciplinas, estando asociada a la lingüística, a la filosofía y a otras especialidades científico-técnicas; y, en tercer lugar, autoras como Cabré Castellví (2004) conciben la terminología como una materia interdisciplinar con base en otras materias pero con su propio ámbito científico en desarrollo. Cualquiera de estos enfoques sobre la LFE está orientado hacia una lengua llena de terminología específica, vinculando de nuevo la LFE directamente con la terminología.

Con vistas a traducir un texto de manera adecuada, es necesario entender los conceptos de especialidad que, a veces, no vienen expresados exclusivamente por un término específico, sino por dinámicas entre dos o más términos. A veces, un término puede ser considerado como específico (es decir, puede estar terminologizado) debido a las palabras con las que dicho término aparece combinado, no formando parte de este contexto de especialidad en caso de no aparecer en lo que Santamaría-Pérez llama “Unidades de conocimiento especializado” en su obra de 2009. El estudio de las unidades de conocimiento especializado es un enfoque para estudiar el léxico que viene de la mano de la lingüística y que es de gran interés para el mundo de la traducción ya que son estas unidades las que el traductor tendrá que reproducir en el texto meta para que tenga sentido. Para entender en profundidad las “Unidades de conocimiento

especializado”, es importante contextualizar dónde aparecen que, en este caso, es en el lenguaje de especialidad.

Los criterios que recoge Pérez Hernández (2002) para definir o caracterizar un lenguaje de especialidad son seis: especificidad del tema; restricción léxica, semántica y sintáctica; reglas gramaticales específicas y poco frecuentes; frecuencia de aparición de estructuras lingüísticas particulares; macroestructuras de texto homogéneas; y, por último, el uso de símbolos específicos.

Santamaría-Pérez (2006) parte entonces de esa base y habla concretamente de “Unidades de Conocimiento Especializado”, que son términos o grupos de términos de difícil comprensión por su pertenencia a una temática específica. Así, cuanto más específica es la temática de un texto, más difícil es la tarea del traductor no especializado ya que el texto en sí mismo presentará una mayor dificultad de comprensión, puesto que estas unidades de conocimiento específico estarán compuestas de una serie de términos o palabras que existen y/o actúan de manera específica dentro de un área del saber.

En el mundo de la traducción y de la terminología específica, se trabaja con el concepto de *colocación*. Una colocación es, según la obra de Koike (2001), una asociación de dos palabras que, cuando aparecen juntas una a continuación de la otra, adquieren un significado concreto. Muchas veces, identificar qué es una colocación es difícil para figuras como las del traductor, especialmente cuando los traductores no son completamente especialistas en la materia sobre la que versa un texto de especialidad. Haber podido identificarlas a través de la herramienta que se propone en la segunda parte de esta tesis doctoral de manera automática supone un importante avance en la caracterización y estudio del lenguaje de especialidad ya que la metodología que planteamos es exportable a cualquier otro corpus, sea cual sea su especialidad. Así, los pares de palabras concretos o colocaciones dentro del lenguaje de especialidad harán referencia a realidades (o conceptos) concretos.

Dicho en otras palabras: una colocación es la unidad de conocimiento específico básica que se ha de conocer si se quiere entender el sentido de un texto especializado. Las colocaciones pueden considerarse, por tanto, como una sola unidad de léxico especializado compuesta de dos palabras. Un ejemplo de colocación es *red compleja* en el ámbito matemático. En este caso concreto, *red compleja* se refiere a un concepto

distinto al que se alude si se utilizan los términos por separado o fuera del lenguaje de especialidad matemático.

1.3 Las unidades de conocimiento especializado y la traducción.

Importancia de la semiótica

Como se resaltaba al inicio de este capítulo y en apartados anteriores, los términos específicos y, por consiguiente, el estudio del léxico especializado, es un área de investigación que, en castellano, está muy vinculado al aprendizaje de una lengua para fines profesionales ya que muchos científicos encuentran en el inglés el vehículo idóneo para la comunicación de sus resultados de investigación. Por esa razón es importante no perder de vista el vínculo entre la investigación en lenguas para fines específicos (LFE) y la enseñanza del inglés con propósitos específicos (ESP, English for Specific Purposes). Autores como Howatt & Widdowson (2004) reflejan en su obra que este aprendizaje del inglés como lengua extranjera con unos fines específicos (como es, en este caso particular, la divulgación científica) derivó en distintas corrientes metodológicas orientadas en su mayoría a la enseñanza y aprendizaje de lenguas para fines específicos (LFE) de cara al desarrollo de los individuos en el ámbito profesional.

El estudio de las LFE y la didáctica de las lenguas están íntimamente ligados, pero ha sido sobre todo a partir de las últimas décadas del s. XX cuando esta corriente empezó a tomar más relevancia. También existe la comunicación científica en español y, por tanto, produce textos llenos de estos términos llamados específicos que requieren estudio. Este hecho motiva la creación en 1997 de la Asociación Española de Terminología (<http://www.aeter.org/>). También surgió en la década de los 90 la AELFE, la Asociación Europea de Lenguas para Fines Específicos, en 1992 (<http://www.aelfe.org/>), y que acoge diversos trabajos sobre este tipo de lenguaje. Estas líneas pretenden ilustrar que el estudio del léxico especializado no se hace exclusivamente en inglés y que el hecho de poder comparar unas lenguas de especialidad con otras en distintos idiomas ofrece la oportunidad de estudiar el lenguaje con mayor profundidad.

De nuevo, hay que destacar que esta curiosidad por los elementos que se ven alterados en lo que se refiere a aquello que transmiten los términos cuando se utilizan

o provienen de un área del saber específica también en castellano es algo de lo que también se ha servido el mundo de la traducción, que pretende trasladar todos los elementos de una lengua a otra. Situando la perspectiva del lenguaje de especialidad en el ámbito de la enseñanza, Banks (2008, pág. 11) destaca que el conocimiento se transmite a través el lenguaje y que el conocimiento refleja tanto las experiencias vitales y los valores como las biografías y las comunidades culturales de los historiadores y científicos de la sociedad que lo crean; por eso el ideal de traducción es un texto meta capaz de evocar todo esto en la figura del lector. Es interesante ver cómo todos estos factores se reflejan en los distintos textos de especialidad (los llamamos textos de especialidad ya que su intencionalidad es la de transmitir un conocimiento relativo a un área muy determinada).

Estos factores caen en el ámbito de la lingüística, donde es relevante hablar de la semiótica. Según la semiótica, una persona puede hacer lecturas "opuestas" del mismo material en un contexto y lecturas "dominantes" en otros contextos (Morley, 1992, pág. 135).

Para un traductor, es fundamental tener cierta familiaridad con la semiótica si se quiere llegar a trasladar la totalidad de un mensaje de una lengua a otra. Esta familiaridad con la semiótica⁵ presentará dificultades cuando se trata de un texto redactado en lenguaje de especialidad. El estudio de la terminología (y, por consiguiente, de la semiótica) son de aplicabilidad debido a todos los rasgos comunes que presentan la terminología y la traducción que Cabré Castellví destaca en su obra de 2004: "a) su origen práctico; b) su reciente consideración disciplinar; c) su interdisciplinariedad constitutiva; d) su imbricación y justificación como materias relacionadas con la información y la comunicación; y e) su base esencialmente lingüística" (Cabré Castellví, 2004, pág. 91)

La coincidencia de estos tres campos (terminología, traducción, y lenguas para fines específicos) se basa en el común interés por, de un lado, las categorías de conocimiento; y, de otro, su interés por las unidades expresivas que permiten expresar

⁵ La semiótica es la rama de la lingüística que estudia los signos concebidos como un elemento dual compuesto de un significante y un significado.

estas categorías. El conocimiento que expresan y transmiten estas unidades es el punto básico más interesante que tienen en común.

Para un traductor es fundamental trasladar de una lengua a otra lo que llamamos *unidades de conocimiento específico* (que coinciden con lo que acabamos de nombrar unidades expresivas). Estas unidades parten de la terminología específica y mezclan otros elementos lingüísticos como la interactuación entre los distintos términos. Como señala Cabré Castellví (2004), la terminología no es el único factor –aunque es el factor privilegiado– de representación de conocimiento especializado. En este punto, su obra habla de que una de las características lingüísticas más destacables de los textos científico-técnicos es precisamente la presencia de unidades específicas que solo existen en ese ámbito. Estas unidades son de varios tipos (morfemas, nominales, preposicionales, correspondientes al lenguaje natural o correspondientes al lenguaje artificial, integradas por partes en las palabras, por palabras, por grupos de palabras o incluso por conjuntos más amplios) y cuanto mayor sea el grado de especialización de los textos, mayor será lo que la autora llama su “espesor terminológico”.

La función primigenia de estas unidades de conocimiento especializado es, precisamente, la de representar dichos conocimientos especializados (y todo lo que hay detrás de sus signos). La autora considera que cada unidad terminológica corresponde a un nudo cognitivo dentro de un campo de especialidad (que el traductor ha de conocer o al menos, identificar para transmitir todo el mensaje de manera adecuada) y, el conjunto de dichos nudos, conectados mediante relaciones específicas (causa-efecto, contigüidad, todo-parte, anterioridad-posterioridad, etc.) constituye la representación conceptual de dicha especialidad. Hay que tener en cuenta que cuando hablamos de un uso correcto del lenguaje de especialidad en textos de especialidad, no solo hablamos de terminología específica, sino que también hablamos de otros factores como los “movimientos lingüísticos” descritos en Peacock (2002, pág. 480), factores que han de ser tenidos en cuenta cuando se intenta trasladar el mensaje de una lengua a otra, así como cuando se redacta o revisa un texto científico. Un movimiento es *“a segment of text that is shaped and constrained by a specific communicative function”*.

A modo de conclusión de estos primeros apartados, hay que señalar que para realizar una tarea de traducción o de redacción especializada de manera correcta, el traductor tendrá que considerar los elementos terminológicos resaltados por

Santamaría Pérez, (2009), donde se menciona la existencia de unidades del lenguaje natural (compuestas por unidades morfológicas especializadas como prefijos y sufijos; unidades léxicas especializadas o sustantivos, verbos, adjetivos y adverbios; y unidades fraseológicas especializadas que divide en nominales, verbales, adjetivales y adverbiales) enfrentadas a las unidades del lenguaje artificial (entendidas como símbolos, nomenclaturas, etc.) dentro de las unidades de conocimiento especializado. Se trata de unidades de conocimiento relevantes ya que son precisamente las unidades de conocimiento especializado las que caracterizan principalmente el lenguaje de especialidad o el lenguaje para fines específicos tal y como se afirma en Estopà (2003).

1.4 El lenguaje científico como lenguaje de especialidad

Para una correcta traducción de textos con conceptos de especialidad, al igual que para poder expresar adecuadamente estos conceptos de especialidad, es necesario aprenderlos. Sin un dominio total en la materia, es necesaria algún tipo de asesoría sobre la especialidad como tal en la que se está escribiendo o hacia la que se está traduciendo de alguien que sea capaz de contemplar todos los elementos lingüísticos del mensaje.

Este dominio de la especialidad se puede alcanzar de muchas formas, entre las que destaca el aprendizaje de una lengua extranjera para fines profesionales, como resaltaba Congost-Maestre ya en 1994 en sus trabajos sobre la traducción técnica. De nuevo, el que el inglés sea el idioma en el que se redactan o hacia el que se traduce la gran mayoría de publicaciones científicas hace que el enfoque didáctico haya sido muy importante para el mundo de los lenguajes de especialidad y que tanto los profesionales como los académicos necesitarán una serie de competencias para expresar adecuadamente sus ideas dentro de sus ámbitos de actuación (Congost Maestre, 1994).

En cualquier caso, tal como señalamos, la mayoría de las publicaciones científicas están escritas en inglés y este inglés científico o inglés para fines específicos tiene una serie de características o variantes respecto al inglés “general” que no está orientado a la comunicación académica de una determinada área del saber. Autores como Reeves (2005), plantearon que una buena forma de entender el lenguaje científico (nuestro objeto de estudio debido a su terminología), es contrastarlo con el lenguaje poético:

ambos lenguajes parten de bases casi opuestas ya que el lenguaje científico debe captar lo general, lo habitual, más que lo individual y lo único, que es lo que trata de reflejar o captar el poético (Reeves, 2005, págs. 6-7).

El lenguaje científico y los términos utilizados en él obedecen directamente a la creación de nuevos campos de conocimiento. El lenguaje científico está siendo especialmente prolífico en las últimas décadas debido a la aparición de conceptos nuevos (y, a veces, disciplinas nuevas).

Ya se ha mencionado en el apartado anterior que cuanto más específica es la temática de un texto, en dicho texto habrá más unidades de conocimiento específico (UCE) y encontraremos una mayor presencia de unidades de lenguaje artificial. Definir las características del lenguaje de las ciencias, lenguaje en el que se enmarcan las matemáticas globalmente –y a todos los niveles, no sólo léxico, sino también gramatical, discursivo, textual y pragmático–, resulta muy interesante con la vista puesta en perfeccionar las competencias de traducción, redacción (escritura) o corrección. El lenguaje matemático está presente, sin lugar a duda, en una cantidad ingente de artículos de investigación publicados en revistas científicas de diferentes ámbitos de la ciencia.

Las publicaciones científicas están llenas de términos de especialidad ya que se centran en un área del saber (tal y como se ha resaltado en los apartados anteriores), por lo que podríamos decir que se trata de publicaciones hechas en lenguaje de especialidad y que existe un nexo entre el lenguaje científico y el lenguaje con fines específicos. Estas revistas científicas basan sus publicaciones en un campo del saber en concreto y merece la pena echar un vistazo a los rasgos del lenguaje que los autores emplean en sus artículos.

Son varios los estudios que nos hablan sobre las distintas variaciones que sufre el lenguaje científico. Por ejemplo, Burrough-Boenisch (2003, pág. 229) describe como *“go beyond correcting grammar and spelling to amend the author's discourse and rhetoric”* al proceso de revisión de un texto para su posterior publicación. El lenguaje científico (con fines específicos) empleado en estas publicaciones también cuenta con otra característica señalada asimismo por Burrough-Boenisch (2003) que es la de que estos textos casi siempre son fruto de una serie de revisiones, bien por parte de los coautores de los textos co-firmados, o bien por un colega del mismo campo. Sin

embargo, en muchos casos, las revisiones de otros investigadores no nativos refuerzan las anomalías del texto con respecto a las características que presentan los textos producidos por autores nativos de inglés (Burrough-Boenisch, 2003, pág. 227).

Este es un factor importante a la hora de considerar el resultado del análisis de cualquier conjunto o corpus de artículos de investigación. También se ve un nexo claro entre el lenguaje científico y la figura del asesor lingüístico o revisor familiarizado con la terminología y elementos lingüísticos característicos de un área del saber. Cuanto más específica sea la materia de un artículo, aparecen más elementos del lenguaje artificial asociados al lenguaje matemático, y más elementos verbales aparecen en torno al mismo. El estudio de estos elementos verbales o lenguaje matemático verbalizado es lo que ha motivado la redacción de esta tesis doctoral.

1.5 El lenguaje matemático de especialidad

Hay que señalar que, aunque el lenguaje matemático no haya sido catalogado como “de especialidad” por los lingüistas, sí que cumple con los criterios que mencionábamos en los apartados anteriores recogidos por Pérez-Hernández (2002) y Santamaría-Pérez (2009). El lenguaje matemático cumple con todas estas características y, además, añade una: los símbolos específicos utilizados en el lenguaje simbólico-numérico matemático son, a su vez, un lenguaje en sí mismo, que puede ser comprendido sin necesidad de utilizar el lenguaje común en diversas ocasiones (Σ , ∞ y la expresión $2+2=4$ son solo un ejemplo de expresiones matemáticas que no necesitan del lenguaje verbalizado para ser comprendidas).

Por otra parte, dado que las matemáticas son una disciplina transversal, la terminología matemática se aplica a muchas áreas y ramas del saber. El hecho de que exista más de un tipo de lenguaje matemático (o distintos sub-lenguajes matemáticos) dificulta aún más la tarea del traductor que desea especializarse en este tipo de lenguaje. Serrano Gómez (2005) ya sienta el precedente de que se necesita hablar de algún tipo de sub-categorización añadida para poder analizar y extraer conclusiones referentes a este lenguaje de especialidad.

Al reto de acotar las sub-áreas del lenguaje matemático, hay que añadir la dificultad ya mencionada referente a que a menudo el lenguaje de especialidad

matemático mezcla dos tipos de lenguaje; a saber, el lenguaje propio de los símbolos matemáticos y el lenguaje matemático expresado con palabras que, en adelante, llamaremos lenguaje de especialidad matemático natural puesto que es el que se utiliza para expresar las unidades de conocimiento específico de las que nos habla Santamaría Pérez (2009) y que se han descrito en el apartado anterior.

Contamos con documentos normativos sobre cómo verbalizar el lenguaje de especialidad matemático, pero están más cercanos a cómo verbalizar el lenguaje matemático que, según Santamaría Pérez (2009) se corresponde con lenguaje artificial. Además, hay que tener en cuenta lo siguiente:

Los símbolos matemáticos no son abreviaciones, sino entidades escritas con valor completo y autónomo. No quedan por tanto sujetos a normativas de carácter lingüístico o gramatical, sino que siguen la propia lógica del lenguaje formal matemático para combinarse en expresiones y fórmulas según ciertas reglas establecidas, ya sea por tradición, ya sea por convenios internacionales, nacionales, locales o personales. Bezos (2007, pág. 1)

Por consiguiente, estos símbolos constituyen unidades de lenguaje artificial y este lenguaje se tiene que considerar como un lenguaje combinado con el lenguaje de especialidad en el que se redactan los textos. El lenguaje artificial siempre requerirá un tratamiento especial como unidad de traducción particular y específica a la hora de plantear una propuesta de traducción. El lenguaje artificial se mantiene en los textos redactados en lenguaje de especialidad, independientemente de la lengua o idioma en que estos estén redactados (inglés, francés, castellano o cualquiera que sea la lengua desde y hacia la que se traduce el lenguaje artificial se mantendrá siempre salvo en contadas ocasiones). Varía el lenguaje verbal en torno a las distintas expresiones del lenguaje artificial o matemático, pero la expresión simbólico-numérica como tal se mantiene siempre de texto origen a texto meta.

Las expresiones verbales matemáticas que rodean a la misma expresión matemática no son siempre similares y existen tentativas para tender a su homogeneización; tal es el caso, por ejemplo, de proyectos como el BIPM [*Bureau International des Poids et Mesures*] (<https://www.bipm.org/en/about-us/>) que, entendemos, es un intento de cuidar la expresión del lenguaje y tender a la homogeneización en esta sub-área con intención normativa.

El de los pesos y las medidas no es el único ámbito básico de la matemática en el que hay disparidad a la hora de verbalizar expresiones. Podemos extraer un ejemplo de esta disparidad a la hora de expresar verbalmente el lenguaje matemático en niveles tan básicos de esta disciplina como una suma. La expresión $2+2=4$ se puede verbalizar de muchas y variadas maneras. Por ejemplo: dos y dos son cuatro; dos más dos es cuatro; dos y dos es igual a cuatro; dos más dos equivale a cuatro o, incluso, dos y dos equivale a cuatro. Aquí podemos ver cómo los elementos que se mantienen son las verbalizaciones numéricas (2: dos y 4: cuatro), y los elementos que varían son los relativos a la acción y el resultado que se quería reflejar en la expresión numérica matemática. Diversos autores han argumentado en sus estudios que existen diferencias demostrables entre artículos escritos por autores nativos del inglés y sus colegas no nativos, por ejemplo, Wood (2001), pero sin llegar a establecer una pauta sobre cómo redactar en lo que hemos denominado antes “lenguaje científico” (plagado de términos matemáticos). Del mismo modo, también contamos con los escritos de Bezos (2007), donde se ofrecen al lector descripciones útiles sobre la ortotipografía y notaciones matemáticas, pero no se trata de un análisis lingüístico en profundidad.

Así pues, es sabido que el lenguaje matemático (o lenguaje de especialidad matemático, ya que cumple con las características típicas del que ya se ha descrito como lenguaje de especialidad), está presente en prácticamente todas las áreas del saber. Para estudiar la mejor manera de verbalizar el lenguaje matemático resulta adecuado iniciar el análisis desde una rama concreta de las matemáticas (álgebra, cálculo, geometría, estadística o topología entre otras). La American Mathematical Society (AMS) tiene catalogados mediante la asignación de códigos concretos las distintas temáticas sobre las que se publican los artículos científicos. Cabe señalar que en la propia catalogación de la AMS (<https://www.ams.org/>) aparecen descriptores específicos comunes a diferentes áreas de especialización que pueden hacer que un mismo texto trate temas comunes a dos áreas se relacione con dos o más sub-apartados.

1.6 Lenguaje matemático de especialidad del área de redes complejas

Un área que resulta de gran interés por la enorme cantidad de artículos científicos e investigaciones que se desarrollan en la actualidad es el área que tiene que ver con la Teoría de las redes complejas. Esta teoría es aplicable a muchas disciplinas y se encuentra una gran variedad textual sobre ella, circunstancia que tiene como consecuencia que la asignación de códigos de especialidad de la AMS no permita adscribir la teoría de las redes complejas a una única área: así, esta teoría tiene asignados varios códigos de especialidad según la aplicación o el tipo de desarrollo realizado: 05C82, 05C90, 90B10, 90C35 o 91D30 (ver: <https://cran.r-project.org/web/classifications/MSC-2010.html>, donde se pueden comprobar todas las áreas matemáticas donde la Teoría de redes complejas tiene cabida).

Con el objetivo de describir en concreto esta área del lenguaje matemático, nos hemos centrado en un corpus compuesto en su gran mayoría por documentos que tratan sobre distintos aspectos de la Teoría de Grafos y de Redes Complejas debido, entre otras cosas, a que esta es un área de especialidad que tiene la peculiaridad de que, además de ser una de las más prolíficas en matemáticas a tenor del número de artículos publicados en las últimas décadas y de sus múltiples aplicaciones, existe una enorme variedad de sistemas reales de interés en ciencia y tecnología que pueden ser descritos y analizados en términos de esta teoría. Así, el estudio de redes complejas no es solo uno de los paradigmas de la ciencia de la complejidad sino también una rama de investigación fascinante en matemáticas, en ingeniería y en la ciencia en general.

Debemos señalar, por otra parte, que tanto la teoría de grafos como la teoría de redes complejas están construidas sobre el mismo tipo de elementos matemáticos, pero la diferencia más importante entre ambas teorías estriba en el tipo de herramientas matemáticas empleadas para resolver los problemas, ya que en el caso de las redes complejas, compuestas por un número mucho mayor de nodos y aristas que en el caso de los grafos, es necesario emplear herramientas de carácter estadístico –y no sólo combinatorio– como sucede en el caso de la teoría de grafos.

Aunque el concepto de red se remonta a Pitágoras (siglos V-VI a.C.) en su teoría del cosmos (κοσμος), el primer libro en el que aparece explícitamente el concepto de

red vio la luz en 1936 (D. König, *Theory of Finite and Infinite graphs* [véase König, 1990]). El análisis de redes tuvo uno de sus momentos más críticos y emocionantes en 1999, con el descubrimiento de nuevos tipos de grafos (redes de tipo “mundo pequeño” y redes de escala libre) denominados *redes complejas*. Desde ese momento, las redes complejas se han utilizado para modelar diferentes sistemas del mundo real y han despertado un gran interés debido, sobre todo, a que su estudio ha resultado ser muy productivo en la ciencia y la tecnología.

Muchos sistemas complejos del mundo real pueden modelarse mediante redes complejas en las que los nodos representan los distintos elementos o componentes del sistema y las aristas representan las interacciones entre ellos. A partir de diferentes estudios desarrollados desde entonces se ha podido constatar que diferentes sistemas, como las redes de transporte (metro, redes aéreas, redes de carreteras), las redes de comunicación (servidores informáticos, Internet), las redes bioquímicas (redes metabólicas, proteicas y genómicas), las redes sociales, las redes de infraestructuras (redes de energía eléctrica, redes de suministro de agua) y algunas otras (incluida la World Wide Web) tienen características comunes en su comportamiento y estructura. Estudios como Bocaletti et.al (2006); Cohen et al (2001) o Newman (2010), hacen énfasis en esta idea.

Es sabido que Paul Erdős (1913-1996) es el matemático más prolífico de la historia, después de Euler, a tenor del número de páginas de matemáticas originales escritas por cada uno (Erdős publicó alrededor de 1.500 artículos a lo largo de su vida sobre muchas ramas de la ciencia, entre otras, teoría de grafos, teoría de la probabilidad, teoría de conjuntos, análisis clásico, teoría de la aproximación y teoría de números). Destaca, sobre todo, el desarrollo de la teoría de Ramsey y la aplicación del método probabilístico a distintos problemas. Sin embargo, Erdős es famoso también por sus excentricidades: toda su vida estuvo viajando entre conferencias científicas y visitando las casas de los colegas de todo el mundo a las que acudía con una maleta, permaneciendo un tiempo limitado que normalmente desembocaba en nuevos trabajos científicos escritos en colaboración. Así, Erdős y su colega Alfred Rényi empezaron a trabajar en grafos para entender la estructura de las redes sociales. En otras palabras, Erdős y Rényi, teniendo en mente la motivación original de analizar, mediante métodos probabilísticos, las propiedades de las redes en función del número creciente de

conexiones aleatorias, fueron los primeros en estudiar este tipo de objetos matemáticos, los grafos aleatorios (Rényi, 1959), en los que la aparición y existencia de una arista entre un par de nodos depende de una cierta probabilidad.

Más allá de la teoría de grafos clásica, que ha desembocado en el área científica conocida como análisis de redes complejas y que involucra no sólo herramientas matemáticas (incluyendo la probabilidad, el análisis de sistemas dinámicos, la teoría de grafos, el análisis matricial y otras), sino también técnicas procedentes de otros campos (como la mecánica estadística o las ciencias de la computación, por nombrar un par de ellas), el éxito de la teoría de redes complejas se debe fundamentalmente a sus numerosas aplicaciones prácticas. Así, la investigación sobre redes complejas ha atraído cada vez más la atención de numerosos campos de investigación debido a su más que probada utilidad para modelar sistemas complejos del mundo real. En cualquier caso, hay que reconocer que en el trasfondo de las descripciones elegantes y eficientes de sistemas complejos muy diferentes se encuentra el marco general de la teoría de redes complejas, cuyo origen está en la teoría de grafos iniciada por Erdős en 1959, aunque muy recientemente, en Benson et al. (2021) se dice, de forma explícita, que “una red es un grafo que representa un sistema complejo”.

Los sistemas complejos pueden estudiarse utilizando modelos matemáticos no lineales y enfoques de modelado informático. De hecho, las redes complejas ofrecen una alternativa natural para representar, caracterizar y modelar la estructura y la dinámica no lineal de todos los sistemas complejos discretos.

Para comprender las funciones de un sistema complejo es esencial estudiar las propiedades estructurales de su red subyacente, así como también, por ejemplo, para cuantificar la importancia estratégica de un nodo (o conjunto de nodos) con el fin de preservar el mejor funcionamiento de la red y consecuentemente del sistema complejo que representa en su conjunto. Hay que resaltar, en cualquier caso, que han sido las sucesivas mejoras en el rendimiento y capacidad de cómputo de los ordenadores en las últimas décadas las que nos han permitido realizar el análisis de redes complejas de enorme tamaño.

En el siguiente capítulo explicaré de manera algo más pormenorizada la teoría de redes complejas con objeto de poder avanzar en cuestiones de índole teórica y matemática que serán necesarias para el desarrollo de la segunda parte práctica de esta

tesis doctoral. A lo largo de dicho capítulo se señala asimismo bibliografía complementaria para profundizar en diferentes aspectos que, por su complejidad, no serán traídos aquí.

1.7 La teoría de redes complejas

En este capítulo se van a esbozar algunos de los conceptos básicos de la teoría de redes complejas puesto que el objetivo de esta tesis doctoral es analizar el lenguaje de especialidad matemático empleado por la comunidad científica que publica sobre redes complejas a partir de un modelo matemático y del procesamiento computacional de un corpus lingüístico concreto. Estos conceptos básicos se presentan con la intención de que ilustren la segunda parte de esta tesis doctoral, donde se presenta el nuevo modelo matemático desarrollado a partir de la teoría de redes complejas, que ha permitido identificar particularidades lingüísticas propias de este lenguaje de especialidad como las colocaciones y otros elementos y giros lingüísticos propios de la misma, todo ello con la finalidad de facilitar la construcción de una herramienta que permita ofrecer pautas para una correcta traducción y/o redacción de textos propios de este área matemática. Dicho con otras palabras, se ha utilizado la teoría de redes complejas para analizar el lenguaje de especialidad de los artículos científicos que versan sobre la misma y, por eso, conviene hacer un repaso a los conceptos básicos asociados a esta teoría.

1.7.1 Introducción

La Teoría de redes complejas surge desde la Teoría de grafos, cuya base radica en observar las relaciones que existen entre los distintos elementos de un sistema. Como se resaltaba en apartados anteriores, para estudiar el lenguaje de especialidad matemático es necesario entender y observar las relaciones existentes entre los distintos términos de especialidad o palabras que lo componen. Para ello, en la parte práctica de esta tesis doctoral se concebirá el lenguaje como una red compleja donde cada palabra es un nodo de tal forma que se podrá hacer un estudio del léxico haciendo énfasis en las relaciones existentes entre las palabras. Con vistas a alcanzar una mayor

claridad sobre la parte práctica y el método que se ha empleado para estudiar el lenguaje de especialidad, es necesario exponer algunos conceptos básicos sobre redes complejas.

Desde un punto de vista esquemático, una red compleja (o un grafo) es un objeto matemático $G = (X, E)$ compuesto por un conjunto de nodos o vértices $X = \{1, \dots, n\}$, y un conjunto de enlaces o aristas (orientadas o no) establecidas entre dichos vértices. El conjunto de aristas puede entenderse como un conjunto de parejas de nodos, ordenadas o no.

Este tipo de representación puede parecer simple, pero, como veremos, tiene un enorme potencial, ya que algunos problemas se vuelven más sencillos y tratables si se representan como un grafo. La teoría de grafos, el andamiaje matemático que sustenta la ciencia de los grafos, tiene su origen en un problema, resuelto por Euler en 1736, conocido como el "Problema de los puentes de Königsberg". Este problema se planteó al tratar de averiguar si los siete puentes de la ciudad de Königsberg sobre el río Pregel pueden recorrerse en un solo viaje sin tener que volver atrás, con el requisito adicional de que el viaje pase por cada uno de los puentes una única vez.

Este problema fue respondido negativamente por Euler, y representó el comienzo de la teoría de los grafos. Euler tuvo la idea de utilizar cuatro nodos para representar cada una de las cuatro áreas de tierra separadas por el río como nodos, distinguiéndolos con las letras A, B, C y D. A continuación, conectó con líneas cada pedazo de tierra que tenía un puente entre ellos. De este modo, construyó un gráfico cuyos nodos eran los terrenos y, los enlaces, los puentes.

Una vez representada la ciudad como un grafo, el problema de los puentes de Königsberg puede reformularse como sigue: ¿es posible encontrar un camino entre dos nodos en el que cada enlace aparezca exactamente una vez? En la teoría de grafos, este camino se denomina paseo de Euler. En un grafo, dos vértices se llaman adyacentes (o vecinos) si están conectados por una arista. El número de vecinos de un nodo es el grado de dicho nodo. Así, podemos decir también que el grado de un nodo es el número de enlaces que el nodo comparte con sus vecinos. Estudiando el grado de las distintas palabras vecinas se puede ver qué términos son más proclives a formar una colocación lingüística como las que se describen en los apartados anteriores.

1.7.2. Conceptos básicos sobre redes y grafos

No hay que olvidar que la Teoría de redes complejas trata de ofrecer modelos para el mundo real. Los conceptos básicos sobre redes complejas que se consideran para elaborar el modelo matemático presentado en la parte práctica (que sirve para estudiar el lenguaje matemático e identificar las colocaciones presentes en el mismo, entre otros usos potenciales que se describirán en la parte dedicada a este fin) son los siguientes:

- Redes y grafos
- Redes “mundo pequeño” (*small world*)
- Redes de escala libre (sin escala)
- Redes asortativas
- *Line Graph*
- Red multicapa
- Centralidad del *PageRank*

a) Redes y grafos

El término "red compleja", o simplemente "red", suele referirse a sistemas reales, mientras que el término "grafo" suele considerarse la representación matemática de una red. Algunos autores han puesto de manifiesto algunas de las diferencias que existen entre la teoría de grafos y la de redes complejas. Estas diferencias radican principalmente en el tamaño (menor para el grafo y mayor para la red) y en los parámetros y las herramientas empleadas para analizar ambas estructuras. Siguiendo esta idea, hablaremos, por ejemplo, de un grafo de doce nodos y de una red de quinientos nodos. En concreto, el hecho de que el tamaño sea menor para los grafos y mayor para las redes obliga a tener un cuidado extra con la complejidad de los algoritmos que se emplearán.

En esta tesis se concibe el lenguaje como un sistema, por lo que hablamos en todo momento de redes complejas (y no de grafos).

b) Redes “mundo pequeño” (*small world*)

En los últimos años, se han propuesto varios modelos de redes complejas tras la idea pionera del modelo de grafos aleatorios de Erdős-Rényi (1994), como el modelo de mundo pequeño de Watts D., y Strogatz (1998). La razón principal fue el descubrimiento de que las redes reales, como por ejemplo el grafo de Internet, tienen características que no se explican por una conectividad aleatoria uniforme entre sus nodos. En concreto, la propiedad de mundo pequeño indica que el diámetro de la red es mucho menor que el número de nodos o, en otras palabras, que se puede llegar a la mayoría de los vértices desde los demás a través de un pequeño número de aristas, como en las redes sociales.

El término mundo pequeño fue introducido por Watts C. y Strogatz S., (1998) en su estudio de varias redes del mundo real, como la red de actores de cine de Hollywood. En Watts y Strogatz (1998) se introducen la distancia media (o longitud del camino característico), que mide las distancias globales en una red, y el coeficiente de agrupamiento (o coeficiente de “clustering”), que viene a ser la proporción media de pares de vecinos de un nodo que también son vecinos entre sí. Watts y Strogatz (1998) propusieron un modelo de mundo pequeño basado en un procedimiento de recableado de las aristas de una red regular con una probabilidad una cierta probabilidad. El algoritmo de Watts-Strogatz comienza con una red regular con n nodos dispuestos en un anillo. Cada arista se recablea aleatoriamente con una probabilidad p . Variando esta probabilidad entre los valores de $p = 0$ a $p = 1$ se puede controlar la transición entre la red regular y la aleatoria y, por ejemplo, para una probabilidad de recableado pequeña el coeficiente de agrupamiento de la red apenas difiere de su valor inicial, pero la longitud de la trayectoria característica desciende rápidamente y es del mismo orden que la de las redes aleatorias. Así, Watts y Strogatz crearon el primer modelo que conciliaba la existencia de un gran coeficiente de agrupamiento con un pequeño diámetro o longitud de camino característica.

Si consideramos un lenguaje como un sistema representable a través de una red, este concepto es importante a la hora de establecer la distancia entre palabras desde una nueva perspectiva.

c) Redes de escala libre (sin escala)

En muchas redes del mundo real, la distribución de grados no sigue una distribución similar a la de Poisson (por ejemplo, la web mundial, la red eléctrica o la red de aeropuertos mundiales), sino que sigue una ley de potencia, en la que aparecen unos pocos nodos (“hubs”) muy conectados y muchos nodos “poco conectados”.

De hecho, se conoce una ley empírica que afirma que la distribución de grados de la mayoría de las redes del mundo real sigue una “ley de potencia” cuyo exponente positivo está comprendido entre dos y tres. Así, si la distribución de grados tiene una cola decreciente en la ley de potencia significa que la gran mayoría de los nodos tienen un grado bajo y que existen pocos nodos, los que se denominan *hubs*, que tienen una conectividad extremadamente alta. Una red con una distribución de grados de este tipo de ley de potencia de grado se denomina red “de escala libre”. Este modelo fue introducido en 1999 por Barabasi y algunos de sus colaboradores en 1999, tras comprobar que la distribución de grados de algunos sistemas complejos sigue este tipo de “leyes de potencia” en lugar de seguir una distribución de tipo Poisson y que, además, muchos de estos sistemas están fuertemente agrupados con un número muy grande de caminos cortos entre los nodos, es decir, obedecen a la propiedad del mundo pequeño.

Las redes de escala libre suelen surgir en el contexto de una red en crecimiento en la que los nuevos nodos prefieren conectarse a los nodos altamente conectados de la red.

d) Redes asortativas

El modelo propuesto por Barabasi y Albert (1999) se basa en dos hechos observados en las redes reales: las redes se expanden continuamente mediante la adición de nuevos vértices, y los nuevos vértices se unen preferentemente a sitios que ya están bien conectados. Señalaremos, en relación con esto, que es importante introducir el concepto de asortatividad de la red como otra propiedad fundamental de las redes. Esta propiedad mide la correlación de grado entre los nodos. En las redes asortativas, la mayoría de las aristas conectan nodos que presentan grados similares (aristocracia de nodos). Por el contrario, en las redes disasortivas los nodos de alto grado están normalmente conectados a nodos de bajo grado.

e) *Line Graph*

La elección de la representación de redes adecuada para un problema determinado y sus herramientas asociadas puede determinar nuestra capacidad para utilizar la teoría de redes con éxito. En algunos casos existe una representación natural del problema mediante redes. En otros casos, la representación elegida puede ayudarnos a resolver el problema planteado. Por ejemplo, a veces parece adecuado dar más importancia a las aristas de un grafo que a sus nodos en el contexto de ciertas redes y grafos. Un ejemplo de ello procede del urbanismo, donde a veces las calles de una ciudad se representan como los nodos del grafo, y las intersecciones entre ellas como los enlaces (Crucitti et al 2006a 2006b y 2006c, y Porta et al, 2006).

Las redes de distribución constituyen otro ejemplo de esta situación. El objeto apropiado para apoyar este punto de vista y estudiar este tipo de problemas es el grafo lineal $L(G)$ o "*Line Graph*" (también llamado grafo dual), como se ha demostrado en el contexto ya mencionado del urbanismo, de las redes de transporte (Anez et al, 1996, Volchenkov-Blanchard, 2018), e incluso del tráfico urbano (Hua et al, 2009).

El grafo lineal asociado a $G = (V,E)$ es el grafo $L(G) = (E,L)$ cuyo conjunto de nodos es el conjunto inicial de aristas del grafo G , y el conjunto L de las aristas de $L(G)$ está formado por pares de aristas de G que comParten un nodo.

Es destacable que los grafos lineales se han estudiado durante más de ochenta años, aunque este concepto ha sido redescubierto varias veces a lo largo de este periodo de tiempo, con diferentes nombres como grafo adjunto, grafo derivado o dual de arista a vértice (Hemminger and Beinek, 1979). Señalaremos que la primera vez que este concepto aparece en la literatura fue en 1932 (Whitney, S. 1932).

Obviamente, muchas propiedades de un grafo G que dependen sólo de la adyacencia entre aristas pueden traducirse en propiedades equivalentes en $L(G)$ que dependen de la adyacencia entre vértices. Por ejemplo, el grafo lineal de un grafo conexo es conexo, ya que si G es conexo, contiene un camino que conecta dos aristas cualesquiera, lo que se traduce en un camino en $L(G)$ que contiene dos vértices cualesquiera de $L(G)$. Más relevante es la siguiente propiedad: Si un grafo G tiene un ciclo de Euler, es decir, si G está conectado y tiene un número par de aristas en cada vértice, entonces el grafo lineal de G es hamiltoniano, es decir, un grafo que posee un ciclo hamiltoniano (un ciclo hamiltoniano en el grafo G es un ciclo que pasa por todos

sus nodos exactamente una vez). En este punto es destacable que no se conocen condiciones (no triviales) que sean necesarias y suficientes para la existencia de un ciclo hamiltoniano en un grafo.

Por otra parte, el grafo lineal ha sido considerado para redes (grafos con muchos nodos y aristas) sólo en un número reducido de estudios y aplicaciones. En particular, los grafos se utilizan habitualmente en el análisis de casos urbanos y territoriales, por ejemplo, si se buscan los lugares o calles dentro de una ciudad que son más importantes -o más frecuentados- que otros debido a su situación geográfica. Como se demostró en Porta et al, 2006, el estudio de la centralidad en redes complejas dentro del contexto del diseño urbano basado en la representación gráfica primal nos da resultados similares que el correspondiente análisis realizado utilizando su grafo lineal asociado.

f) Red Multicapa

En los últimos años, los científicos de redes han dirigido su atención al diferente tipo de conexiones entre nodos que aparecen en los sistemas del mundo real, y han considerado explícitamente la naturaleza multicapa de las redes complejas (véase, por ejemplo, Boccaletti et al 2014, Criado et al 2012 o Kivela et al 2014). En Boccaletti et al 2014 se puede encontrar una revisión exhaustiva de la organización estructural y dinámica de una red hecha de diversas relaciones (capas) entre sus componentes y en Chapela et al 2015 una síntesis de todos estos conceptos. La revisión que podemos encontrar en Boccaletti et al 2014 cubre varios temas relevantes, desde una redefinición completa de las medidas estructurales básicas, hasta la comprensión de cómo la naturaleza multicapa de la red afecta a los procesos y a la dinámica. Los sistemas complejos incorporan múltiples canales de conectividad y comunicación. Las redes multicapa incorporan explícitamente múltiples canales de conectividad y constituyen el entorno natural para describir sistemas interconectados a través de diferentes categorías de conexiones, de forma que cada canal (relación, actividad, categoría) estará representado por una capa y un mismo nodo o entidad puede tener diferentes tipos de conexiones (diferentes conjuntos de vecinos en cada capa).

Por ejemplo, en las redes sociales, se pueden considerar varios tipos de relaciones de los distintos actores: amistad, vecindad, pertenencia a la misma sociedad cultural, etc. En este caso, la representación adecuada es una red multicapa, en la que

los nodos interactúan a través de múltiples capas de enlaces para abarcar adecuadamente aquellas propiedades topológicas de los sistemas de tipo heterogéneo que no pueden ser capturadas por la representación clásica de red de una sola capa. La red compleja a la que nos referimos en la parte práctica es una red multicapa (de tal manera que encontramos términos clasificados en distintas capas de la red en función de una serie de decisiones lingüísticas que se expondrán en el apartado correspondiente).

g) Centralidad del *PageRank*

El algoritmo *PageRank* (Brin y Page, 1998, Page et al 1999), que es el “motor” que está detrás de Google para ordenar las páginas web en cada una de las búsquedas que se realizan, nos proporciona una forma de medir con precisión y cuantitativamente la relevancia de no sólo de una página web en la *World Wide Web*, sino que este método se puede utilizar también en otro tipo de redes complejas.

Así, este algoritmo se utiliza para señalar los nodos o elementos más relevantes de un sistema complejo en una amplia gama de aplicaciones de todo tipo. En este punto, cabe señalar que *PageRank* y sus posteriores extensiones han demostrado su eficacia y versatilidad en una amplia gama de aplicaciones de todo tipo en las que es necesario señalar los nodos o elementos más relevantes de un sistema complejo utilizando la estructura de la red y el concepto de “paseante aleatorio” sobre la misma (Albert et al 2004, Chung 1997, Criado et al 2012, Chapela et al 2015). Este concepto nos va a permitir en la parte III señalar los elementos o palabras más relevantes de un corpus desde una perspectiva diferente pudiendo hacer hincapié en las relaciones sintagmáticas entre las palabras (habiéndose concebido previamente como nodos).

En este punto es importante destacar que, junto con los avances y las nuevas herramientas relacionadas con este algoritmo y sus múltiples aplicaciones, han ido apareciendo progresivamente nuevos desarrollos y conceptos en la teoría de redes complejas.

Según hemos afirmado, la forma en que funciona el algoritmo *PageRank* puede asociarse a la idea de considerar un paseante aleatorio en la red, empezando en un nodo seleccionado al azar y siguiendo por las diferentes aristas de la red, según el grado de salida de cada nodo, teniendo en cuenta que, además de poder navegar por las aristas

de una red, el paseante aleatorio a veces puede teletransportarse a otros nodos de la red que no están conectados directamente con el nodo actual por una arista.

Concretamente, cuando el paseante aleatorio se encuentra en el nodo i , con una probabilidad q puede desplazarse a uno de los nodos adyacentes a i , y con una probabilidad $(1-q)$ puede teletransportarse a cualquier otro nodo de la red al que no esté directamente conectado por una arista. Teniendo en cuenta que cuando el paseante aleatorio se teletransporta aleatoriamente a cualquier nodo de la red, esta teletransportación debe realizarse a un nodo u otro de acuerdo con una cierta probabilidad que se establece previamente mediante el llamado vector de personalización. Esto último tiene que ver, llevado a la práctica, con dónde vamos a considerar el inicio de una frase.

1.8. Recapitulación

Antes de pasar a la parte que tiene que ver con la metodología de análisis novedosa a través de las redes complejas, conviene hacer una serie de recapitulaciones de las conclusiones a las que se han llegado hasta ahora. En este primer capítulo se ha hablado sobre el papel clave que juega la terminología específica a la hora de enfrentarse a la traducción de un texto englobado dentro de una especialidad científica enfatizando en las unidades de conocimiento específico o unidades de traducción y el papel que el traductor desempeña si lo consideramos como un terminólogo. Se muestra la dificultad que plantean los textos redactados en lenguajes de especialidad en el mundo de la traducción ya que, muchas veces, el traductor tiene que ejercer el papel de terminólogo de especialidad (por ejemplo, dependiendo del área de especialidad, la palabra *Kernel* a veces se mantendrá como tal y, en otras ocasiones, habrá de traducirse como “Núcleo”).

Una vez determinado qué es un lenguaje de especialidad y en qué consiste la terminología específica, se demuestra que el lenguaje matemático de especialidad cumple con los requisitos necesarios para entrar en el grupo de lenguajes de especialidad del ámbito científico. Por último, en este primer capítulo, se hace una presentación del lenguaje de especialidad matemático que tiene que ver con la teoría de redes complejas y se hace una introducción a la propia teoría de redes complejas,

que, en posteriores apartados de la tesis, se utilizará como herramienta de análisis de un corpus lingüístico. También se sientan las bases de los motivos que llevaron a utilizar un corpus como objeto de análisis de este tipo de lenguaje de especialidad, aunque este hecho es algo sobre lo que se profundiza en el siguiente capítulo de esta tesis doctoral.

Capítulo 2. Lingüística de corpus y lingüística computacional.

Descripción del corpus. Análisis con *Wconcord*, *Flair* y *Voyant*

Una vez introducido el lenguaje de especialidad matemático y la teoría de redes complejas, antes de plantear el nuevo enfoque de estudio a través de la propia teoría de redes complejas, a fin de ilustrar mejor el avance que el uso de este nuevo enfoque supone, pretendemos definir el lenguaje matemático de especialidad utilizado por la comunidad científica que publica sobre la teoría de redes complejas mediante las herramientas de las que ya se disponían previamente a la publicación de esta tesis doctoral. Asimismo, la descripción de este lenguaje de especialidad en concreto arrojará cierta luz sobre los rasgos que el traductor tendrá que imitar si se enfrenta a un texto de estas características (es decir, si tiene que traducir un texto redactado en lenguaje de especialidad matemático de la teoría de redes complejas) ya que estos rasgos serán los mismos que encontrarán en textos similares o paralelos (en otros textos que tratan sobre la teoría de redes complejas), y las conclusiones sobre este lenguaje de especialidad matemático se extraen a partir de un corpus o conjunto de textos utilizado como herramienta de análisis. Se parte de la base de que para formar parte de un corpus textual, un texto tiene que cumplir con una serie de características que lo vinculan al resto de textos de los que se compone un corpus; en este caso concreto, el elemento o característica común de los textos que integran el corpus utilizado como objeto de análisis es que todos los textos tratan sobre la teoría de redes complejas (aplicando esta teoría distintos fenómenos del mundo real). En cualquier caso, se describe el corpus objeto de análisis en profundidad en los siguientes apartados de esta tesis doctoral.

En este segundo capítulo, se describe en profundidad el lenguaje de especialidad matemático que tiene que ver con la teoría de redes complejas utilizando únicamente las que se han llamado herramientas de análisis textual “tradicionales” (ya que estaban disponibles previamente al desarrollo de la herramienta de traducción que se presenta en el capítulo 3 de esta tesis doctoral, de elaboración propia) y que son las herramientas de análisis textual *Wconcord*, *Flair* y *Voyant*. Para llegar a estas primeras conclusiones sobre el lenguaje matemático de redes, en este segundo capítulo también se presentará y describirá el corpus que ha servido como objeto de análisis. Asimismo, nos

acercaremos a los enfoques característicos de análisis procedentes de la lingüística de corpus y justificaremos el hecho de utilizar esta disciplina como forma de acercarnos a las características lingüísticas de este lenguaje de especialidad matemático.

Además, en este segundo capítulo presentaremos unas primeras conclusiones sobre el lenguaje matemático de especialidad relativo a la teoría de redes complejas a las que se llegan tras el análisis realizado con las herramientas tradicionales *Wconcord*, *Flair* y *Voyant*, a fin de establecer diferencias entre los diferentes métodos y subrayar la mejora que supone la metodología que presentamos en el capítulo 3 de esta tesis.

2.1 Introducción a la lingüística de Corpus. Herramienta para analizar los lenguajes de especialidad

Uno de los objetivos de este segundo capítulo es presentar y caracterizar el lenguaje matemático de especialidad que trata sobre la teoría de redes complejas. Para ello vamos a utilizar un conjunto significativo de textos de esta área de especialidad que se va a ser procesado con determinadas herramientas de análisis de corpus lingüísticos ya existentes y que, en adelante, llamaremos “herramientas tradicionales”; estas son *Flair*, *WConcord* y *Voyant*. Para esta investigación vamos a considerar la lingüística de corpus (concebida como se describe en los siguientes apartados) como la metodología de análisis apropiada para este tipo de análisis, ya que un corpus lingüístico, por definición, es un conjunto de documentos que proporciona una muestra amplia y estructurada de ejemplos reales del uso de la lengua.

Con vistas a obtener una primera caracterización del lenguaje de especialidad matemático y, más específicamente, del lenguaje empleado en el ámbito de la teoría de redes complejas, hemos creado y refinado un corpus de textos de esta especialidad para, posteriormente, analizarlo con diferentes herramientas. Aunque en esta parte nos centraremos en los programas de análisis más habituales y tradicionales ya mencionados, introduciremos brevemente la nueva herramienta que se propone en la parte práctica de esta tesis doctoral con el objetivo de establecer los criterios sobre los que se ha compuesto el corpus de trabajo (y que también vamos a tratar como una red compleja matemática). Es necesario destacar el hecho de que la herramienta que se propone en la segunda parte práctica de la presente tesis doctoral tendrá asociado un

nuevo modelo matemático, por lo que los resultados obtenidos con vistas a la caracterización del lenguaje de especialidad son complementarios y constituyen, de hecho, una mejora de los presentados en esta segunda parte, en la que se ofrecen las conclusiones sobre este lenguaje de especialidad a las que se llegaría de no haber desarrollado una nueva herramienta de análisis.

En este punto conviene recordar que numerosos estudios léxicos de lingüística de corpus, por lo general, han estudiado las relaciones paradigmáticas entre los distintos términos o palabras de un corpus determinado: es decir, las relaciones que se establecen entre las palabras en aquellas ocasiones en las que se pueden sustituir unas palabras por otras; por ejemplo, en el sintagma “la casa”, se puede sustituir el determinante artículo “la” por el determinante demostrativo “esta” sin que se altere la percepción del sintagma como un conjunto de palabras correctamente enunciado. Una característica relevante de esta tesis doctoral es que, además, con el desarrollo de la herramienta de análisis que propondremos en la segunda parte práctica, podremos estudiar las relaciones sintagmáticas entre las palabras de cualquier corpus, entendiendo por relaciones sintagmáticas aquellas que permiten observar palabras y expresiones que deben aparecer juntas porque forman una unidad de conocimiento – por ejemplo, la concordancia es otra relación sintagmática, aunque en nivel diferente del léxico– que en la primera parte de esta tesis hemos llamado *unidad de traducción* o *unidad de conocimiento específico*.

Por lo tanto, en el contexto de la presente tesis doctoral, esta segunda parte nos servirá para, en primer lugar, determinar el motivo por el que establecemos un corpus de especialidad matemática, así como también el motivo por el que se utiliza la lingüística de corpus para estudiar el lenguaje de especialidad presente en el mismo; en segundo lugar, servirá asimismo para realizar un repaso a las aportaciones que tradicionalmente han hecho este tipo de estudios lingüísticos; y servirá, en tercer y último lugar, para establecer una primera caracterización o definición del lenguaje matemático que tiene que ver con la teoría de redes complejas.

Para poder definir el lenguaje matemático de especialidad utilizado en el ámbito de la teoría de redes complejas se ha recopilado un número significativo de textos de esta área hasta crear un corpus lingüístico.

Son diversas las definiciones que se han dado de los corpus como tales, pero nos interesa concebir un corpus lingüístico en la línea de autores como Llamazares (2008, págs. 329-349), que definen un corpus lingüístico como un conjunto de textos, relativamente grande, creado independientemente de sus posibles formas o usos. En este sentido, siempre que se tenga una muestra textual compuesta por diferentes documentos, de manera que sea lo suficientemente grande como para poder ser procesada computacionalmente, ésta se puede considerar como un corpus.

Lo fundamental es que la representación de la lengua usada en los textos de los que está compuesto el corpus sea significativa y real (ya que la manera en la que se procese la información del corpus depende, en cualquier caso, del usuario). Los corpus guardan ciertas similitudes con los textos de los que están compuestos, pero los textos no tienen que tener necesariamente coherencia ni cohesión entre sí –así, analizar la totalidad del corpus como una unidad de significado no tiene ningún sentido–, aunque sí proporcionan información sobre el uso del lenguaje en su conjunto. De un mismo corpus se pueden obtener diferentes resultados, todos ellos complementarios, dependiendo de qué se esté buscando y la herramienta que se utilice.

La necesidad de trabajar las muestras recogidas de forma eficaz y económica, teniendo en cuenta su enorme extensión, ha alentado el desarrollo de una de las ramas con más futuro de la lingüística contemporánea: la lingüística computacional. Actualmente los corpus se recogen y almacenan de manera electrónica (Llamazares, 2008, págs. 329-349). A nadie se le escapa que, debido entre otras particularidades, a su tamaño, accesibilidad, muy alta fiabilidad e información lingüística y enciclopédica, la compilación de los corpus ha llegado a ser uno de los principales –si no el principal– método e instrumento de la investigación de la lengua en la lingüística general y en muchas de sus disciplinas derivadas.

Partiendo de la base de que un corpus lingüístico es un conjunto amplio y estructurado de ejemplos reales de uso de una lengua, a través de su estudio, y entre otras muchas áreas que tienen que ver con la lingüística, se pueden obtener conclusiones muy relevantes sobre los lenguajes de especialidad. Además, el uso de herramientas informáticas nos permite procesar cantidades de información con mucha rapidez para así ver las relaciones que efectivamente existen entre las diferentes palabras o términos que forman un corpus.

En este punto es importante señalar que, por lo general, el estudio de corpus suele desarrollarse sobre textos escritos, aunque también se dispone –pero en menor medida– de corpus de muestras orales (pues también existen corpus textuales compuestos de transcripciones de hablantes). De nuevo, autores como Llamazares (2008, pág. 349), afirman que el uso de corpus para el estudio de la lengua se considera una metodología empírica de trabajo, basada en el empleo de datos reales y de muestras de uso de la lengua. Esta es la base sobre la que se sustenta la decisión de utilizar la lingüística de corpus para extraer conclusiones sobre el lenguaje de especialidad matemático tras haber elaborado un corpus textual con documentos pertenecientes a esta área científica.

Como veremos en los siguientes apartados, la lingüística de corpus no es el único método que se utiliza para estudiar en profundidad, y de una forma cuantitativa, el lenguaje, aunque, tal y como se justifica en esta tesis doctoral, es el enfoque que nos interesa contemplar. En este sentido, Pérez Hernández (2002, pág. *online*), cita a Fillmore (1992, pág. 35) y su planteamiento de entender el estudio de la lengua de dos formas diferentes, que bien puede asociarse al debate que ha habido en torno a la lingüística de corpus, y que, asimismo, puede ayudar a entender la perspectiva y la motivación por la que, para este caso concreto, hemos optado por el uso de la lingüística de corpus para el estudio del lenguaje matemático de especialidad:

Armchair linguistics does not have a good name in some linguistic circles. A caricature of the armchair linguist is something like this. He sits in a deep soft comfortable armchair, with his eyes closed and his hands clasped behind his head. Once in a while he opens his eyes, sits up abruptly shouting, "Wow, what a neat fact!", grabs his pencil, and writes something down. Then he paces around for a few hours in the excitement of having come still closer to knowing what language is really like. (There isn't anybody exactly like this, but there are some approximations.)

Corpus linguistics doesn't have a good name in some linguistic circles. A caricature of the corpus linguist is something like this. He has all the primary facts that he needs, in the form of approximately one zillion running words, and he sees his job as that of deriving secondary facts from his primary facts. At the moment he is busy determining the relative frequencies of the eleven parts of speech as the first word of a sentence. (There isn't anybody exactly like this, but there are some approximations.) These two don't speak to each other very often, but when they do the corpus linguist says to the armchair linguist, "Why

should I think that what you tell me is true?", and the armchair linguist says to the corpus linguist, "Why should I think that what you tell me is interesting?"⁶ (Fillmore, 1992, pág. 35)

En este caso entendemos por "lingüística de sillón" la lingüística que no está basada en corpus o en el uso real de la lengua. En cualquier caso, hay que decir que lingüistas teóricos como Chomsky se han valido de la lingüística aplicada (entendiendo como lingüística aplicada la rama de la lingüística encargada de la resolución de incógnitas relativas al lenguaje en un determinado ámbito), y han realizado grandes aportaciones a la misma a través de lo que se conoce como modelos simbólicos (utilizados en la lingüística computacional), tal y como puede verse en Moreno Sandoval (2019). Los sistemas simbólicos son objetos matemáticos definidos a partir de un conjunto de expresiones (los axiomas) y un conjunto de reglas (las reglas de derivación). Estas reglas transforman los axiomas en expresiones nuevas, los teoremas. Dicho de otra manera, los teoremas se deducen de los axiomas. En el caso de Chomsky, se aprecia cómo a través de los símbolos asociados a cada estructura y jerarquía dentro de la gramática generativa se llega a un correcto análisis de la frase (el principal argumento de Chomsky en contra de la lingüística de corpus tiene que ver con que no coinciden la

⁶ Esta afirmación de Fillmore (1992, pág. 35), uno de los lingüistas más influyentes de las últimas décadas, citada por Pérez Hernández (2002, pág. *online*), en castellano es:

La lingüística de sillón no tiene buena fama en algunos círculos lingüísticos. Una caricatura del lingüista de sillón es algo así. Se sienta en un sillón profundamente mullido y cómodo, con los ojos cerrados y las manos juntas detrás de la cabeza. De vez en cuando abre los ojos, se levanta bruscamente y grita: "¡Vaya, qué dato tan claro!", coge el lápiz y escribe algo. Luego se pasea durante unas horas con la emoción de haber estado aún más cerca de saber cómo es realmente el lenguaje. (No hay nadie exactamente así, pero hay algunas aproximaciones). La lingüística de corpus no tiene buena fama en algunos círculos lingüísticos. Una caricatura del lingüista de corpus es algo así. Tiene todos los hechos primarios que necesita, en forma de aproximadamente un millón de palabras corrientes, y considera que su trabajo consiste en derivar hechos secundarios de sus hechos primarios. En este momento está ocupado en determinar las frecuencias relativas de las once partes de la oración como primera palabra de una frase. (No hay nadie exactamente así, pero hay algunas aproximaciones.) Estos dos no se hablan muy a menudo, pero cuando lo hacen el lingüista de corpus le dice al lingüista de sillón: "¿Por qué debería pensar que lo que me dices es cierto?", y el lingüista de sillón le dice al lingüista de corpus: "¿Por qué debería pensar que lo que me dices es interesante?". Pérez Hernández (2002, pág. *online*)

cantidad de frases ni la estructura de las frases que un hablante es capaz de producir con la que efectivamente produce).

En realidad, los sistemas simbólicos son actualmente muy necesarios en contextos muy diferentes. De hecho, según se aprecia en estudios como los de Moreno Sandoval (2019), en la lingüística teórica (aunque no en todas sus vertientes), y en la lingüística computacional se entiende como requisito obligado el uso de gramáticas formales para dar explicaciones a los fenómenos lingüísticos. La utilidad práctica que conlleva tener un modelo teórico bien definido y formalizado deriva, según dicho autor en que, por un lado, se pueden hacer predicciones sobre la estructura y la gramaticalidad de las oraciones y, por otro, en que se pueden desarrollar aplicaciones para resolver problemas concretos, tales como por ejemplo la traducción automática o la extracción de información de documentos.

En nuestro caso, pretendemos observar el lenguaje matemático de especialidad desde un punto de vista basado en el hablante que utiliza este lenguaje para ofrecer conclusiones reales. Los lenguajes de especialidad son utilizados por un grupo de hablantes muy determinado y se emplean en contextos muy específicos, de modo que, para poder describir los lenguajes de especialidad, la lingüística de corpus (siempre que se usen las adecuadas herramientas computacionales) es el método más pertinente, puesto que se basa en ejemplos o usos reales del habla. Siempre que se observe un enunciado, o una forma de hablar recurrente, se puede decir que en un determinado ámbito se habla específicamente de esa manera y con esos términos concretos. Esto no es algo intuitivo, como lo que Fillmore llamaba “lingüística de sillón” (Fillmore, 1992), sino que se trata de conclusiones basadas en la experiencia. A esto hay que añadir que las herramientas informáticas avanzan cada vez más y tienen mayor alcance cada año (es obvio que no es comparable la tecnología informática de la que disponemos hoy en día con la tecnología existente en 1992, fecha en la que Fillmore hizo la afirmación anterior). Nos basamos, pues, en un corpus textual por tratarse de una herramienta empírica de trabajo tal y como la concibe Llamazares en 2008 (y de esto hace ya más de diez años).

Por otra parte, es muy difícil separar la lingüística computacional de la lingüística de corpus, ya que los conjuntos de textos son procesados, hoy en día, mediante programas informáticos. Se puede hablar incluso de lingüística computacional teórica y

de lingüística computacional aplicada (Moreno Ortiz, Pérez Hernández & Faber, 1999). La primera tiene como objetivos la elaboración de modelos lingüísticos en términos formales e implementables, la aplicación de dichos modelos a los diferentes niveles de descripción lingüística y la comprobación computacional de la congruencia del modelo y de sus predicciones. La lingüística computacional aplicada, a la que se supone una orientación más tecnológica que a la lingüística computacional, se centra, a grandes rasgos, en el diseño de sistemas informáticos capaces de gestionar, comprender, producir y traducir enunciados orales y escritos en lenguaje natural. Para ello desarrolla aplicaciones informáticas que pueden agruparse en cuatro grandes categorías:

1) Sistemas de consulta a bases de datos a través del lenguaje natural; 2) aplicaciones de las tecnologías del habla, como los sistemas de conversión de texto a voz; 3) herramientas para el procesamiento de textos para la elaboración, gestión y revisión de documentos (que se incluyen en la mayoría de los procesadores de textos actuales), o los programas de generación automática de resúmenes, los sistemas de extracción de información y los de catalogación documental automatizada; y, por último, 4) las herramientas orientadas al procesamiento de más de una lengua o de una lengua extranjera, como son las aplicaciones didácticas para la enseñanza de las lenguas, las herramientas de traducción (semi)automática, las memorias de traducción y las bases de datos terminológicas. (Moreno Ortiz, Pérez Hernández & Faber, 1999, pág. 177)

El campo de trabajo que se caracteriza por la aplicación de los ordenadores a la investigación lingüística, es decir, al estudio científico del lenguaje, es lo que Guinovart (2000, pág. 221) denomina “informática aplicada a la lingüística” o lingüística informática, expresión que puede aplicarse en un sentido amplio a todas las disciplinas de la lingüística que utilizan herramientas informáticas para sus estudios, sobre todo en el caso de la lingüística de corpus o la lingüística histórica computacional. Es en este término más amplio donde tiene cabida la segunda parte práctica de la presente tesis doctoral.

2.2 Lingüística de corpus

Tal como se señalaba en el apartado anterior, un corpus es un conjunto de textos o documentos escritos en una lengua determinada. Según Llamazares (Llamazares, 2008, pág. 329), “el conjunto de datos es lo que se denomina ‘corpus’ en un sentido general del término”.

El término “lingüística de corpus” –entendida a rasgos generales como la parte de la lingüística centrada en analizar determinados conjuntos de textos o corpus lingüísticos– es relativamente reciente, pese a que la recopilación de textos es una tarea que se ha llevado a cabo desde que existe el concepto de *corpus* como tal. Llamazares (2008) resalta algunos de los principales acontecimientos que han sido parte en el proceso de desarrollo y consolidación de la lingüística de corpus y ha prestado especial atención a la evolución del concepto de corpus en sí mismo. Así, con anterioridad al siglo XIX (previamente al uso de la informática), un corpus se definía por ser un conjunto de textos escritos (datos) y tener como finalidad el estudio de lenguas muertas (latín, sánscrito...). Desde esta perspectiva, los corpus eran necesarios para llevar a cabo los estudios lingüísticos. Según Llamazares (2008), los datos que eran contemplados en estos corpus antiguos eran el único acercamiento posible a lenguas como el latín o el sánscrito ya que estas lenguas no contaban con hablantes vivos.

Para entender la lingüística de corpus, se puede tomar la definición dada por el Instituto Cervantes (Centro Virtual Cervantes, s.f.), que habla de la lingüística de corpus como una rama de la lingüística que basa sus investigaciones en datos obtenidos a partir de corpus, esto es, muestras reales de uso de la lengua tal y como se mencionaba en la introducción a esta segunda parte de la presente tesis doctoral.

En Moreno Ortiz, Pérez Hernández & Faber (1999), se indica que el comienzo de la lingüística computacional como tal se puede situar en los últimos años de la segunda guerra mundial, debido a la rivalidad científico-técnica entre Estados Unidos y la Unión Soviética, cuyos especialistas comenzaron a trabajar en diversos proyectos para elaborar programas de traducción entre el inglés y el ruso. En este repaso sobre la aplicación de herramientas informáticas a conjuntos de textos, los autores destacan en su obra de 1999 que durante los años cuarenta y cincuenta se produjeron importantes avances en dos áreas que resultaron claves para las tecnologías de procesamiento del

lenguaje natural⁷: la teoría de autómatas, que se originó en los trabajos de Alan Turing (uno de los padres de la informática), y los modelos de teoría de la información, que surgieron de los trabajos de Claude Shannon (1948), quien aplicó la teoría de la probabilidad de procesos de Markov para desarrollar autómatas que procesaran el lenguaje humano.

De cualquier modo, tal como se señala en estudios como (Moreno Ortiz, Pérez Hernández, & Faber, 1999), a partir de la década de los noventa es cuando aparece una mayor urgencia por perfeccionar las tecnologías de procesamiento del lenguaje gracias a la entrada de la era de internet o la revolución de la *World Wide Web*. A día de hoy, se trabaja en programas informáticos para el análisis automático de la fonética, la fonología, la morfología, la sintaxis y la semántica de diferentes lenguas (y sus textos).

Otro campo de relevancia dentro de la lingüística computacional, que se encuentra unido al de la inteligencia artificial, es el de la representación formalizada del conocimiento por medio de la construcción de ontologías, que, como decíamos en apartados anteriores, son representaciones semánticas independientes de la lengua (pero con aplicaciones a ella) y que actualmente tiene un amplio campo de aplicación en el desarrollo de la Web 3.0 o Web semántica (Moreno Ortiz, Pérez Hernández, & Faber, 1999). En una ontología, definida como la representación explícita de una conceptualización, se especifica un conjunto de tipos de conceptos y de sus relaciones, organizados y representados formalmente para su uso computacional, y que describen formalmente el conocimiento de un ámbito de especialidad determinado (Gruber, 1995).

Llamazares (2008) también resalta que, con el avance del siglo XIX y hasta mediados del XX, se siguió empleando esta forma de trabajar basada en la recopilación de una gran cantidad de datos escritos (corpus) para tareas tales como dar cuenta del proceso de adquisición del lenguaje infantil a través de la transcripción de las

⁷ En inglés, natural language processing, NLP. Se trata un campo de las ciencias de la computación, de la inteligencia artificial y de la lingüística que estudia las interacciones e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas y máquinas por medio del lenguaje natural, es decir, de las lenguas del mundo. En este caso es importante resaltar que entendemos lenguaje natural como el utilizado por las personas en contraposición al lenguaje artificial “de las máquinas”.

interacciones de los niños con sus padres; establecer convenciones ortográficas; obtener listas de vocabulario para la enseñanza de segundas lenguas; y realizar estudios comparativos de lenguas. De nuevo, cabe resaltar en este punto que el uso que se le da a un conjunto de datos depende de la perspectiva que quiera explotar el investigador en cuestión, por eso la lingüística de corpus tiene actualmente tantas aplicaciones.

En la actualidad, un corpus se concibe de manera muy diferente precisamente porque se emplea como recurso de investigación. Para poder considerarse como una herramienta lingüística, un corpus ha de presentarse en formato electrónico o informatizado de manera que el investigador (o usuario) pueda buscar de forma rápida tanto palabras como secuencias de palabras (lo que le permitirá asignar categorías y esquemas gramaticales). Además, al estar informatizado, un corpus permitirá recuperar la información en segundos sobre todos los casos en los que aparece una palabra, una secuencia de palabras y las concordancias entre ellas (mediante el contexto inmediato anterior y posterior). Tratándose de conjuntos de textos digitales, se puede examinar la frecuencia de aparición de las palabras o, incluso, se puede pedir al sistema que ordene los términos del corpus por orden alfabético, por procedencia geográfica o por temática. Todo depende de qué tipo de información se quiera extraer de ese conjunto de textos o corpus.

En cualquier caso, una característica con la que ha de cumplir un corpus es la de que, por un lado, sus datos han de ser reales en cuanto al uso de la lengua objeto de estudio, ya que a partir de estas muestras es a partir de lo que se van a construir las teorías empíricas que tratan de dar explicación al uso de la lengua. Y, por otro lado, los textos que integran el corpus tendrán que haberse seleccionado siguiendo una serie de criterios de tal manera que, para que un texto pueda integrarse dentro de un corpus, tendrá que cumplir una serie de requisitos previamente definidos y comunes al resto de textos que ya lo integran (normalmente los criterios de un corpus dependen de la finalidad que persiga la investigación).

Un corpus ha de ser representativo de la lengua que se está analizando, por eso podemos encontrar actualmente corpus de diversos tamaños, ya que no es lo mismo analizar la obra de un determinado autor que analizar el estilo poético de un determinado siglo. Un corpus confeccionado para el segundo propósito será, por lógica, mucho mayor que el primero. Llamazares (2008) hace énfasis sobre una idea muy

interesante que acaba con muchas críticas hacia los corpus textuales basadas en la idea de que estos estaban sesgados, exponiendo ejemplos de algunos corpus abiertos o monitor, como el del proyecto COBUILD dirigido por J. Sinclair en la Universidad de Birmingham, (de gran interés para la lexicografía), o el propio CREA de la Real Academia Española. En palabras de Llamazares (2008 pág. 332):

En el pasado se pensaba que el tamaño era muy importante: mientras mayor fuera el corpus, más posibilidades tenía de reflejar el funcionamiento real de la lengua en todas sus variedades, pero en la actualidad se priman los criterios de diseño, es decir, el tamaño solo es importante en la medida en que así lo exija la finalidad del corpus.

Una de las áreas básicas de actuación de la lingüística computacional, y cuyo desarrollo es un requisito indispensable para que las demás aplicaciones computacionales que se diseñen para el procesamiento del lenguaje natural sean útiles, es precisamente la del estudio del uso y de la estructura lingüística: lo que Gómez Guinovart (2000, pág. 201) denomina lingüística informática y cuyo máximo exponente es la lingüística de corpus, entendida como el estudio empírico de la lengua a partir de los datos que proporciona el análisis de ejemplos reales de producciones lingüísticas (orales o escritas) almacenadas en un ordenador. Cuanto mayor sea el conocimiento que tengamos de la estructura lingüística (en el nivel morfológico, sintáctico, semántico y pragmático) y del uso que los humanos hacemos de la lengua, mejores serán las aplicaciones informáticas que podremos diseñar para comunicarnos, para manejar, almacenar o extraer información o para interactuar con los ordenadores; por eso es importante seleccionar qué elementos son representativos de lo que se quiere estudiar a través del corpus.

En el apartado anterior señalamos que la lingüística de corpus como tal es una rama de la lingüística que plantea ciertos problemas en su definición; siendo así, resulta relevante repasar brevemente la discusión que sigue en estas líneas, puesto que afecta directamente a la investigación de esta tesis doctoral.

Los problemas de definición de la lingüística de corpus tienen que ver con dos posturas: se discute, desde el principio, si se trata de una teoría o de una metodología de análisis (Pérez Hernández, 2002). Pérez Hernández relaciona una amplia bibliografía

que trata sobre este asunto, de entre los que destaca los trabajos de McEnery & Wilson (1996); Kennedy (1998); Stubbs (1995); o los de Tognini-Bonelli (1996). En cualquiera de los casos, tiene sentido pensar que el análisis de textos en conjunto permite comparar unos textos con otros y extraer así conclusiones más relevantes que las que se obtendrían si se analizan los textos por separado.

La lingüística de corpus no solo ha planteado algunos problemas a la hora de definirse; también ha ido evolucionando en lo que a popularidad se refiere –también ha recibido muchas críticas–. Hay que tener en cuenta que la capacidad de procesamiento de datos no es la misma ahora que hace cincuenta años. De hecho, hoy en día, gracias a la lingüística computacional, algunos autores hablan, por ejemplo, de lexicografía computacional y lexicografía computacional aplicada como buenas herramientas lexicográficas o métodos de recopilación de términos (Moreno Ortiz, Pérez Hernández & Faber, 1999). La lexicografía computacional se refiere al uso de medios técnicos computacionales en los diferentes procesos que se siguen en la elaboración de un diccionario: desde que la primera idea parte del equipo editorial, pasando por decisiones que atañen a la macro-estructura del diccionario (lista de lemas, orden, etc.) y a su microestructura (el almacenamiento de la información durante el proceso de compilación de las entradas o los complejos medios de maquetado y edición en las fases posteriores). Cuando se habla de lexicografía computacional aplicada, en cambio, nos referimos la rama de la lingüística computacional que se encarga de la elaboración de software y de modelos de análisis o programas de inteligencia artificial para la mejora y afinamiento de los parámetros de búsqueda de términos o palabras para la elaboración de listas y bases de datos terminológicas.

Pérez Hernández (2002) también menciona varias críticas que se han hecho a la lingüística de corpus, como las de Noam Chomsky, aunque para el caso que nos ocupa – el lenguaje especialidad matemático– no serían aplicables. Las críticas de Chomsky tenían su base en el rechazo al paradigma estructuralista, oponiéndose radicalmente al uso de cualquier metodología descriptiva en la teoría lingüística. Chomsky propuso la ya conocida distinción entre *competence* (competencia), el conocimiento interiorizado de una lengua, y *performance* (actuación), la evidencia externa de la competencia lingüística (que, en el caso que nos ocupa, es lo que podremos ver en un corpus textual de especialidad).

Pérez Hernández (2002) señala que esta diferenciación es similar a la dicotomía propuesta por Saussure entre *langue* (lengua) y *parole* (habla) en su *Cours de Linguistique Générale* (1916). Establece así una comparación entre ambos lingüistas ya que, en ambos casos, *langue* (en el caso de Saussure) y competencia lingüística (en el caso de Chomsky) se conciben como algo sistemático y, por ello, debe ser el único objeto de estudio de la ciencia lingüística, aunque sea una entidad abstracta, y por tanto, no observable. En el otro extremo se sitúan la *parole* (Saussure) o *actuación* (Chomsky), a la que se considera idiosincrásica y asistemática y, por ello, carente de interés para el lingüista teórico.

En esta línea de pensamiento, Pérez Hernández (2002) señala que tal y como apuntan autores como Tognini-Bonelli (1996) o Stubbs (1995), el concepto de *parole* o actuación lingüística es observable, por definición, solo de forma fragmentaria, puesto que como un todo es inobservable y, en cualquier caso, no refleja la competencia lingüística innata del hablante-oyente ideal. En otras palabras, es prácticamente imposible observar la totalidad de los enunciados lingüísticos que puede emitir un hablante, pero en nuestro caso del lenguaje de especialidad, sí que podemos observar una buena muestra de cómo se utiliza el lenguaje y cómo interactúan los términos en este caso.

Con el paso del tiempo, se ha ido perfilando la distinción entre *competence* y *performance* a través de lo que se conoce como I-language (*internalized language*) y E-language (*externalized language*) (Chomsky 1986, 1992), aunque se sigue manteniendo que la *E-language* posee una condición muy difícil de caracterizar en términos teóricos y sólo la competencia lingüística o *I-language* puede ser igualada a la noción de lengua, por lo que la separación entre *competence* y *performance* se mantiene en lo esencial. Es destacable que tal y como señala Pérez-Hernández (2002), en su obra disponible *online*, a través de una cita de Tognini-Bonelli (1996, pág. 23) estas dos dicotomías pueden considerarse en muchos sentidos paralelas, aunque difieren en un aspecto fundamental: para Saussure *langue* y *parole* son interdependientes, mientras que Chomsky deja muy claro que *competence* y *performance* no lo son.

En el otro extremo, y como defensa de la lingüística de corpus –más bien, de la posibilidad de trabajar con el lenguaje en acción–, Pérez Hernández (2002), resalta la

obra de Firth (1957), quien habla de la separación saussureana al subrayar la importancia de la función social del lenguaje y del hecho de que tenga lugar en un contexto situacional. Pérez Hernández (2002) deduce del trabajo de Firth que el lenguaje debe ser estudiado teniendo en cuenta la situación en la que este se produce y en concordancia con los participantes, las acciones verbales y no verbales que llevan a cabo y las consecuencias o efectos de la acción verbal. En esta misma línea, la autora resalta el hecho de que Firth toma el postulado básico de Wittgenstein de que *“the meaning of words lies in their use”* y lo desarrolla en su teoría contextual del significado: para él, el uso lingüístico no existe de forma aislada, ya que ha de ser estudiado de forma contextual y de acuerdo con la función social que posee.

Estos argumentos utilizados por Firth (1957) que tienen que ver con la importancia del contexto situacional y la función social que cumple el lenguaje están directamente relacionados con la lingüística de corpus ya que, si el significado de las palabras no puede separarse de su uso, cualquier estudio lingüístico deberá partir del estudio detallado de ejemplos reales del mismo –que es precisamente el tipo de información que encontramos en los corpus: ejemplos reales de uso–. A nadie llama la atención –teniendo en cuenta su formación funcionalista– que autores como Halliday o Sinclair (Sinclair & Kirby, 1990), denominados “neofirthianos”, hayan puesto el foco en la importancia de estudiar las lenguas a través del análisis sistemático del uso lingüístico.

En el caso que nos ocupa, que es el de estudiar el lenguaje de especialidad, es especialmente relevante observar cómo varía el significado de las palabras según su contexto situacional y las unidades de significado que se forman en dichos contextos de especialidad, por lo que, para este estudio, la lingüística de corpus ofrece resultados reales frente a los hipotéticos que podría ofrecer la lingüística tradicional o teórica. Si comparamos la lingüística teórica con la lingüística de corpus, algunos autores como Sinclair (1990) recalcan las posibles inconsistencias o inexactitudes de las intuiciones lingüísticas considerando incluso algunos casos en los que el hablante nativo simplemente puede no poseer el conocimiento intuitivo suficiente para postular una parte de la teoría. Como señala Stubbs (1996 pág. 27), “en cualquier ámbito científico se da por sentado que el científico desarrolla una teoría para describir y explicar un fenómeno que ya existe, partiendo de una serie de datos o experimentos externos”.

En las últimas décadas la lingüística de corpus y la lingüística computacional se han convertido en disciplinas muy populares y conocidas. Cuéllar (2015) (Cuéllar, 2015) establece una diferencia clara entre la lingüística computacional y la lingüística de corpus afirmando que, generalmente, la lingüística de corpus centra su interés en la utilización de los corpus disponibles o en la confección de corpus mediante el empleo de las herramientas que han diseñado para este fin los profesionales de la lingüística computacional. De este modo, el lingüista de corpus no es necesariamente un programador (el lingüista computacional sí que lo es). Aunque se trata de actividades parecidas, tienen objetivos distintos: uno de ellos se centra en desarrollar la herramienta de análisis en sí (el lingüista computacional), mientras que el lingüista de corpus trabajará para extraer conclusiones exclusivamente lingüísticas aprovechándose de los diseños del primero. A la lingüística computacional le compete el diseño de las herramientas informáticas para un óptimo procesamiento automático del lenguaje natural.

Esta tesis doctoral es, por tanto, de interés desde el punto de vista tanto de la lingüística de corpus como desde el punto de vista de la lingüística computacional, ya que se presenta una nueva metodología para extraer resultados lingüísticos que bien puede tener cabida en esta área.

Cuéllar (2015), al igual que Moreno Ortiz y Pérez Hernández & Faber (1999), realiza un recorrido por todas las aportaciones que hace la lingüística computacional a la lingüística de corpus. No obstante, deseo destacar –asimismo es relevante para esta tesis– el hecho de que habla de la traducción automática, área de trabajo con la que se reconoce que comenzó la lingüística computacional en la década de los años sesenta del siglo pasado. Por otra parte, a la lingüística de corpus le interesa realizar análisis lingüísticos léxicos, gramaticales, semánticos y pragmáticos (discursivos) mediante el uso de las herramientas informáticas diseñadas para este fin, aspectos relevantes para la tarea de la traducción.

Así, antes de presentar la nueva metodología que utiliza la teoría de redes complejas, partiremos de la observación de un conjunto de textos (o corpus) y de cómo interactúan las palabras para poder extraer conclusiones como palabras clave en base a parámetros tales como la frecuencia absoluta, el tiempo verbal o las relaciones sintagmáticas y paradigmáticas entre los distintos términos.

Hoy en día contamos con varias herramientas de software y programas que, a partir de una modelación estadística o basada en reglas, permiten, por ejemplo, la caracterización y posterior búsqueda de categorías léxicas o combinación de palabras en textos por medio de los etiquetadores (conocidos como *taggers*), así como la descripción y búsqueda de ciertas estructuras sintácticas mediante el uso de los analizadores sintácticos (conocidos como *parsers*). Para ello es necesario realizar un gran trabajo de etiquetado y de preparación del corpus por parte del lingüista. Este proceso es a veces conocido como proceso de limpieza del corpus, y es una de las principales dificultades que se presentan al trabajar con la lingüística de corpus, puesto que en muchas ocasiones es necesario el ojo humano para decidir, entre otros aspectos, y a modo de ejemplo, qué es una palabra y qué no lo es, o si la frase se considera entre los dos puntos o entre el punto y la coma.

Esta es la razón por la que, en nuestra investigación, la lingüística de corpus nos va a ser realmente útil ya que, en palabras de Parodi (2008 pág. 95):

“la lingüística de corpus no se entiende como una rama o un área de la lingüística tal como son la fonología, la semántica, la sintaxis, sino como un método de investigación que puede ser empleado en todas las ramas o áreas de la lingüística, en todos los niveles de la lengua y desde enfoques teóricos diferentes. Sus aplicaciones son múltiples y no limitan las posibilidades de indagación.”

La lingüística de corpus permite, en definitiva, estudiar el lenguaje tal y como se utiliza –o tal y como aparece–, y cuestionar muchos postulados lingüísticos que se daban por ciertos. Para nuestro estudio es especialmente relevante hablar de unidades de conocimiento específico tal y como se mencionaba en la primera parte de esta tesis doctoral, ya que estas unidades son de vital importancia para el mundo de la traducción. Sinclair (1991), por ejemplo, demuestra cómo en ocasiones, formas diferentes de un mismo lema deben considerarse como unidades léxicas independientes, ya que su comportamiento sintáctico o su significado es diferente. La tradicional noción por la que se asigna un significado (o significados) a todas sus formas posibles no se corresponde a veces con la frecuencia y distribución que se encuentra en un corpus (especialmente si nos encontramos ante un caso que trate de lenguaje de especialidad). Un buen ejemplo de la afirmación anterior es el uso del término *primo* fuera del ámbito de las matemáticas; la palabra o término *primo* altera totalmente su significado dentro de esta

área, especialmente si aparece tras la palabra *número*, formando la colocación *número primo*. Lo mismo sucede con la colocación *raíz cuadrada*, en la que los términos, analizados por separado, tienen un significado muy diferente.

Por consiguiente, es necesario analizar cómo funcionan los términos o palabras en contexto. Renouf & Sinclair (1991) introducen el concepto de *collocational framework* (marcos colocacionales). Dentro de estos marcos colocacionales no es posible separar el estudio léxico del estudio gramatical, ya que en la mayoría de los casos las estructuras sintácticas y las léxicas son interdependientes, de forma que no es posible separar el estudio léxico del sintáctico. Esta línea de pensamiento es ampliamente desarrollada y aplicada al español por Butler (1998). También destacan los estudios realizados por Louw (1993) sobre *semantic prosodies* que evidencian que también es necesario replantearnos el uso tradicional de la palabra como unidad básica de significado, ya que, en ocasiones, puede ser la palabra la unidad básica de significado que le interesa al traductor y sin embargo, en otras, se puede estar buscando un conjunto de palabras que aluda a un significado concreto.

Los trabajos sobre equivalencia de traducción que presentan autores como Sinclair, Payne, & Pérez Hernández (1996), y Tognini-Bonelli (1996) muestran que, con más frecuencia de lo que un diccionario bilingüe parece indicar, no es posible asignar un equivalente de traducción apropiado sin tener en cuenta el contexto situacional y el contexto lingüístico en el que las palabras aparecen, por lo que es necesario ampliar la noción tradicional de equivalencia de traducción. Para un traductor es interesante estar familiarizado con la lexicografía (entendida como la técnica para elaborar o confeccionar diccionarios) para saber cuándo estamos hablando de una sola palabra como unidad de conocimiento, o unidad de traducción, o cuándo el significado de esa palabra o término variará si viene acompañada de alguna otra, máxime si se trata de términos cuyo significado cambia si se utilizan en un determinado ámbito de especialidad. Para este fin, un corpus es una herramienta lexicográfica fundamental para el estudio de las diferentes acepciones de las entradas léxicas y para el estudio de las colocaciones y la fraseología (Baugh, Harley & Jellis 1996; Sánchez et al. 1995; Sinclair 1987b, 1992a; Sinclair & Kirby 1990, Moon 1998). Un ejemplo de la afirmación anterior es la variación en cuanto a significado que sufren las palabras "*primo*" o "*entero*" en castellano cuando estas vienen precedidas de la palabra "*número*", formando las colocaciones "*número*

primo” y “*número entero*” respectivamente (pertenecientes al área de especialidad matemática). Dependiendo de en qué contexto (o área de especialidad en este caso) se utilicen estas palabras y del resto de términos que las acompañen, tendrán un significado u otro.

2.3 Descripción del corpus utilizado

En los apartados anteriores se ha justificado la necesidad de utilizar la lingüística de corpus y la lingüística computacional para realizar la investigación que se presenta en esta tesis. Estas razones quedarán más claras a través de las líneas y apartados que siguen, puesto que, en ellas, se describen las características del corpus empleado en esta investigación. Así, en la primera parte se ha visto como, actualmente, la lingüística de corpus y la lingüística computacional constituyen en conjunto la metodología más eficaz para el estudio de los lenguajes de especialidad. El poder contar con un corpus de análisis permite ver cómo se emiten enunciados en una determinada área del saber, así como dar respuesta a una serie de incógnitas lingüísticas (entre otras, el significado que adquieren los términos dentro de ese lenguaje de especialidad; o también si se combinan en grupos de dos o más palabras para adquirir ese significado específico; el uso de frases complejas, subordinadas o simples; o, incluso, el uso del aspecto verbal) frente a las especulaciones que podrían hacerse desde sólo la teoría. Se parte, por tanto, del método científico basado en la observación de los datos que ya se tienen para extraer conclusiones. Este es el motivo por el que se ha de seleccionar un corpus que sea representativo de aquello que se quiere estudiar. No hay que perder de vista que en esta tesis doctoral se pretende analizar el lenguaje matemático de especialidad que tiene que ver con la teoría de redes complejas y extraer conclusiones sobre el mismo.

1. Criterios de representatividad y tipología de corpus

En el estudio realizado, para elegir los textos que integran el corpus de análisis, se han tenido en cuenta los criterios de representatividad de los que hablan Atkins, Clear y Ostler (1992), ya que se ha tomado una buena muestra de este lenguaje de especialidad, en el que la mayoría de las publicaciones están hechas en inglés por hablantes no nativos. Estos autores distinguen entre diversos tipos de corpus o

colecciones textuales (todos ellos digitalizados): los archivos (conjuntos de textos sin relacionar ni coordinar de forma alguna, como el Oxford Text Archive); bibliotecas de texto en formato magnético (o *ETL en inglés*, *Electronic Text Library*), que son una colección de textos en formato electrónico con formato estandarizado que siguen ciertas convenciones en cuanto al contenido pero sin rigurosas limitaciones de selección; corpus *ETL*, creados siguiendo unos criterios de selección explícitos y con un propósito específico (como por ejemplo el corpus de Cobuild o el de Longman/Lancaster); y, por último, los subcorpus, que son selecciones de un corpus mayor o más complejo, hechas de forma dinámica *online* mientras se está consultando el corpus principal (Atkins, Clear, & Ostler, 1992).

Pérez Hernández (2002) señala que, entre los posibles tipos de corpus, destacan los *reference corpus*, creados para que sean una muestra representativa de las variedades más importantes de una lengua, así como de sus estructuras y vocabulario generales, de forma que ofrezcan información lo más amplia posible sobre una lengua y puedan servir de base para la construcción de gramáticas, diccionarios y obras de referencia. El *British National Corpus*, el *Bank of English* y el CREA son ejemplos de *corpora* de referencia. Otro modelo de corpus diferente es el de «corpus monitor» (*monitor corpus*); en él se propone la creación de un corpus con un tamaño constante, en el que se van añadiendo continuamente materiales nuevos a la vez que se van eliminando cantidades equivalentes de material antiguo para ofrecer así al lingüista la posibilidad de observar cambios recientes en el uso de la lengua.

La composición de nuestro conjunto de textos obedece tanto a la intención de cumplir con los requisitos de especialidad y representatividad como a la de que pueda ser considerado como una herramienta específica de análisis para un lenguaje de especialidad muy concreto.

2. Área de especialidad matemática: catálogo AMS para Mathematical Reviews Data Base

Es necesario volver a subrayar el hecho de que acotar el lenguaje matemático es una tarea muy complicada ya que las matemáticas son una disciplina presente prácticamente en todas las áreas del saber. Pocas son las cosas que no se pueden medir o cuantificar de un modo u otro, por lo que, para estudiar la mejor manera de verbalizar

el lenguaje matemático, resulta más apropiado iniciar el análisis desde una rama concreta de las matemáticas (álgebra, cálculo, geometría, estadística o topología, entre tantísimas otras). La *American Mathematical Society* (AMS) asigna diferentes códigos a las distintas áreas temáticas sobre las que se publican los artículos científicos; este listado es, esencialmente, una ontología⁸ de conocimiento acerca de las disciplinas en las que las matemáticas tienen cabida o pueden considerarse como áreas de aplicación matemática.

Cabe señalar además que la propia catalogación de la AMS (<https://www.ams.org/>) sufre continuas revisiones ya que, debido a la aparición de nuevas áreas y sub-áreas matemáticas obliga a ello. A través de este enlace <https://mathscinet.ams.org/mathscinet/msc/pdfs/classifications2020.pdf>, se puede descargar la catalogación más reciente. La dificultad de catalogar las distintas áreas matemáticas queda visible en la propia página web de la AMS <https://mathscinet.ams.org/mathscinet/msc/msc2020.html> los nuevos campos y cambios relacionados con las distintas áreas y subtipos que se han introducido desde el año 2010 hasta el año 2020. Destaca el hecho de que aparecen descriptores específicos comunes a diferentes áreas de especialización que pueden hacer que un mismo texto trate temas comunes a dos áreas diferentes, de forma que dicho texto aparezca relacionado con dos o más sub-apartados.

⁸ Las ontologías de conocimiento (además de para catalogar) se han utilizado tradicionalmente como bases de datos lingüísticas ya que una ontología es una representación formal del conocimiento donde los conceptos, las relaciones y las restricciones conceptuales son explicitadas mediante formalismos en un determinado dominio (Fensel, McGuinness, Vam Harmelen, & Patel-Schneider, 2001). De este modo, cada contexto y a veces cada término queda asociado a un campo o área concreta. Su función más frecuente en la lingüística es de apoyo para sistemas de Traducción Automática Basada en el Conocimiento y para la Terminografía (práctica de la Terminología) (Fensel, McGuinness, Vam Harmelen, & Patel-Schneider, 2001). En ambos casos, la ontología es una representación formal y explícita de la estructura conceptual del campo sobre el que se trabaja (o campo de especialidad). Por ejemplo, la palabra “raíz” en una ontología catalogada dentro del área de la botánica tiene una acepción diferente de la palabra “raíz” cuando aparece junto a “raíz cuadrada” en una ontología dentro del área de las matemáticas (esto último se considera una colocación de especialidad).

Lo que se utiliza en la AMS para catalogar es el *Mathematics Subject Classification scheme* o esquema de Clasificación de Materias Matemáticas, cuyo principal objetivo es ayudar a los usuarios a encontrar los artículos de interés actual o potencial para ellos con la mayor facilidad posible, empleada en la *Mathematical Reviews Database* (también conocida como MRDB, fuente disponible a través de la propia <https://www.ams.org/>) y en otras fuentes de conocimiento y divulgación matemática de gran importancia como *Zentralblatt MATH* (<https://zbmath.org/>) o en cualquier otro lugar donde se utilice este esquema de clasificación mencionado y conocido como *Mathematics Subject Classification scheme* o MSC.

Un artículo científico de especialidad matemática debe clasificarse de manera que atraiga la atención de todos los posibles interesados en él. Puede tratarse de un tema que se inscriba en un área clara del MSC, o puede abarcar varias áreas. Lo ideal es que los códigos del MSC asociados a un artículo representen los temas a los que dicho artículo contribuye. La clasificación debería servir tanto a los que están estrechamente relacionados con áreas temáticas específicas, como a los que están lo suficientemente familiarizados con los temas como para aplicar sus resultados y métodos en otros lugares, dentro o fuera de las matemáticas. Será muy útil para los usuarios y los clasificadores familiarizarse con todo el sistema de clasificación y conocer así todas las clasificaciones que puedan interesarles.

Cada elemento de la *Mathematical Reviews Data Base* (MRDB) recibe precisamente una clasificación primaria, que es simplemente el código MSC que describe su principal contribución. Cuando un artículo contiene varias contribuciones principales a diferentes áreas, la clasificación primaria debe abarcar la más importante de ellas. A un artículo o libro se le pueden asignar uno o varios números de clasificación secundaria para cubrir cualquier otra contribución principal, resultados auxiliares, motivación u origen de los asuntos tratados, campo de aplicación previsto o potencial, u otros aspectos significativos que merezcan ser destacados.

Se entiende que la contribución principal es la que incluye la parte más importante del trabajo realmente realizado en el artículo. Por ejemplo, un artículo cuyo contenido general principal es la solución de un problema de teoría de grafos, que surgió en la informática y cuya solución es (quizás) en la actualidad sólo de interés para los informáticos, tendría una clasificación primaria en 05C (Teoría de Grafos) con una o más

clasificaciones secundarias en 68 (Informática); a la inversa, un artículo cuyo contenido general radica principalmente en la informática debería recibir una clasificación primaria en 68, incluso si hace un uso intensivo de la teoría de grafos y demuestra varios resultados nuevos de la teoría de grafos en el camino.

Hay dos tipos de referencias cruzadas al final de muchas de las entradas del MSC. El primer tipo está entre corchetes: “[Para A, véase X]”; si esto aparece en la sección Y, significa que a las contribuciones descritas por A se les debe asignar normalmente el código de clasificación X, no Y. El otro tipo de referencia cruzada se limita a señalar las clasificaciones relacionadas; está entre corchetes: “[Véase también ...]”, “[Véase principalmente ...]”, etc., y los códigos de clasificación que aparecen entre paréntesis pueden, aunque no es necesario, incluirse en los códigos de clasificación de un documento, o pueden utilizarse en lugar de la clasificación en la que se da la referencia cruzada. El clasificador debe juzgar qué clasificación es la más apropiada para el documento en cuestión

Para la confección de este corpus nos hemos centrado en textos que versan sobre la Teoría de Grafos y Redes Complejas debido, entre otras cosas, a que esta es un área de especialidad que tiene la peculiaridad de que, además de ser una de las más prolíficas en matemáticas (a tenor del número de artículos publicados en las últimas décadas y de sus múltiples aplicaciones), existe una enorme variedad de sistemas reales de interés en ciencia y tecnología que pueden ser descritos y analizados en términos de esta teoría.

La elección de esta área es interesante porque a través de las redes complejas se puede estudiar el propio lenguaje de especialidad que tiene que ver con la teoría de las redes complejas (si se tiene un corpus adecuado). Como se afirmaba en la primera parte de esta tesis, la teoría de redes complejas es aplicable a muchas disciplinas y se encuentra una gran variedad textual sobre la misma, circunstancia que tiene como consecuencia que la asignación de códigos de especialidad de la AMS no permita adscribir la teoría de las redes complejas a una única área.

Así, esta teoría tiene asignados varios códigos de especialidad según la aplicación o el tipo de desarrollo realizado: 05C82, 05C90, 90B10, 90C35, 91D30. Estos son los códigos a los que quedarían adscritos los documentos que se han incluido en el corpus de trabajo. Por tanto, todos los documentos que se han analizado y las expresiones

utilizadas en los mismos pertenecen al área de la matemática aplicada en general y, dentro de ella –y de manera abrumadoramente dominante– a la teoría de redes complejas.

3. Textos seleccionados

Los documentos textuales utilizados para la confección del presente corpus son los *extended abstracts*⁹ publicados en la revista IJCS (International Journal on Complex Systems) entre los años 2011 al 2016 (ambos incluidos) cuyas áreas de especialidad matemática están centradas en la matemática aplicada en general y muy en particular en la teoría de redes complejas. Se trata de un corpus redactado en inglés. La propuesta de traducción de los términos clave se hará sobre la base de las conclusiones obtenidas tras el correspondiente análisis textual del corpus y el estudio de la descripción y del comportamiento de estos términos.

Es importante señalar que estos textos pertenecen al área de especialidad de las redes complejas y que esta disciplina matemática en la mayoría de las ocasiones se aplica a otra área del saber (es decir, se aplican las herramientas y algoritmos de la teoría de redes complejas para describir los sistemas presentes en el mundo real, pudiendo dichos sistemas pertenecer a cualquier otra disciplina o área de conocimiento). En términos de lenguaje de especialidad, nos encontraremos con palabras del lenguaje de especialidad matemático de redes complejas como tal, así como terminología específica del área de investigación sobre el que se está aplicando la teoría de redes complejas. Por esto último, estudiar las palabras clave del lenguaje de especialidad matemático relacionado con las redes complejas conlleva una dificultad añadida. Se trata de ver cuáles son los términos de esta área matemática concreta y cuáles son los términos de especialidad que quedan fuera de este campo.

⁹ Los *extended abstracts* son versiones reducidas de los artículos. Tienen un número mayor de caracteres que los *abstracts* como tal y son particularmente relevantes para este estudio, ya que son una buena muestra de lenguaje matemático de especialidad independiente del lenguaje matemático intrínseco de las fórmulas y demás notaciones representativas de esta disciplina.

Los documentos del corpus se pueden consultar en el lector de documentos de la herramienta de análisis Voyant (a través del enlace <https://voyant-tools.org/?corpus=717e576737b1ba0d82d9b9cb197af784>).

Hablaremos de Voyant en los siguientes apartados de esta tesis (aunque cada uno de los textos que constituyen el corpus está también disponible en la página web de los editores). En el enlace anterior se pueden ver los textos del corpus previamente al proceso de etiquetado, es decir, sin procesar, por lo que resulta evidente el esfuerzo que se ha hecho para desarrollar la nueva herramienta de traducción en base a la teoría de redes complejas que se presenta en la parte tres de esta tesis doctoral. Un ejemplo de *extended abstract* de uno de los documentos incluidos en el corpus objeto de análisis es el siguiente (Benito et al., 2011):

Optical Communication Networks

R.M. Benito^{1,†}, J.P. Cárdenas¹, J.C. Losada¹, A. Santiago¹ and M.L. Mouronte²

¹ Grupo de Sistemas Complejos, Departamento de Física, Universidad Politécnica de Madrid, 28040 Madrid (Spain).

² Telefónica I+D, c/Emilio Vargas 6, 28043 Madrid (Spain).

Abstract.

Two closely related protocols are the standard technology for information transmission in broadband optical networks: the Synchronous Digital Hierarchy (SDH) and its North American equivalent SONET. Unlike systems with unplanned growth, such as those of natural origin or the Internet, these telecommunication networks are strictly planned as rings, meshes, stars or tree-branches structures designed to connect different equipments. In spite of this planification, we have found that the SDH network operated by Telefónica in Spain shares remarkable topological properties with other complex networks with different origins. In particular, the telecommunication network displays power-law scaling in the degree distribution of SDH equipments and small-world networks properties. In this work, the complexity observed in the Spanish SDH system is reproduced by two complex networks models. One of them, the ad hoc model, considers real planning directives that take into account geographical and technological variables with the aim to predict the growth of the system or generate policies for future system designs. The other model has a completely different philosophy. Based on a single evolutive rule, named compatibility, the Compatibility Attachment Model (CAM) is a generic approach to the complexity of the networks, including the one observed in the SDH system.

Keywords: complex networks, communication networks, optical communications, SDH network, model

MSC 2000: 05C82, 90B15

1. Introduction

During the last decade many complex systems have been studied from a network theoretical perspective. The results of these studies have shown the presence of particular, common and non intuitive statistical properties in these systems when they are abstracted to a network [1]. The so-called complex Optical Communication Networks networks display interesting topological properties such as scale-free carácter (i.e., scaling in the distribution of connectivities), properties of small-world networks [2] and modularity. Under this perspective we have studied the Spanish Synchronous Digital Hierarchy (SDH) telecommunication network operated by Telefónica España. This system is a multiplexing protocol for transferring multiple digital bit streams over the same optical fiber [3]. Structurally is a end-to-end circuit strictly designed as ring, mesh, bus and other motifs connecting equipments with different link capacities (Mbps/Gbps). In our research, the telecommunication system, introduced in Spain in 1992, was abstracted to a graph $G = (E, S)$ composed by a set E of nodes corresponding to SDH Equipments (as well as any equipment belonging to other technologies connected by optical systems to SDH equipments) and a set S of links corresponding to Synchronous Transport Module (STM-N) optical system that interconnects two SDH equipments or SDH equipments with others belonging to other technologies. Our results [4] have shown that the Spanish SDH system is a scale-free network that displays properties of small-world networks. With the aim to understand the origin of that complexity two network models were proposed: the ad hoc model and the compatibility attachment model (CAM).

2. Modeling the Topology of SDH Systems

The ad hoc model [4] takes into account real planning directives used by Telefónica España in order to simulate the network growth. Through a stochastic process, in each iteration a new node (equipment) with particular characteristics is added to the network. The new added node will be linked to existing ones depending on technical and engineering rules. In this way, the ad hoc model works according to the following general characteristics of the real system. Assumes that the SDH network is built as hierarchical networks of equipments interconnected through bidirectional links. In this way, adopts the hierarchy of the real SDH system: a lower layer that corresponds to the Access Network and three upper layers that correspond to the Carrier Network. Based on this hierarchy the model distinguishes five SDH equipment classes: A (equipment at national level), B (equipment at regional level), C (equipment at interconnection level), D (equipment at access level) and E (client equipment connected to SDH equipment by SDH connections). In real SDH networks each equipment is characterized by a number of installed ports that serve to establish SDH connections, as well as by a maximum number of ports that can be installed. In the ad hoc model all the equipments belonging to a given class are equivalent in these respects, thus the model characterizes each equipment class by a number of initially installed ports and a number R . M. Benito et al 35 Figure 1: Distribution of the number of nodes with k connections, $N \cdot P(k)$, obtained from the entire Spanish SDH network (left) and from the Sevilla province (right) in comparison with the distributions generated by the models of maximally supported ports. Concerning geographical location, each equipment in real SDH systems posses an ID that signals its address. In the ad hoc model the region where the network is deployed is divided into territorial units denominated provinces, and each province is divided into lower-level territorial units called sectors. Although traffic demand is an important factor behind the growth of real SDH networks the model

considers, for simplicity, that the Spanish territory has a uniform demand and therefore the probability that an equipment is added to any region will be considered uniformly distributed. The other model, CAM [5], is a generic proposal completely different. The model also simulates the network growth through a stochastic process similar to the ad hoc model. However, the link between the added equipment and those already present is governed by a single evolutive rule named compatibility that indicates the compatibility between their characters defined by a certain probability density function (PDF). Thus, the new added equipment is compatible with others when the difference between their characters is less than the compatibility threshold $C_m(\tau) = d \tau$, where d is a constant called compatibility distance and τ represents the time of arrival of the new added equipment.

3. Results and Discussions

Fig. 1 provides the degree distribution $N \cdot P(k)$ obtained from the networks generated by the ad hoc model (left) and CAM (right) compared to the real SDH systems. Remarkably, both models reproduce the distribution observed in real networks. Moreover, the models reproduce the small-world properties of the real system. The ad hoc model generates networks with values of the mean clustering $h_{Ci} = 0.08$ and the average path length $h_{li} = 10.4$ similar to the ones obtained empirically in the entire SDH network (0.004 and 11.3 respectively). Values of $h_{Ci} = 0.07$ and $h_{li} = 4.3$ obtained by the CAM are close to the ones obtained from the province of Sevilla (0.07 and 6.2 respectively).

The complexity found in the Spanish SDH network depends on a strict planning, but also on factors such as user demand, costs associated to the connections of a new equipment and complex intrinsic constrictions. The fact that the ad hoc model reproduces the empirically observed properties of the real SDH network evidences the possibility to generate an algorithm that captures the events associated to the network evolution and predicts the future of the system. However, all these events can be "compressed" into the compatibility concept. Our results suggest that compatibility between two SDH equipments is sufficient to generate the complex topology observed.

Acknowledgments

This work was supported by the Spanish Ministry of Science, under Projects No. MTM2009-14621 and 'Ingenio Mathematica (i-MATH)' No. CSD2006-00032, and Telefonica I+D Spain.

References

- [1] A.-L. Barabasi, Science 325, 412-413 (2009).
- [2] D. Watts and S. Strogatz, Nature, 393, 440-442 (1998).
- [3] H.G. Perros, Connection-oriented networks : SONET/SDH, ATM, MPLS and optical networks, New York: Wiley (2005).
- [4] A. Santiago, J.P. Cardenas, M.L. Mouronte, V. Feliu and R.M. Benito, International Journal of Modern Physics C, 19, 1809-1820 (2008).
- [5] J.P. Cardenas, M.L. Mouronte, R.M. Benito and J.C. Losada, Physica A, 389, 1789-1798 (2010)

Los documentos no son demasiado extensos (al tratarse de *extended abstracts*), pero son lo suficientemente significativos para considerarse como muestras de análisis ya que mezclan el lenguaje verbal con el simbólico numérico matemático y están aplicando la teoría de las redes complejas sobre otra área del saber.

El documento más extenso cuenta con un total de 7022 palabras y el más corto consta de 557 palabras diferentes. Son solamente cinco los textos que cuentan con menos de mil palabras en total. En cualquier caso, son muy pocos los documentos del corpus que tienen menos de mil palabras diferenciadas y todos los documentos cuentan con una alta densidad de vocabulario (es decir, en todos ellos se utiliza una gran variedad de palabras diferentes).

4. Limpieza y etiquetado

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. De hecho, al consultar el corpus en Voyant sin procesar (<https://voyant-tools.org/?corpus=717e576737b1ba0d82d9b9cb197af784>) se puede ver cómo difieren las palabras clave que se detectan automáticamente de las que efectivamente son las palabras clave (que se exponen al final de esta segunda parte de la tesis doctoral). Lo primero que salta a la vista es que Voyant considera “div” como una palabra o término, y no lo es. El elemento div es un contenedor genérico o una etiqueta de hipertexto que se utiliza en lenguajes como html sin un significado semántico en particular pero que permite crear secciones o agrupar contenidos para que los textos, después de procesados, tengan una determinada apariencia. Se utiliza comúnmente en el desarrollo de documentos con propósitos estilísticos, en conjunto con los atributos *style* y *class* (cuando forma parte del hipertexto). Si se sigue el enlace correspondiente se puede ver cómo *Voyant* identifica estos elementos como términos dentro del propio corpus. También se convierten en palabras o términos otro tipo de elementos matemáticos como, por ejemplo, la variable x , la variable y o la variable j , entre otros símbolos matemáticos que aparecen seguidos o precedidos de otros signos de puntuación, como pueden ser los dos puntos, el punto, la coma o el punto y coma. *Voyant* permite añadir y quitar a mano lo que se considera palabra y lo que no a través

de una lista llamada “lista de palabras excluidas”, pero este proceso ha de hacerse a mano, y el sistema no lo guarda una vez se refresca la página:

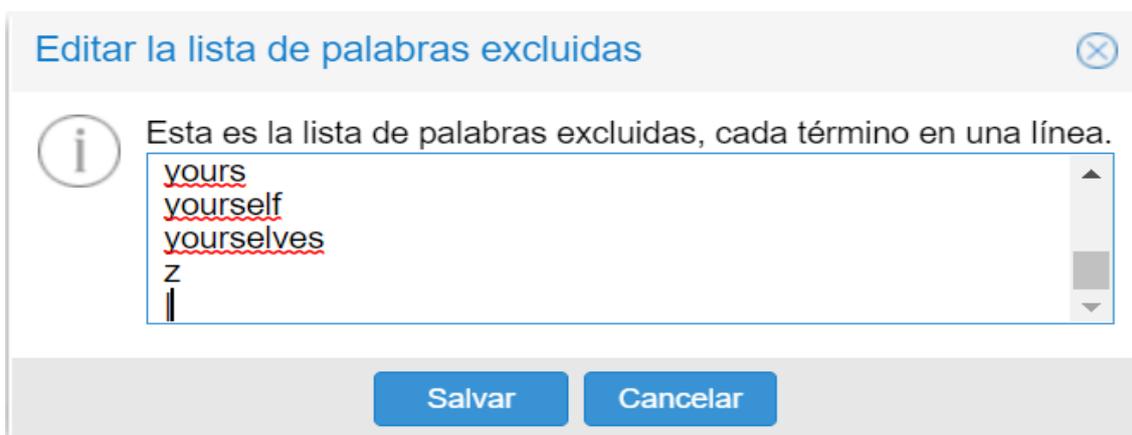


Figura 1. Imagen extraída de *Voyant* a modo de ejemplo con una lista de palabras que no forman parte del lenguaje de especialidad matemático.

En el proceso que se ha utilizado para analizar el lenguaje de especialidad matemático –que pasa por crear una red compleja multicapa, tal y como se describe en la parte tres de esta tesis– se muestra cómo *Voyant* no es una herramienta de análisis eficaz para este lenguaje concreto. Es importante resaltar que se han realizado dos procesos de etiquetado del corpus: uno clásico, para poder obtener resultados con herramientas de análisis ya existentes; y, posteriormente, tal y como se describe en la parte tres de esta tesis doctoral, se ha realizado un segundo proceso en el que se han dividido los términos en capas para poder crear una nueva herramienta de análisis. En un primer momento, centrándonos en emplear las herramientas tradicionales de análisis, se han eliminado o no se han considerado palabras los elementos del lenguaje artificial contemplados en las distintas fórmulas matemáticas presentes en los *extended abstracts*. Posteriormente, tras haber eliminado todos los elementos que no pueden considerarse palabras, hemos obtenido un corpus de 147.637 palabras y 25.210 oraciones (todas aquellas palabras que aparecen entre punto y punto).

Si tomamos en cuenta el ejemplo de *extended abstract* incluido en el apartado anterior, vemos cómo durante el proceso de limpieza se han tenido que eliminar no sólo aquellos términos relativos al lenguaje simbólico-numérico matemático, sino también los contenidos relativos a los nombres de los autores, los símbolos incluidos en las referencias y otros elementos como las fechas. Además, con vistas al desarrollo de la herramienta de análisis, se han clasificado las palabras una a una para incluirlas en las

distintas capas de la red compleja multicapa que se presenta en la siguiente parte de esta tesis doctoral. Antes de clasificar los términos en capas, se ha hecho este primer cribado de los elementos simbólico-numéricos y los términos que tienen que ver con los autores de los distintos artículos.

En los primeros apartados de esta segunda parte se hablaba de que se han mejorado mucho los sistemas de búsqueda gracias a las herramientas y programas informáticos, y de que la lingüística de corpus y la lingüística computacional necesitan la una de la otra para poder extraer resultados. El proceso de limpieza de nuestro corpus objeto de estudio y los resultados que se han obtenido son una buena muestra de cómo los avances en la informática permiten estudiar el lenguaje en profundidad. De hecho, se van a utilizar tres herramientas de análisis textual: *Wconcord*, *Flair* y *Voyant*, todos ellos descritos exhaustivamente en el siguiente apartado, en el que se podrá comprobar que dichas herramientas ofrecen resultados complementarios entre ellas.

La primera dificultad que hubo que afrontar para elaborar la presente tesis fue la de efectivamente seleccionar un área de conocimiento lo suficientemente específica como para poder extraer conclusiones relevantes porque, como hemos señalado, las matemáticas se aplican a prácticamente todas las áreas del saber y, por consiguiente, hablamos de un lenguaje de especialidad que va a aparecer siempre en otros contextos de especialidad. La dificultad de separar los términos del área de especialidad a la que se ha aplicado la teoría de las redes complejas, de los términos que efectivamente pertenecen al lenguaje matemático (de especialidad) de las redes complejas fue, en parte, lo que motivó a crear una nueva herramienta de análisis.

El corpus ha seguido distintos procesos. Por un lado, en la primera fase de análisis con herramientas tradicionales, se han considerado todos los términos del documento y se ha observado cómo interactúan entre sí para extraer conclusiones. Por otro lado, en una segunda etapa (tal y como se expondrá en la tercera parte práctica de esta tesis¹⁰), se ha hecho una catalogación posterior o división de los términos en capas lingüísticas de tal manera que se pudiera crear una red compleja multicapa dirigida

¹⁰ Durante la segunda fase de análisis, el corpus se procesa con Python 3.7 y se divide en cuatro capas diferentes. Este procesamiento del corpus forma parte de la creación de la herramienta que se presenta en la segunda parte práctica, por lo que se desarrolla con exhaustividad en ella.

(Criado-Alonso, A. et al. 2020, y Criado-Alonso, A. et al. 2021). Sobre estos aspectos se darán más detalles más adelante.

Por consiguiente, en esta primera fase de análisis, se han utilizado las herramientas de análisis conocidas como *Wconcord*, *Flair* y *Voyant*. Para la primera presentación de resultados, el corpus no se ha subdividido aún en capas y se ha trabajado como un conjunto (léxico) único. De esta manera, y tras una primera limpieza de términos y después de haber eliminado todos los elementos que no pueden considerarse palabras, nuestro corpus está formado por 147.637 palabras y 25.210 oraciones, escrito en inglés por hablantes nativos y no nativos, y que está compuesto por los *extended abstracts* que tratan de la especialidad de la teoría de redes complejas y que se publicaron en la *International Journal on Complex Systems* entre los años 2011 al 2016 (ambos incluidos).

Normalmente, las herramientas de procesamiento textual cuentan con una lista de palabras llamadas *stopwords* que ayudan a determinar el lenguaje de especialidad presente en los distintos documentos. En el caso de este corpus en concreto, ha habido que meter a mano muchos términos tales como las variables matemáticas y otros elementos como div o p que tienen que ver con el formato visual inherente a los documentos de los que se compone el corpus y que no forman parte ni del lenguaje de especialidad matemático ni de la disciplina a la que este se aplica ni del lenguaje vehicular (en este caso, inglés).

Existe una gran variedad de herramientas de procesamiento textual, la mayoría de ellas se presentan en esta web: <https://corpus-analysis.com/>, muchas de ellas son gratuitas y presentan resultados similares. Para un primer análisis y extracción de resultados del lenguaje matemático de especialidad sobre la teoría de redes complejas se han seleccionado *Wconcord*, *Flair* y *Voyant*. En los casos de las herramientas de *Flair* y *Voyant*, se puede acceder a ellas a través de la web (no es necesario descargarse ningún programa), pero *Wconcord* sí que requiere descarga.

2.4 Estudio tradicional del corpus mediante las herramientas *WConcord*, *Flair* y *Voyant*. Primeros resultados del estudio

Para justificar la mejora que supone el uso de la nueva metodología que se propone en esta tesis, así como para fundamentar las conclusiones de este estudio, nos apoyaremos en los resultados obtenidos a partir del estudio de nuestro corpus mediante tres conocidas herramientas de análisis textual ya existentes: *Wconcord*, *Flair* y *Voyant*. Más adelante podremos comparar los resultados de estos análisis con los resultados que obtendremos a partir de nuestro propio sistema desarrollado sobre el *Enriched Line Graph* (cuya base está en las redes complejas; véase Criado-Alonso 2021).

Se suele hacer una distinción entre dos tipos generales de análisis del corpus: el análisis cualitativo, en el que se hace una descripción detallada y completa de un fenómeno lingüístico o del comportamiento de una palabra o grupo de palabras, y el análisis cuantitativo, en el que se asignan índices de frecuencia a los fenómenos lingüísticos observados en el corpus y que pueden servir para construir modelos estadísticos más complejos que expliquen la evidencia hallada en el texto.

Estos dos tipos de análisis no deben considerarse excluyentes, sino más bien complementarios, ya que el análisis cualitativo, por un lado, ofrece una gran riqueza y precisión en las observaciones realizadas; así, por ejemplo, los fenómenos poco frecuentes pueden recibir igual atención que los muy frecuentes. Por otro lado, el análisis cuantitativo puede ofrecer información que sea estadísticamente significativa y resultados que pueden considerarse generalizables (McEnery y Wilson, 1996), por lo que es hoy muy frecuente que se combinen ambos tipos de análisis. La mayoría de los paquetes informáticos que se han desarrollado en los últimos años ofrecen la posibilidad de llevar a cabo ambos tipos de análisis, y en este sentido se han hecho enormes progresos y han aparecido diversas herramientas y publicaciones a este respecto.

Wconcord, *Flair* y *Voyant* son buenas herramientas de análisis textual y sus respectivos interfaces de análisis de corpus ofrecen parámetros y resultados que son complementarios entre sí. Como decíamos, la investigación que se presenta en esta segunda parte también servirá para apreciar las novedades que ofrece la nueva herramienta de análisis (fundamentada en la teoría de redes complejas) y su perspectiva novedosa en comparación con las herramientas anteriores.

En principio, y dado que nos enfrentamos al procesamiento textual a través de la programación en todas las herramientas tradicionales de análisis descritas en este estudio (*Wconcord*, *Flair* y *Voyant*), cabe concebir la hipótesis de que los resultados obtenidos en todas las herramientas deberían ser análogos, según autores como Vallez (2007). Cada una de estas herramientas de procesamiento textual está concebida con vistas a alcanzar un objetivo de búsqueda diferente: *Wconcord* es la herramienta más básica y es capaz de hacer un recuento de palabras y de poner palabras en contexto, cosa realmente útil para traductores y lingüistas; *Flair* ofrece resultados más orientados al tipo de oración ya que sus resultados son mucho más afinados, como veremos a continuación; y *Voyant*, que es realmente útil para analizar un corpus e incluso representar gráficamente qué término se une con otro, entre otras opciones. En las siguientes líneas describiré pormenorizadamente cada una de estas herramientas y los resultados principales que ofrecen a partir de nuestro corpus.

2.4.1 *Wconcord*, *Flair* y *Voyant*

a) *Wconcord*

Wconcord es una herramienta 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).

El análisis que ofrece *Wconcord* nos devuelve un resultado sobre la totalidad de los textos. De todas las herramientas que van a ser analizadas, esta es seguramente la más básica, pero la única que ofrece resultados de búsqueda que no han sido procesados lingüísticamente. Este motivo hace de esta herramienta la ideal para extraer resultados que se refieren a los términos del documento, independientemente de la especialidad a la que este lenguaje pertenezca, ya que este programa es incapaz de diagnosticar por sí mismo los términos clave del lenguaje de especialidad del corpus analizado. No hay que olvidar que *Wconcord* es un software de 1998 y los patrones de reconocimiento de similitud de palabras, colocaciones y cercanía entre palabras hacen que las herramientas más actuales desprecien este primer análisis que, en un principio,

parece poco aplicable a la extracción de conclusiones. Sin embargo, ayuda a ver cómo interactúan los términos de especialidad con el lenguaje común.

A favor de esta herramienta hay que destacar también que su interfaz de usuario es muy intuitiva. Además, *Wconcord* trabaja con texto plano, por lo que los términos que se extraigan de las distintas búsquedas que haga el usuario se pueden utilizar más tarde en memorias de traducción. En el caso de *Wconcord*, cuando se procesa todo el corpus, nos dice que la palabra más frecuente es *the*, seguida de *of*, *a*, *an* (en ese orden). En otras palabras, nos sirve para ver qué hay cerca de cada término, pero, de nuevo, *Wconcord* no es capaz de determinar los pares de palabras que han de aparecer juntos, ni es capaz de determinar cuáles son las palabras que van más allá del lenguaje común. Por el contrario, las otras dos herramientas que presentamos a continuación (*Flair* y *Voyant*), sí que cuentan con esa lista de palabras de lenguaje común y, con mayor o menor acierto, saben determinar las palabras clave de un texto ya que tienen incorporada de manera automática una lista de palabras frecuentes que el sistema determina automáticamente que no pueden ser palabras clave (un ejemplo de este tipo de palabras que vienen incluidas entre las que no pueden ser consideradas como palabras clave lo constituye la lista de pronombres personales o los determinantes artículo).

b) *Flair*

Esta herramienta es, en realidad, una interfaz web que permite realizar un análisis textual desde un punto de vista lingüístico y sintáctico. El usuario de esta herramienta debe subir a la página web <http://sifnos.sfs.uni-tuebingen.de/FLAIR/> los textos que desee analizar. La interfaz es muy intuitiva y sencilla de manejar. *Flair* devuelve, como resultado, el nivel de complejidad de los enunciados, pudiendo, incluso, analizar los textos más cortos, obteniéndose resultados que permiten ya una primera extracción de conclusiones respecto al tipo de expresiones que se utilizan en el corpus analizado. Además, *Flair* permite clasificar los textos analizados de acuerdo con el Marco Común Europeo de Referencia de las Lenguas (publicado por el Consejo Europeo en el año 2001). Esto implica que, entre otras cosas, si se desea imitar el estilo de los autores, se deberán utilizar enunciados sencillos.

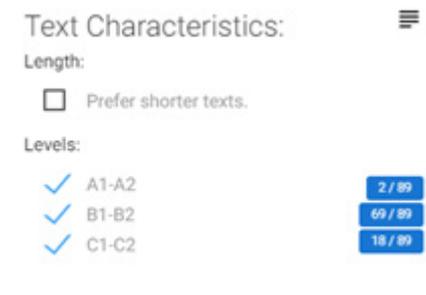


Figura 2: Imagen de la interfaz de Flair agrupando los 89 documentos de los que se compone el corpus objeto de análisis según su nivel lingüístico determinado por el MCERL (Marco Común Europeo de Referencia de las Lenguas).
Fuente: <http://sifnos.sfs.uni-tuebingen.de/FLAIR/>

La imagen anterior muestra cómo Flair puede, de manera automática, extraer información mucho más elaborada sobre los textos de los que está compuesto el corpus. En concreto, como hemos dicho, este sistema permite diagnosticar el nivel de complejidad de los enunciados según el Marco Común Europeo de Referencia. Según esta herramienta de análisis, una de las características del lenguaje matemático verbalizado empleado en nuestro corpus objeto de análisis es precisamente que los enunciados son más simples que los que emitiría un hablante nativo de habla inglesa cuando utiliza el lenguaje común, pero con un vocabulario mucho más específico. De hecho, la mayoría de los textos del corpus analizado presenta una complejidad a medio camino entre el nivel B1 y B2. Más allá del nivel de idioma empleado por los autores (que, como ya hemos dicho, no tienen por qué ser nativos), Flair nos ofrece información sobre el tipo de construcciones gramaticales presentes en los textos del corpus, así como del tipo de frases utilizadas por los autores (afirmativas, negativas, interrogativas, etc.). Este tipo de información es muy útil en el mundo de la traducción para saber de qué manera se ha de redactar un texto meta, pues no ha de perderse de vista que, en lenguaje de especialidad, una buena traducción tratará de emular en el lector meta lo mismo que percibió el lector que recibió el texto escrito en la lengua original.

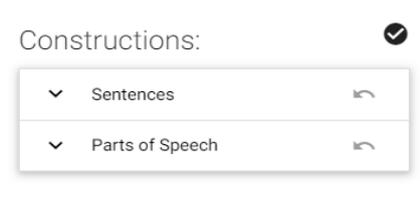


Figura 3: Imagen de la interfaz de Flair donde se aprecia que se pueden agrupar los textos del corpus según el tipo de construcciones lingüísticas presentes en cada uno de ellos. Fuente: <http://sifnos.sfs.uni-tuebingen.de/FLAIR/>

En la figura 4, se puede apreciar la cantidad de pestañas diferentes que ofrece Flair, todas ellas destinadas a obtener información adicional muy variada sobre los textos del corpus que Flair nos brinda una vez que se despliegan las pequeñas secciones de *sentences* y *parts of speech*. Es importante destacar que Flair es una herramienta muy útil a la hora de extraer conclusiones formales sobre los patrones que se deben seguir en el nivel fraseológico (en el apartado *Sentences*) y el tipo de sintagmas utilizado a lo largo del corpus (en el apartado *Parts of Speech*), según se ilustra en las imágenes adjuntas.

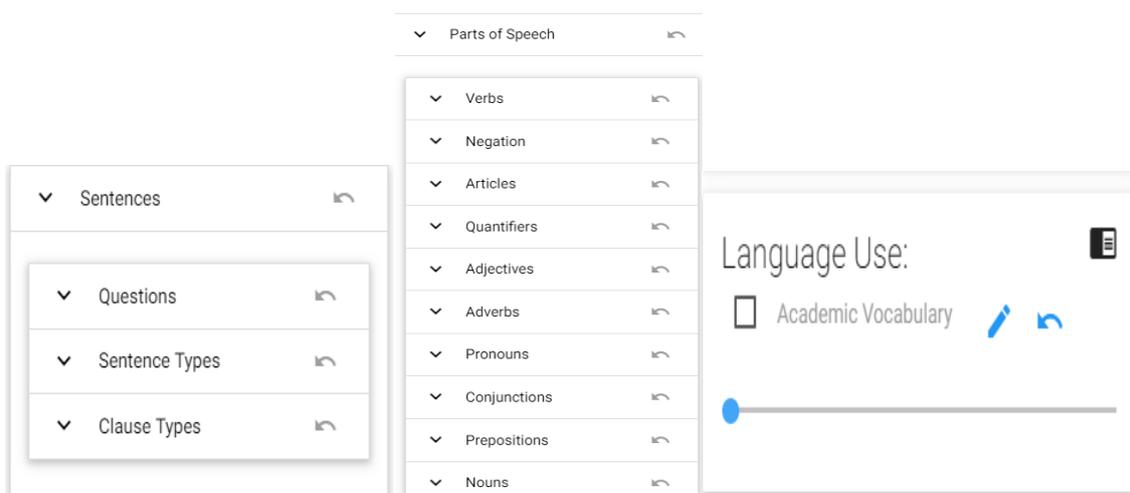


Figura 4: Muestra de la perspectiva de análisis que ofrece Flair. Fuente de las imágenes: <http://sifnos.sfs.uni-tuebingen.de/FLAIR/>

La imagen de la figura 4 ilustra que Flair puede dar pistas, tal y como se desarrollará más adelante de manera específica, sobre el estilo que deberá imitar el traductor si quiere seguir la línea de los autores de este lenguaje de especialidad en todo lo que se refiere a qué tipo de frase utilizar (interrogativa, afirmativa, subordinada, coordinada, etc.); también se puede apreciar el tipo de construcciones que utilizan los autores en el uso de los artículos, así como las conjunciones, las preposiciones, los

pronombres y, en definitiva, todos los apartados que se pueden ver en la imagen. Además, tal y como se aprecia en la imagen de la figura 4, Flair también cuenta con la capacidad de determinar si se utiliza o no el vocabulario académico en los textos que el usuario ha analizado. Para un traductor, estas son características muy importantes de cara a redactar un texto meta.

Una buena traducción de Lenguaje de Especialidad es aquella que se encuentra próxima al original en términos de sintaxis y significado, aquella en la que el traductor es invisible. Autores como Levý (2011) afirman que el lector meta debería creer que está leyendo al autor original de la misma forma que los lectores de la lengua de partida leerían a dicho autor, tal y como da a entender a lo largo de toda su obra publicada en 2011. Sin Flair, sería mucho más complicado llegar a conclusiones a nivel sintáctico, circunstancia que dejaría al traductor trabajando únicamente con el significado, el significante y el referente. Por ejemplo, si un traductor entiende una unidad de significado específico como puede ser *complex network* y lo traduce como red compleja, pero engloba este concepto en una frase muy larga o enrevesada, el documento meta no quedaría acorde a los estándares del lenguaje de especialidad matemático ya que normalmente utiliza enunciados más sencillos.

c) *Voyant*

La tercera herramienta de análisis empleada es *Voyant* (www.voyant-tools.org; se trata de un software de 2016 con código abierto publicado en GitHub <https://github.com/sgsinclair/Voyant>), cuya interfaz de usuario es también muy amigable. Los resultados obtenidos (parámetros y gráficas) pueden variar en función de los criterios y variables de búsqueda que introduzca el usuario.

2.5 Palabras clave según las herramientas tradicionales

En el mundo de la traducción, es fundamental tener una lista de palabras o términos clave. Normalmente se espera que el traductor domine la materia de los textos que va a traducir, pero a veces esto no sucede cuando el traductor se enfrenta a textos técnicos. Se habla mucho de traducción automática y de memorias de traducción para ayudar a los traductores en sus tareas, sobre todo cuando no son expertos en la materia. Una memoria de traducción ayuda al traductor a saber, por ejemplo, cuándo dos términos forman una única unidad de conocimiento específico o cuándo se ha de utilizar una expresión concreta (Candel-Mora, 2011). Un sistema de memoria de traducción es un producto de software diseñado para acelerar el proceso de traducción de un texto incorporando, entre otras, las funciones de almacenar y recuperar unidades de traducción anteriores (o segmentos paralelos) y términos equivalentes, garantizando la coherencia y la calidad de las traducciones, junto con un aumento de la productividad. Es algo parecido a traducir basándose en la experiencia; por eso, utilizar la lingüística de corpus a modo de herramienta de análisis de documentación paralela es un recurso muy popular entre los traductores.

Como decíamos, *Flair*, *Voyant* y *Wconcord* ofrecen resultados complementarios entre sí de manera que, si no se cuentan con traducciones previas o memorias de traducción, el traductor puede identificar el estilo que tiene que imitar y las palabras clave que tiene que saber para buscar las unidades de conocimiento específico dentro del texto de origen de manera que consiga redactar un texto meta sin falsos sentidos. Identificar las palabras clave en un texto es algo básico para el traductor o para una persona que no es experta en la materia y que quiere entender qué está escrito en un documento. Tal y como se decía en la primera parte de esta tesis, a veces, la dificultad radica en identificar qué elementos constituyen una unidad de conocimiento específico. Esto puede observarse viendo qué términos aparecen juntos y con qué frecuencia, y viendo cómo se relacionan estos conjuntos de términos entre ellos (de hecho, tal y como veremos en la parte tres de esta tesis doctoral, el lenguaje matemático de especialidad de la teoría de redes complejas se caracteriza sobre todo a través de las unidades de significado compuestas por dos términos).

En el supuesto de que no se tenga una memoria de traducción previa, ni ninguna otra directriz sobre cómo redactar, el primer paso es identificar las palabras clave de un texto para poder ver las unidades de conocimiento específico que se enarbolan en torno a ellas. Las pistas sobre cómo redactar se pueden extraer a través de herramientas como Flair, o también a través de textos paralelos. Tratar de imitar el estilo de los textos de origen es una buena idea, pero esto no siempre se puede hacer en la lengua meta. Para identificar las palabras clave, como se ha dicho, este lenguaje de especialidad concreto plantea una dificultad añadida porque es un lenguaje de especialidad que se suele entremezclar con el de otra disciplina (a la que se esté aplicando la teoría de las redes complejas), por lo que las palabras que aparecen en Voyant y en Flair como *Academic Language*, no van a pertenecer exclusivamente a esta área.

Normalmente, la lógica dicta que la palabra clave tendrá que ver con la palabra más frecuente, pero esto no siempre es así ya que la palabra más frecuente en inglés es *the* y, las herramientas de análisis textual más básicas como *Wconcord* únicamente son capaces de alinear textos, ver qué término va seguido de qué otro término y hacer recuento de palabras. De cualquier modo, estas herramientas básicas sirven para ver propiedades del lenguaje más difíciles de observar a partir de otras herramientas más desarrolladas (por ejemplo, a través de *Wconcord*, se puede ver con más facilidad cómo interactúan los términos del lenguaje vehicular con los términos o palabras de especialidad).

Para el caso concreto de este lenguaje de especialidad, donde conviven cuatro subtipos de lenguaje (lengua vehicular inglés, lenguaje simbólico-numérico matemático, lenguaje de especialidad relativo a la teoría de redes complejas y lenguaje de especialidad de la disciplina a la que se está aplicando la teoría de las redes complejas), conviene aplicar una técnica ligeramente distinta a las habituales para identificar una lista de palabras clave que sea de términos que pertenezcan exclusivamente al lenguaje matemático de especialidad (que es, al fin y al cabo, el objeto de estudio de esta tesis doctoral).

Con esta finalidad, se ha utilizado una combinación de *Wconcord* y *Voyant* teniendo en cuenta una búsqueda a partir del término *keywords* (que son las palabras clave que consideran los propios autores de los artículos). Una vez identificadas las palabras clave, se hace una propuesta de traducción de unidades de conocimiento específico en torno a ellas teniendo también en cuenta las pautas de traducción deducidas a partir del análisis del corpus obtenido desde Flair antes de presentar las conclusiones parciales de este lenguaje de especialidad y pasar a describir la nueva herramienta de traducción desarrollada en la segunda parte de esta tesis doctoral.

2.5.1 *Voyant* y *Wconcord* para obtener las palabras clave

Los resultados que muestra *Wconcord* en lo que se refiere a la tabla de frecuencias (o el número de veces que una palabra se repite) son los siguientes: *the*, *of*, *and*, *a*, *in* y *to*.

```
1. the|9707|
2. of|5609|
3. and|3904|
4. a|3787|
5. in|3181|
6. to|2639|
```

Figura 6: Imagen extraída de *Wconcord* tras procesar el corpus. Fuente: elaboración propia a través de *Wconcord*

Este resultado, tal y como señalamos antes, obedece exclusivamente al número de veces que se repite un término. No podemos considerar estos términos como palabras-clave del lenguaje matemático relativo a la teoría de grafos ya que son términos típicos del lenguaje común. De hecho, estos términos ya han sido catalogados como los más frecuentes en lengua inglesa por diversos autores, entre ellos Zipf (Zipf, 1965).

Como se puede observar, la relación anterior se ajusta bastante bien a lo establecido por la ley de Zipf (1965)¹¹ para la frecuencia de aparición de las palabras en

¹¹ A tenor de la ley de Zipf veríamos cómo el término o palabra más frecuente se repetirá alrededor del doble de veces que el segundo término más frecuente, tres veces más que el tercer término más frecuente y así sucesivamente.

los textos escritos en inglés. Esta ley es coherente con las palabras que aparecen en los textos del corpus con mayor frecuencia que, según *Wconcord*, son: *the, of, and, a, in, to*.

Según estos resultados, y en lo que se refiere a la ley de Zipf, el lenguaje de especialidad matemático se comporta de manera casi análoga al lenguaje común, por lo que es complicado establecer cuáles son los términos clave (*keywords*) que caracterizan a este Lenguaje de Especialidad (LE) en concreto, ya que los términos que aparecen con mayor frecuencia son términos del lenguaje común. Así pues, las afirmaciones de Bergenholtz (1995) que describimos en apartados anteriores en relación con el lenguaje de especialidad y cómo este interactúa con el lenguaje vehicular cobran todo el sentido ya que, pese a tratarse de un corpus de Lenguaje de Especialidad, las palabras más frecuentes siguen siendo las del lenguaje común (en este caso inglés). Los términos clave del lenguaje de especialidad, entonces, no tienen que ver con la palabra que más se repite en los textos; de hecho, volviendo a la primera parte del proyecto de tesis, se ve que, en realidad, el lenguaje de especialidad se integra dentro del lenguaje vehicular, pero con una terminología específica y tomando prestados muchos otros términos del lenguaje común de manera que estos aparecen terminologizados o con su significado alterado.

Un ejemplo notable de la diferencia en cuanto al proceso de análisis que realizan *Wconcord* y *Voyant* se ejemplifica mediante la búsqueda de las palabras *network* y *networks* diagnosticadas a lo largo de este estudio como palabras clave tanto por los expertos como por Flair que no resultan ser significativas según *Wconcord*. Esta última herramienta de análisis nos presenta la siguiente lista parcial en la que el primer dígito indica el orden relativo del término según su frecuencia de aparición:

11.	network	1093
12.	with	979
13.	by	907
14.	this	901
15.	are	897
16.	n	891
17.	as	840
18.	m	807
19.	j	797
20.	networks	793

Figura 7: Imagen de las palabras clave una vez procesado el corpus con *Wconcord*. Fuente: elaboración propia a través de *Wconcord*

Aquí vemos como *Wconcord* identifica como palabra cualquier término que se encuentre entre dos espacios, incluidas las variables matemáticas n (línea 16), m (línea 18) o j (línea 19) junto al número de veces (o frecuencia) con la que esa palabra se repite a lo largo del corpus. No muestra si hay relación o cómo se relacionan los términos *network* y *networks*.

Voyant, por su parte, emplea una estrategia capaz de diagnosticar los términos clave de especialidad más repetidos a lo largo del corpus y, además, muestra la gráfica con la tendencia (o frecuencia de aparición en los textos del corpus) que sigue el término (tanto en cada texto del corpus como en el total de textos). Así, en Voyant, *Network* y *Networks* sí que figuran como palabras clave. Estas dos palabras se repiten y aparecen en todos los documentos del corpus.

	Términos	Contar	Tendencia
1	network	959	
2	networks	693	
3	nodes	444	
4	systems	373	
5	complex	364	
6	model	356	

Figura 8: Imagen del Corpus de análisis procesado con *Voyant Tools* donde se muestran los que son los seis términos principales según esta herramienta de análisis Fuente: Retrieved June 21, 2019 from <http://voyant-tools.org>

Concretamente, según Voyant, los términos (o palabras) clave del corpus son: *network, networks, nodes, systems, complex y model*.

Para obtener estas palabras clave, la estrategia de Voyant se apoya en una lista de términos o palabras prohibidas *stopwords* automáticamente excluidas de las susceptibles de ser consideradas como palabras clave, junto a la aplicación de una fórmula TF-IDF (*Term Frequency-Inverse Document Frequency*) para cada término dentro de cada documento.

Voyant cuenta con la ventaja extra de mostrar cierta relación o cercanía entre *network* y *networks* (términos diagnosticados como *términos clave* tras la aplicación de las estrategias anteriores).

Es importante resaltar que, hoy en día, se ofrecen herramientas capaces de procesar un corpus tanto cualitativa como cuantitativamente y que las distintas herramientas ofrecen resultados complementarios entre ellas (Llamazares, 2008). Esta

es la motivación que ha llevado a extraer conclusiones sobre este tipo de lenguaje combinando varias herramientas de análisis tradicionales. No hay que perder de vista que todas estas herramientas trabajan sobre un listado de *stopwords* para detectar las palabras más frecuentes –en los casos de Voyant y Flair–.

Nuestro caso, que plantea analizar el lenguaje de especialidad matemático de la teoría de redes complejas, se topa con un corpus en el que están presentes cuatro tipos distintos de lenguaje (el lenguaje vehicular, el lenguaje de la especialidad a la que se aplica la teoría de las redes complejas, el simbólico-numérico matemático y el lenguaje de especialidad de las redes complejas como tal). Por ello, las listas de palabras clave ofrecidas por las herramientas tradicionales necesitaban un nuevo filtro, hecho que motivó en parte la creación de la nueva herramienta de análisis que se presenta en la segunda parte del proyecto de tesis.

Wconcord (la herramienta más básica), al no aplicar ningún tipo de filtro lingüístico, no distingue entre términos específicos del lenguaje de especialidad y términos del lenguaje común, por lo que no tiene sentido considerar las palabras diagnosticadas como más frecuentes por esta herramienta como términos clave. FLAIR por su parte no diagnostica las palabras clave (su utilidad principal es la de detectar patrones y analizar sintácticamente los documentos); Voyant, sin embargo, diagnostica una serie de palabras clave en base al lenguaje de especialidad y al número de veces que estos términos se repiten.

La imagen del apartado anterior replicada abajo nos habla de las siguientes palabras clave, sin diferenciar entre mayúsculas y minúsculas: *network*, *networks*, *nodes*, *systems*, *complex* y *model*. Vemos cómo la interfaz muestra un resultado que diferencia entre singular y plural.

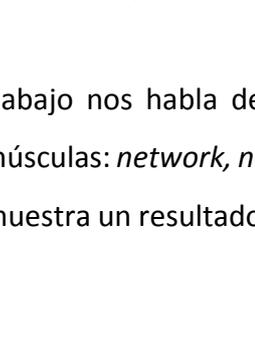
	Términos	Contar	Tendencia
1	network	959	
2	networks	693	
3	nodes	444	
4	systems	373	
5	complex	364	
6	model	356	

Figura 9. Imagen del Corpus de análisis procesado con *Voyant Tools* donde se muestran los que son los seis términos principales según esta herramienta de análisis. Fuente: Retrieved June 21, 2019 from <http://voyant-tools.org>

Como se puede apreciar, el tipo de resultados gráficos obtenidos por Voyant es muy diferente del tipo de resultados obtenidos a través de la interfaz que nos ofrece *Wconcord*. Voyant está programado de forma que representa una red compleja con las palabras clave localizadas de manera automática. Sin embargo, las relaciones establecidas entre los términos contemplan el corpus en la totalidad, en lugar de contemplar la relación que se establece dentro de un solo documento concreto. Así, los términos representados por un nodo con fondo azul se refieren a palabras clave del lenguaje de especialidad, y los que tienen fondo naranja son palabras cercanas por colocación a las palabras clave dentro de los textos, pero no son palabras clave en sí mismas. El tamaño de los nodos de las palabras clave de esta red compleja es proporcional al número de documentos del corpus en los aparece dicha palabra, y el tamaño de las conexiones (“links” o aristas) entre estos términos es proporcional al número de documentos en los que dichos términos son palabras cercanas a las palabras clave por colocación. Obsérvese que no hay conexiones (*links*) entre los nodos con fondo de color naranja.

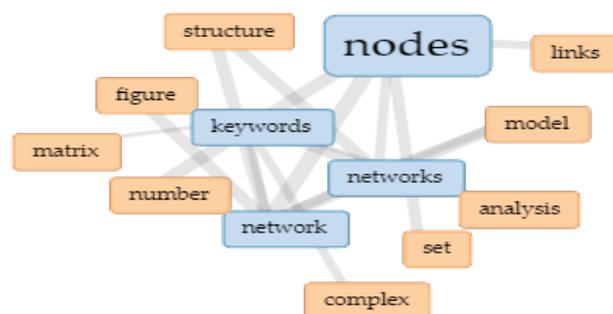


Figura 10. Términos clave de un texto identificados por Voyant dispuestos en forma de grafo Fuente: Sinclair, S. & G. Rockwell. (2019). Términos. *Voyant Tools*. Retrieved June 21, 2019 from <http://voyant-tools.org>

Para obtener resultados similares en *Wconcord* (que sólo puede contar y alinear palabras y contextos sin distinguir ni elementos del lenguaje de especialidad ni palabras funcionales), consideraremos las *palabras clave* que resaltan los autores. En los *extended abstracts* son estas palabras clave las que nos hablan de la temática de la que trata el artículo. Hemos de ser conscientes, por tanto, de que obtendremos sobre todo palabras sustantivas, pero todas ellas dentro del campo de especialidad matemático relativo a la teoría de grafos y teoría de redes complejas. Todos los documentos tienen un apartado denominado *Keywords*, o “palabras clave” y son los propios autores los que

seleccionan los términos que consideran que van a dar las claves de significado al lector sobre la temática de su artículo.

Una vez realizada la búsqueda en *Wconcord*, y ejecutando dicha búsqueda a partir de los términos incluidos en cada documento del corpus dentro del apartado “*keywords*” –apartado que aparece en todos los documentos del corpus–, hemos adoptado en nuestra aproximación como palabras clave aquellos términos relacionados con la Teoría de Grafos que se repiten más de una vez (dentro de este apartado) despreciando los conectores y los determinantes que no son intrínsecamente propios del lenguaje de especialidad matemático¹² en concreto. Se han diagnosticado las siguientes palabras por ser las únicas que se repiten más de una vez en el listado de palabras clave que los autores proponen: *Networks*, *Complex*, *Cluster* (*Clustering*, *Clustered*) y *System* (*Systems*).

Término	Repeticiones	Complemento (S/N) (A/D)	Variaciones
<u>networks</u>	6	S A6	0
<u>Complex</u>	4	S A1 D3	0
<u>Cluster*</u>	3	S A3	2 <u>1.clustering</u> 1. <u>clustered</u>
<u>community</u>	2	N	0
<u>System</u>	3	S A3	2 1. <u>System</u> 2. <u>Systems</u>

Figura 11. Fuente: Elaboración propia a partir de los datos extraídos mediante *Wconcord* donde se muestra el número de veces que se repite una palabra, su tipo de complemento (si va antes o después del verbo principal) y las variaciones presentes sobre el lexema de los términos identificados como más repetidos.

En una primera aproximación, puede sorprender el número tan bajo de repeticiones de estas palabras a lo largo de todo el corpus, pero tiene sentido ya que, como señalamos en apartados anteriores, la teoría de grafos (o teoría de redes complejas) es aplicable a prácticamente todos los ámbitos de la ciencia. Estos textos tienen, por tanto, hasta cuatro tipos de lenguaje (catalogados como diferentes entre

¹² Estos términos aparecen a menudo y son muy importantes en la estructura del lenguaje y a la hora de transmitir el mensaje, pero con vistas a que este trabajo sea útil para el traductor, nos hemos centrado en palabras sustantivas, adjetivas y/o verbales para poder ofrecer soluciones viables a los problemas de traducción. Por ejemplo, a través de *clustered*, se puede llegar al verbo *clustered together* que es exclusivo de este lenguaje de especialidad y se traduce por agrupar. Este verbo es especial ya que se refiere a los nodos de una red compleja.

ellos) interactuando entre sí. Por un lado, tenemos el lenguaje común (inglés); en segundo lugar, nos aparece el lenguaje específico derivado de la verbalización de las distintas expresiones matemáticas; en tercer lugar, encontramos el lenguaje simbólico-gráfico-numérico matemático como tal y, por último, nos encontraremos con la terminología específica de la disciplina a la que se esté aplicando la Teoría de Grafos.

Flair, por su parte, no puede ser utilizada para este tipo de análisis ya que no devuelve resultados cuantitativos relacionados con este punto de vista. Únicamente obtenemos resultados sintáctico-lingüísticos, pues esta herramienta se centra en lo que denomina *términos de especialidad* o *vocabulario académico* como veremos más tarde.

Parece importante destacar que ambas herramientas de análisis coinciden en que nos permiten obtener qué términos hay alrededor de cualquier palabra que el usuario desee buscar, aunque en cualquier caso hay que resaltar que, tanto por el resto de las características de la herramienta como por la interfaz de usuario, Voyant es la herramienta más útil a la hora de definir las palabras clave del lenguaje de especialidad.

an image by using complex	network	analysis. We associate a weighted geometrical
geometrical and fast-computable complex	network	to each image that gives some
use local measures of the associated	network	and we discuss the use
and properties of the weighted geometrical	network	. 2. Interest points: A complex

Figura 12. Imagen de una tabla extraída de Voyant que refleja los términos que se muestran cercanos a *network*
 Fuente: Sinclair, S. & G. Rockwell. (2019). Términos. *Voyant Tools*. Retrieved June 21, 2019 from <http://voyant-tools.org>

Como ejemplo ilustrativo, la anterior es una imagen extraída a partir de la interfaz de *Voyant*, y la que aparece a continuación está extraída de los resultados obtenidos tras la ejecución de *Wconcord*:

1.	analyze them using complex	network	properties.	ij-css01_01-83.txt
2.	algorithm on a	network	with the aim detect the position	ij-css01_01-83.txt
3.	spatial and weighted complex	network		ij-css01_01-83.txt

Figura 13 Fuente: Elaboración propia a partir de *Wconcord* que muestra los términos que aparecen antes y después de *network* en un documento del corpus

Ambas herramientas nos permiten diagnosticar qué tipo de palabra es “*Network*” (sustantivo, adjetivo, verbo, ...) atendiendo a su situación dentro de la oración y a las palabras que rodean ese término. Normalmente este es el tipo de análisis (basado en la frecuencia de aparición) en los que se basan los lingüistas para establecer los marcos colocacionales o las unidades de conocimiento específico dentro del campo de la lexicografía computacional para la creación de memorias de traducción.

Faltaría hablar en este apartado, aunque sea de manera colateral, de la tercera herramienta de análisis contemplada: Flair. Hay que decir que Flair no nos permite establecer cuáles son las palabras clave. Por el contrario, utiliza lo que denomina *Academic Vocabulary*, obtenido a partir del pre-procesamiento de ciertas bases de datos ya programadas para este fin. Flair hace un análisis morfosintáctico y macrotectual de los textos del corpus (o de los textos que el usuario introduzca). Esta herramienta tiene características específicas que la distinguen de las dos antes descritas, aunque ciertamente permite obtener resultados que complementan tanto a *Wconcord* como a *Voyant*, ya que ofrece la opción de catalogar los textos ordenándolos de acuerdo con su grado de especialización (dado por el número de términos técnicos que aparecen en un texto). Flair arroja luz sobre otra serie de características morfosintácticas que se presentan a continuación.

Según Flair, el documento que tiene un registro con un número mayor de términos técnicos de especialidad es el documento etiquetado en nuestro corpus como *ij-css01_01-83* o *Interest Point Detection in Images by a Local Centrality Algorithm on Complex Networks*

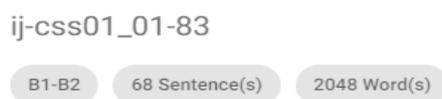
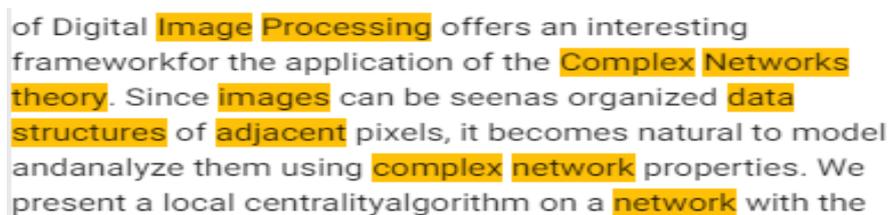


Figura 14. Imagen extraída de Flair en la que se identifican las características del texto identificado automáticamente como el más representativo del corpus objeto de análisis.

En una primera aproximación se puede comprobar el nivel de competencias lingüísticas desarrolladas a través de los textos del corpus analizado corresponde en su mayoría a un nivel B1-B2 (hablante intermedio) según el MCERL. El documento *ij-css01_01-83*, consta de 68 frases y 2048 palabras. Flair permite visualizar el texto

seleccionado resaltando los términos técnicos que ha diagnosticado como lenguaje de especialidad, pero, de nuevo, no nos habla de porcentajes o frecuencias.



of Digital **Image Processing** offers an interesting framework for the application of the **Complex Networks theory**. Since **images** can be seen as organized **data structures** of **adjacent** pixels, it becomes natural to model and analyze them using **complex network** properties. We present a local centrality algorithm on a **network** with the

Figura 15. Imagen obtenida de Flair con los términos clave resaltados del texto identificado automáticamente como el más característico del corpus objeto de análisis. Fuente: <http://sifnos.sfs.uni-tuebingen.de/FLAIR/>

Llama la atención el hecho de que sean únicamente 18 de 89 los artículos que tienen enunciados complejos equiparables a los que emitiría un hablante anglófono, es decir, con un nivel C2. En un principio podríamos decir que este lenguaje de especialidad matemático presente en este corpus se caracteriza por una sintaxis simple, con enunciados característicos de un hablante intermedio (B1-B2), pero este hecho tendrá que ser analizado más en profundidad a futuro ya que se trata de un corpus de textos redactados en inglés por una gran mayoría de hablantes no nativos.

El hecho de que haya en el corpus tantos documentos de nivel B1-B2 puede deberse a que, como hemos señalado, la mayoría de los autores no sean anglófonos o a que, como señalamos en apartados anteriores y como aparece en Criado-Alonso et al. 2020, en este lenguaje de especialidad se mezclan a la vez cuatro tipos de lenguaje (el de notación matemática o lenguaje matemático como tal, el lenguaje de especialidad matemático, el lenguaje natural y el lenguaje de especialidad de la disciplina sobre la que se está utilizando la teoría de redes) y, al estar reflejado el lenguaje de especialidad en el lenguaje simbólico-gráfico-numérico, el lenguaje de especialidad no se complica tanto como el lenguaje de especialidad de otras áreas que no tienen otro lenguaje extra como punto de apoyo. En otras palabras, quizá el nivel B2 de media sea una consecuencia directa del uso del lenguaje simbólico matemático en el texto, ya que el lenguaje matemático verbalizado es una descripción de este.

Como señalamos en los párrafos en los que se hablaba de *Wconcord*, llama la atención el hecho de que esta herramienta no distingue de manera automática cuáles términos son *términos de especialidad* y cuáles no. Muestra cómo los términos más

frecuentes no corresponden a terminología específica matemática (*the, of, and, a, in, to*). Estos resultados se ajustan a lo descrito por la ley de Zipf (1965) sobre el lenguaje natural. En relación con esta ley, es muy interesante ver cómo se relacionan las palabras del lenguaje natural entre sí. La ley de Zipf describe un patrón de repetición de palabra desde la más frecuente a la menos frecuente siguiendo una ley de potencias. Fue Zipf quien dijo que las palabras más frecuentes son las *palabras funcionales*, como ocurre en el caso de nuestro corpus con *the, of, and, a, in, to*.

Palabras como *the, of, and, a, in, to* cumplen una función más gramatical que léxica (de alguna manera, describen las palabras que están a su alrededor), cabrían dentro del concepto de palabras funcionales establecido a partir de la ley de Zipf.

Se puede afirmar, en cualquier caso, que el lenguaje matemático utiliza términos típicos del lenguaje natural como los más comunes: *the, of, and, a, in, to* que determinan las palabras con mayor carga léxica que los rodean por lo que la sintaxis o estructura de las frases es muy similar a la utilizada en el lenguaje común: [determinante+adjetivo (si lo hubiera)+nombre (palabra con carga de significado léxico)+verbo (palabra con significado léxico)+sintagma complemento (compuesto a su vez de preposiciones, adjetivos, sustantivos, verbos o incluso oraciones subordinadas)]. La estructura anterior es un esquema desarrollado sobre Koike (2001), donde se habla de *colocaciones léxicas*. En esta línea, haciéndonos cargo de la diferencia entre "significado léxico" y "significado gramatical", vemos que las palabras relativas a la Teoría de Redes Complejas (o grafos) deberían ser aquellas con mayor carga léxica dentro del corpus.

En el caso de las redes complejas, falta un tercer término que ayude a determinar el significante exacto. Por ejemplo, "redes complejas neuronales" es una colocación que tiene un referente concreto; sin embargo, una "red compleja" (sin determinar), puede aludir a casi cualquier cosa que se pueda representar gráficamente mediante el uso de vértices, nodos y aristas. Este hecho pone aún más de manifiesto que, en estos textos, nos enfrentamos a un total de cuatro lenguajes de especialidad: el lenguaje artificial, el lenguaje de especialidad exclusivo de la teoría de redes complejas, el lenguaje de la especialidad a la que se está aplicando la teoría de redes complejas y, por último, el lenguaje vehicular.

Además de las características sintácticas que acabamos de describir, ya hemos visto, gracias a FLAIR, que la mayor parte de los textos matemáticos del corpus analizado

correspondiente a esta área de especialidad muestran enunciados simples característicos de hablantes de la lengua con un nivel B1-B2 por lo que, en principio, la redacción de este lenguaje no debería ser complicada. El problema fundamental para los traductores es la falta de especialización y la no comprensión del lenguaje matemático expresado en forma de fórmulas.

En resumen, para poder comprender el significado del texto de origen, se han de buscar las unidades de conocimiento específico que hay en él y, para eso, identificar las palabras clave es el primer paso. Al estar mezclados los cuatro lenguajes de especialidad, no podemos basarnos únicamente en los resultados computacionales; por eso, se ha utilizado, por un lado, la frecuencia absoluta con los filtros habituales a través de Voyant (que filtra automáticamente las palabras funcionales) y, por otro lado, filtrando a través de *keywords* con *Wconcord*, se han encontrado las palabras clave que consideran los autores de los documentos que integran el corpus y se han seleccionado aquellas que tienen que ver con la teoría de redes complejas y que aparecen más de una vez.

Los resultados finales del estudio de palabras clave obtenidas para este corpus, si combinamos los resultados de *Voyant* y *Wconcord*, son:

- *Networks, Complex, Cluster (Clustering, Clustered) y System (Systems)* -Según *Wconcord*, a partir de la búsqueda *keywords*
- *network, networks, nodes, systems, complex y model*- según Voyant. Cabe destacar que estos últimos términos se han obtenido sin considerar las palabras clave que han destacado los expertos. Así, dichas palabras son las palabras clave obtenidas a través del procesamiento tecnológico de estos textos con la herramienta de análisis de Voyant. Podemos afirmar que estas palabras clave son más utilizadas que las que los propios autores señalan como sus *keywords*, ya que la herramienta las ha seleccionado a partir de su uso real en todo el corpus. Así pues, las *keywords* que señalan los autores como tales no tienen por qué ser las más repetidas en el corpus.

A continuación, vamos a ver si ciertamente estas palabras clave que hemos identificado como tales, a través de *Wconcord* y *Voyant*, resultan ser las más adecuadas,

haciendo para ello una prueba sobre el texto que Flair ha identificado como el más típico de la teoría de redes complejas.

En otras palabras, para realizar un análisis de las palabras clave del corpus en mayor profundidad, hemos comprobado si el documento diagnosticado por Flair como el de mayor especialización incluye las palabras clave señaladas por *Wconcord* y *Voyant* como tales; también hemos comprobado el comportamiento de dichas palabras clave dentro de dicho texto señalado por Flair como el de mayor especialización y, finalmente, si en dicho texto aparece alguna otra palabra o construcción que no coincida con los resultados extraídos de la búsqueda a gran escala.

Analizando el texto de mayor especialización en exclusividad, *Voyant* nos devuelve unos resultados que reflejan el hecho de que se mezclan cuatro lenguajes de especialidad al analizar este único texto considerado ya que identifica términos que tienen que ver con la teoría de redes complejas y términos que tienen que ver con el ámbito de aplicación de esta teoría sobre otro área de especialidad. Varían ciertos términos clave (que en este caso concreto referido al texto de mayor especialización serían *image*, *interest*, *points*, *network*, *complex* y *point* si se desprecian los plurales). Por otra parte, seguimos encontrándonos con *complex* y con *networks* entre los términos más repetidos.

Este fenómeno se debe a que la teoría de las redes complejas es aplicable a muchísimos ámbitos, y a que la búsqueda que nos está devolviendo el sistema obedece a los términos específicos de un ámbito concreto al que ha sido aplicada dicha teoría, por lo que podemos desechar los términos extra diagnosticados, ya que serían una mezcla entre el lenguaje de las redes complejas y la terminología específica del campo de la visión artificial, ya que es precisamente en esta disciplina específica en la que se centra el artículo: *Interest Point Detection in Images by a Local Centrality Algorithm on Complex Networks* (documento ij-css01_01-83). Vemos cómo los propios autores han incluido en el título palabras que ya aparecen repetidas analizando en profundidad el Corpus: *complex* y *networks* que sí aparecen con frecuencia en el resto del Corpus.

Por su parte, el resto de los términos que podríamos diagnosticar como palabras clave (*Images*, *point detection*, *interest*, etc.) aparecen terminologizados, entendiendo por *terminologización* el proceso mediante el cual un término adquiere un significado diferente dentro de un área de especialidad concreta al que adquiere dentro de los

parámetros del lenguaje común (Tallová L. 2004:39)] –aquí observamos que se utiliza en el contexto de la visión artificial–. Así pues, si nos centramos en la teoría de redes complejas en general, únicamente seguirían formando parte de los términos clave las palabras *complex* y *networks*. El resto tienen más que ver con la visión artificial.

Sin embargo, este análisis sobre un único texto señalado como representativo nos permite concluir que, al ser la teoría de las redes complejas una teoría de aplicación interdisciplinar, nos enfrentamos en este tipo de textos, según venimos afirmando, a cuatro lenguajes de especialidad (el matemático, el matemático verbalizado de especialidad, el natural y el lenguaje de especialidad del área de la aplicación concreta de la teoría de las redes complejas). Esto hace que la representación de las palabras clave de la red varíe, pero no así su comportamiento: siguen siendo palabras-sustantivo inter-relacionadas entre sí. De esta forma, en la siguiente ilustración, vemos cómo Voyant detecta como palabras clave las correspondientes a la disciplina de Visión Artificial (en la figura sobre fondo azul), y quedan las palabras clave de la teoría de redes complejas como palabras cercanas (en la figura en color naranja).

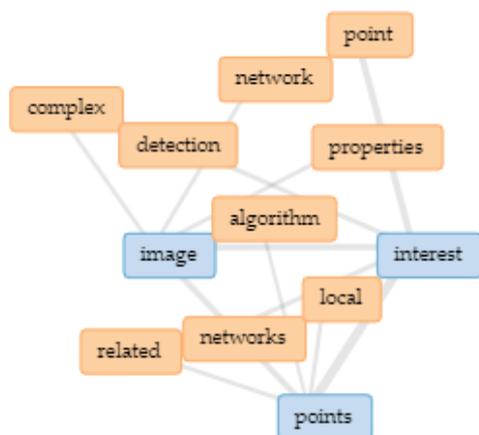


Figura 16. Imagen de red de términos extraída de Voyant

La imagen anterior muestra un gráfico en forma de red de los términos con una frecuencia más alta que aparecen por proximidad. Las palabras clave se muestran en azul y las colocaciones (o palabras próximas) se muestran en color naranja. Esta imagen está extraída de Voyant y se parece a un grafo, pero la red compleja que se desarrolla en la segunda parte de este proyecto de tesis tiene que ver con considerar el corpus como una red compleja dirigida y pesada. De momento, lo que hace Voyant de manera

análoga a lo que hace la herramienta que se presenta en la parte tres de esta tesis es convertir las palabras del corpus en los nodos de una red, aunque Voyant lo hace como una mera representación gráfica. Voyant es una herramienta realmente útil para observar un corpus a través de distintas representaciones gráficas. Por ejemplo, tiene otra forma de representar el corpus muy interesante: junto al lector de documentos, el usuario puede seleccionar la opción “términos Berry”, donde aparece el número de veces que se repite el término y en cuántos documentos se repite.

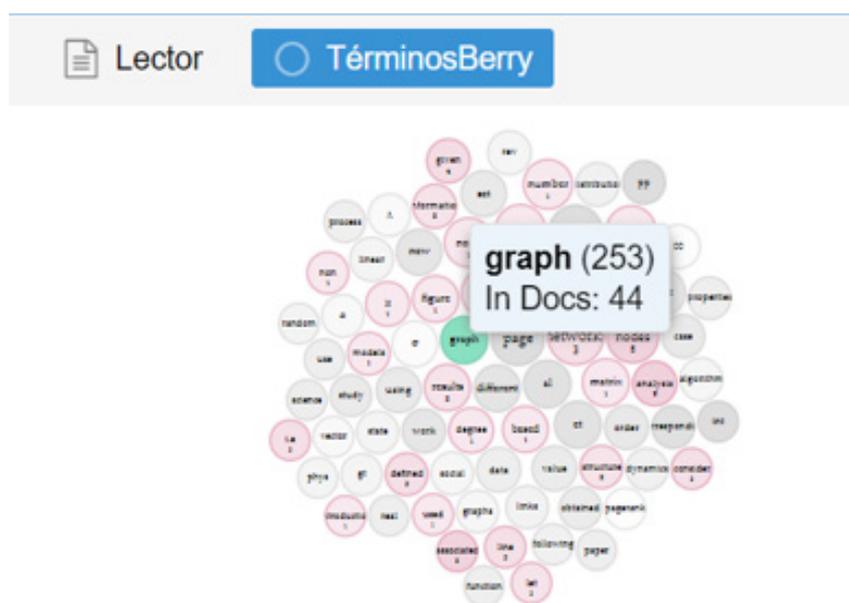


Figura 17. Términos del corpus representados como TérminosBerry en Voyant.

Las representaciones gráficas son una buena forma de ver cómo funciona el lenguaje. Ya hemos avanzado que, en la parte tres de esta tesis doctoral, se desarrolla una nueva herramienta de análisis que pasa por convertir el corpus en una red compleja pesada y dirigida en la que cada uno de los términos o palabras de la red será un nodo. Como veremos, se hablará de una red compleja multicapa, concepto éste que ofrece muchas más posibilidades para realizar el tratamiento computacional de textos y que no se limita a representar gráficamente un corpus.

A modo ilustrativo, se ha utilizado Python 3 (un lenguaje de programación) para representar a modo de red las *palabras clave* identificadas en los apartados anteriores como nodos (*matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering y clustered*), ya que se ha llegado a esta conclusión combinando *Wconcord* y *Voyant* y no disponemos de ninguna gráfica automática

generada por Voyant con las palabras clave reales. En la figura 18 se realiza una primera representación gráfica *ad hoc* (utilizando el co-texto) de ciertas palabras entre las que existe cercanía en el corpus, entendiendo que dos palabras están cercanas si aparecen en la misma frase con cierta frecuencia, lo que se representa estableciendo un enlace (o *link*, o “segmento”) entre ellas a partir de su posición dentro de la frase. Hay que decir que la siguiente figura no forma parte de la herramienta de traducción que se presenta en la parte tres de esta tesis doctoral, se trata únicamente de una representación gráfica que no se ha podido generar a partir de ninguna otra herramienta de análisis automático debido a las características intrínsecas del lenguaje de especialidad matemático (las herramientas existentes no pueden distinguir las palabras clave reales de este lenguaje de especialidad en concreto, ya que aparecen mezcladas con otros lenguajes de especialidad).

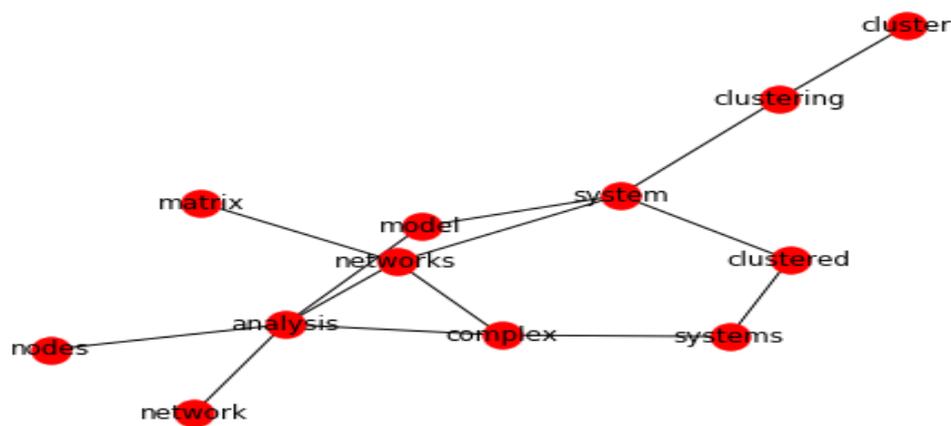


Figura 18. Fuente: Elaboración propia con código Python 3 en Spyder en base a los resultados obtenidos en combinación de las herramientas de análisis *Wconcord* y *Voyant*.

2.6 Conclusiones. Resumen y rasgos de la morfosintaxis de este tipo de textos. Presentación del estilo que se debe seguir en la redacción de textos

Buscando que este trabajo resulte de utilidad en el mundo de la traducción, a continuación, se describe el lenguaje de especialidad matemático de la teoría de redes complejas y se ofrece una propuesta de traducción de algunas frases construidas en torno a las que han sido identificadas como palabras clave en estos textos. Como se ha señalado en los primeros apartados de la segunda parte de esta tesis doctoral, normalmente la lingüística computacional y la lingüística de corpus utilizan criterios que

tienen que ver con la frecuencia de aparición de palabras y con listas de palabras funcionales para determinar cuáles son las palabras clave de un texto o de un corpus, pero, en este caso concreto, no es un método efectivo. Por eso, para empezar con una propuesta de traducción que parta de las palabras clave, vamos a comprobar si Voyant es capaz de determinar las mismas palabras clave que *Wconcord* a partir del mismo patrón de búsqueda: la palabra *keywords*. Nuestra lista se amplía con dos términos sobre la gráfica que ya teníamos: *matrix* y *analysis*.

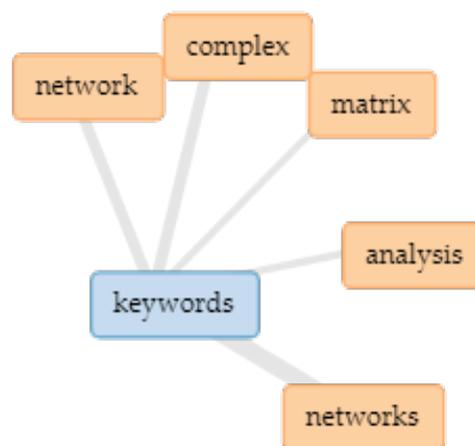


Figura 19. Imagen de Voyant con términos clave dispuestos en forma de grafo. Fuente: Sinclair, S. & G. Rockwell. (2019). Términos. *Voyant Tools*. Retrieved June 21, 2019 from <http://voyant-tools.org>

Así pues, ya tenemos un total de 12 *términos clave* característicos de este lenguaje de especialidad apoyándonos en la opinión del experto matemático que son exclusivos del lenguaje matemático como tal: *matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering y clustered*. El hecho de realizar una búsqueda a través del término *keywords* hace que se tenga en cuenta un filtro que va más allá de la frecuencia absoluta, y es curioso ver cómo efectivamente son términos que se repiten mucho a lo largo de un corpus. Un segundo cribado de detección humano es el que efectivamente permite clasificar cuáles son los términos que efectivamente tienen que ver con la teoría de redes complejas y cuáles son los que tienen que ver con la otra área matemática en la que se aplica dicha teoría.

Si tuviéramos en cuenta únicamente la frecuencia de aparición y la no pertenencia a la lista de palabras funcionales, la lista de las 24 palabras clave del corpus (ordenadas de acuerdo a su frecuencia de aparición) sería:

1) <i>network</i>	2) <i>networks</i>	3) <i>nodes</i>	4) <i>system</i>	5) <i>information</i>	6) <i>matrix</i>
7) <i>model</i>	8) <i>structure</i>	9) <i>systems</i>	10) <i>properties</i>	11) <i>new</i>	12) <i>node</i>
13) <i>complex</i>	14) <i>data</i>	15) <i>graph</i>	16) <i>dynamics</i>	17) <i>study</i>	18) <i>process</i>
19) <i>distribution</i>	20) <i>analysis</i>	21) <i>models</i>	22) <i>algorithm</i>	23) <i>links</i>	24) <i>problem</i>

Tabla 1. Tabla de elaboración propia donde aparece la lista de las 24 palabras clave del corpus ordenadas de acuerdo a su frecuencia de aparición.

La tabla anterior se basa en términos únicos (compuestos de una sola palabra). Este es el resultado obtenido después de programar a *Wconcord* para que excluyera listas de palabras funcionales predeterminadas junto con la lista de palabras que incluimos en la capa que denominamos “capa resto” de la red compleja multicapa que se presentará en la parte tres de esta tesis. De cualquier modo, lo que se presenta en la tabla son términos frecuentes, mientras que las unidades de conocimiento específico del lenguaje matemático suelen estar compuestas por dos términos o palabras (ya que se trata de colocaciones), por lo que se sugiere al lector que, para una correcta comprensión del lenguaje matemático, se centre en la parte tres y no en esta lista de palabras, ya que se demuestra que las unidades de conocimiento específico vienen dadas por los pares de palabras en el caso concreto que se refiere a este lenguaje de especialidad. Los 12 términos clave que podríamos haber identificado sin ayuda del experto matemático y sin ayuda del recurso de la red multicapa que presentamos en la parte 3 son los que decíamos anteriormente: *matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering y clustered*.

Una vez que tenemos estos doce términos clave (que serían los que obtendríamos sin ayuda del experto a través del corpus con herramientas tradicionales, ya que en la Tabla 1 se ha utilizado una capa de la red multicapa como lista de palabras excluidas en *Wconcord*), para hacer la propuesta de traducción se utiliza de nuevo Flair, con objeto de observar qué tipo de frases y construcciones gramaticales se deben utilizar en los textos meta.

Estudios como los de Mogorrón (2004) ponen de manifiesto la importancia de hacer un análisis fraseológico y morfosintáctico eficiente para poder comprender una mayor parte del significado de las palabras. Un análisis en profundidad siguiendo los principios de la semiótica descritos en los apartados anteriores, es necesario y de gran ayuda si se quiere entender la totalidad de un mensaje.

Flair no elabora gráficos, pero es capaz de llegar a conclusiones relevantes sobre este tipo de lenguaje con respecto a su morfología y sintaxis. Estos datos son de especial interés para el traductor si desea imitar el estilo del Texto Origen en el Texto Meta.

A) Frases interrogativas. Estilo que debe imitar el traductor o quien desee redactar textos de este tipo.

Según Flair, solo hay doce del total de ochenta y nueve textos analizados que plantean preguntas o frases interrogativas, por lo que podemos decir que no es una estructura sintáctica común en este tipo de textos. De cara al traductor, es interesante saber que, ante la duda, es mejor escribir un enunciado de sintaxis afirmativa.

Respecto a las frases interrogativas que la herramienta diagnostica llama la atención que en ninguno de los textos se han encontrado *have questions*, *yes-no questions* ni tampoco *tag questions*. Las *have questions* preguntan por acciones ya pasadas en inglés, por lo que podemos afirmar que los textos de esta área parten de bases o *referentes* considerados como verdaderos. El hecho de que estemos ante la que es, según el dicho popular, la ciencia exacta que no da lugar a error, puede explicar el hecho de que no consten preguntas. No se espera una opción distinta a la planteada y, por supuesto, tampoco existen *tag questions*, que en inglés sirven para verificar la información (parecido al, *¿verdad?* que aparece detrás de algunas afirmaciones orales en castellano). Únicamente aparecen preguntas en 5 textos.

B) Frases afirmativas/simples. Estilo que debe imitar el traductor.

En cuanto a los tipos de frases, podemos ver frases simples, coordinadas, subordinadas e incompletas en todos los documentos.

Aparecen prácticamente todos los tipos de complemento, pero el tipo de complemento adverbial se da en absolutamente todos los textos. Podemos, por tanto, decir que se trata de un lenguaje que gira más en torno a caracterizar o describir una acción que a caracterizar a los sustantivos o conceptos (ya que, para describir a los sustantivos, lo común es usar adjetivos y, para caracterizar acciones –o verbos-, se

utilizan los adverbios). Los conceptos como tales a los que se aplica la Teoría de Grafos integran el cuarto lenguaje de especialidad propio de la disciplina de aplicación.

El párrafo anterior es una prueba de la importancia que adquiere el lenguaje simbólico matemático ya que el texto que lo acompaña está centrado en describir el proceso expresado mediante el lenguaje simbólico. No hay que olvidar que la Teoría de Redes complejas describe y analiza diferentes aspectos del mundo real.

Este hecho se pone de manifiesto si nos fijamos en los sustantivos y adjetivos diagnosticados como palabras clave de la Teoría de Redes Complejas: *matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering y clustered* (matriz, análisis, red, redes, nodos, sistemas, sistema, complejo, modelado, agrupación, proceso de agrupación y agrupado). Todas estas palabras necesitan valerse de una palabra extra para establecer un referente claro. Por ejemplo, un sistema como tal no ofrece un referente concreto; sin embargo, si caracterizamos “sistema”, el término cobra sentido de manera similar a cómo lo hace en “sistema de información”.

C) Comportamiento verbal que han de tener los textos de especialidad matemática.

Respecto al comportamiento verbal, el resultado más llamativo es que se usan mucho más las formas verbales regulares que las irregulares a lo largo del corpus. Las formas regulares aparecen en todos los textos. Sin embargo, se ve como los verbos irregulares aparecen solo en 87 de los 89 textos. Profundizando en este hecho vemos cómo hay dos documentos que no presentan ningún verbo irregular entre sus frases. Este hecho responde a textos de nivel medio-bajo de competencia lingüística MCERL (véanse los enlaces de la bibliografía). De todos modos, Flair solo mide la competencia escrita por lo que el establecimiento del nivel puede no ser el adecuado.

El tiempo verbal muestra asimismo rasgos particulares cuando habla de redes complejas o grafos. El presente simple es, sin duda, el favorito de los autores ya que aparece en absolutamente todos los textos de la muestra. El matemático es, por tanto, un lenguaje descriptivo para el que los autores utilizan el presente simple. El uso del presente simple responde a la atemporalidad de las matemáticas: dos más dos siempre son cuatro. Además, como desarrollamos a continuación hablamos sobre todo de textos redactados en tiempos de aspecto perfecto tal como muestra Flair.

El segundo tiempo verbal favorito, el presente perfecto, nos habla de una acción que acaba de suceder y que ha finalizado, es decir, un tiempo de aspecto perfecto. La conclusión del párrafo anterior sigue siendo válida ya que el presente perfecto caracteriza al presente simple a la hora de situar una acción en el tiempo dotando así de cierta atemporalidad al presente simple y a la totalidad del texto.

El presente continuo aparece en tercer lugar, presente en treinta y dos de los ochenta y nueve textos del corpus. Vemos la gran diferencia entre uno y otro.

En el extremo opuesto tenemos el pasado perfecto (un tiempo anterior al presente perfecto). Al igual que el presente perfecto caracteriza al presente simple en el tiempo, el presente perfecto ayuda a situar la acción descrita a través del presente simple.

El matemático es, por tanto, un lenguaje descriptivo para el que los autores utilizan el presente simple. El uso del presente simple responde de nuevo a la atemporalidad de las matemáticas. A esta conclusión sobre los verbos se le puede añadir una extra: los autores no son amigos de los *Phrasal Verbs*, que aparecen solo en treinta y tres de los ochenta y nueve textos. El traductor debería, por tanto, optar por la opción simple de raíz latina más cercana al referente (por ejemplo, para *analizar* sería mejor *analyse* que *work over*)

D) Voz

Otro rasgo muy determinante es el uso de la voz pasiva en ochenta y ocho de los ochenta y nueve textos, apareciendo únicamente un solo texto escrito exclusivamente en voz activa. Este rasgo también es importante si se quiere imitar a la hora de traducir textos de este lenguaje de especialidad.

E) Sintagmas

E.1) Artículos

El uso de los artículos muestra que este Lenguaje de Especialidad se integra perfectamente en el lenguaje natural ya que, según Flair, “the”(el/la), “a”(un/una) y “and” (y) aparecen como los más frecuentemente utilizados, estando presentes (y repetidos más de una vez) en los 89 textos.

E.2) Sustantivos

Llama la atención la etimología de los sustantivos de este lenguaje de especialidad. Vemos como en ochenta y cuatro de los ochenta y nueve textos analizados aparecen sustantivos derivados del verbo con su forma terminada en “ing”. Estamos, por tanto, ante sustantivos que describen acciones. Un ejemplo de lo anterior lo da el término *learning*. Esta palabra viene del verbo *to learn* pero, expresada en *ing* y en este lenguaje de especialidad ha de traducirse por el sustantivo *aprendizaje*, no por el verbo en gerundio *aprendiendo*. La teoría de grafos tiene como uno de sus ámbitos de aplicación en la recientemente creada disciplina conocida como *machine learning*. La traducción correcta de esa colocación, siguiendo las pautas anteriores, es “aprendizaje máquina” no “máquinas aprendiendo” como podrían pensar los ojos de alguien no familiarizado con la materia. También es característico el hecho de que la mayoría de los sustantivos aparecen en plural (por ejemplo, los autores prefieren “sistemas” a “sistema”).

2.6.1 Propuestas de traducción basadas en los doce términos identificados como importantes y en las directrices de redacción del apartado anterior.

En el apartado anterior, se ofrecían pautas para resolver posibles problemas de traducción que plantea el lenguaje de especialidad matemático que se refiere a las redes complejas. Para facilitar aún más la tarea a los traductores, en este apartado analizaremos las principales palabras clave de la teoría de redes complejas ofreciendo una alternativa de traducción en base al significado y comportamiento de dichas palabras. Hemos seleccionado las palabras clave según los autores de los textos que aparecen diagnosticadas a través de Voyant por ser una mezcla entre el resultado lingüístico real y la opinión del autor-experto. Es destacable el hecho de que, por el momento, las unidades de conocimiento específico con las que podemos trabajar se componen en su mayoría de un único término (pues para poder identificar las colocaciones de este lenguaje de especialidad, se necesita o bien la opinión de un experto en el área de conocimiento, o bien un análisis en mayor profundidad con la técnica que se presenta en Criado-Alonso et. al 2020). Este factor es uno de los motivos que impulsó a continuar con esta investigación en un intento de descubrir cómo se unen

las palabras y qué pares de palabras formarán una colocación. En la parte 3 de esta tesis se presentarán nuevas unidades de conocimiento específico a las que no se habría podido llegar haciendo uso únicamente de las herramientas tradicionales disponibles.

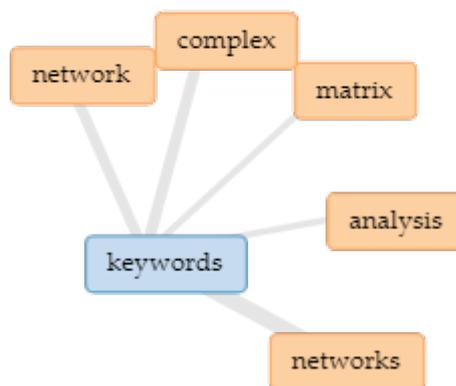


Figura 20. Imagen de red extraída de Voyant cuando se buscan las palabras clave a partir de las keywords que identifican los autores de los artículos.

La imagen anterior evidencia que se cumple la afirmación respecto al uso de las palabras sustantivas en estos textos que ya aventuraba Flair y que especificábamos en apartados anteriores: predomina el plural. La línea *networks* aparece de mayor tamaño ilustrando el número de conexiones que existen hacia y desde esa palabra. *Complex* y *matrix* no son términos que admitan forma plural en inglés.

La palabra más frecuente es “*Networks*” y vemos cómo esta palabra tiene que venir acompañada por otras para dotarse de significado y remitirse a un referente concreto. Es un sustantivo que incluso en castellano, suele aparecer determinado por las palabras próximas. Para ilustrar la afirmación anterior usaremos la frase “esta es una red perfecta”. La anterior es una frase sintácticamente hablando, pero si buscamos el referente sin más contexto, queda evidente la necesidad de determinar qué tipo de red es. Voyant nos da a entender que *Networks* ha de aparecer acompañado, pero no nos dice exactamente qué par de palabras forman la unidad de conocimiento específico (y la red no representa el orden de la colocación).

El término *Networks* es, por tanto, un término abstracto que no solo se determina a partir del lenguaje propio de la notación matemática, sino que también se determina lingüísticamente en función de las palabras que aparecen a su alrededor siguiendo un patrón de sintaxis común.

1.	Algorithm on Complex	Networks		ij-css01_01-83.txt
2.	the application of the Complex	Networks	theory.	ij-css01_01-83.txt
3.	s, Image Analysis, Geometrical	Networks	, Local Centrality	ij-css01_01-83.txt
4.	rest Point Detection in Complex	Networks		ij-css01_01-83.txt
5.	The use of complex	networks	with a spatial structure are used	ij-css01_01-83.txt

Figura 21. Fuente: Elaboración propia a partir de *Wconcord* donde se muestra qué términos aparecen antes y después de *networks* en un documento del corpus las cinco primeras veces que aparece este término (sin diferenciar si la primera letra se muestra en mayúscula o minúscula)

Otra conclusión acerca de la palabra *Networks* viene determinada por los ejemplos anteriores y el uso que hacen los autores de la mayúscula y de la minúscula. Vemos como cuando este término aparece en mayúscula hace referencia a la Teoría de Grafos (Diestel, 2000) o Teoría de las Redes Complejas (Boccaletti et al, 2006). Sin embargo, cuando vemos *networks* -en minúscula- observamos que el término habla de redes que sí vienen caracterizadas. No se refiere a la Teoría de Redes Complejas como tal, sino a un cierto tipo de Red.

1.	Algorithm on Complex	Networks		ij-css01_01-83.txt
2.	the application of the Complex	Networks	theory.	ij-css01_01-83.txt
3.	analyze them using complex	network	properties.	ij-css01_01-83.txt
4.	algorithm on a	network	with the aim detect the position	ij-css01_01-83.txt
5.	spatial and weighted complex	network		ij-css01_01-83.txt
6.	measure of the corresponding	network	, is proposed.	ij-css01_01-83.txt
7.	s, Image Analysis, Geometrical	Networks	, Local Centrality	ij-css01_01-83.txt
8.	rest Point Detection in Complex	Networks		ij-css01_01-83.txt
9.	s of an image by using complex	network	analysis.	ij-css01_01-83.txt
10.	l and fast-computable complex	network	to each image	ij-css01_01-83.txt

Figura 22. Fuente: Elaboración propia a partir de *Wconcord* donde se muestra qué términos aparecen antes y después de *networks* en un documento del corpus las diez primeras veces que aparece este término (sin diferenciar si la primera letra se muestra en mayúscula o minúscula)

La propuesta de traducción para los casos reflejados en la figura anterior evidencia que la afirmación del párrafo inmediatamente anterior respecto a la palabra “networks” se cumple. Se observa también que cuando hablamos de las redes concretas, es decir, de la Teoría de las Redes Complejas aplicada se habla de *red* y no de *redes*. La siguiente enumeración muestra las propuestas de traducción y, entre corchetes, las pautas o los comentarios de cara al traductor.

1. Algoritmos en las Redes Complejas [Aquí se refiere a la teoría de las redes complejas].
2. Aplicación de la Teoría de Redes Complejas.

3. [...] analizarlos utilizando las propiedades de una red compleja [“utilizando la red”, no tiene sentido. Vemos cómo “red” necesita ser determinada en cuanto a significativo por las palabras más próximas: *compleja* y *propiedades*. Aún así, se trata de un concepto abstracto].
4. [...] el algoritmo en una red para detectar la posición [...]. [En este caso vemos un fallo leve gramatical cometido por parte del autor, característico de hablantes de nivel intermedio. Faltaría una preposición detrás de “aim”. Cabe destacar que este problema en concreto sería imposible de resolver sin acceso al lenguaje matemático representado y al contexto].
5. *Weighted Complex Network*. Se puede traducir por “Red Compleja con Pesos”. Este concepto hace referencia a una red compleja con valores numéricos (pesos) en las aristas.
6. [...] medición de la red correspondiente [aquí vemos una vez más como *red* necesita de más contexto y más palabras para hacer comprender al lector el significativo al que el concepto hace referencia].
7. [...] Redes Geométricas, Centralidad Local [las redes geométricas son un tipo de red compleja]
8. [...] Detección de puntos en Redes Complejas [vemos como en esta frase la palabra *puntos* necesita ser determinada más a fondo de alguna manera. Es típico del lenguaje matemático utilizar conceptos abstractos determinados (como Redes Complejas), seguidos inmediatamente de otros conceptos vacíos que se determinan, a su vez, por su propio contexto].
9. [...] de una imagen a través del análisis de la red compleja [aquí vemos como la red compleja es, en cierto modo, la representación gráfica concreta tras haber aplicado la Teoría de Redes Complejas, en este caso, a una imagen.
10. [...] una red compleja rápidamente computable para cada imagen.

Dado que se han analizado los términos *Network* y *Networks* en las propuestas de la figura y el párrafo anterior, ahora le llega al turno al término *complex*. Realizando la búsqueda en *Wconcord*, el sistema nos devuelve lo siguiente:

1.		Complex	Systems in Science	ij-css01_01-83.txt
2.	Algorithm on	Complex	Networks	ij-css01_01-83.txt
3.	for the application of the	Complex	Networks theory.	ij-css01_01-83.txt
4.	analyze them using	complex	network properties.	ij-css01_01-83.txt
5.	An spatial and weighted	complex	network	ij-css01_01-83.txt
6.	Interest Point Detection in	Complex	Networks	ij-css01_01-83.txt
7.	est points of an image by using	complex	network analysis.	ij-css01_01-83.txt
8.	ometrical and fast-computable	complex	network to each image	ij-css01_01-83.txt
9.	The use of	complex	networks with a spatial structure	ij-css01_01-83.txt
10.	Interest points: A	complex	networks approach	ij-css01_01-83.txt

Figura 23. Fuente: Elaboración propia a partir de *Wconcord* donde se muestra qué términos aparecen antes y después de *complex* en un documento del corpus las diez primeras veces que aparece este término (sin diferenciar si la primera letra se muestra en mayúscula o minúscula)

Vemos cómo, en todos los casos, *complex* es traducible por complejo o compleja, y que esta palabra aparece combinada con sustantivos, en este caso sistemas y red o redes dependiendo de la línea. Vemos como hay algunas líneas análogas a si se realiza la búsqueda a partir de *network*.

1. Sistemas Complejos en la Ciencia
2. Algoritmo en Redes Complejas
3. Para la aplicación de la Teoría de las Redes Complejas
4. Analizarlos usando las propiedades de la red compleja
5. Una red compleja espacial y pesada
6. Detección de los puntos de Interés en las Redes Complejas
7. De una imagen usando el análisis de la red compleja
8. Red compleja rápidamente computable para cada imagen
9. El uso de redes complejas con estructura especial [...]
10. Puntos de Interés: Aproximación desde las redes complejas [aquí, sin embargo, vemos cómo se utiliza en minúscula y este simple hecho arroja luz sobre el término. No hablamos de la Teoría de las Redes Complejas sino de varias redes descritas a partir de esta teoría que seguramente aparecerán comparadas de alguna manera a lo largo del artículo del que se extrajo esa frase.

En el caso del término *Analysis*, obtenemos los siguientes resultados:

Se combina con los resultados de la Teoría de Grafos aplicada, es decir, con las redes o *networks* que lingüísticamente hablando han de ser determinados porque no nombran la teoría en sí, sino sus resultados.

1.	Keywords: Interest Points, Image	Analysis	, Geometrical Networks, Local	ij-css01_01-83.txt
2.	Image	Analysis	and Computer Vision systems	ij-css01_01-83.txt
3.	Image by using complex network	analysis	.	ij-css01_01-83.txt
4.		analysis	in problems related to Comput	ij-css01_01-83.txt
5.	Images using complex network	analysis	, to appear in	ij-css01_01-83.txt
6.		analysis	, Cambridge University Press,	ij-css01_01-83.txt
7.	planning (modeling of policies,	analysis	of the consequences of new ur	ijcss01_02-096.txt
8.	A Network	Analysis	of Frontier Capital Markets	ijcss01_02-107.txt
9.	A social network	analysis		ijcss01_02-107.txt
10.	Network	analysis	provides	ijcss01_02-107.txt

Figura 24. Imagen extraída de *Wconcord* una vez procesado el corpus. Fuente: Elaboración propia a partir de *Wconcord* donde se muestra qué términos aparecen antes y después de *analysis* en un documento del corpus las diez primeras veces que aparece este término (sin diferenciar si la primera letra se muestra en mayúscula o minúscula)

1. Análisis de Imagen.
2. Análisis de Imagen y Sistemas de Visión Computadora [aquí vemos como el término computadora es prueba de la aparición de la cuarta lengua de especialidad a la que se está aplicando la Teoría de Grafos. Es un término que aparece en textos de alto nivel de especialidad de la informática y la ciberseguridad.
3. [...] usando el análisis de las redes complejas. [esta traducción corrobora la afirmación que aparece al principio de este apartado.
4. Análisis en problemas relativos a...
5. Imágenes usando el análisis de las redes complejas que [...].
6. Análisis [es el enunciado de un párrafo].
7. Modelos de políticas, análisis de las consecuencias [...].
8. Análisis desde la perspectiva de red de Mercados de Capital Fronterizos [vemos como aquí aparecen términos de lenguaje económico, la cuarta lengua de especialidad].
9. Análisis de la red social.
10. El análisis a través de la red ofrece [...].

Con vistas a su utilización por parte de un traductor, hay que destacar el hecho de que *Analysis* encuentra prácticamente siempre su traducción concreta en forma sustantiva “análisis” en lugar de en su forma verbal “analizar”.

La última palabra clave de la muestra extraída a partir de la opinión del autor-experto es *matrix* (o matriz en castellano). Realizando la búsqueda en *Wconcord*, encontramos las siguientes interacciones:

1.	s the locally averaged moment	matrix	using the image gradients, and	ij-css01_01-83.txt
2.	the eigenvalues of the moment	matrix	to compute	ij-css01_01-83.txt
3.		matrix	$A(G) = (a_{ij})$ determined by the	ijcss01_02-100.txt
4.	E, E(B(G)) whose adjacency	matrix	is given by	ijcss01_02-100.txt
5.	where $I(G)$ is the incidence	matrix	of G.	ijcss01_02-100.txt
6.	where $A(G) + gr$ denotes the	matrix	obtained by adding to $A(G)$ the	ijcss01_02-100.txt
7.		matrix	(b_{ij}) and b_{ii} is the degree of the	ijcss01_02-100.txt
8.	ow $A(G) = (a_{ij})$ is the adjacency	matrix	of the directed network G	ijcss01_02-100.txt
9.	E, E(B(G)) whose adjacency	matrix	is given by	ijcss01_02-100.txt
10.	where $H = H(G)$ is the incidence	matrix	of heads of G defined by	ijcss01_02-100.txt

Figura 25. Fuente: Elaboración propia a partir de *Wconcord* donde se muestra qué términos aparecen antes y después de *matrix* en un documento del corpus las diez primeras veces que aparece este término (sin diferenciar si la primera letra se muestra en mayúscula o minúscula)

En la mayoría de las veces que aparece *matrix*, se observa que la palabra inmediatamente anterior es *adjacency* o *incidence*. Estos términos y su posible traducción nos llevaron hacia las colocaciones de este lenguaje de especialidad, así como hacia el hecho de que es necesario ser más preciso tanto en la localización de los términos de especialidad como en su definición; por eso es especialmente útil para el traductor el desarrollo y la utilización de la herramienta que se presenta en la parte tres de esta tesis.

Vemos cómo, al igual que “red”, la palabra “matriz” necesita ser determinada por el resto de las palabras del contexto debido a que, por sí misma, tiene un significado demasiado abstracto. En la figura anterior, la línea 3 muestra cómo el lenguaje simbólico-numérico de notación matemática, o lenguaje matemático como tal, es el apoyo que define al significante real de conceptos tan abstractos como “red” o “matriz”.

1. La matriz de momentos localmente promediados utilizando los gradientes de las imágenes.
2. La matriz de los autovalores del momento para computar.
3. La matriz $A(G) = (a_{ij})$ determinada por el
4. E, E(B(G)) cuya matriz de adyacencia está dada por
5. donde $I(G)$ es la matriz de incidencia de G.
6. donde $A(G)+gr$ denota la matriz obtenida añadiendo a $A(G)$ el

7. la matriz (b_{ij}) y b_{ii} es el grado de los
8. ... $A(G)=(a_{ij})$ es la matriz de adyacencia de la red dirigida G
9. $E, E(B(G))$ cuya matriz de adyacencia está dada por
10. Donde $H=H(G)$ es la matriz de incidencia de las cabezas de G definida por [nodos-cabeza (o principio) y nodos-cola "tail" (o finales) de una arista dirigida).

2.6.2 Recapitulaciones generales sobre el lenguaje de especialidad matemático de la teoría de redes complejas

De nuevo, a tenor del análisis previo y de las propuestas de traducción, se puede concluir que los textos que versan sobre la Teoría de Redes Complejas y sus aplicaciones mezclan cuatro lenguajes que podrían ser considerados de especialidad: 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. Se ve, por tanto, una gran variación de los términos o palabras clave en función del texto que se escoja. Sin embargo, usando el método descrito en este estudio, se han podido diagnosticar un total de doce términos clave exclusivos de la Teoría de Redes Complejas: *matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering y clustered*. Estos términos necesitan del cuarto lenguaje de especialidad o ámbito de aplicación para establecer un referente claro en la mente del lector. Los términos clave varían de texto a texto.

También encontramos enunciados muy simples, típicos de los que emiten los hablantes B2 en una lengua. Esto contradice la idea que se tiene sobre la dificultad del lenguaje matemático, aunque, sin embargo, se observa cómo el lenguaje verbalizado se apoya en el simbólico-gráfico-numérico para acercarse al concepto (o referente) expresado a través de palabras como "red" (*network*) o "matriz" (*matrix*). Normalmente, cuanto más bajo sea el nivel MCERL, más sencillo debería ser el texto con vistas a su comprensión. La dificultad de traducción en este tipo de textos no reside únicamente en los términos de especialidad en sí mismos, sino que también viene dada por la falta de comprensión por parte de los traductores del lenguaje matemático.

Una vez probada la relación y la existencia de estos cuatro sub-lenguajes 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 Teoría de Redes Complejas en sí misma para extraer conclusiones lingüísticas, ya que es necesario perfilar qué pares de palabras han de aparecer juntas y analizar cómo funciona este lenguaje de especialidad a la hora de, por ejemplo, crear verbos de especialidad. Todo esto, junto con la generación automática de textos, se puede llevar a cabo gracias al concepto de *Enriched Line Graph* (“Grafo Lineal Enriquecido”) que presentamos en la segunda parte de esta tesis doctoral.

SEGUNDA PARTE PRÁCTICA

Capítulo 3: Estudio del corpus mediante Redes complejas

En este tercer capítulo de esta tesis doctoral, se presenta y describe exhaustivamente la metodología que nos ha permitido caracterizar en profundidad el lenguaje matemático de especialidad de redes complejas. En relación con lo expuesto en capítulos anteriores, haremos a continuación una serie de recapitulaciones sobre el lenguaje matemático de especialidad.

Hemos podido concluir que los textos que versan sobre la Teoría de Redes Complejas y sus aplicaciones mezclan cuatro lenguajes que podrían ser considerados de especialidad: el lenguaje simbólico-gráfico-numérico matemático, el lenguaje de especialidad matemático verbalizado, el lenguaje natural en el que se exprese el texto y el lenguaje relativo al área de especialidad al que se esté aplicando la teoría de redes complejas (estas áreas de conocimiento a las que se aplica la teoría de redes complejas varían entre los distintos *extended abstracts* que componen el corpus). Se ve, por tanto, una gran variación de los términos o palabras clave en función del texto que se escoja ya que muchas veces las herramientas de análisis tradicionales identificarán como palabras clave dentro de un *extended abstract* algunos términos relativos al área a la que se está aplicando la teoría de redes complejas (sin ser parte de la teoría de redes complejas como tal).

Utilizando las herramientas de análisis tradicionales, hemos podido identificar un total de doce términos clave exclusivos de la Teoría de Redes Complejas: *matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering* y *clustered*. Es notable el hecho de que para haber podido identificar estos términos importantes o palabras clave, se ha tenido que seguir un procedimiento de análisis particular ya que no han aparecido directamente identificados en ninguna de las tres herramientas de análisis tradicionales descritas en el capítulo anterior.

Para ejemplificar la dificultad que plantea este corpus de especialidad en concreto para identificar los términos clave, no hay que perder de vista que *matrix, analysis, network, networks, nodes, systems, system, complex, model, cluster, clustering*

y *clustered* necesitan del cuarto lenguaje de especialidad o ámbito de aplicación para establecer un referente claro en la mente del lector. Es muy destacable el hecho de que los términos clave varían de texto a texto.

Una vez que se ha visto la relación y la existencia de estos cuatro sub-lenguajes 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 Teoría de Redes Complejas en sí misma para extraer conclusiones lingüísticas, de tal manera que podamos discernir el ámbito al que pertenecen cada uno de los términos del corpus.

Esta metodología de análisis es lo suficientemente sensible para perfilar qué pares de palabras han de aparecer juntas y analizar cómo funciona este lenguaje de especialidad a la hora de, por ejemplo, crear verbos de especialidad. Todo esto, junto con la generación automática de textos, se puede llevar a cabo gracias al concepto de *Enriched Line Graph* (“Grafo Lineal Enriquecido”) presentado en las próximas líneas de este tercer capítulo.

3.1 Presentación del enfoque para la caracterización del lenguaje matemático

Para entender el funcionamiento de este lenguaje de especialidad más allá del que devuelve el estudio realizado con herramientas tradicionales, tenemos que volver nuestra mirada hacia algunos conceptos interdisciplinarios, como, por ejemplo, el concepto de sistema lingüístico que se presenta en trabajos como los de Cong & Liu (2014), Martincic et al. 2010; o Hudson (2010). Ya se ha visto en capítulos anteriores que el lenguaje de especialidad se caracteriza por el uso de terminología específica y por el uso de unidades de conocimiento específico. A partir de ahora vamos a ver cómo estas unidades de conocimiento específico obedecen a un sistema determinado, por lo que se pueden identificar y estudiar. El hecho de que estas unidades de conocimiento específico obedezcan a un sistema identificado, permite la identificación de las colocaciones.

Entendemos las combinaciones de dos palabras como colocaciones –una unidad de significado y no la suma de sendos significados de la combinación–. Las colocaciones son sin duda unidades de conocimiento especializado, que son el vehículo de las lenguas de especialidad. El factor que hace de las colocaciones un elemento de estudio de especial interés para los traductores es precisamente que la suma de dos palabras o términos van a resultar en una única unidad de significado (Koike, 2001) y esto, llevado al contexto de los lenguajes de especialidad, resulta de gran relevancia si tenemos en cuenta que normalmente, para el mundo de la traducción, la disciplina matemática es una gran desconocida (y es necesario entender el significado de un texto para poder traducirlo). De momento ya hemos visto cómo funcionan las palabras individualmente, pero es interesante ver qué sucede cuando contemplamos unidades de conocimiento específico compuestas por dos palabras (y su proceso de formación). De hecho, una vez realizado el estudio computacional, parece ser que este lenguaje de especialidad en concreto está mucho más caracterizado por unidades de conocimiento formadas por dos palabras que por unidades de conocimiento específico formadas por una sola palabra. En este caso aplicaremos la teoría y las herramientas computacionales de las redes complejas al estudio de las lenguas.

3.2 Proceso de creación de la herramienta: convertir el corpus en una red compleja multicapa

Para analizar las colocaciones que forman parte de la terminología de la lengua, se ha considerado una red multicapa y algunas aristas concretas que formarán parte de cierto *Line Graph* (Criado-Alonso et al., 2020).

Cada una de las palabras del corpus se convierte en un nodo (o vértice). Cuando una palabra aparece a continuación de otra se coloca una arista dirigida entre dichos nodos (considerando que las dos palabras están relacionadas), de manera que queda establecido qué término va delante y qué término va detrás. No es la primera vez que se crea una red compleja para analizar una lengua. De hecho, el modelo de red lingüística considerado en este trabajo surge de la naturaleza variable del texto, que “se lee hacia adelante” de manera que los enlaces o aristas dirigidos y ponderados resultan especialmente adecuados para representar las relaciones entre las unidades lingüísticas

como en Cong & Liu (2014), Martincic S et al. (2016) o Masucci A & Rodgers (2006). De este modo, la relación de co-ocurrencia se establece entre dos palabras o unidades lingüísticas adyacentes en una frase, en la que la dirección del enlace refleja la secuencia de las palabras, y el peso en el enlace refleja la frecuencia de aparición de esa secuencia de dos palabras enlazadas (se trata, pues, de una red compleja dirigida y pesada). Cuando dos palabras están conectadas por una arista que tiene un peso elevado, los dos términos o nodos se convierten en uno solo y son estos nodos formados por pares de palabras los que forman las colocaciones que, en este caso, interesa estudiar.

La razón por la que se ha elegido este sistema es que los últimos avances en el estudio de los sistemas complejos han permitido pasar del análisis y el estudio de las propiedades aisladas de un sistema a una modelización más realista en la que interactúan y coexisten múltiples fenómenos de naturaleza diferente. Este hecho fundamenta el uso de redes multicapa para el análisis de la estructura y caracterización de los lenguajes. Es importante mencionar que, aunque el modelo proporcionado por las redes multicapa ha aparecido recientemente en la literatura, este modelo ya ha sido considerado como una estructura adecuada para el análisis estructural de sistemas lingüísticos [véase Martincic S. et al. (2016) o Mehler A. et al. (2016)]. En cualquier caso, la aplicación de la teoría y las herramientas computacionales de las redes complejas al estudio de las lenguas no es una tendencia científica nueva [Borge-Holthoefer J (2010), Cárdenas JP. et al. (2014), Criado-Alonso et al. (2020)].

Las redes multicapa constituyen un marco unificado que permite modelar 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. Para nuestro análisis nos interesa considerar una red multicapa, tal y como quedó establecido en Criado-Alonso et al. (2020), compuesta en nuestro caso por cuatro capas:

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.

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. Y, 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).

3.3 Asociaciones entre capas y palabras: colocaciones

El estudio de la formación de asociaciones de estas unidades lingüísticas (o colocaciones) que forman unidades de conocimiento específico se ha realizado utilizando el concepto de *Line Graph* [(Crucitti et al. (2006)a, Crucitti et al. (2006)b y (Crucitti, Latora, & Porta, 2006)c, y Porta et al. (2006)]. Como se señalaba en el primer apartado, este concepto proporciona una buena representación de las propiedades de la red cuando es más apropiado dar más importancia a las aristas de una red que a sus nodos y, en nuestro contexto, es especialmente adecuado para estudiar y analizar las colocaciones en un texto, ya que cualquier secuencia de dos unidades lingüísticas o palabras colocadas consecutivamente en un texto puede ser representada por el enlace entre esas dos palabras. Dicho de otro modo, si nos interesa estudiar las relaciones entre palabras, este concepto, con pesos en las aristas o relaciones entre palabras, es el que nos interesa tener presente.

Es importante señalar que, al utilizar esta metodología, no ponemos el foco en las palabras como tales, sino en las relaciones entre palabras. Las palabras son importantes porque aparecen muchas veces y porque están relacionadas con otras que

también lo son: si una palabra muy importante se relaciona directamente con otra que también lo es, es lógico pensar que ambas dos formen una unidad de conocimiento especializado. Uno de los resultados especialmente relevantes de este estudio es que la estructura de las asociaciones de palabras (es decir, de colocaciones como "red compleja") es mucho más sensible a las variaciones en la intensidad de su aparición que las mismas palabras aisladas o en otros contextos: por ejemplo, la palabra "red" en "red compleja", "red" no significa lo mismo que en "red ferroviaria". El primer término pertenece al lenguaje de la especialidad matemática y el segundo no. Con vistas a la traducción (automática o no), y al análisis lingüístico, es relevante saber cuáles son los pares de palabras que efectivamente forman parte de esta terminología específica del lenguaje matemático y que han de ser consideradas como unidades de conocimiento especializado independientes. El uso del concepto de *Line Graph* junto con la red compleja multicapa pesada y dirigida es lo suficientemente sensible como para determinar cuáles son los nodos que formarían estas colocaciones de especialidad. Puede decirse que una de las mayores dificultades en la tarea de un traductor es encontrar traducciones apropiadas para las colocaciones que encuentra en la lengua hacia la que traduce, tal y como se señalaba en los primeros apartados.

Se pueden distinguir diferentes tipos de colocaciones que forman unidades de conocimiento específico; entre otros, adjetivo-sustantivo, sustantivo-sustantivo, verbo-adverbio y sustantivo-adverbio, tal y como muestran los estudios de Byrne (2014) sobre traducción de especialidad o incluso los de Sinclair (1991), cuando ya hablaba de colocaciones. Otros autores como Varó (2000), que es un referente en lo que a inglés académico se refiere, también hablan de los diferentes tipos de colocaciones y de la importancia que estas adquieren en la comunicación de especialidad.

Otra característica de las colocaciones, que es importante mencionar aquí (además de su importancia), es que el significado de las colocaciones es *compositivo* –a diferencia de las expresiones idiomáticas–. En el primer caso, nos referimos a secuencias de dos o más unidades lingüísticas que representan inequívocamente un concepto ya que es posible identificar qué parte del significado aporta cada uno de sus componentes individuales; en el segundo –las expresiones idiomáticas–, no todas las unidades lingüísticas que las componen tienen significados parciales (Sinclair, 1991).

En este trabajo, a partir de textos escritos en inglés en el lenguaje especializado de las redes complejas, nos centraremos en los tipos de colocaciones adjetivo-sustantivo y sustantivo-sustantivo. En todo caso, debemos subrayar que las colocaciones son un indicador sensible de la afinidad de un texto con un área de especialidad y que, por tanto, este indicador puede utilizarse para clasificar los textos: los textos muy especializados de un área específica de conocimiento tienen muchas colocaciones.

3.4 Descripción formal del corpus usando Python

Ya hemos visto que las herramientas de análisis tradicionales ofrecen cierta disparidad al analizar el mismo documento y, para obtener resultados claros, tomaremos los siguientes datos obtenidos al procesar el corpus utilizado también en *Wconcord*, *Flair* y *Voyant* con Python 3.7. Recordemos que nuestro corpus es un corpus lingüístico bruto de 86 artículos y resúmenes ampliados y un total de 147.637 palabras y 25.210 frases. La unidad de análisis considerada en este caso es la frase, es decir, las palabras encerradas entre dos puntos (Halliday & Matthiessen, 2004). Además, es importante tener en cuenta que las comas y otros signos de puntuación dentro de la frase han sido filtrados para el análisis realizado. Este corpus lingüístico está formado de los volúmenes 1 a 6 de la revista *International Journal of Complex Systems in* (IJCSS) publicados entre abril de 2011 y noviembre de 2016 (<http://www.ij-css.org>).

Una vez procesado el corpus, obtenemos la representación gráfica de la red, que se puede ver en la imagen a continuación:

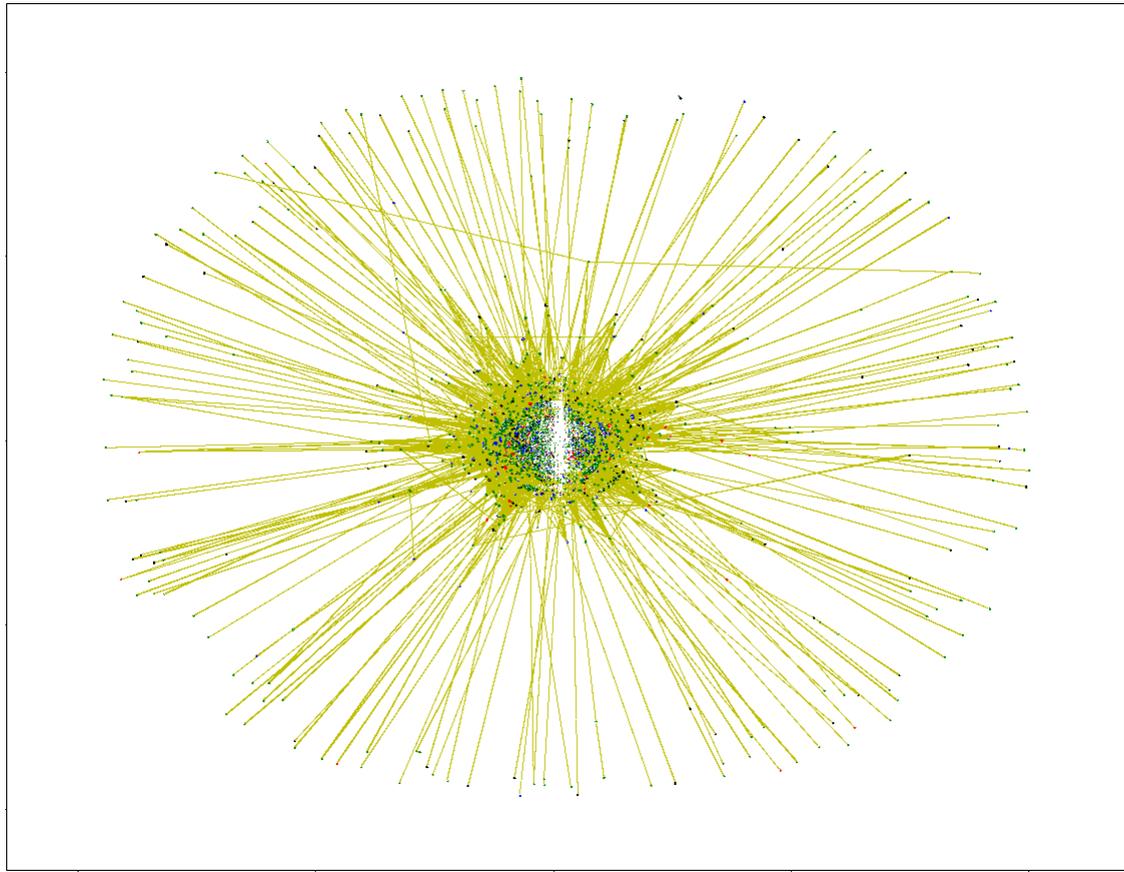


Imagen 1. Corpus de análisis representado como una red compleja. Fuente: Elaboración Propia

Mirando la red un poco más en detalle, se aprecia que los nodos están expuestos en diferentes colores y otra serie de características (véase imagen a continuación):

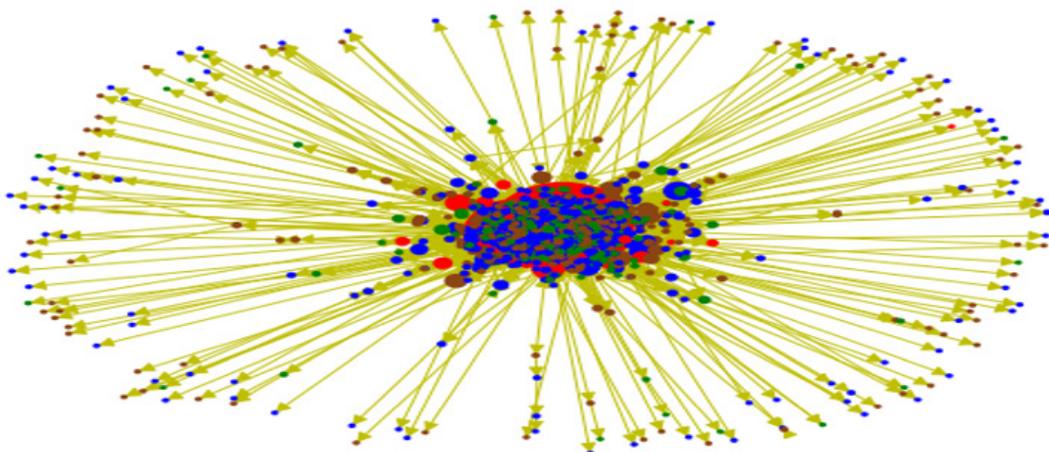


Figura 26. Red multicapa en detalle. Fuente: Criado-Alonso et al. (2020), Social Networks Analysis and Mining 10:69, p. 69

En la red han quedado divididos los nodos por colores (dependiendo de la capa en la que estén) de tal modo que hay un total de 829 verbos resaltados en color verde, 1487 palabras de especialidad (en color azul) 150 palabras hilo (en rojo) y 754 palabras que no se corresponden al lenguaje de especialidad matemático de redes interactuando entre sí. Haciendo una ampliación de la imagen (observándola con más profundidad), se pueden obtener representaciones gráficas como la de la Fig. 27, donde partiendo de un nodo (o palabra) se llega a una secuencia lógica dentro del lenguaje matemático que forma una frase. En el caso de la Fig. 27, podemos ver, cómo se llega a la frase “*this algorithm may be of interest in other contexts*” (este algoritmo puede tener interés en otros contextos), con sentido.

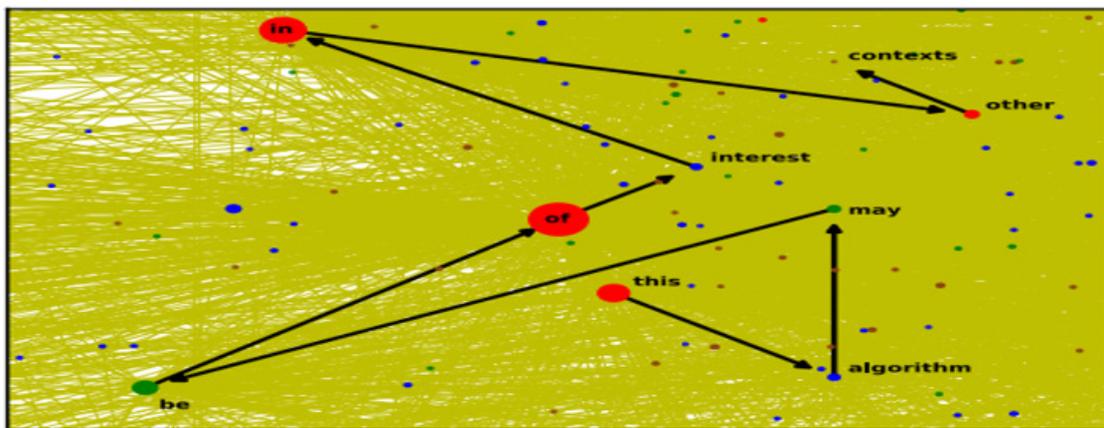


Figura 27. Zoom de la red multicapa considerada. Fuente: Criado-Alonso et al. (2020), Social Networks Analysis and Mining 10:69, p.69

Es importante destacar que el tamaño de los nodos es proporcional a su grado de salida. El modelo que presentamos es bastante diferente de otras aproximaciones al estudio de una lengua bajo la perspectiva de las redes complejas ya que considera, entre otras, las relaciones de dependencia sintáctica y las relaciones e interacciones entre unidades lingüísticas y subsistemas (sílabas y grafemas), con el fin de analizar los mecanismos de influencia estructural entre ellos [ver, por ejemplo, (Martincic, Margan, & Mestrovic, 2016)].

Las figuras anteriores muestran el primer resultado de esta metodología novedosa que parte de considerar una red multicapa dirigida y ponderada construida a partir del corpus estudiado: Los nodos de la red son las palabras que aparecen en

cualquiera de los textos que componen el corpus, y se establece un enlace (dirigido) que conecta dos palabras si aparecen consecutivamente en el texto (co-ocurrencia dirigida).

Como hemos resaltado, la mezcla del concepto de red multicapa dirigida y pesada es novedoso porque nos permite localizar las colocaciones más características del lenguaje de especialidad, y esto nos permite ir más allá de los resultados que aportan los enfoques tradicionales de análisis léxico basado en frecuencias.

3.5 Uso de *Line Graph* y Centralidad del *PageRank* en nuestra Red

Con la mirada puesta de manera específica en realizar este estudio, el concepto de *Line Graph* se utilizará en este contexto concreto para calcular el *PageRank* de las colocaciones. Intuitivamente, para una red dirigida y ponderada $G = (X, E)$, el *PageRank* de una arista $i \rightarrow j$ está relacionado de manera directa con la frecuencia con la que un caminante aleatorio pasa por esa arista (i, j) . Sin embargo, es importante recordar este concepto: dado un grafo dirigido $G = (X, E)$ el *Line Graph* de G es el grafo $L(G) = (E, \tilde{E})$, en el que los nodos son las aristas del grafo original, y dos nodos de este nuevo grafo están conectados directamente por una arista si en el grafo original G es posible “recorrer” las aristas correspondientes a dichos nodos, una a continuación de la otra. Es decir:

$$\tilde{E} = \{((i, j), (j, k)); (i, j), (j, k) \in E\}$$

El *Line Graph* del grafo original es la estructura adecuada para estudiar y analizar las colocaciones de un texto, ya que cualquier secuencia de dos unidades lingüísticas o palabras colocadas consecutivamente en un texto puede representarse mediante el enlace entre esas dos palabras, que pasarían a ser un nodo en el *Line Graph* asociado. Por lo tanto, una colocación puede identificarse con el enlace que une las palabras que la componen y, en consecuencia, con el nodo correspondiente del *Line Graph* de la red considerada. Es importante mencionar aquí que existe una conexión, y en algunos casos una relación directa, entre las propiedades de un grafo y las de su *Line Graph* asociado [véase Criado et al. (2012), o Criado et al. (2018)]. Está claro, intuitivamente hablando, que el *PageRank* de una arista de un grafo dirigido y ponderado $G = (X, E)$, es decir, la frecuencia con la que el caminante aleatorio pasa por la arista $(i, j) \in E$ está relacionada

con el *PageRank* del propio nodo i , ya que cada vez que un caminante aleatorio pasa por la arista $i \rightarrow j$ este pasa también por el nodo i . En el artículo Criado et al. (2018), se presenta un método para calcular el *PageRank* de las aristas de un grafo ponderado y dirigido a partir de su *Line Graph* asociado. Así, siguiendo a Criado et al. (2018), podemos obtener el *PageRank* de la arista $i \rightarrow j$ mediante la expresión

$$\text{LPR}(i \rightarrow j) = \text{PR}(i) p_{ij}$$

en la que p_{ij} es la probabilidad de que el paseante aleatorio siga su camino desde el vértice i hacia el vértice j en lugar de ir hacia otro nodo directamente conectado con él. Teniendo en cuenta que el número medio de palabras que componen una frase dentro del corpus que hemos considerado es 18,44, en este contexto, el factor de amortiguación correspondiente a esta situación es $q=0,94$, tal y como se puede obtener repitiendo los cálculos realizados en Criado et al. (2018).

En otras palabras, si un paseante aleatorio se mueve sobre la red de forma aleatoria siguiendo las posibles direcciones que marcan las aristas entre los nodos, está claro que pasará más a menudo por los nodos más accesibles. Para modelar matemáticamente esta idea se utiliza un concepto matemático conocido como *cadena de Markov*: para calcular la probabilidad buscada es necesario considerar una extensión del clásico algoritmo *PageRank* que utiliza Google, de manera que el paseante aleatorio se mueve de un nodo a otro siguiendo la red moviéndose por sus aristas con una cierta probabilidad $q \in (0, 1)$ que depende, en nuestro caso, de la media del número de términos que tiene una frase del corpus (en el caso del algoritmo de Google es $q = 0.85$). Este valor es la probabilidad de que un caminante aleatorio no cambie su trayectoria saltando a otro nodo de la red no conectado con el anterior, en lugar de desplazarse a un nodo conectado directamente por una arista con el nodo actual. Este valor se denomina factor de amortiguación, *damping factor*, y en nuestro caso toma el valor $q=0,94$. Este valor de $q=0,94$ es la probabilidad de que un caminante aleatorio no cambie su trayectoria saltando a otro nodo de nuestra red no directamente conectado con el anterior, en lugar de desplazarse a un nodo conectado directamente por una arista con el nodo actual. En nuestro contexto, este salto puede interpretarse como el final de la frase actual y el punto de partida de una nueva frase, hecho que resulta de gran

relevancia tanto para el análisis de la frase en sí como para tareas tales como la creación automática de textos.

3.6 Nuestro modelo multicapa en detalle. Resultados.

La red multicapa dirigida y pesada en la que se basa nuestro modelo se construye a partir del corpus que hemos descrito, procesando sus datos y elementos mediante el lenguaje de programación Python 3.7. De este modo, los nodos de la red son las palabras que aparecen en cualquiera de los textos que componen el corpus, y se coloca un enlace (dirigido) que conecta dos palabras si aparecen consecutivamente en algún lugar de un texto de los que está compuesto el corpus. Si estas palabras están conectadas más de una vez, se asigna un peso a la arista que las une, indicando la frecuencia de aparición de esta conexión. En la red consideramos cuatro capas: La capa léxica, cuyos nodos son principalmente los sustantivos y adjetivos, la capa verbal cuyos nodos son los verbos, la capa formada por las palabras de enlace y una cuarta capa formada por el resto de las palabras (no caracterizadas). La asignación de las palabras del corpus a cada una de estas capas se ha realizado mediante un proceso de supervisión manual con ayuda de un experto del área de especialidad propia del lenguaje. En esta red, los enlaces sustantivo-sustantivo serían conexiones intra-capas dentro de la capa léxica; las conexiones verbo-verbo serían también conexiones intra-capas dentro de la capa verbal, y, por ejemplo, las conexiones verbo-enlace o enlace-verbo, serían conexiones inter-capas entre las capas correspondientes. Es importante destacar que dentro de la capa léxica se han incluido también los términos polisémicos con referente en el mundo de las matemáticas (por ejemplo, "*limit*" y "*limits*" pueden actuar como verbo o como sustantivo en inglés (traducidos como límite, limitar, o, a veces incluso limitación). El sustantivo es más específico en las matemáticas que el verbo, que está más cerca del lenguaje general, por lo que este tipo de palabras se incluyen en la capa léxica).

Además, los nodos de cada una de estas capas tienen diferentes colores según la capa a la que pertenecen, tal y como se muestra tanto en las Fig. 26. y 2. como en la Fig. 28; En la Figura 28 se muestra una red lingüística compuesta por 147.637 palabras de las que 1.487 son palabras incluidas en la capa léxica. El tamaño de los nodos es

proporcional a su grado de salida, y el color del nodo indica la capa a la que dicho nodo pertenece.

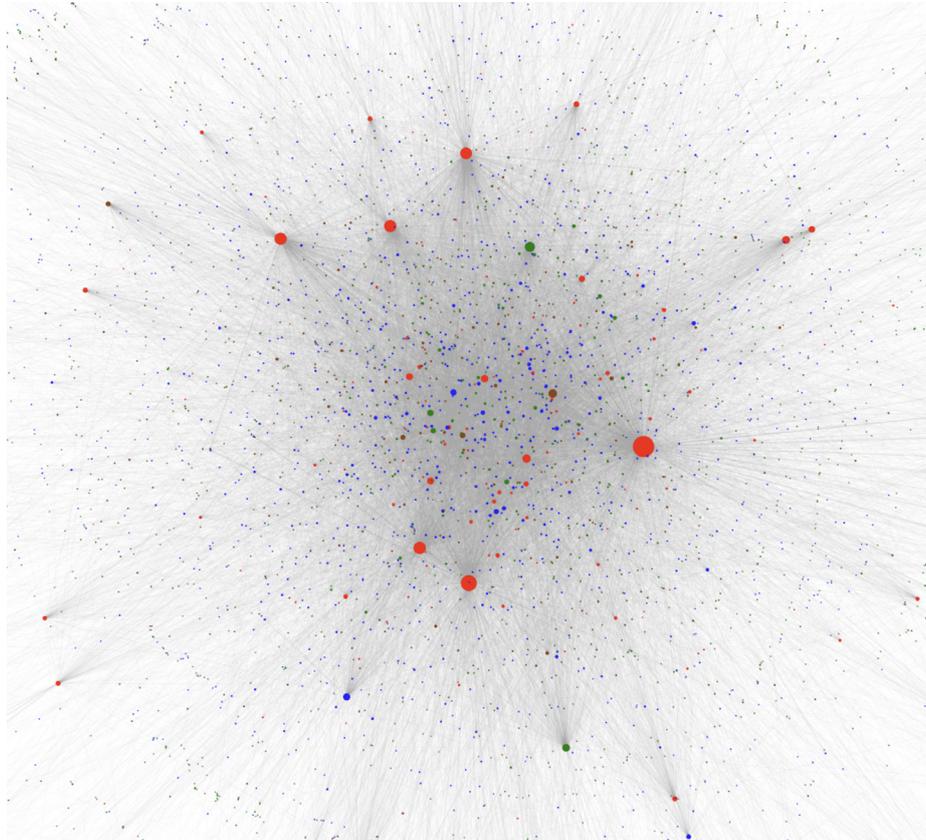


Figura 28. Red lingüística compuesta por 147.637 palabras de las que 1.487 son palabras incluidas en la capa léxica. El tamaño de los nodos es proporcional a su grado de salida, y el color del nodo indica la capa a la que dicho nodo pertenece.: Fuente: A. Criado-Alonso et. al (2021, p. 110509)

Así, los nodos de la capa léxica son azules, los verbos son verdes, las palabras de enlace son rojas, y las palabras restantes (no caracterizadas) son de color marrón. En la Fig. 29, podemos seguir los enlaces dirigidos entre las palabras para completar la frase "Many complex systems have been studied by complex network theory" (en español, "Muchos sistemas complejos han sido estudiados por la teoría de redes complejas", frase que aparece como tal en el corpus). Obsérvese que cada palabra tiene el color de la capa a la que pertenece.

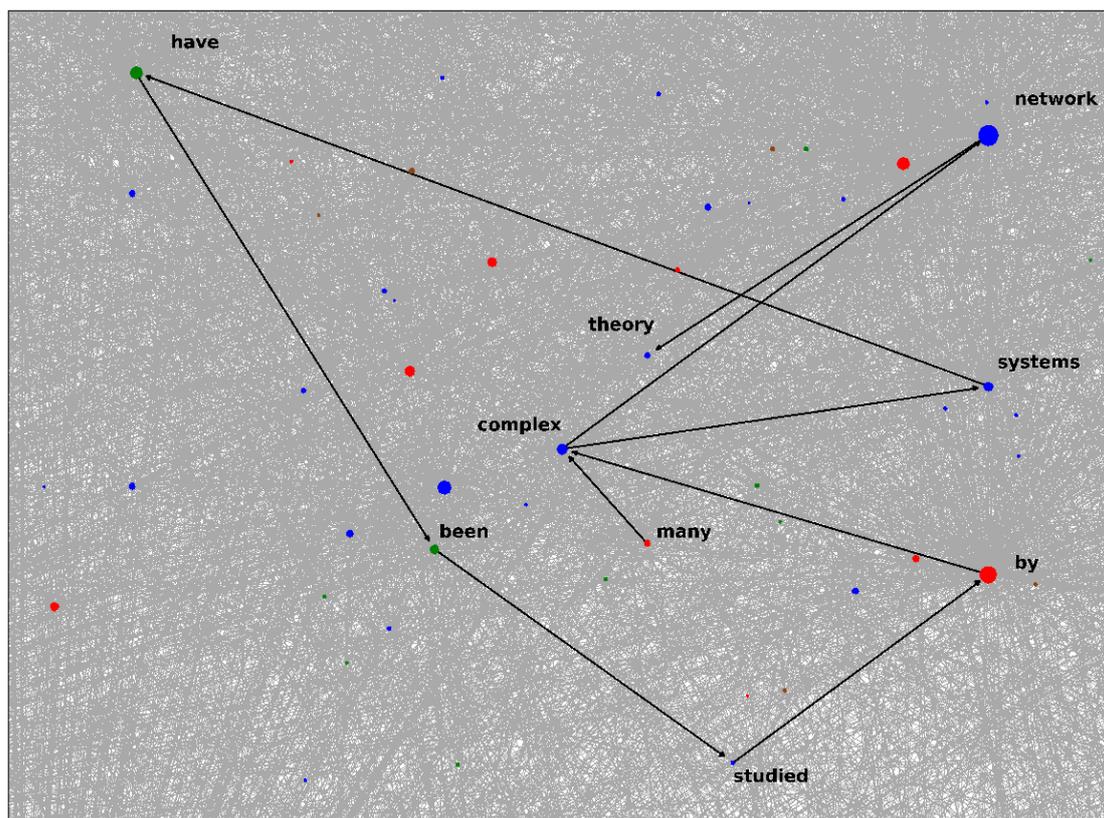


Figura 29.: Ampliación de la red compleja multicapa estudiada. El color de cada nodo indica la capa a la que dicho nodo pertenece. Fuente: A. Criado-Alonso et. al (2021, p. 110509)

Para detectar cuáles son las colocaciones de especialidad, resultan de especial interés las conexiones intra-capa de la capa léxica (que, como decíamos antes, son las que han sido caracterizadas como de especialidad matemática) ya que nos permitirán experimentar con las colocaciones lingüísticas hasta ver, mediante una técnica que describiremos en el siguiente apartado, cuáles son las que efectivamente caracterizan el lenguaje matemático de especialidad.

3.7 Resultados computacionales sobre el lenguaje de especialidad presente en el corpus

Ya hemos descrito los rasgos generales de los diferentes tipos de palabras que se incluyen en cada capa, pero hay ocasiones en las que el mismo término es susceptible de aparecer simultáneamente en más de una de dichas capas, por lo que hay que establecer la correspondiente prioridad de unos criterios sobre otros. Con el fin de otorgar una mayor contundencia a esta explicación, utilizaremos términos del lenguaje matemático de redes directamente en castellano por ser este el idioma en el que se redacta esta tesis.

Partimos de la base de la que ya hemos hablado en apartados anteriores, de que cada término de especialidad sufre una ligera (o completa) alteración en su significado cuando se refiere a un área de conocimiento determinada o aparece en la misma (por ejemplo, la palabra "caso" no hace referencia al mismo concepto si hablamos en un contexto legal que el concepto de la palabra "caso" cuando hablamos de un "caso clínico" en el ámbito de la medicina). Todos los enfoques sobre el lenguaje de especialidad coinciden en que algunos términos son especiales en una lengua porque se manifiestan únicamente en un área de conocimiento determinada y/o no se refieren al mismo concepto cuando se utilizan en el lenguaje común, como en el caso anteriormente descrito de la palabra "caso".

Así pues, el primer paso para describir un lenguaje de especialidad consiste en identificar los términos de especialidad dentro del conjunto de términos y observar sus interacciones. Además, el modelo también incluye los términos típicos de la lengua en la que está escrito el texto (o términos estándar que pertenecen al lenguaje común, sin alterar su significado). De este modo, y tal y como hemos mencionado anteriormente, los términos o palabras de los de los textos de muestra del corpus con el que trabajamos se han distribuido en diferentes capas para representar las palabras como nodos en una red multicapa de la siguiente manera:

1. Capa léxica. En esta capa hemos incluido tanto los términos que pertenecen exclusivamente al lenguaje matemático de la especialidad como los términos sujetos a referirse a otros diferentes conceptos cuando no aparecen en el contexto matemático.

Consideramos que todos estos términos son "unidades léxicas" o "palabras específicas relevantes". Las palabras incluidas en esta capa son, por ejemplo: *networks, cluster, node, nodes, ...*

2. Capa de verbos (o capa verbal). En esta capa hemos incluido todos los verbos independientemente de su conjugación. Es muy importante resaltar que cada tiempo del mismo verbo que aparece en el corpus tiene asociado un nodo diferente en esta capa. Algunas palabras incluidas como nodos en esta capa son, por ejemplo: *are, is, using, be, may, ...*

3. Capa de enlaces (o capa de "linking words"). En esta capa hemos incluido palabras con relevancia gramatical o sintáctica en lo que respecta a la construcción de oraciones y a su vinculación en lo que se refiere tanto a la corrección de oraciones como a la corrección gramatical. Las preposiciones, los conectores, los adverbios de frecuencia y los determinantes responden a esta descripción. Algunas palabras incluidas como nodos en esta capa son, por ejemplo: *en (in), otro (other), para (for), dentro de (within), además (besides) ...*

4. Capa resto (o capa de palabras restantes). En esta capa hemos incluido las palabras que no sufren ninguna alteración cuando aparecen en el texto matemático porque, o bien son palabras del lenguaje cotidiano (no matemático), o son palabras de otro campo de especialidad al que se aplica la teoría de redes complejas.

Las capas descritas anteriormente se clasifican según su relevancia respecto a cuáles son más útiles a la hora de caracterizar las unidades de conocimiento especializado presentes en el corpus de la siguiente forma: 1-Lexical, 2-Verbo, 3-Enlace, 4-Palabras restantes. Por eso, en caso de confusión (consideramos que hay confusión cuando una palabra es susceptible de aparecer en dos o más capas, pudiendo ser considerada bien como un verbo o bien como adjetivo, por ejemplo, *study*-sustantivo y *study*-verbo), el criterio seguido consiste en dar prioridad a la capa léxica, excepto en el caso de las palabras determinadas por el experto asesor lingüístico como exclusivamente útiles como verbos o exclusivamente útiles como unidades léxicas. En

cualquier caso, el orden de prioridad para asignar una palabra (como nodo) a una capa concreta es el orden descrito anteriormente.

Una vez que se han dividido las palabras en las diferentes capas, podemos aplicar el algoritmo *PageRank* para ordenar los términos de la capa léxica y ver, de este modo, cuáles son las palabras que caracterizan este lenguaje de especialidad:

1) <i>network</i>	2) <i>networks</i>	3) <i>nodes</i>	4) <i>system</i>	5) <i>information</i>	6) <i>number</i>
7) <i>model</i>	8) <i>structure</i>	9) <i>systems</i>	10) <i>properties</i>	11) <i>new</i>	12) <i>node</i>
13) <i>complex</i>	14) <i>data</i>	15) <i>graph</i>	16) <i>dynamics</i>	17) <i>study</i>	18) <i>process</i>
19) <i>distribution</i>	20) <i>analysis</i>	21) <i>models</i>	22) <i>algorithm</i>	23) <i>links</i>	24) <i>problem</i>
25) <i>only</i>	26) <i>state</i>	27) <i>case</i>	28) <i>approach</i>	29) <i>degree</i>	30) <i>connected</i>
31) <i>use</i>	32) <i>paper</i>	33) <i>graphs</i>	34) <i>interest</i>	35) <i>pagerank</i>	36) <i>interactions</i>
37) <i>value</i>	38) <i>way</i>	39) <i>linear</i>	40) <i>matrix</i>	41) <i>multiplex</i>	42) <i>associated</i>
43) <i>topology</i>	44) <i>phase</i>	45) <i>example</i>	46) <i>theory</i>	47) <i>important</i>	48) <i>real</i>
49) <i>critical</i>	50) <i>main</i>	51) <i>extinction</i>	52) <i>topological</i>	53) <i>method</i>	54) <i>behavior</i>
55) <i>functions</i>	56) <i>proposed</i>	57) <i>local</i>	58) <i>random</i>	59) <i>patterns</i>	60) <i>evolution</i>
61) <i>population</i>	62) <i>processes</i>	63) <i>present</i>	64) <i>size</i>	65) <i>applied</i>	66) <i>related</i>
67) <i>high</i>	68) <i>weighted</i>	69) <i>agents</i>	70) <i>dynamical</i>	71) <i>large</i>	72) <i>clustering</i>

73) <i>measures</i>	74) <i>line</i>	75) <i>function</i>	76) <i>layer</i>	77) <i>space</i>	78) <i>measure</i>
79) <i>controllability</i>	80) <i>account</i>	81) <i>image</i>	82) <i>step</i>	83) <i>section</i>	84) <i>communities</i>
85) <i>control</i>	86) <i>similar</i>	87) <i>level</i>	88) <i>corresponding</i>	89) <i>interaction</i>	90) <i>positive</i>
91) <i>error</i>	92) <i>structural</i>	93) <i>simple</i>	94) <i>growth</i>	95) <i>performance</i>	96) <i>values</i>
97) <i>application</i>	98) <i>initial</i>	99) <i>shape</i>	100) <i>result</i>		

Tabla 2: Las cien primeras palabras de la capa léxica ordenadas según su importancia o *PageRank* en la red

La tabla 2 nos da una idea de cuáles son los términos o palabras clave más significativas e importantes, pero con vistas a entender, poder traducir y caracterizar el lenguaje matemático, nos interesa analizar las asociaciones entre las palabras de la capa léxica, que es lo que nos ha permitido caracterizar las siguientes colocaciones o unidades de conocimiento específico según su importancia:

1) <i>complex-network</i>	2) <i>topological-properties</i>	3) <i>network-analysis</i>	4) <i>centrality-measures</i>
5) <i>network-structure</i>	6) <i>minimum-number</i>	7) <i>structural-controllability</i>	8) <i>multiplex-network</i>
9) <i>degree-distribution</i>	10) <i>explosive-percolation</i>	11) <i>mutualistic-networks</i>	12) <i>conditional-variance</i>
13) <i>mutualistic-networks</i>	14) <i>linear-system</i>	15) <i>network-topology</i>	16) <i>distribution-networks</i>
17) <i>concept-terms</i>	18) <i>complex-networks</i>	19) <i>community-detection</i>	20) <i>structural-controllability</i>
21) <i>personalization-vector</i>	22) <i>distribution-networks</i>	23) <i>simulated-networks</i>	24) <i>extinction-scenario</i>

25) <i>total-number</i>	26) <i>community-structure</i>	27) <i>pollination-networks</i>	28) <i>random-network</i>
29) <i>data-analysis</i>	30) <i>competitvity-graph</i>	31) <i>multiplex-networks</i>	32) <i>conditional-volatility</i>
33) <i>power-law</i>	34) <i>network-model</i>	35) <i>phase-transition</i>	36) <i>new-friends</i>
37) <i>connected-components</i>	38) <i>shape-classification</i>	39) <i>formal-languages</i>	40) <i>memory-phase</i>
41) <i>synthetic-networks</i>	42) <i>dynamical-system</i>	43) <i>large-networks</i>	44) <i>cayley-tree</i>
45) <i>phase-transitions</i>	46) <i>important-role</i>	47) <i>weighted-networks</i>	48) <i>relevance-measure</i>
49) <i>network-dimensions</i>	50) <i>network-oscillations</i>	51) <i>operational-capacity</i>	52) <i>structural-properties</i>
53) <i>clustering-coefficient</i>	54) <i>smallworld-networks</i>	55) <i>real-networks</i>	56) <i>graph-theory</i>
57) <i>similarity-matrix</i>	58) <i>highly-connected</i>	59) <i>positive-systems</i>	60) <i>positive-linear</i>
61) <i>variance-equation</i>	62) <i>collective-dynamics</i>	63) <i>sustainable-efficiency</i>	64) <i>ordinary-percolation</i>
65) <i>concept-terms</i>	66) <i>multiplex-graphs</i>	67) <i>degree-distributions</i>	68) <i>dynamical-processes</i>
69) <i>strength-distributions</i>	70) <i>finite-size</i>	71) <i>generalist-core</i>	72) <i>phase-diagram</i>
73) <i>upper-bound</i>	74) <i>spatial-scales</i>	75) <i>vital-minimum</i>	76) <i>secondary-bifurcation</i>
77) <i>networks-theory</i>	78) <i>linear-continuous</i>	79) <i>linear-system</i>	80) <i>new-approach</i>
81) <i>new-nodes</i>	82) <i>real-sdh</i>	83) <i>numerical-simulations</i>	84) <i>network-growth</i>

85) <i>algebraic-topology</i>	86) <i>mutualistic-communities</i>	87) <i>leafwing-population</i>	88) <i>square-matrices</i>
89) <i>extinction-sequence</i>	90) <i>boundary-conditions</i>	91) <i>multiagent-systems</i>	92) <i>unconnected-node</i>
93) <i>synaptic-plasticity</i>	94) <i>propagation-algorithm</i>	95) <i>adjacent-nodes</i>	96) <i>random-networks</i>
97) <i>information-flows</i>	98) <i>information-processing</i>	99) <i>consensus-problem</i>	100) <i>percolation-model</i>

Tabla 3: Las cien primeras colocaciones de especialidad ordenadas según su importancia o *PageRank* en la red.

En la lista de colocaciones anterior, hay que destacar que el método utilizado para identificarlas, descrito por primera vez en el artículo publicado recientemente cuya referencia es Criado-Alonso et al. (2021), tiene un enfoque sintagmático, diferente del enfoque paradigmático común a otros trabajos de lingüística computacional. Este tipo de enfoque consiste en estudiar las relaciones entre las palabras dentro del paradigma. Se puede entender por paradigma el conjunto de palabras que pueden aparecer en el mismo lugar de una estructura lingüística. Por ejemplo, si tenemos una construcción (adjetivo + sustantivo) con una función sintáctica (presentes en el sujeto de una frase), el primer elemento del paradigma es el conjunto de palabras que pueden aparecer en esa posición (es decir, los adjetivos en el paradigma adjetivo + sustantivo) y el segundo elemento del paradigma es el conjunto de palabras que pueden aparecer en esa posición (es decir, los sustantivos). De este modo, la relación paradigmática se basa en el valor por oposición que poseen los elementos lingüísticos. Este tipo de enfoque es, como decíamos, el que hasta ahora se había hecho en los estudios estadísticos previos de lingüística computacional.

Para este caso concreto, se ha puesto el foco en las relaciones sintagmáticas, es decir, en las relaciones que podemos encontrar entre las colocaciones que se han expuesto previamente en este apartado. El uso del *Line Graph* nos permite identificar cuándo una asociación entre palabras es lo suficientemente fuerte como para cobrar un sentido único y convertirse en una unidad de conocimiento específico. Dentro de una unidad de conocimiento específico o colocación, no se puede sustituir un elemento por

sentido único y convertirse en una unidad de conocimiento específico. Dentro de una unidad de conocimiento específico o colocación, no se puede sustituir un elemento por otro (como ocurre en el caso de las relaciones paradigmáticas del que se hablaba anteriormente). Son palabras que aparecen juntas y adquieren un significado diferente como conjunto. Así, podemos pensar en el mismo ejemplo de “red compleja”, “red” y “red ferroviaria” para aclarar el concepto una vez más.

Este tipo de asociaciones sintagmáticas son las que presentan las estructuras de los *phrasal verbs* en inglés. Un ejemplo de phrasal verb es: *look up to* (que se traduciría por “admirar”); aquí se ve claramente cómo no se pueden sustituir unos elementos por otros sin que cambie el significado del conjunto entero. Este último ejemplo no es un par de palabras o colocación (ya que las colocaciones son pares de palabras), pero sí una unidad de conocimiento que un traductor debe identificar y comprender. La creación del modelo multicapa establecido es realmente útil ya que la interacción entre capas es lo que facilita la descripción de la formación de los verbos de especialidad. Además, teniendo en cuenta que el sintagma es la unidad sintáctica básica, este modelo permite describir la conversión de estas unidades en otras nuevas según su significado sintáctico (siendo capaz de contemplar unidades o palabras simples, colocaciones y construcciones formadas por más palabras).

No hay que olvidar que se están utilizando conceptos y algoritmos como el *PageRank* en lugar de las metodologías tradicionales utilizadas habitualmente por algunos lingüistas en los que la frecuencia de aparición de una palabra ocupa un lugar central. En la nueva metodología descrita, uno de los aspectos más importantes considerados consiste en analizar las relaciones entre palabras, de manera que una palabra aparecerá en los primeros lugares del *ranking* no tanto por el hecho de que su frecuencia de aparición en el corpus considerado sea alta, como por el hecho de que aparezca relacionada secuencialmente muchas veces con otra palabra importante (es decir, otra palabra que, siguiendo este algoritmo, está destinada a aparecer también en los primeros puestos del *ranking*). El proceso de identificación de estas unidades se hace a través de lo que hemos llamado “*Enriched Line Graph of a Weighted Network*” (Grafo Enriquecido de una Red Pesada) y que permite centrar la búsqueda de estas unidades con relaciones sintagmáticas.

Así, si dos palabras o nodos aparecen fuertemente conectados (es decir, si dichas palabras aparecen muchas veces simultáneamente una seguida de la otra), la arista que los une se convertirá en un nuevo nodo. Si vemos el grafo de la figura 30 con nodos etiquetados con números y pesos en las aristas que representan la frecuencia de la aparición secuencial de los nodos que unen, se puede apreciar en dicha figura un ejemplo en el que la frecuencia de aparición de los grupos formados por dos palabras es distinta. De acuerdo con el estudio, se pueden establecer diferentes umbrales. Para mejor comprensión, en la figura hemos considerado como valores de umbral tres y uno y vemos que solo aquellas aristas cuyos pesos son mayores o iguales que tres se convierten en nuevos nodos. De este modo, tenemos un grafo obtenido del original enriquecido con nuevos nodos que representan a aristas distinguidas, que serán las que darán lugar a la aparición de colocaciones. A partir de este grafo enriquecido podemos eliminar aquellas palabras (o nodos) que únicamente aparecen en colocaciones (es decir, en combinación con otras palabras). Así se obtiene el *Semi-Line Graph* correspondiente al umbral 3 que se representa en la figura donde aparecen únicamente las colocaciones importantes (nodos azules) y aquellos nodos o palabras que no aparecen únicamente en colocaciones para el umbral escogido.

Por ejemplo, la arista que conecta el nodo 1 con el nodo 2 se convertiría en un nodo (el nodo 12) si consideramos que el un peso en las aristas superior a 3 es lo suficientemente fuerte como para formar una colocación o par de palabras separado. Lo mismo sucedería con la arista que conecta el nodo 5 y el nodo 1; quedaría convertida en el nodo 51.

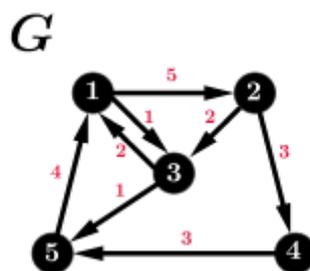


Figura. 30: Grafo con nodos etiquetados mediante números y pesos en las aristas

Estableciendo el umbral en 3, se convertiría asimismo en nodo la arista que va del 2 al 4, y la que va del 4 al 5.

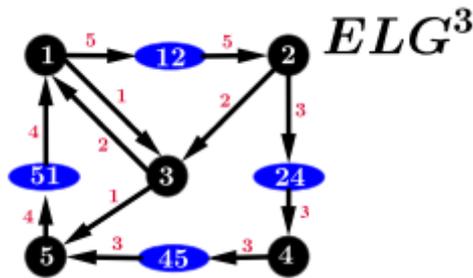


Figura 31: Enriched Line Graph generado a partir de la Figura 30 con el umbral en 3

Tal y como se ve en la figura 31, el peso de la arista que ha dado lugar a un nuevo nodo se repite en la arista que precede al nuevo nodo y en la que sigue a ese nuevo nodo incorporado en el *Enriched Line Graph* pesado que se ha construido según el umbral.

Si consideramos que lingüísticamente hablando hay veces que las palabras aparecen sólo en combinación (o en forma de colocación), se puede establecer de nuevo un umbral para quedarnos solo con las que nos interesan para en este caso, los lenguajes de especialidad. El estudio de los lenguajes de especialidad utilizando la metodología facilitada por esta construcción se puede hacer a partir del establecimiento y el estudio de los *Line Graph* Enriquecidos y *SemiLine Graphs* obtenidos al utilizar diferentes umbrales.

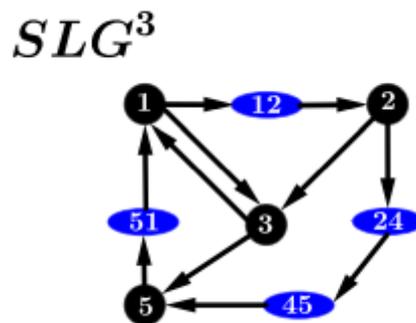


Figura 32: Imagen del Semi Line Graph obtenido con el umbral en 3

En la figura 32, se puede apreciar como el nodo etiquetado como número 4 sería una de esas palabras desaparecería como palabra de interés al considerar el umbral 3. Estableciendo el umbral en 3, vemos como, por ejemplo, el nodo etiquetado con el número 2 no desaparecería puesto que hay aristas que parten del mismo hacia nodos que no son azules. En otras palabras, el nodo etiquetado con el número dos aparece en el corpus en frases en las que no intervienen secuencialmente con ninguna de las colocaciones que aparecen con el umbral considerado. Tal como hemos dicho, este nuevo concepto es aplicable a estructuras lingüísticas de más de dos palabras considerando un nuevo *Line Graph* enriquecido construido sobre un *Line Graph* enriquecido previamente definido. Para ilustrar el ejemplo de selección de términos más allá del lenguaje de especialidad tomaremos el hecho de que en inglés existe el concepto de *black mail*¹³ pero no el concepto de “*white mail*”. La palabra *black* también puede aparecer con *trousers* formando “*black trousers*” (pantalones negros). Existen los “*white trousers*” (pantalones blancos) pero no existe la colocación “*white mail*”. Al aparecer un nuevo nodo correspondiente a una colocación, el nuevo *PageRank* que se le asigna al nodo correspondiente a la primera palabra de la colocación recientemente aparecida es el resultado de “restarle” al *PageRank* original de dicho nodo el *PageRank* que se lleva el nuevo nodo (nueva colocación lingüística). El umbral para los pesos permite controlar qué aristas se convertirán en nodos, evitando así que entren nodos como “*white mail*” en la red compleja.

¹³ Según el diccionario monolingüe de inglés Merriam-Webster (<https://www.merriam-webster.com/>), *blackmail* tiene las siguientes acepciones:

- 1: un tributo antiguamente exigido en la frontera escocesa por los jefes saqueadores a cambio de inmunidad contra el pillaje
- 2a: extorsión o coacción mediante amenazas, especialmente de exposición pública o enjuiciamiento penal
- b: el pago de dinero por extorsión

3.8 Las colocaciones y el lenguaje de especialidad. Presentación del Kendall (τ) para este caso

Para entender qué se ha hecho exactamente para afirmar que este lenguaje de especialidad es más sensible a las colocaciones, es importante hablar del parámetro numérico que en matemáticas se conoce como “tau de Kendall (τ)”. El coeficiente de correlación conocido como la tau (τ) de Kendall es una herramienta matemática estándar utilizada para comparar clasificaciones u ordenaciones diferentes establecidas para el mismo conjunto de elementos. Por lo tanto, si tomamos dos clasificaciones (ordenaciones, en principio diferentes) c_1 y c_2 de un conjunto de n elementos, el coeficiente de correlación “ τ de Kendall” referido a estas ordenaciones concretas, se define de la siguiente manera:

$$\tau(c_1, c_2) = \frac{K(c_1, c_2) - K(c_1, c_2)}{(n^2)}$$

Fuente: Criado-Alonso et al. (2020, p. 69)

En la expresión anterior, $K(c_1, c_2)$ denota el número de pares (i, j) que no cambian su posición relativa mutua con respecto a ambas clasificaciones c_1 y c_2 , y $K(c_1, c_2)$ denota el número de pares (i, j) que cambian su posición relativa mutua (de uno respecto a otro) en esa ordenación (o “*ranking*”). Es importante destacar que

$$(n^2)$$

es el número de todos los pares posibles de nodos (i, j) , y que, teniendo en cuenta la fórmula que lo define, el coeficiente de Kendall varía entre -1 y $+1$. Así, si la ordenación (o “*ranking*”) c_1 coincide (es exactamente igual) a la ordenación (“*ranking*”) dada por c_2 , el valor del coeficiente de correlación $\tau(c_1, c_2) = 1$, si hay muchos pares de nodos que han variado su posición relativa de c_1 a c_2 , el coeficiente de correlación $\tau(c_1, c_2)$ tomará valores cercanos a 0 y, si finalmente la ordenación de los nodos establecida por c_1 es justamente la inversa de la ordenación dada por c_2 el valor del coeficiente $\tau(c_1, c_2) = -1$. Dicho en otras palabras, si los órdenes son totalmente inversos, el valor de este coeficiente sería igual a -1 (lo que está en último lugar pasa a estar en el primero); si los órdenes son iguales, obtendremos un valor igual a 1 y si se trata de un

orden distinto, nos moveremos en valores alejados de 1 y, dependiendo de lo diferentes que sean dichas ordenaciones c_1 y c_2 , puede que en valores cercanos a 0.

3.9 Centralidades *PageRank* de los términos (unidades) del corpus. Mejora del uso del *PageRank* sobre la frecuencia absoluta para detectar palabras importantes

Una primera aproximación de la centralidad (o importancia) de cada uno de los términos del corpus nos la da la frecuencia con la que dicha palabra aparece en el corpus. Sin embargo, al poner nuestro foco de atención sobre las relaciones entre los términos, la centralidad de cada término dependerá no sólo de su frecuencia de aparición, sino también de la frecuencia de aparición de “sus nodos vecinos”, es decir, de aquellos otros términos que aparecen antes o después de cada término en las diferentes frases que componen el corpus. Esto nos lleva a considerar un concepto de centralidad frecuentemente utilizado, que es el que nos ofrece el algoritmo *PageRank* (que es el empleado por Google en la presentación ordenada de los resultados de sus búsquedas ordenadas de páginas web según la “importancia” de dichas páginas web en el contexto de toda la World Wide Web). De ahí que un término será “más importante” y, en consecuencia, aparecerá en los primeros lugares del *ranking* si, no solamente aparece de manera frecuente en las frases del corpus, sino que, además, los términos que le preceden en dichas frases son también “términos importantes”. Los diferentes *rankings* de términos ordenados de acuerdo con su centralidad o importancia son los que se obtienen a partir de la aplicación del algoritmo *PageRank* sobre la red objeto de nuestro análisis. En este sentido los valores de la centralidad (o importancia) obtenidos para dichos términos del corpus se corresponden, intuitivamente, con “la frecuencia” con la que un caminante aleatorio que se mueve aleatoriamente de un nodo a otro de la red a través de las aristas pasa por dichos nodos (o por dichas aristas). El algoritmo *PageRank* también emplea un “salto aleatorio” que, en nuestro caso, se produce al finalizar cada frase del corpus. Al realizar dicho “salto aleatorio”, el caminante aparecerá en otro término, de manera que la probabilidad de que el salto aleatorio tenga como destino un nodo concreto está relacionada con la frecuencia relativa con la que dicho nodo (término) aparece al principio de las frases de los textos que componen el corpus.

Las probabilidades de que cada nodo de la red sea el nodo de destino de un salto aleatorio se explicitan en lo que se denomina “vector de personalización”, es decir, una secuencia de tantos valores numéricos como nodos tiene la red, en la que a cada nodo i se le asocia la probabilidad de que cuando el paseante aleatorio da un salto aleatorio, el nodo de destino sea precisamente dicho nodo i .

Así, para obtener los diferentes *rankings* de acuerdo con la centralidad o importancia de los términos del corpus, hemos utilizado el algoritmo *PageRank* con un vector de personalización para el que hemos calculado la frecuencia de aparición de cada una de las palabras que aparecen al comienzo de cada frase del corpus. Por lo tanto, la probabilidad o valor de la componente del vector de personalización asociado al *PageRank* de un término (o palabra específica) del corpus es el resultado obtenido añadiendo 1 a su frecuencia relativa de aparición al principio de las frases de los textos corpus, dividido por el número de nodos. De este modo, si v es el vector de personalización, sus componentes son, $\forall i \in \{1, \dots, n\}$,

$$v_i = \frac{1 + f_i}{n + \sum_i f_i},$$

donde f_i es la frecuencia descrita anteriormente. Dicho de otra manera, si una palabra (o término) no aparece al principio de cualquier frase en un texto del corpus, el valor de la probabilidad correspondiente a dicha palabra en el vector de personalización será relativamente pequeña.

En todo caso, y tal y como se plantea en Criado-Alonso et al. (2020), el estudio matriz, es importante señalar que para el cálculo de *PageRank* utilizado a lo largo de todo este trabajo se ha utilizado el algoritmo de cálculo descrito en Aleja et al. (2019). En la Tabla 2, se puede contemplar el ranking de las diez palabras con mayor *PageRank* de toda la red multicapa, así como el *ranking* de las diez palabras con mayor *PageRank* dentro de cada una de las capas diferentes en las que se ha dividido el corpus previamente.

	General	Lexical words	Verbs	Linking words	Remaining words
1 st	<i>the</i>	<i>network</i>	<i>is</i>	<i>the</i>	<i>we</i>
2 nd	<i>of</i>	<i>networks</i>	<i>are</i>	<i>of</i>	<i>it</i>
3 rd	<i>to</i>	<i>nodes</i>	<i>be</i>	<i>to</i>	<i>different</i>
4 th	<i>and</i>	<i>system</i>	<i>can</i>	<i>and</i>	<i>one</i>
5 th	<i>in</i>	<i>information</i>	<i>have</i>	<i>in</i>	<i>not</i>
6 th	<i>a</i>	<i>number</i>	<i>has</i>	<i>a</i>	<i>two</i>
7 th	<i>is</i>	<i>model</i>	<i>will</i>	<i>that</i>	<i>social</i>
8 th	<i>that</i>	<i>structure</i>	<i>using</i>	<i>this</i>	<i>based</i>
9 th	<i>are</i>	<i>systems</i>	<i>results</i>	<i>for</i>	<i>them</i>
10 th	<i>we</i>	<i>properties</i>	<i>been</i>	<i>with</i>	<i>species</i>

Tabla 4: Las diez primeras palabras de toda la red separadas en capas. Fuente: Criado-Alonso et. Al 2020 p. 69

Con el fin de apreciar mejor la diferencia obtenida entre la ordenación de los diferentes términos obtenida mediante el algoritmo *PageRank* y mediante la frecuencia, hemos construido la siguiente tabla que muestra las primeras diez palabras ordenadas según su centralidad obtenida al utilizar el algoritmo *PageRank*, en comparación con cómo quedarían si se ordenan según su frecuencia absoluta y su frecuencia relativa (obsérvese que, por ejemplo, *PageRank* considera “más importante” el término “*and*” que los términos “*in*” y “*a*”, pese a que su frecuencia absoluta (y relativa) es menor que la de estos últimos, y lo mismo sucede con los términos “*that*” y “*are*” en relación con el término “*we*”).

	PageRank centrality	Absolute frequency	Relative frequency
the	0.078	1886	0.090
of	0.048	1007	0.048
to	0.028	627	0.029
and	0.026	507	0.024
in	0.025	597	0.028
a	0.024	563	0.026
is	0.018	412	0.019
that	0.012	258	0.012
are	0.010	212	0.010
we	0.009	299	0.014

Tabla 5: Las 10 primeras palabras de toda la red ordenadas según la centralidad de su *PageRank*, su frecuencia absoluta y su frecuencia relativa de aparición. Fuente: Criado-Alonso et al. (2021 p. 110509)

Así pues, en la Tabla 5 se puede observar que la centralidad *PageRank* ordena las palabras de manera diferente a cómo aparecerían ordenadas dichas palabras si las ordenásemos según su frecuencia de aparición en los textos del corpus lingüístico con el que estamos trabajando, y que, a pesar de la especificidad y la especialidad de los textos del corpus, la ley de Zipf se ve satisfecha (aproximadamente) si se tiene en cuenta la ordenación dada por las frecuencias. En todo caso es razonable que se cumpla la Ley de Zipf también textos escritos en lenguaje matemático.

3.10 Resultados computacionales: sobre la estructura de este lenguaje matemático de especialidad y sus colocaciones principales

Una vez establecido el modelo matemático de red multicapa para definir lingüísticamente el corpus considerado, analizamos el lenguaje matemático de especialidad contenido en el corpus comparando la sensibilidad de las variaciones de la estructura de las colocaciones con la sensibilidad de las variaciones palabras aisladas en función de cambios “artificialmente realizados” en la frecuencia o intensidad de su aparición utilizando para la comparación de los diferentes rankings el coeficiente de correlación τ Kendall.

Con el fin de realizar este análisis, modificamos artificialmente el peso de las aristas, en primer lugar, de la capa léxica (Fig. 33) multiplicando cada arista de esta capa por un parámetro $\alpha \in (0, 10)$. Esta modificación simula un cambio en la intensidad de aparición de los términos de la capa léxica (términos especializados). Así, cuando $0 < \alpha < 1$ estamos “disminuyendo artificialmente” la frecuencia de aparición de dichos términos y si $\alpha > 1$ estamos “aumentando artificialmente” la frecuencia de aparición de dichos términos. Obviamente, cuando $\alpha = 1$ obtenemos el *ranking* “real” correspondiente a los valores precisos de la frecuencia real con la que aparecen los términos en la capa léxica. Colocando en el eje vertical (“eje de ordenadas”) los valores correspondientes a la τ de Kendall de cada una de las diferentes capas consideradas, observamos cómo al variar artificialmente los pesos correspondientes a la frecuencia en los términos de la capa léxica se producen variaciones significativas en las ordenaciones de los términos de otras capas, resultando especialmente relevante las variaciones producidas en la ordenación de las “colocaciones” (términos especializados) de la capa léxica.

De igual modo, en la Fig 34. se puede apreciar cómo al variar artificialmente los pesos (frecuencias) de las aristas que enlazan nodos de la capa verbal con nodos de la capa de enlace (“*linking words*”) se producen otro tipo de variaciones en las diferentes capas. Así pues, con este análisis hemos podido medir y comparar la sensibilidad de variaciones en la estructura de las colocaciones (o unidades de conocimiento especializado) con la sensibilidad de variaciones de palabras aisladas.

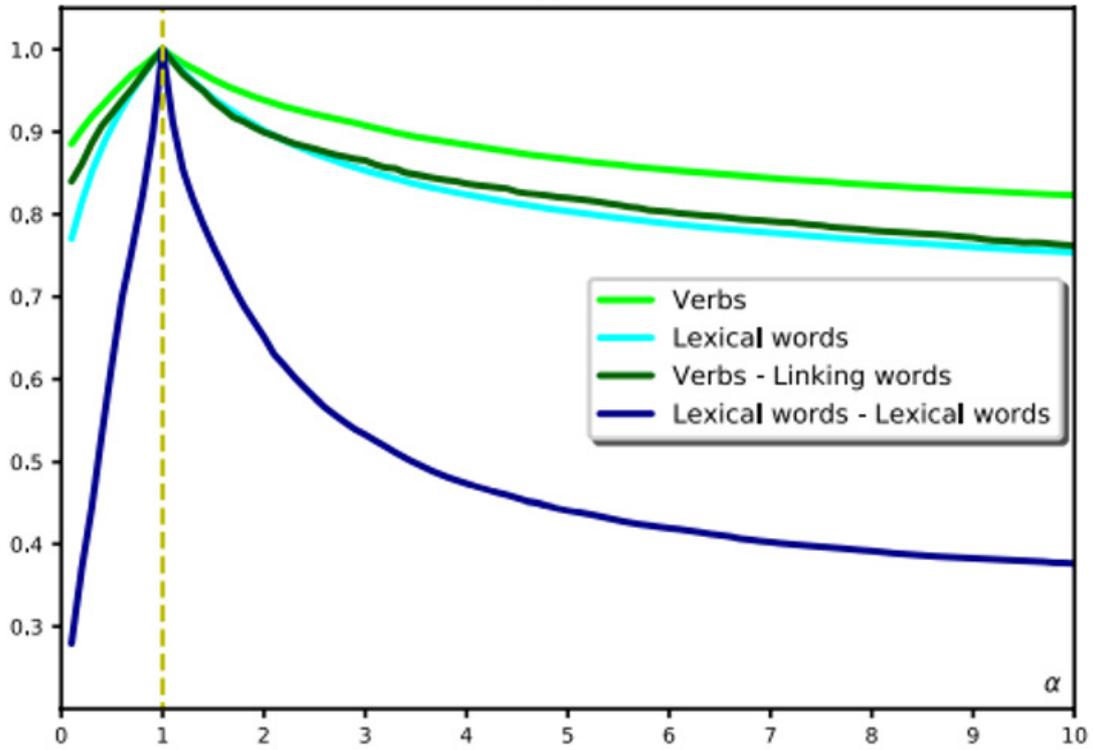


Figura 33. Fuente Criado-Alonso et. al (2020 p.70) coeficiente Kendall τ de los cuatro subgrupos variando los pesos en la capa léxica en términos del parámetro multiplicativo de los pesos denotados por α .

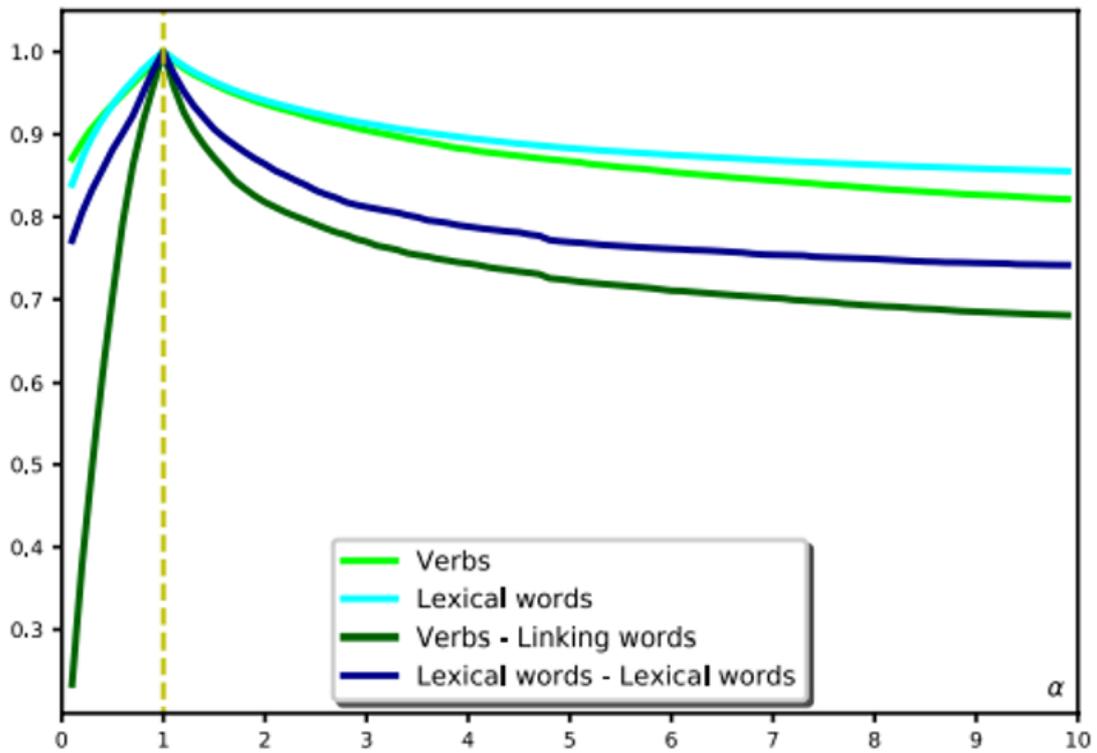


Figura 34. Fuente: Criado-Alonso et al. (2020 p.69) coeficiente Kendall τ de cuatro subgrupos variando los pesos: verbos- palabras hilo en términos del parámetro multiplicativo de los pesos denotados por α .

Para el cálculo de los coeficientes Kendall de las figuras anteriores (Fig.33 y Fig.34) hemos considerado las 500 primeras palabras de cada capa [ver anexo] ordenadas según el algoritmo *PageRank* y también de los dos grupos de colocaciones consideradas (palabra léxica-palabra léxica, verbo-palabra enlace). El *ranking* “real” (es decir, el correspondiente a la red original), en el que las palabras aparecen ordenadas según su *PageRank* se corresponde con el valor $\alpha = 1$. La Figura 34 muestra los cambios importantes que se producen en las ordenaciones de las colocaciones “palabra léxica-palabra léxica” cuando variamos artificialmente la frecuencia de las palabras de la capa léxica (el parámetro α correspondiente).

Es importante resaltar una vez más que para obtener el *PageRank* de las colocaciones entre palabras léxicas (Palabra Léxica-Palabra Léxica) se ha utilizado el método planteado en Criado (2018) originando en el caso de esta red lingüística lo que son unidades de conocimiento específico que se corresponden con colocaciones lingüísticas. También es importante destacar que al variar artificialmente la frecuencia de las aristas “verbo-palabra de enlace” (“*verb-linking word*”), según se muestra en la Fig. 34, los cambios en la ordenación de las colocaciones léxicas (lex-lex) son similares a los cambios producidos en los pares “verbo-palabra enlace”. Sin embargo, la diferencia entre cómo varían estas ordenaciones o rankings al variar artificialmente el parámetro α correspondiente a la frecuencia de los términos de la capa léxica y la variación correspondiente a la capa (lex-lex) en comparación con la variación de la capa “verbo-palabra de enlace” en la Fig.8 es abrumadoramente mayor y significativa. Estas son las colocaciones que caracterizan el lenguaje de especialidad.

La Figura 34 también muestra que los cambios producidos en los rankings al variar los pesos en las aristas correspondientes a los pares “verbos -Palabras de enlace” son similares a los cambios producidos en los rankings correspondientes a verbos y palabras léxicas, siendo las diferencias mucho menos sustanciales que en el caso reflejado en la figura anterior (Fig. 33.), con lo que vemos que el “lenguaje de especialidad de las colocaciones” es mucho más sensible.

A modo ilustrativo, se incluye la Fig. 35 a continuación, donde, si consideramos las 15 colocaciones léxicas que aparecen en primer lugar en la capa léxica (capa de especialidad) según el orden dado por el algoritmo *PageRank* y que corresponde (para $\alpha = 1$), en la Fig. 35 podemos ver las variaciones en la clasificación de estas colocaciones

al cambiar artificialmente el peso de los nodos (términos) de la capa léxica. En la misma Fig. 35 también se puede observar la diferente forma en que varían los rankings en los casos $\alpha \in (0, 1)$ y $\alpha \in (1, 10)$.

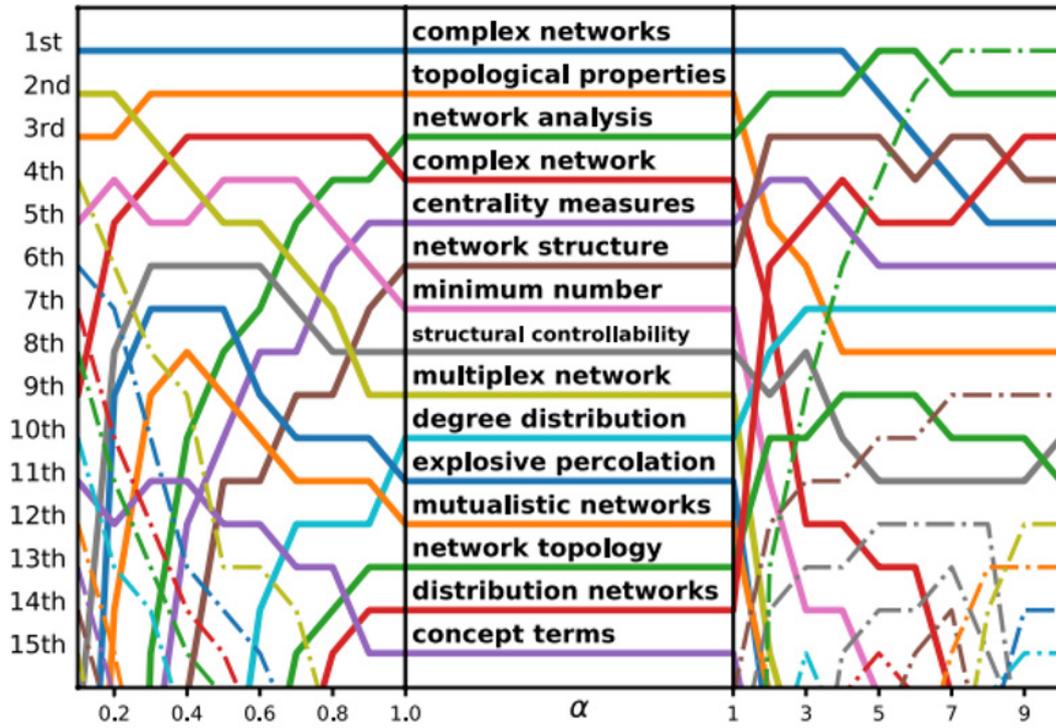


Fig. 35. Ordenación de las colocaciones léxicas variando los pesos de la capa léxica según α . Fuente: Criado-Alonso et al. (2020 p. 69)

Conviene señalar que algunas de estas colocaciones no ven muy alterada su posición en el ranking al variar el parámetro α (por ejemplo, la posición en la clasificación de *complex networks* cambia poco al variar el parámetro α . De hecho, su posición es la primera en el ranking para $\alpha \in (0, 4)$ y termina en la fuera de posición cuando $\alpha > 8$). Por el contrario, hay otras colocaciones que caen drásticamente de estas 15 primeras posiciones del ranking al variar α (por ejemplo, *multiplex network* y *mutualistic network*), e incluso algunas colocaciones aparecen o desaparecen de estas primeras 15 posiciones para valores de α entre 0 y 1 y cuando $\alpha > 1$ (estos casos vienen representados por líneas de puntos en la figura anterior, Figura 35).

3.11 Recapitulaciones sobre el estudio del corpus mediante redes complejas

A partir de la metodología empleada y descrita en este capítulo y de la constatación de la existencia e interrelaciones entre los cuatro lenguajes que configuran el lenguaje de especialidad matemático (el lenguaje simbólico-gráfico-numérico matemático, el lenguaje de especialidad matemático verbalizado, el lenguaje natural en el que se expresa el texto y el lenguaje del área de especialidad al que se aplica la teoría de redes complejas) es posible extraer conclusiones lingüísticas discerniendo el ámbito al que pertenece cada uno de los términos del corpus y haciendo una distribución supervisada de cada uno de estos términos (transformados en nodos) en cada una de las cuatro capas de la red compleja multicapa considerada (la capa léxica, la capa verbal, la capa de enlaces y la capa resto) .

Además, la metodología de análisis empleada es suficientemente sensible para perfilar qué pares de palabras de la capa léxica han de aparecer juntas aludiendo a una única realidad del ámbito de especialidad (cuya combinación constituye lo que se denominan colocaciones) a partir de la fortaleza de su conexión en la capa léxica, e incluso, la observación de las interconexiones entre los términos de la capa de enlaces y la capa verbal para el estudio de los verbos de especialidad. Para realizar este estudio ha sido especialmente relevante la definición del nuevo concepto de *Enriched Line Graph* (“Grafo Lineal Enriquecido”) que incluye en su definición la posibilidad de que sus nodos sean colocaciones.

El análisis realizado a partir del estudio computacional sobre el corpus estudiado muestra que la estructura de colocaciones en este corpus es mucho más sensible a las variaciones en la intensidad de su aparición en los textos de especialidad que la estructura de las palabras aisladas del propio texto de especialidad, apuntando a una característica intrínseca en la estructura lingüística de este tipo de textos. La metodología y el modelo descritos pueden constituir una herramienta muy útil para los traductores que trabajan con textos especializados en el área de conocimiento correspondiente al corpus estudiado, una tarea especialmente difícil si el traductor no

es especialista en el campo concreto al que se refiere el corpus, pues el uso específico de palabras existentes y de combinaciones específicas de palabras constituyen, esencialmente, la terminología específica de un lenguaje de especialidad, lo que resulta, en ocasiones, una barrera difícil de superar para el traductor no especialista en el área de conocimiento a la que pertenece el texto específico a traducir.

Capítulo 4: Conclusiones

Conclusiones parciales

Las gráficas presentadas en los últimos apartados de la parte práctica ponen de manifiesto que el procedimiento utilizado, fundamentado en la teoría y en las herramientas propias de las redes complejas, permite una evaluación más adecuada del comportamiento sintagmático (léxico, sintaxis,...) que el de otro tipo de enfoques que, basados en la frecuencia, están orientados a la relación paradigmática. Las herramientas de análisis que proporcionan las redes complejas abren nuevas posibilidades de uso de los datos lingüísticos para lograr una mayor precisión semántica. Además, el uso en el futuro de estas herramientas proporcionadas por las redes complejas en corpus más generales –independientemente la lengua en la que estén escritos los textos– se contempla para procesos automáticos como pueden ser la verificación, la identificación, la traducción o la generación automática de textos.

Los resultados de nuestro estudio computacional revelan que la estructura de las colocaciones en este corpus de textos de especialidad matemática es mucho más sensible a las variaciones en la intensidad de su aparición en los textos de especialidad que las palabras aisladas del propio texto de especialidad, lo que apunta a una característica intrínseca emergente en la estructura lingüística de este tipo de textos. Así pues, como conclusión relevante, cabe destacar que, dado que las colocaciones son mucho más sensibles que las palabras aisladas a las variaciones en la intensidad de su aparición dentro de la lengua de especialidad, la estructura lingüística y la expresión de su significado adjunto son coherentes y forman una unidad de conocimiento especializado de la que nos habla Castellví (2004) y que se corresponde con una colocación.

Esta tesis arroja nueva luz sobre enfoques anteriores descritos por autores como Parodi (2008), cuya orientación está enfocada a desentrañar los sistemas del lenguaje a través de conjuntos de textos, y cuya metodología se puede aplicar a cualquier otro corpus utilizando el texto paralelo, lo que en el caso de corpus de textos de lenguaje matemático no parece efectivo. El trabajo desarrollado ha demostrado que el estudio del lenguaje de especialidad matemático tiene muchas unidades de conocimiento

específico formadas por dos palabras y que, en muchas ocasiones, son más significativas que las formadas por una sola palabra. Por ello es muy importante tener muy presente qué palabra va en el texto antes o después de la palabra considerada. Este hecho se inscribe en la línea de afirmaciones sobre colocaciones como las que hace Bosque (2001, pág. 10): “el hecho de que entre varias combinaciones compatibles de una serie de piezas léxicas, una o dos entre ellas predominen marcadamente sobre las demás es un hecho que no debe de ser desestimado”.

Una de las líneas de investigación futuras hacia la que se puede dirigir este trabajo es el estudio de conjuntos de términos de más de dos nodos como nuevos vértices de una red compleja, de modo que se identifiquen no únicamente pares de palabras, sino también conjuntos de ellas como unidades de conocimiento específico. La teoría de redes complejas permite la observación desde un punto de vista no basado en la frecuencia absoluta de los términos, sino en las asociaciones entre términos establecidas mediante un valor alto del *PageRank* de su correspondiente nexo, de tal manera que se pueda afirmar con seguridad que lo identificado (o asociado) como un nuevo nodo de la red constituirá un elemento singular característico del lenguaje de especialidad analizado.

Conclusiones sobre el estudio del lenguaje de especialidad matemático a través de las redes complejas. Trabajo futuro.

La primera conclusión es que el lenguaje matemático es catalogable como un lenguaje de especialidad en sí mismo según los criterios de Pérez Hernández (2002) ya que posee las siguientes características: especificidad del tema; restricción léxica, semántica y sintáctica; reglas gramaticales específicas y poco frecuentes; frecuencia de aparición de estructuras lingüísticas particulares; macroestructuras de texto homogéneas; y, por último, el uso de símbolos específicos. Asimismo, hemos discutido el motivo por el cual la mayor parte del lenguaje matemático en el ámbito de la divulgación científica se muestra en inglés y, partiendo de la base de que se creó en enseñanza el concepto de la enseñanza del “inglés para fines específicos” (ESP, *English for Specific Purposes*), se crearon las bases para la investigación referente al funcionamiento de lo que se denomina LFE (Lenguas para Fines Específicos).

Los criterios que recoge Pérez Hernández (2002) para definir o caracterizar un lenguaje de especialidad son seis: especificidad del tema; restricción léxica, semántica y sintáctica; reglas gramaticales específicas y poco frecuentes; frecuencia de aparición de estructuras lingüísticas particulares; macroestructuras de texto homogéneas; y, por último, el uso de símbolos específicos. Todos estos criterios permiten caracterizar el lenguaje matemático como un lenguaje de especialidad siendo esta una de las primeras conclusiones que se pueden extraer de esta tesis doctoral.

Tal como resaltan Alcalá & Antuña (2011), se ha utilizado el inglés como lengua base para la comunicación del conocimiento especializado, ya que en los años sesenta era la lengua inglesa la que se utilizaba para la comunicación científica y esto dio lugar a que los científicos se interesaran en aprender ese “inglés para fines específicos” (ESP, *English for Specific Purposes*).

Santamaría-Pérez (2006) parte de la base de que las LFE (Lenguas para Fines Específicos) se caracterizan por el uso de términos y conjuntos de términos de manera especializada y habla concretamente de “Unidades de Conocimiento Especializado”, que son términos o grupos de términos de difícil comprensión por su pertenencia a una temática específica. Así, cuanto más específica es la temática de un texto, más difícil es la tarea del traductor no especializado ya que el texto en sí mismo presentará una mayor dificultad de comprensión, puesto que estas unidades de conocimiento específico estarán compuestas de una serie de términos o palabras que existen y/o actúan de manera específica dentro de un área del saber.

El estudio de la terminología específica en las LFE pone de manifiesto el fenómeno conocido como *terminologización*, que es el proceso mediante el cual un término adquiere un significado diferente dentro de un área de especialidad concreta al que adquiere dentro de los parámetros del lenguaje común (Tallová 2004 pág. 39). Hasta el momento, la metodología más eficaz para el análisis de las LFE y su terminología es el uso de la lingüística de corpus. El uso de la lingüística de corpus viene avalado por autores como Llamazares (2008, pág. 349), que afirman que el uso de corpus para el estudio de las lenguas se considera una metodología empírica de trabajo ya que esta se basa en el empleo de datos reales y de muestras de uso de la lengua.

Entre las aportaciones de esta tesis doctoral destaca la presentación de una herramienta novedosa concebida a partir de la Teoría de Redes Complejas y consistente

en la representación y creación de un modelo de red multicapa dirigida y pesada, cuyos componentes se generan a partir de los elementos del corpus considerado, facilitando su procesamiento informático (mediante el uso de Python 3.7), herramienta que no solo permite analizar los componentes paradigmáticos de una unidad de conocimiento específico a la hora de realizar una traducción, sino que también permite analizar las relaciones sintagmáticas que existen entre los términos de un corpus. Esto facilita la identificación de las colocaciones léxicas, de pares de palabras (por ejemplo, verbo-palabra enlace) o, incluso, de un conjunto de ellas. Así, la herramienta desarrollada es capaz de diferenciar entre *network* y *complex network* estableciendo *complex network* como una unidad de significado.

Sus resultados pueden aplicarse al mundo de la traducción (automática o no) y a la redacción de textos relativos a este lenguaje de especialidad. Sin esta perspectiva de análisis basada en la teoría de redes complejas y el uso de la red compleja dirigida y pesada que se describe en los apartados relativos a la metodología, la lingüística de corpus podía identificar mediante el uso de textos paralelos palabras que aparecían de forma consecutiva una delante de la otra, pero no se podía afirmar con la necesaria precisión que estas fueran unidades de conocimiento específico del ámbito considerado. Convirtiendo los términos en nodos y utilizando conceptos como el de *Line Graph* asociado a una red, el algoritmo *PageRank* y el novedoso concepto propio de este modelo matemático presentado como *Enriched Line Graph* en Criado-Alonso et. al (2021), se pueden identificar unidades de significado del lenguaje de especialidad analizado y elaborar una lista de palabras y conjuntos de palabras clave que caracterizan, en este caso, al lenguaje de especialidad matemático propio del corpus descrito (aunque hay que resaltar que este novedoso enfoque es aplicable a cualquier otro lenguaje de especialidad siempre que esté debidamente acotado).

También cabe resaltar que este procedimiento es aplicable a cualquier corpus de otra especialidad en caso de que se aplique el mismo etiquetado para extraer los datos intrínsecos a los distintos lenguajes de especialidad. El convertir los términos en nodos y usar la teoría de las redes complejas aporta una mayor precisión semántica en el análisis. Además, esta metodología que convierte un corpus en red, deriva en diferentes puntos de vista o enfoques a través de los que se puede llegar a la generación de textos automáticos, la identificación del estilo de un autor. Incluso puede contabilizarse el

grado de especialidad de un texto si se le asigna un valor numérico a los términos utilizados; si un texto tiene “x” palabras en la capa léxica y “x” colocaciones extraídas a partir del *LineGraph*, por comparación con el resto de los documentos de los que conste el corpus, es susceptible de que se le asigne un valor numérico hablando así del grado de especialización.

De cara al futuro, esta metodología supone aportar una nueva perspectiva de análisis dentro del contexto de los lenguajes de especialidad. El modelo matemático que se presenta en la parte práctica de esta tesis doctoral es una perspectiva de análisis que va más allá de simplemente concebir el corpus como una red compleja, puesto que permite la extracción de conclusiones semánticas mucho más precisas que las que aportan hasta ahora las perspectivas de análisis relacionadas con la lingüística de corpus.

Las redes y sistemas complejos del mundo real suscitan un gran interés. La investigación que emplea las redes complejas desde el punto de vista de las lenguas es capaz de generar nuevos modelos matemáticos de interés para la propia comunidad interesada en la propia teoría de redes complejas y en la comunidad interesada en el estudio de las lenguas (y es de especial ayuda para la lingüística de corpus, área que ha arrojado mucha luz sobre los lenguajes de especialidad).

Para concluir esta tesis doctoral, debemos detenernos en varias cuestiones acerca del uso del enfoque de redes complejas para extraer conclusiones acerca del lenguaje matemático de especialidad.

Hemos ofrecido argumentos que permiten sostener que los modelos y medidas de las redes complejas proporcionan una metodología necesaria y funcional para concebir las lenguas como un sistema –por ejemplo, Cong & Liu (2014)–. Así, tal y como resaltan estos autores, en algunas teorías que conciben la lengua o el lenguaje como un sistema se hace hincapié en que el lenguaje es una red, tal y como sucede en Lamb (1999) o Hudson (2010). A diferencia de estos trabajos, en el modelo desarrollado en esta tesis doctoral, en el que se ha empleado una red multicapa formada por cuatro capas distintas, las relaciones entre palabras se desarrollan tanto entre aquellas palabras pertenecientes a la misma capa (conexiones intra-capa) como entre palabras pertenecientes a diferentes capas (conexiones inter-capa). Es decir, en ninguno de estos trabajos citados se había empleado el concepto de red multicapa introducido recientemente en la bibliografía (noviembre de 2014). Por ejemplo, Hudson (2010)

argumenta que el léxico, junto con las relaciones entre las palabras en varios niveles de análisis, es una red (compleja) en la que, obviamente, sus elementos no aparecen distribuidos en capas según su ámbito lingüístico de especialidad y/o función gramatical.

En este caso, hemos usado esa premisa, la de que el lenguaje es una red compleja, para estudiar el léxico de especialidad ya que, con los métodos de análisis lingüístico tradicionales, la lengua como sistema (o concebida como una red) apenas puede examinarse en términos empíricos para obtener conclusiones relevantes como han sido, en nuestro caso, las palabras más importantes según el *PageRank* o las unidades de significado o colocaciones de especialidad del lenguaje matemático. Asimismo, se ha podido observar que el lenguaje matemático de especialidad usado por la comunidad científica que publica sobre redes complejas se caracteriza precisamente por el uso de unidades de significado compuestas de dos términos y esto es una conclusión relevante ya que normalmente cuesta distinguir qué es una unidad de significado en términos de lenguaje de especialidad. Así pues, en esta tesis, se ha seguido la línea descrita por Ferrer-i-Cancho (2014), que afirma que la estadística no desempeña un papel central en la obra de Saussure y de muchos de sus sucesores.

El estudio de las redes complejas aplicado a las lenguas contribuye a cambiar esta situación y permite examinar las propiedades cuantitativas de la lengua cuando esta se concibe como un sistema. Así, las unidades de conocimiento específico, se basan en el uso observable del lenguaje por parte de la comunidad científica de redes complejas. Tal y como señala Cong & Liu (2014) aunque aceptemos la distinción entre el sistema lingüístico ideal y el uso real de la lengua, no hay ninguna buena razón para creer que ambos estén totalmente desconectados. Las redes lingüísticas basadas en el uso real del lenguaje pueden, al menos, arrojar luz sobre el sistema lingüístico ideal y abstracto.

Por otra parte, una de las principales ventajas del uso de las redes complejas en la investigación de los lenguajes de especialidad es la capacidad de captar las propiedades holísticas (o globales) del lenguaje que vienen dadas a través de las medidas topológicas de los modelos de redes lingüísticas. Las propiedades topológicas de las redes complejas lingüísticas caracterizan la macroestructura del lenguaje para la que el todo es mayor que la suma de las partes. La caracterización global de la lengua con el enfoque de las redes complejas no puede lograrse con los métodos de análisis

lingüísticos tradicionales que suele centrarse en los rasgos lingüísticos a nivel más básico.

Gracias al uso del enfoque de redes complejas, también hemos podido analizar características lingüísticas más allá del nivel de la frase Chen (2014). De cara a futuro, ya se están haciendo investigaciones prometedoras como las que se mencionan a lo largo de esta tesis doctoral y las que se citan en Cong & Liu (2014). De hecho, con modelos y medidas más allá de los rasgos lingüísticos detallados extraídos en los análisis lingüísticos tradicionales, el enfoque de redes complejas puede utilizarse para la caracterización del lenguaje humano y aplicarse como metodología potencial para ramas lingüísticas como la tipología lingüística o la adquisición de la primera lengua, donde una caracterización global del lenguaje puede desempeñar un papel relevante.

La utilización de una red de cuatro capas permite ajustarnos a la máxima destacada por Chen (2014), que apunta a que las propiedades topológicas de las redes lingüísticas pueden integrar la información relativa a los rasgos lingüísticos en diferentes niveles. Por su parte, Köhler (2014) compara cómo se tratan los datos lingüísticos extraídos a partir de la investigación de las redes lingüísticas y los datos extraídos a partir de la lingüística cuantitativa tradicional, poniendo de manifiesto las ventajas asociadas a la utilización del primer tipo de tratamiento de datos. Además, también considera acertadamente la diferencia entre las medidas de las redes complejas y las de la lingüística cuantitativa tradicional como la distinción entre datos relacionales y datos de carácter atributivo. Al centrarse en los textos en red, el enfoque de las redes complejas emplea medidas relativas a las relaciones entre unidades lingüísticas, en lugar de las relativas a las propiedades de las unidades lingüísticas individuales, como la frecuencia de aparición absoluta y la longitud de frase.

Las ventajas de las redes complejas en la investigación lingüística no imponen un análisis “privilegiado” sobre los demás métodos de investigación lingüística. La modelización adecuada de la lengua depende de la pregunta de investigación que se quiera abordar. Las redes complejas son apropiadas cuando se desea caracterizar un lenguaje en términos de unidades lingüísticas y sus relaciones. En este caso, además, se han identificado las unidades de conocimiento específico compuestas por pares de palabras o colocaciones a través del *Line Graph*.

Sin embargo, hay casos para los cuales los modelos y medidas que ofrecen las redes complejas pueden no ser de utilidad; por ejemplo, en Cong & Liu (2014), se especifica que Macutek (2014) no está del todo de acuerdo con la afirmación de que las redes complejas pueden captar el lenguaje en su totalidad o de manera global tal y como decíamos en párrafos anteriores. Macutek (2014) acierta en que cualquier modelo es una reducción de la realidad, con los elementos y relaciones relevantes considerados y los demás elementos y relaciones descartados. Por "lenguaje en su totalidad" no se pretende afirmar que el enfoque de redes complejas pueda abarcar todos los aspectos de una lengua, sino que los modelos y medidas de redes pueden arrojar luz sobre las propiedades globales de las unidades lingüísticas y sus relaciones de un tipo particular como un todo en red. Como cualquier modelo de red lingüística, el modelo de red desarrollado en esta tesis doctoral, cubre una parte o enfoque de los muchos métodos de análisis o enfoques desde los que se puede estudiar una lengua. El análisis de redes y la comparación de diferentes niveles lingüísticos como subsistemas pueden reflejar mejor la naturaleza multinivel del lenguaje humano. Un adecuado análisis de redes complejas del lenguaje depende de la selección de modelos de redes lingüísticas. Es decir, el modelo de red adecuado depende de lo que se quiera estudiar.

Tal y como se expone en esta tesis doctoral, las redes complejas son una herramienta para entender el lenguaje humano, y, además, son de gran ayuda para entender el lenguaje de especialidad. Los resultados del análisis que aportan las redes complejas o los resultados de las redes del lenguaje sobre el propio lenguaje necesitan una interpretación en términos de la lingüística y sus disciplinas cercanas. La comprensión de los resultados del análisis de redes del lenguaje depende en gran medida de la comprensión de los modelos de redes lingüísticas.

Según Hudson (2014), la lingüística proporciona formalismos detallados para el análisis de las relaciones entre unidades lingüísticas, que determinan los vértices y las aristas en varios tipos diferentes de redes lingüísticas. Por ejemplo, en esta tesis doctoral consideramos las redes lingüísticas como redes de valencia interdependiente donde la valencia se entiende como cuantificada por la importancia de la palabra (el *PageRank* o centralidad) dado por el contexto en el que aparece. En estas redes, cada unidad lingüística o palabra se considera como un vértice o nodo y se encuentra en el contexto de otras unidades con las que puede combinarse para formar relaciones de un tipo

determinado. La conectividad de cada vértice es un indicador directo o indirecto de la capacidad combinatoria de la unidad lingüística correspondiente. La propiedad libre de escala de una red lingüística dinámica, por tanto, significa la heterogeneidad en la distribución de capacidad combinatoria de las unidades lingüísticas en el subsistema lingüístico correspondiente.

Además de este trabajo, otros autores como Hudson (2014), hacen uso de la teoría de las redes complejas para describir el lenguaje afirmando que las redes lingüísticas dinámicas relacionan los subsistemas lingüísticos con el uso real de la lengua. De hecho, tal como se señala en Cong & Liu (2014), el enfoque basado en el uso del lenguaje humano (Tomasello, 2009) considera que la representación mental del lenguaje se basa en los ejemplos. En otras palabras, el conocimiento del lenguaje surge sobre la base del uso real del lenguaje al que está expuesto.

Con este supuesto, las propiedades topológicas de una red lingüística compleja dirigida pesada como la descrita en este trabajo arrojan luz sobre los patrones de organización de la representación mental del sistema lingüístico que se sigue al utilizar el lenguaje de especialidad matemático relativo a las redes complejas.

No obstante, hay que destacar que las redes complejas están lejos de ser una metodología lingüística madura. La investigación de los lenguajes de especialidad a través de las redes complejas concebidas como redes lingüísticas necesita un esfuerzo metodológico adicional ya que los especialistas en redes complejas no suelen ser especialistas en lenguaje, lenguas o lingüística. Los modelos y las medidas de las redes complejas necesitan una mayor evaluación para mejorar su aplicabilidad a las diferentes ramas lingüísticas en las que podrían desempeñar un papel importante. Hay que considerar la posibilidad de utilizar un conjunto coherente de modelos y medidas de red para el mismo tipo de cuestiones de investigación.

En cualquier caso, se pueden generar modelos de red exportables a cualquier corpus de análisis. Parece que la relación entre las macro y microestructuras del lenguaje humano es una línea de investigación de especial interés y trascendencia. Por ejemplo, Zhao (2014) considera esta línea como una de las direcciones más importantes de investigación lingüística desde el enfoque de las redes complejas, apuntando en la misma dirección que el trabajo desarrollado en esta tesis doctoral, y, Chen (2014), por su parte, reclama una mayor conexión entre los estudios de redes lingüísticas y los de la

lingüística dominante fundamentados en estudios realizados a través de enfoques más tradicionales. Además, Köhler (2014) considera que las redes lingüísticas son un campo joven de la lingüística cuantitativa y sugiere que los estudios de las redes lingüísticas se combinen con los de la lingüística cuantitativa tradicional. Uno de los retos a los que se enfrenta este intento, según él, es cómo interpretar las leyes del lenguaje a nivel de sistema que revelan los patrones estadísticos de las redes lingüísticas. Estas leyes a nivel de sistema son diferentes de las leyes descubiertas en la lingüística cuantitativa tradicional.

En Cong & Liu (2014) se menciona que otra posible aplicación de la teoría de redes para el estudio de las lenguas bastante destacable es, por ejemplo, la defendida por Gong et al. (2014) donde sugieren que la dinámica evolutiva de una lengua se rastree mediante la comparación topológica de redes lingüísticas basadas en datos lingüísticos de diferentes periodos de tiempo. Además, según Gong et al. (2014), el análisis de las redes lingüísticas basado en datos lingüísticos diacrónicos puede revelar nuevos datos sobre el contacto entre diferentes lenguas a lo largo de la historia.

Como ha demostrado Ferrer-i-Cancho (2014), la ciencia de las redes también proporciona herramientas más allá de la descripción de las redes del mundo real, es decir, modelos de la dinámica de las redes. Los modelos ayudan a explicar por qué las redes del mundo real son como son. También pueden desempeñar un papel en la reconstrucción de la dinámica evolutiva del lenguaje humano (Liang, Shi & Huang, 2014). En futuras investigaciones, es necesario determinar si los modelos de dinámica de redes pueden tener en cuenta múltiples factores relacionados con el lenguaje (por ejemplo, lingüísticos, cognitivos y socioculturales) para poder desarrollar mecanismos plausibles de evolución del lenguaje a partir del enfoque de redes complejas. Además de esta investigación teórica, las redes complejas también constituyen una herramienta útil para campos prácticos como el procesamiento del lenguaje natural. Amancio (2014) estudia brevemente cómo pueden utilizarse los modelos de redes para el procesamiento del lenguaje natural y ofrece sugerencias para futuras investigaciones. Aunque el enfoque de redes complejas del lenguaje humano, como tema emergente, se enfrenta a problemas, su futuro es prometedor si se tienen en cuenta sus ventajas y los nuevos conocimientos que ya ha aportado.

En lo que a la disciplina de la traducción se refiere, esta tesis doctoral aporta una metodología de análisis a un lenguaje de especialidad que hasta ahora no se había estudiado en profundidad y, gracias a la esta nueva perspectiva de análisis, se han identificado las principales unidades de significado compuestas de dos palabras o colocaciones fundamentales para constituir ese lenguaje de especialidad. Además, gracias a las propiedades de la red compleja multicapa pesada y dirigida se ha podido ver que son precisamente los pares de palabras los que caracterizan el lenguaje de especialidad matemático que tiene que ver con la teoría de redes complejas.

También hemos llegado a la conclusión de que este lenguaje es en sí mismo un lenguaje de especialidad –no se había catalogado previamente como tal– y que este lenguaje, cuando se expresa en un documento escrito, está a su vez dividido en cuatro registros o sub-lenguajes: lengua vehicular inglés, lenguaje simbólico-numérico matemático, lenguaje de especialidad relativo a la teoría de redes complejas y lenguaje de especialidad de la disciplina a la que se está aplicando la teoría de las redes complejas como tal. Esto sucede porque en términos de lenguaje de especialidad, nos encontraremos con palabras del lenguaje de especialidad matemático de redes complejas como tal, así como terminología específica del área de investigación sobre el que se está aplicando la teoría de redes complejas. La dificultad añadida que supone la existencia e interacción de estos subtipos de lenguaje matemático de especialidad de redes hace que el estudio de este lenguaje de especialidad a través de un corpus de textos sea el enfoque pertinente.

En definitiva, se ha caracterizado el lenguaje de especialidad de redes complejas a través de listas de palabras y colocaciones que no se habrían podido obtener de otra manera y se ha observado cómo interactúan los subtipos de lenguaje de especialidad matemático de redes complejas para obtener estas listas de palabras y colocaciones más importantes. Asimismo, se ha desarrollado un modelo matemático de red compleja multicapa único para este fin de identificar y ordenar las colocaciones de especialidad este corpus según su *PageRank* que es una metodología exportable a cualquier corpus de análisis, llegando más allá de los métodos de análisis lingüísticos de corpus que se hacían tradicionalmente basados en la frecuencia de aparición.

La demostración de que el lenguaje matemático de redes complejas es un lenguaje de especialidad en sí mismo caracterizado por unidades de significado

compuestas de dos palabras y compuesto de cuatro subtipos de lenguaje es un resultado relevante si se quiere, por un lado, facilitar la tarea de traducción de un texto de esta área de especialidad; y, por otro, si lo que se desea abrir una nueva ventana a la investigación de las lenguas desde este nuevo punto de vista. Por último, también hemos tratado de contribuir al desarrollo de cierta normalización en lo que a la redacción de este tipo de textos se refiere cuando el traductor actúa como terminólogo de especialidad.

Bibliografía

- Albert, R., Albert, I., & Nakarado, G. (2004). Structural vulnerability of the North American power grid. *Phys. Rev. E* 69, 025103.
- Alcalá, A., & 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*, 907-932.
- Alcaraz Varó, E. (2000). *El inglés profesional y académico*. Alianza.
- Aleja, D., Criado, R., del Amo, A., Pérez, Á., & Romance, M. (2019). Non-backtracking PageRank: from the classic model to Hashimoto matrices. *Chaos, Solitons & Fractals*(126), 283-291.
- Amancio, D. (2014). A perspective on the advancement of natural language processing tasks via topological analysis of complex networks: Comment on “Approaching human language with complex networks” by Cong and Liu. *Physics of Life Reviews*, 641-643.
- Anez, J., De La Barra, T., & Perez, B. (1996). Dual graph representation of transport networks. *Trans. Res. B* 30, 209-216.
- Atkins, B., Clear, J., & Ostler, N. (1992). Corpus Design Criteria. *Literary and Linguistic Computing*, 7(1), 1-16.
- Banks, J. (2008). *An introduction to multicultural education*.
- Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.
- Benito R. et al (2011) *Optical Communication Networks* en Int. J. Complex Systems in Science Vol 1. Págs 33-36
- Benson, A., Gleich, D., & Higham, D. (2021). Higher-order Network Analysis Takes Off, Fueled by Classical Ideas and New Data.
- Bezos, J. (2007). Ortotipografía y notaciones matemáticas. *BuenasTareas.com*.
- Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C., Romance, M., Sendiña-Nadal, I., . . . Zanin, M. (2014). The structure and dynamics of multilayer networks. *Phys. Rep.* 544 (1), 1-122.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Phys.Rep.* 424, 75-308.
- Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: structure and dynamics. *Entropy* 12:1264–130. *Entropy*(12), 1264-130.
- Bosque, I. (2001). Sobre el concepto de ‘colocación’ y sus límites. *Lingüística española actual*, 23(1), 9-40.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Comput.Netw.* 30, 107.
- Burrough-Boenisch, J. (2003). Shapers of published NNS research articles. *Journal of Second Language Writing*(12), 223-243.
- Butler, C. S. (1998). Collocational Frameworks in Spanish. *International Journal of Corpus Linguistics*, 3 Issue 1, 1-32. doi:<https://doi.org/10.1075/ijcl.3.1.02but>
- Byrne, J. (2014). Framework for the Identification and Strategic Development of Translation Specialisms. *Meta*, 1(59), 125.

- Cabré Castellví, M. (2004). *La terminología en la traducción especializada. Manual de documentación y terminología para la traducción especializada*. Madrid: Arco Libros, 89-125. (C. Gonzalo García, & V. García Yebra, Edits.) Madrid: Arco Libros.
- Cárdenas, J., Olivares, G., & Alfaro, R. (2014). Clasificación automática de textos usando redes de palabras. *Revista signos, estudios de lingüística* 86, 346-364.
- Centro Virtual Cervantes. (s.f.). *Centro Virtual Cervantes*. Obtenido de https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/linguisticacorporus.htm
- Chen, X. (2014). Language as a whole – A new framework for linguistic knowledge integration: Comment on “Approaching human language with complex networks” by Cong and Liu. *Physics of Life Reviews*, 11(4), 628-629. doi:<https://doi.org/10.1016/j.plrev.2014.07.011>.
- Cohen, R., Erez, K., Ben-Avraham, D., Havri, & Havril, S. (2001). Breakdown of the Internet under intentional attacks. *Physical Review Letters*(86), 3682.
- Cong, J., & Liu, H. (2014). Approaching human language with complex networks. *Physics of Life Reviews*, 11(4).
- Cong, J., & Liu, H. (2014). Linguistic complex networks: rationale, application, interpretation, and directions: reply to comments on "approaching human language with complex networks". *Physics of life reviews*, 11(4), 644-649.
- Congost Maestre, N. (1994). *Problemas de la traducción técnica: los textos médicos en Inglés*. Alicante: Publicaciones de la Universidad de Alicante.
- Criado, R., Flores, J., García del Amo, A., Gómez-Gardeñes, J., & Romance, M. (2012). A mathematical model for networks with structures in the mesoscale. *Int. J. of Computer Math.* 89, 3. *International Journal of Computer Math.*, 89(3), 291-309.
- Criado, R., Moral, S., Pérez, Á., & Romance, M. (2018). On the edges' PageRank and line graphs. *Chaos* 28, 075503.
- Criado-Alonso, Á. (2021). “Teoría de las redes complejas como herramienta para el análisis lingüístico de un corpus textual y la caracterización del lenguaje de especialidad matemático”. En: *Aplicaciones de las Humanidades Digitales en el ámbito de la Lingüística* (C.V. Herranz-Llacer, A. Segovia, L. Arroyo (eds.)). Ediciones CINCA S.A., pp 17-35.
- Criado-Alonso, Á., Battaner Moro, E., Aleja, D., Romance, M., & Criado, R. (2020). “Using complex networks to identify patterns in specialty mathematical language: a new approach”. *Social Network Analysis and Mining*, 10(1), pp. 1-10. (Springer-Nature).
- Criado-Alonso, Á., Battaner Moro, E., Aleja, D., Romance, M., & Criado, R. (2021a). “Enriched Line Graph: A new structure for searching language collocations”. *CHAOS. Chaos, Solitons and Fractals*, 142, 110509.
- Criado-Alonso, Á., Battaner Moro, E., Aleja, D., Romance, M., & Criado, R. (2021b). “A new mathematical structure with applications to computational linguistics and specialized text translation”. *Modelling for Engineering & Human Behaviour* 2021, pp. 60-66 [imm-Instituto de Matemática Multidisciplinar, UPV \(https://imm.webs.upv.es/jornadas/proceedings/Modelling2021.pdf\)](https://imm.webs.upv.es/jornadas/proceedings/Modelling2021.pdf)
- Crucitti, P., Latora, V., & Porta, S. (2006). Centrality in networks of urban streets. *Chaos*, 015113.
- Crucitti, P., Latora, V., & Porta, S. (2006). Centrality Measures in Spatial Networks of Urban Streets. *Physics Review*, E 73, 03612(5).

- Crucitti, P., Latora, V., & Porta, S. (2006). Network analysis of urban streets. *Physica A*, 0411241.
- Cuéllar, S. B. (2015). La lingüística de corpus: perspectivas para la investigación lingüística contemporánea. *Forma y Función*, 28(1), 31-54.
- Estopà, R. (2003). "La unidad de conocimiento especializado". *Curso de postgrado online de Introducción a la terminología*. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra.
- Europe, C. o. (2001). *Common European Framework of Reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press and the Council of Europe.
- Ferrer-i-Cancho, R. (2014). Beyond description: Comment on "Approaching human language with complex networks" by Cong and Liu. *Physics of Life Reviews*, 11(4), 621-623. doi:<https://doi.org/10.1016/j.plrev.2014.07.014>.
- Fillmore, C. (1992). *Corpus Linguistics or Computer-aided Armchair Linguistics*. J. Svartvik.
- Firth, J. R. (1957). A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis, Special Volume*, 1-32.
- Gómez Guinovart, X. (2000). Lingüística computacional. (E. X.-D. F. G. Ramallo, Ed.) *Manual de ciencias da linguaxe*, 221-268.
- Gong, T., Shuai, L., & Yicheng, W. (2014). Extending network approach to language dynamics and human cognition: Comment on "Approaching human language with complex networks" by Cong and Liu. *Physics of Life Reviews*, 11(4), 639-640.
- Gruber, T. R. (1995). Toward Principles for the Design of Ontologies used for Knowledge Sharing. *International Journal of Human and Computer Studies*, 43(5-6), 907-928.
- Guinovart, X. (2000). Aspectos lingüísticos de la neología terminológica informática en castellano, catalán y gallego. *Panorama actual de la lingüística aplicada [Recurso electrónico]: conocimiento, procesamiento y uso del lenguaje*, 1045-1502.
- Halliday, M., & Matthiessen, C. (2004). *Introduction to Functional Grammar* (3 ed.). London: Routledge.
- Hoffmann, L. (1998). *Llenguatges d'especialitat. Selecció de textos*. Barcelona: IULA-UPF.
- Howatt, A., & Widdowson, H. (2004). *History of ELT*. Oxford University Press.
- Hudson, R. (2010). *An introduction to word grammar*. Cambridge University Press.
- Hudson, R. (2014). The view from linguistics. Comment on "Approaching human language with complex networks" by Cong and Liu. *Physics of Life Reviews*, 11(4), 619-620. doi:10.1016/j.plrev.2014.06.013
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London. London: New York: Longman.
- Köhler, R. (2014). Linguistic complex networks as a young field of quantitative linguistics: Comment on "Approaching human language with complex networks" by J. Cong and H. Liu. *Physics of Life Reviews*, 11(4), 630-631. doi:<https://doi.org/10.1016/j.plrev.2014.07.016>.
- Koike, K. (2001). *Colocaciones léxicas en el español actual: análisis formal y léxico-semántico*. Universidad de Alcalá.
- König, D. (1990). Theory of finite and infinite graphs. In Theory of Finite and Infinite Graphs. *Birkhäuser Boston*, 45-421.
- Lamb, S. M. (1999). *Pathways of the brain: The neurocognitive basis of language* (Vol. 170). John Benjamins Publishing.

- Liang, W., Shi, Y., & Huang, Q. (2014). Modeling the Chinese language as an evolving network. *Physica A: Statistical Mechanics and its Applications*, 393, 268-276.
- Llamazares, M. V. (2008). Lingüística con corpus (I). (E. H. Filología., Ed.) *Estudios Humanísticos. Filología.*, 30, 329-349. Obtenido de <http://revpubli.unileon.es/ojs/index.php/EEHFFilologia/article/view/2847>
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. (G. F.-B. M. Baker, Ed.) *Text and Technology: In Honour of John Sinclair*, 157-176.
- Mačutek, J. (2014). Complex networks are not (so much) privileged: Comment on “Approaching human language with complex networks” by Cong and Liu. *Physics of Life Reviews*, 11(4), 635-636. doi:<https://doi.org/10.1016/j.plrev.2014.07.015>.
- Martín Camacho, J. (2004). *Los procesos neológicos del léxico científico. Esbozo de clasificación. Anuario de Estudios Filológicos* (Vol. XXVII). AEF. Obtenido de <http://hdl.handle.net/10662/846>
- Martincic, S., Margan, D., & Mestrovic, A. (2016). *Multilayer network of language: a unified framework for structural analysis of linguistic subsystems*. *Phys Rev E* 74:026102.
- Masucci, A., & Rodgers, G. (2006). Network properties of written human language. *Physica A*(457), 117-128.
- McEnery, T., & Wilson, A. (1996). *Corpus Linguistics. Edinburgh Textbooks in Empirical Linguistics*. Edinburgh: Edinburgh University Press.
- Mehler, A., Lücking, A., Banisch, S., Blanchard, P., & Frank-Job, B. (2016). *Towards a theoretical framework for analyzing complex linguistics networks*. Berlin: Springer.
- Moreno Ortiz, A., Pérez Hernández, C., & Faber, P. (1999). Lexicografía computacional y lexicografía de corpus. *Revista española de lingüística aplicada*, monográfico(1), 175-214.
- Moreno Sandoval, A. (2019). *Lenguas y computación*. (J. C. Cabrer, Ed.) Madrid: Editorial Síntesis S.A.
- Morley, D. (1992). Populism, revisionism and the ‘new’ audience research. *Poetics*, 4(21), 339-344.
- Newman, M. (2010). *Networks: an introduction*. Oxford University Press.
- Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System*, 30(4), 479-497.
- Pérez Hernández, C. (2002). Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística Española*, 18. Obtenido de <http://elies.rediris.es/elies18/>
- Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design*, 33(5), 705-725.
- Reeves, C. (2005). *The language of science*. Routledge.
- Rondeau, G. (1983). *Introduction à la terminologie*. Chicoutimi: Gaëtan Morin.
- Santamaría Pérez, I. (2006). *La terminología: definición, funciones y aplicaciones*. Liceus, Servicios de Gestió.
- Santamaría-Pérez, I. (2009). La terminología y los términos. *Terminología I*.
- Schleiermacher, F., & Yebra, V. (2000). *Sobre los diferentes métodos de traducir*. Madrid: Gredos.

- Serrano Gómez, W. (2005). ¿Qué constituye a los lenguajes natural y matemático? *Sapiens. Revista Universitaria de Investigación*, 6(1), 47-59.
- Sinclair, J. (1991). *Corpus, concordance, collocation. Describing English language*. Oxford: Oxford University Press.
- Sinclair, J. M., & Kirby, D. M. (1990). Progress in Computational Lexicography. *World Englishes*, 9(1), 233-257.
- Sinclair, J., Payne, J., & Pérez Hernández, C. (1996). Corpus to Corpus: A Study Of Translation Equivalence. *International Journal of Lexicography*, 9(3).
- Stubbs, M. (1995). Collocations and Semantic Profiles: On the Cause of the Trouble with Quantitative Studies. *Functions of Language*, 2(1), 23-55.
- Tognini-Bonelli, E. (1996). *Corpus Theory and Practice*. Birmingham: TWC.
- Tognini-Bonelli, E. (1996). Towards Translation Equivalence from a Corpus Linguistics Perspective. *International Journal of Lexicography*, 9(3), 197-217.
- Tomasello, M. (2009). *Constructing a language*. Harvard university press.
- Vallez, M., & Pedraza-Jimenez, R. (2007). *Natural language processing in textual information retrieval and related topics*. (Vol. 5). Hipertext: net.
- Wood, A. (2001). International scientific English: the language of research scientists around the world. (J. P. Flowerdew, Ed.) *Research Perspectives on English for Academic Purposes*, 71-89.
- Zhao, Y. (2014). Three lines to view language networks: Comment on “Approaching human language with complex networks” by Cong and Liu. *Physics of Life Reviews*, 11(4), 637-638. doi:<https://doi.org/10.1016/j.plrev.2014.07.012>.
- Zipf, G. L. (1965). *Human behavior and the principle of least effort*. Hafner.

Anexos

A continuación, se presentan las listas de palabras de las que está compuesto el corpus utilizado para esta tesis doctoral. Se presentan las primeras 500 palabras divididas en capas ordenadas según el *PageRank*. Las unidades de conocimiento específico o unidades de traducción referentes al lenguaje matemático de especialidad de redes complejas son los términos que se encuentran en la lista llamada lexical (que son los términos que se encuentran en la capa léxica) y las colocaciones que se encuentran en la lista Lexical-lexical y que forman las unidades de conocimiento específico.

WORDS- Lista de las 500 primeras palabras del corpus ordenadas según su *PageRank*

1) <i>the</i>	2) <i>of</i>	3) <i>to</i>	4) <i>and</i>	5) <i>in</i>	6) <i>a</i>
7) <i>is</i>	8) <i>that</i>	9) <i>are</i>	10) <i>we</i>	11) <i>network</i>	12) <i>for</i>
13) <i>with</i>	14) <i>this</i>	15) <i>as</i>	16) <i>on</i>	17) <i>networks</i>	18) <i>by</i>
19) <i>be</i>	20) <i>an</i>	21) <i>can</i>	22) <i>it</i>	23) <i>between</i>	24) <i>nodes</i>
25) <i>system</i>	26) <i>which</i>	27) <i>from</i>	28) <i>each</i>	29) <i>or</i>	30) <i>model</i>
31) <i>different</i>	32) <i>have</i>	33) <i>one</i>	34) <i>these</i>	35) <i>has</i>	36) <i>not</i>
37) <i>at</i>	38) <i>systems</i>	39) <i>number</i>	40) <i>its</i>	41) <i>information</i>	42) <i>new</i>
43) <i>their</i>	44) <i>results</i>	45) <i>more</i>	46) <i>complex</i>	47) <i>will</i>	48) <i>two</i>
49) <i>such</i>	50) <i>some</i>	51) <i>other</i>	52) <i>based</i>	53) <i>also</i>	54) <i>structure</i>
55) <i>node</i>	56) <i>our</i>	57) <i>been</i>	58) <i>properties</i>	59) <i>all</i>	60) <i>time</i>
61) <i>using</i>	62) <i>both</i>	63) <i>data</i>	64) <i>when</i>	65) <i>into</i>	66) <i>graph</i>
67) <i>dynamics</i>	68) <i>order</i>	69) <i>them</i>	70) <i>study</i>	71) <i>process</i>	72) <i>distribution</i>
73) <i>analysis</i>	74) <i>used</i>	75) <i>may</i>	76) <i>models</i>	77) <i>algorithm</i>	78) <i>same</i>
79) <i>links</i>	80) <i>where</i>	81) <i>problem</i>	82) <i>than</i>	83) <i>most</i>	84) <i>species</i>
85) <i>social</i>	86) <i>but</i>	87) <i>neural</i>	88) <i>only</i>	89) <i>obtained</i>	90) <i>considered</i>
91) <i>through</i>	92) <i>were</i>	93) <i>state</i>	94) <i>first</i>	95) <i>case</i>	96) <i>there</i>
97) <i>approach</i>	98) <i>they</i>	99) <i>degree</i>	100) <i>connected</i>	101) <i>set</i>	102) <i>work</i>
103) <i>three</i>	104) <i>was</i>	105) <i>very</i>	106) <i>use</i>	107) <i>paper</i>	108) <i>graphs</i>

109) <i>interest</i>	110) <i>over</i>	111) <i>those</i>	112) <i>points</i>	113) <i>market</i>	114) <i>pagerank</i>
115) <i>interactions</i>	116) <i>value</i>	117) <i>neurons</i>	118) <i>among</i>	119) <i>show</i>	120) <i>if</i>
121) <i>rev</i>	122) <i>way</i>	123) <i>linear</i>	124) <i>matrix</i>	125) <i>multiplex</i>	126) <i>given</i>
127) <i>associated</i>	128) <i>topology</i>	129) <i>phase</i>	130) <i>example</i>	131) <i>theory</i>	132) <i>important</i>
133) <i>real</i>	134) <i>any</i>	135) <i>so</i>	136) <i>several</i>	137) <i>how</i>	138) <i>us</i>
139) <i>found</i>	140) <i>critical</i>	141) <i>about</i>	142) <i>main</i>	143) <i>users</i>	144) <i>extinction</i>
145) <i>topological</i>	146) <i>non</i>	147) <i>then</i>	148) <i>method</i>	149) <i>point</i>	150) <i>behavior</i>
151) <i>functions</i>	152) <i>proposed</i>	153) <i>many</i>	154) <i>local</i>	155) <i>since</i>	156) <i>random</i>
157) <i>patterns</i>	158) <i>evolution</i>	159) <i>population</i>	160) <i>processes</i>	161) <i>present</i>	162) <i>size</i>
163) <i>possible</i>	164) <i>following</i>	165) <i>global</i>	166) <i>propose</i>	167) <i>well</i>	168) <i>applied</i>
169) <i>related</i>	170) <i>according</i>	171) <i>high</i>	172) <i>weighted</i>	173) <i>agents</i>	174) <i>must</i>
175) <i>dynamical</i>	176) <i>large</i>	177) <i>consider</i>	178) <i>general</i>	179) <i>clustering</i>	180) <i>obtain</i>
181) <i>measures</i>	182) <i>pipes</i>	183) <i>line</i>	184) <i>function</i>	185) <i>layer</i>	186) <i>space</i>
187) <i>while</i>	188) <i>analyze</i>	189) <i>measure</i>	190) <i>link</i>	191) <i>controllability</i>	192) <i>account</i>
193) <i>pipe</i>	194) <i>image</i>	195) <i>step</i>	196) <i>section</i>	197) <i>communities</i>	198) <i>allows</i>
199) <i>represent</i>	200) <i>whole</i>	201) <i>control</i>	202) <i>similar</i>	203) <i>level</i>	204) <i>corresponding</i>
205) <i>interaction</i>	206) <i>positive</i>	207) <i>error</i>	208) <i>structural</i>	209) <i>simple</i>	210) <i>user</i>
211) <i>growth</i>	212) <i>performance</i>	213) <i>find</i>	214) 2	215) <i>trade</i>	216) <i>l</i>
217) <i>values</i>	218) <i>fact</i>	219) <i>within</i>	220) <i>application</i>	221) <i>water</i>	222) <i>initial</i>
223) <i>no</i>	224) <i>shape</i>	225) <i>result</i>	226) <i>signal</i>	227) <i>edges</i>	228) <i>markets</i>
229) <i>specific</i>	230) <i>research</i>	231) <i>observed</i>	232) <i>existence</i>	233) <i>flow</i>	234) <i>robustness</i>
235) <i>however</i>	236) <i>conditions</i>	237) <i>centrality</i>	238) <i>authors</i>	239) <i>due</i>	240) <i>required</i>
241) <i>here</i>	242) <i>higher</i>	243) 3	244) <i>percolation</i>	245) <i>natural</i>	246) <i>presented</i>
247) <i>because</i>	248) <i>defined</i>	249) <i>methods</i>	250) <i>terms</i>	251) <i>parameters</i>	252) <i>without</i>
253) <i>frequency</i>	254) <i>rate</i>	255) <i>particular</i>	256) <i>percentage</i>	257) <i>distributions</i>	258) <i>effect</i>
259) <i>matrices</i>	260) <i>tree</i>	261) <i>better</i>	262) <i>behaviour</i>	263) <i>therefore</i>	264) <i>core</i>
265) <i>layers</i>	266) <i>out</i>	267) <i>above</i>	268) <i>usually</i>	269) <i>field</i>	270) <i>after</i>
271) <i>determine</i>	272) <i>even</i>	273) <i>mutualistic</i>	274) <i>does</i>	275) <i>activity</i>	276) <i>community</i>
277) <i>form</i>	278) <i>provide</i>	279) <i>small</i>	280) <i>methodology</i>	281) <i>features</i>	282) <i>attributes</i>
283) <i>whose</i>	284) <i>called</i>	285) <i>economic</i>	286) <i>threshold</i>	287) <i>memory</i>	288) <i>components</i>
289) <i>do</i>	290) <i>rules</i>	291) <i>distance</i>	292) <i>type</i>	293) <i>classification</i>	294) <i>consensus</i>
295) <i>part</i>	296) <i>external</i>	297) <i>lines</i>	298) <i>individual</i>	299) <i>useful</i>	300) <i>algorithms</i>
301) <i>view</i>	302) <i>complexity</i>	303) <i>known</i>	304) <i>resources</i>	305) <i>synchronization</i>	306) <i>framework</i>

307) <i>give</i>	308) <i>noise</i>	309) <i>differences</i>	310) <i>thus</i>	311) <i>including</i>	312) <i>provides</i>
313) <i>having</i>	314) <i>shows</i>	315) <i>selected</i>	316) <i>influence</i>	317) <i>every</i>	318) <i>stages</i>
319) <i>total</i>	320) <i>certain</i>	321) <i>taking</i>	322) <i>risk</i>	323) <i>efficiency</i>	324) <i>could</i>
325) <i>financial</i>	326) <i>edge</i>	327) <i>equation</i>	328) <i>considering</i>	329) <i>detection</i>	330) <i>appears</i>
331) <i>plant</i>	332) <i>under</i>	333) <i>solution</i>	334) <i>average</i>	335) <i>elements</i>	336) <i>connections</i>
337) <i>necessary</i>	338) <i>next</i>	339) <i>like</i>	340) <i>types</i>	341) <i>simulations</i>	342) <i>means</i>
343) <i>make</i>	344) <i>infected</i>	345) <i>highest</i>	346) <i>less</i>	347) <i>correlation</i>	348) <i>cases</i>
349) <i>conditional</i>	350) <i>ie</i>	351) <i>compared</i>	352) <i>become</i>	353) <i>minimum</i>	354) <i>sets</i>
355) <i>spreading</i>	356) <i>allow</i>	357) <i>describe</i>	358) <i>bipartite</i>	359) <i>variables</i>	360) <i>should</i>
361) <i>low</i>	362) <i>grid</i>	363) <i>developed</i>	364) <i>shown</i>	365) <i>lower</i>	366) <i>communication</i>
367) <i>resulting</i>	368) <i>studied</i>	369) <i>depends</i>	370) <i>computing</i>	371) <i>procedure</i>	372) <i>addition</i>
373) <i>equations</i>	374) <i>follows</i>	375) <i>class</i>	376) <i>identify</i>	377) <i>relevant</i>	378) <i>rankings</i>
379) <i>problems</i>	380) <i>power</i>	381) <i>around</i>	382) <i>strength</i>	383) <i>search</i>	384) <i>context</i>
385) <i>performed</i>	386) <i>complete</i>	387) <i>analyzed</i>	388) <i>countries</i>	389) <i>individuals</i>	390) <i>science</i>
391) <i>vector</i>	392) <i>appear</i>	393) <i>robust</i>	394) <i>capital</i>	395) <i>until</i>	396) <i>stability</i>
397) <i>computational</i>	398) <i>further</i>	399) <i>active</i>	400) <i>maximum</i>	401) <i>compare</i>	402) <i>transition</i>
403) <i>others</i>	404) <i>would</i>	405) <i>weight</i>	406) <i>amount</i>	407) <i>hand</i>	408) <i>understand</i>
409) <i>apply</i>	410) <i>previous</i>	411) <i>pattern</i>	412) <i>mean</i>	413) <i>similarity</i>	414) <i>times</i>
415) <i>tools</i>	416) <i>condition</i>	417) <i>urban</i>	418) <i>stable</i>	419) <i>during</i>	420) <i>states</i>
421) <i>changes</i>	422) <i>stock</i>	423) <i>learning</i>	424) <i>returns</i>	425) <i>objects</i>	426) <i>mathematical</i>
427) <i>relations</i>	428) <i>highly</i>	429) <i>single</i>	430) <i>swarm</i>	431) <i>best</i>	432) <i>definition</i>
433) <i>component</i>	434) <i>need</i>	435) <i>introduce</i>	436) <i>infection</i>	437) <i>targets</i>	438) <i>demand</i>
439) <i>future</i>	440) <i>include</i>	441) <i>characteristics</i>	442) <i>aim</i>	443) <i>nature</i>	444) <i>implemented</i>
445) <i>compute</i>	446) <i>follow</i>	447) <i>neuron</i>	448) <i>area</i>	449) <i>bus</i>	450) <i>years</i>
451) <i>gas</i>	452) <i>importance</i>	453) <i>solutions</i>	454) <i>identification</i>	455) <i>voting</i>	456) <i>along</i>
457) <i>few</i>	458) <i>dynamic</i>	459) <i>increases</i>	460) <i>tool</i>	461) <i>propagation</i>	462) <i>involved</i>
463) <i>last</i>	464) <i>weights</i>	465) <i>cells</i>	466) <i>linked</i>	467) <i>finite</i>	468) <i>stochastic</i>
469) <i>stimuli</i>	470) <i>scenario</i>	471) <i>signals</i>	472) <i>images</i>	473) <i>spatial</i>	474) <i>capacity</i>
475) <i>4</i>	476) <i>languages</i>	477) <i>schemes</i>	478) <i>being</i>	479) <i>collective</i>	480) <i>take</i>
481) <i>family</i>	482) <i>second</i>	483) <i>variance</i>	484) <i>strategy</i>	485) <i>topologies</i>	486) <i>define</i>
487) <i>requires</i>	488) <i>pairs</i>	489) <i>before</i>	490) <i>development</i>	491) <i>difference</i>	492) <i>supply</i>
493) <i>applications</i>	494) <i>property</i>	495) <i>elections</i>	496) <i>5</i>	497) <i>series</i>	498) <i>independent</i>
499) <i>made</i>	500) <i>close</i>	501) <i>presence</i>	502) <i>introduced</i>	503) <i>needed</i>	504) <i>event</i>

Lexical- Lista de las 500 primeras palabras del corpus incluidas en la capa léxica ordenadas según su PageRank

1) <i>network</i>	2) <i>networks</i>	3) <i>nodes</i>	4) <i>system</i>	5) <i>information</i>	6) <i>number</i>
7) <i>model</i>	8) <i>structure</i>	9) <i>systems</i>	10) <i>properties</i>	11) <i>new</i>	12) <i>node</i>
13) <i>complex</i>	14) <i>data</i>	15) <i>graph</i>	16) <i>dynamics</i>	17) <i>study</i>	18) <i>process</i>
19) <i>distribution</i>	20) <i>analysis</i>	21) <i>models</i>	22) <i>algorithm</i>	23) <i>links</i>	24) <i>problem</i>
25) <i>only</i>	26) <i>state</i>	27) <i>case</i>	28) <i>approach</i>	29) <i>degree</i>	30) <i>connected</i>
31) <i>use</i>	32) <i>paper</i>	33) <i>graphs</i>	34) <i>interest</i>	35) <i>pagerank</i>	36) <i>interactions</i>
37) <i>value</i>	38) <i>way</i>	39) <i>linear</i>	40) <i>matrix</i>	41) <i>multiplex</i>	42) <i>associated</i>
43) <i>topology</i>	44) <i>phase</i>	45) <i>example</i>	46) <i>theory</i>	47) <i>important</i>	48) <i>real</i>
49) <i>critical</i>	50) <i>main</i>	51) <i>extinction</i>	52) <i>topological</i>	53) <i>method</i>	54) <i>behavior</i>
55) <i>functions</i>	56) <i>proposed</i>	57) <i>local</i>	58) <i>random</i>	59) <i>patterns</i>	60) <i>evolution</i>
61) <i>population</i>	62) <i>processes</i>	63) <i>present</i>	64) <i>size</i>	65) <i>applied</i>	66) <i>related</i>
67) <i>high</i>	68) <i>weighted</i>	69) <i>agents</i>	70) <i>dynamical</i>	71) <i>large</i>	72) <i>clustering</i>
73) <i>measures</i>	74) <i>line</i>	75) <i>function</i>	76) <i>layer</i>	77) <i>space</i>	78) <i>measure</i>
79) <i>controllability</i>	80) <i>account</i>	81) <i>image</i>	82) <i>step</i>	83) <i>section</i>	84) <i>communities</i>
85) <i>control</i>	86) <i>similar</i>	87) <i>level</i>	88) <i>corresponding</i>	89) <i>interaction</i>	90) <i>positive</i>
91) <i>error</i>	92) <i>structural</i>	93) <i>simple</i>	94) <i>growth</i>	95) <i>performance</i>	96) <i>values</i>
97) <i>application</i>	98) <i>initial</i>	99) <i>shape</i>	100) <i>result</i>	101) <i>edges</i>	102) <i>research</i>
103) <i>existence</i>	104) <i>flow</i>	105) <i>robustness</i>	106) <i>conditions</i>	107) <i>centrality</i>	108) <i>required</i>
109) <i>higher</i>	110) <i>percolation</i>	111) <i>natural</i>	112) <i>methods</i>	113) <i>terms</i>	114) <i>parameters</i>
115) <i>frequency</i>	116) <i>particular</i>	117) <i>percentage</i>	118) <i>distributions</i>	119) <i>effect</i>	120) <i>matrices</i>
121) <i>tree</i>	122) <i>behaviour</i>	123) <i>core</i>	124) <i>layers</i>	125) <i>field</i>	126) <i>mutualistic</i>
127) <i>activity</i>	128) <i>community</i>	129) <i>small</i>	130) <i>methodology</i>	131) <i>features</i>	132) <i>attributes</i>
133) <i>threshold</i>	134) <i>memory</i>	135) <i>components</i>	136) <i>rules</i>	137) <i>distance</i>	138) <i>classification</i>
139) <i>consensus</i>	140) <i>part</i>	141) <i>lines</i>	142) <i>individual</i>	143) <i>algorithms</i>	144) <i>complexity</i>
145) <i>resources</i>	146) <i>synchronization</i>	147) <i>framework</i>	148) <i>noise</i>	149) <i>stages</i>	150) <i>total</i>
151) <i>certain</i>	152) <i>efficiency</i>	153) <i>edge</i>	154) <i>equation</i>	155) <i>detection</i>	156) <i>solution</i>
157) <i>average</i>	158) <i>elements</i>	159) <i>connections</i>	160) <i>next</i>	161) <i>types</i>	162) <i>simulations</i>
163) <i>infected</i>	164) <i>highest</i>	165) <i>correlation</i>	166) <i>cases</i>	167) <i>conditional</i>	168) <i>compared</i>

169) <i>minimum</i>	170) <i>bipartite</i>	171) <i>variables</i>	172) <i>grid</i>	173) <i>developed</i>	174) <i>lower</i>
175) <i>communication</i>	176) <i>studied</i>	177) <i>computing</i>	178) <i>procedure</i>	179) <i>addition</i>	180) <i>equations</i>
181) <i>class</i>	182) <i>relevant</i>	183) <i>rankings</i>	184) <i>problems</i>	185) <i>power</i>	186) <i>strength</i>
187) <i>individuals</i>	188) <i>vector</i>	189) <i>robust</i>	190) <i>stability</i>	191) <i>computational</i>	192) <i>active</i>
193) <i>maximum</i>	194) <i>transition</i>	195) <i>amount</i>	196) <i>pattern</i>	197) <i>similarity</i>	198) <i>times</i>
199) <i>tools</i>	200) <i>condition</i>	201) <i>stable</i>	202) <i>states</i>	203) <i>objects</i>	204) <i>mathematical</i>
205) <i>relations</i>	206) <i>highly</i>	207) <i>single</i>	208) <i>definition</i>	209) <i>component</i>	210) <i>future</i>
211) <i>characteristics</i>	212) <i>nature</i>	213) <i>implemented</i>	214) <i>area</i>	215) <i>importance</i>	216) <i>solutions</i>
217) <i>identification</i>	218) <i>dynamic</i>	219) <i>tool</i>	220) <i>propagation</i>	221) <i>weights</i>	222) <i>linked</i>
223) <i>finite</i>	224) <i>stochastic</i>	225) <i>stimuli</i>	226) <i>scenario</i>	227) <i>images</i>	228) <i>spatial</i>
229) <i>capacity</i>	230) <i>languages</i>	231) <i>schemes</i>	232) <i>collective</i>	233) <i>family</i>	234) <i>variance</i>
235) <i>topologies</i>	236) <i>pairs</i>	237) <i>development</i>	238) <i>difference</i>	239) <i>applications</i>	240) <i>property</i>
241) <i>series</i>	242) <i>independent</i>	243) <i>presence</i>	244) <i>comparison</i>	245) <i>standard</i>	246) <i>idea</i>
247) <i>distributed</i>	248) <i>weak</i>	249) <i>criteria</i>	250) <i>role</i>	251) <i>directly</i>	252) <i>areas</i>
253) <i>equivalent</i>	254) <i>sectors</i>	255) <i>relationship</i>	256) <i>strong</i>	257) <i>friends</i>	258) <i>numerical</i>
259) <i>volume</i>	260) <i>intrinsic</i>	261) <i>organization</i>	262) <i>output</i>	263) <i>input</i>	264) <i>wsn</i>
265) <i>digital</i>	266) <i>probability</i>	267) <i>pollinator</i>	268) <i>synaptic</i>	269) <i>rest</i>	270) <i>security</i>
271) <i>nonlinear</i>	272) <i>range</i>	273) <i>connectivity</i>	274) <i>formed</i>	275) <i>consequence</i>	276) <i>processing</i>
277) <i>probabilities</i>	278) <i>implementation</i>	279) <i>coefficient</i>	280) <i>determined</i>	281) <i>marked</i>	282) <i>groups</i>
283) <i>directed</i>	284) <i>recommendation</i>	285) <i>academic</i>	286) <i>competitiveness</i>	287) <i>stationary</i>	288) <i>theoretical</i>
289) <i>parameter</i>	290) <i>regions</i>	291) <i>metrics</i>	292) <i>nestedness</i>	293) <i>volatility</i>	294) <i>positions</i>
295) <i>items</i>	296) <i>correlations</i>	297) <i>roles</i>	298) <i>people</i>	299) <i>source</i>	300) <i>factors</i>
301) <i>mechanisms</i>	302) <i>accuracy</i>	303) <i>simulated</i>	304) <i>classical</i>	305) <i>combination</i>	306) <i>insights</i>
307) <i>length</i>	308) <i>prediction</i>	309) <i>described</i>	310) <i>asymmetry</i>	311) <i>oscillations</i>	312) <i>stratification</i>
313) <i>statistical</i>	314) <i>clusters</i>	315) <i>tested</i>	316) <i>accurate</i>	317) <i>heterogeneous</i>	318) <i>inputs</i>
319) <i>intentional</i>	320) <i>formal</i>	321) <i>representation</i>	322) <i>internal</i>	323) <i>stage</i>	324) <i>connectance</i>
325) <i>sdh</i>	326) <i>direct</i>	327) <i>test</i>	328) <i>smallworld</i>	329) <i>paths</i>	330) <i>design</i>
331) <i>common</i>	332) <i>flows</i>	333) <i>path</i>	334) <i>recent</i>	335) <i>mechanism</i>	336) <i>coupled</i>
337) <i>unique</i>	338) <i>invariant</i>	339) <i>dependent</i>	340) <i>characteristic</i>	341) <i>perspective</i>	342) <i>cluster</i>
343) <i>scenarios</i>	344) <i>upper</i>	345) <i>vital</i>	346) <i>preferences</i>	347) <i>response</i>	348) <i>continuous</i>
349) <i>analytical</i>	350) <i>produced</i>	351) <i>temporal</i>	352) <i>sequence</i>	353) <i>spectrum</i>	354) <i>choice</i>
355) <i>explosive</i>	356) <i>reduce</i>	357) <i>underlying</i>	358) <i>periodic</i>	359) <i>resilience</i>	360) <i>scalefree</i>
361) <i>ways</i>	362) <i>errors</i>	363) <i>functional</i>	364) <i>options</i>	365) <i>basic</i>	366) <i>largest</i>

367) <i>pressure</i>	368) <i>scheme</i>	369) <i>original</i>	370) <i>routing</i>	371) <i>rule</i>	372) <i>metric</i>
373) <i>sufficient</i>	374) <i>populations</i>	375) <i>multiple</i>	376) <i>huge</i>	377) <i>approaches</i>	378) <i>synchronized</i>
379) <i>organizations</i>	380) <i>frame</i>	381) <i>boundary</i>	382) <i>phases</i>	383) <i>realistic</i>	384) <i>generalist</i>
385) <i>singular</i>	386) <i>measurement</i>	387) <i>pair</i>	388) <i>relevance</i>	389) <i>levels</i>	390) <i>reduced</i>
391) <i>reservations</i>	392) <i>operational</i>	393) <i>relation</i>	394) <i>density</i>	395) <i>characterization</i>	396) <i>mutualism</i>
397) <i>binary</i>	398) <i>sources</i>	399) <i>recognition</i>	400) <i>personalization</i>	401) <i>connection</i>	402) <i>final</i>
403) <i>neighbors</i>	404) <i>environment</i>	405) <i>reachability</i>	406) <i>identical</i>	407) <i>generation</i>	408) <i>degrees</i>
409) <i>coupling</i>	410) <i>represented</i>	411) <i>survival</i>	412) <i>procedures</i>	413) <i>removal</i>	414) <i>product</i>
415) <i>phenomenon</i>	416) <i>tracked</i>	417) <i>ranking</i>	418) <i>potential</i>	419) <i>clear</i>	420) <i>figure</i>
421) <i>period</i>	422) <i>top</i>	423) <i>regular</i>	424) <i>key</i>	425) <i>partition</i>	426) <i>task</i>
427) <i>exponents</i>	428) <i>constant</i>	429) <i>examples</i>	430) <i>aspects</i>	431) <i>strongly</i>	432) <i>impact</i>
433) <i>chaotic</i>	434) <i>sense</i>	435) <i>invariants</i>	436) <i>engineering</i>	437) <i>management</i>	438) <i>free</i>
439) <i>experiment</i>	440) <i>interconnected</i>	441) <i>sum</i>	442) <i>numbers</i>	443) <i>rich</i>	444) <i>law</i>
445) <i>dimensions</i>	446) <i>efficiently</i>	447) <i>popular</i>	448) <i>feature</i>	449) <i>protocols</i>	450) <i>objective</i>
451) <i>dependence</i>	452) <i>concepts</i>	453) <i>conclusions</i>	454) <i>pollination</i>	455) <i>branches</i>	456) <i>modular</i>
457) <i>typical</i>	458) <i>recurrence</i>	459) <i>algebraic</i>	460) <i>meaning</i>	461) <i>oscillators</i>	462) <i>cycles</i>
463) <i>significant</i>	464) <i>unsupervised</i>	465) <i>pareto</i>	466) <i>compatibility</i>	467) <i>decision</i>	468) <i>homogeneous</i>
469) <i>option</i>	470) <i>clearly</i>	471) <i>added</i>	472) <i>iteration</i>	473) <i>valid</i>	474) <i>explored</i>
475) <i>layouts</i>	476) <i>remote</i>	477) <i>fundamental</i>	478) <i>backpropagation</i>	479) <i>sensory</i>	480) <i>phenomena</i>
481) <i>experts</i>	482) <i>effective</i>	483) <i>literature</i>	484) <i>smaller</i>	485) <i>relative</i>	486) <i>failures</i>
487) <i>transitions</i>	488) <i>origin</i>	489) <i>scales</i>	490) <i>query</i>	491) <i>novel</i>	492) <i>effects</i>
493) <i>calculations</i>	494) <i>betweenness</i>	495) <i>requirements</i>	496) <i>belonging</i>	497) <i>experiments</i>	498) <i>limit</i>
499) <i>convergence</i>	500) <i>sequences</i>	501) <i>straight</i>	502) <i>incoming</i>	503) <i>negative</i>	504) <i>unknown</i>

Verbs- Lista de las 500 primeras palabras incluidas en la capa verbal ordenadas según su PageRank

1) <i>is</i>	2) <i>are</i>	3) <i>be</i>	4) <i>can</i>	5) <i>have</i>	6) <i>has</i>
7) <i>results</i>	8) <i>will</i>	9) <i>been</i>	10) <i>using</i>	11) <i>order</i>	12) <i>used</i>
13) <i>may</i>	14) <i>were</i>	15) <i>state</i>	16) <i>set</i>	17) <i>work</i>	18) <i>was</i>
19) <i>points</i>	20) <i>show</i>	21) <i>rev</i>	22) <i>given</i>	23) <i>found</i>	24) <i>point</i>
25) <i>following</i>	26) <i>propose</i>	27) <i>according</i>	28) <i>must</i>	29) <i>consider</i>	30) <i>obtain</i>
31) <i>analyze</i>	32) <i>link</i>	33) <i>allows</i>	34) <i>represent</i>	35) <i>find</i>	36) <i>research</i>
37) <i>observed</i>	38) <i>due</i>	39) <i>defined</i>	40) <i>rate</i>	41) <i>determine</i>	42) <i>does</i>
43) <i>form</i>	44) <i>provide</i>	45) <i>called</i>	46) <i>do</i>	47) <i>type</i>	48) <i>view</i>
49) <i>known</i>	50) <i>give</i>	51) <i>including</i>	52) <i>provides</i>	53) <i>having</i>	54) <i>shows</i>
55) <i>selected</i>	56) <i>influence</i>	57) <i>taking</i>	58) <i>risk</i>	59) <i>could</i>	60) <i>considering</i>
61) <i>appears</i>	62) <i>plant</i>	63) <i>like</i>	64) <i>means</i>	65) <i>make</i>	66) <i>become</i>
67) <i>sets</i>	68) <i>spreading</i>	69) <i>allow</i>	70) <i>describe</i>	71) <i>should</i>	72) <i>low</i>
73) <i>shown</i>	74) <i>resulting</i>	75) <i>depends</i>	76) <i>follows</i>	77) <i>identify</i>	78) <i>search</i>
79) <i>complete</i>	80) <i>analyzed</i>	81) <i>appear</i>	82) <i>compare</i>	83) <i>would</i>	84) <i>weight</i>
85) <i>hand</i>	86) <i>understand</i>	87) <i>apply</i>	88) <i>mean</i>	89) <i>changes</i>	90) <i>learning</i>
91) <i>returns</i>	92) <i>swarm</i>	93) <i>need</i>	94) <i>introduce</i>	95) <i>targets</i>	96) <i>demand</i>
97) <i>include</i>	98) <i>aim</i>	99) <i>compute</i>	100) <i>follow</i>	101) <i>voting</i>	102) <i>increases</i>
103) <i>involved</i>	104) <i>weights</i>	105) <i>being</i>	106) <i>take</i>	107) <i>define</i>	108) <i>requires</i>
109) <i>supply</i>	110) <i>made</i>	111) <i>close</i>	112) <i>introduced</i>	113) <i>needed</i>	114) <i>increase</i>
115) <i>correspond</i>	116) <i>lead</i>	117) <i>focus</i>	118) <i>might</i>	119) <i>assigned</i>	120) <i>had</i>
121) <i>get</i>	122) <i>help</i>	123) <i>located</i>	124) <i>depend</i>	125) <i>seen</i>	126) <i>organized</i>
127) <i>consists</i>	128) <i>note</i>	129) <i>tracking</i>	130) <i>included</i>	131) <i>scaling</i>	132) <i>provided</i>
133) <i>cannot</i>	134) <i>increasing</i>	135) <i>done</i>	136) <i>improve</i>	137) <i>build</i>	138) <i>taken</i>
139) <i>classify</i>	140) <i>index</i>	141) <i>finding</i>	142) <i>place</i>	143) <i>change</i>	144) <i>target</i>
145) <i>tend</i>	146) <i>identified</i>	147) <i>works</i>	148) <i>indicates</i>	149) <i>attacks</i>	150) <i>controls</i>
151) <i>generalized</i>	152) <i>characterized</i>	153) <i>cover</i>	154) <i>generated</i>	155) <i>root</i>	156) <i>built</i>
157) <i>switched</i>	158) <i>occurs</i>	159) <i>require</i>	160) <i>fixed</i>	161) <i>map</i>	162) <i>includes</i>
163) <i>detect</i>	164) <i>depending</i>	165) <i>makes</i>	166) <i>perform</i>	167) <i>determines</i>	168) <i>evolving</i>

169) <i>evaluate</i>	170) <i>produce</i>	171) <i>induced</i>	172) <i>affect</i>	173) <i>updated</i>	174) <i>defining</i>
175) <i>plants</i>	176) <i>occur</i>	177) <i>develop</i>	178) <i>developing</i>	179) <i>indexes</i>	180) <i>nested</i>
181) <i>calculated</i>	182) <i>share</i>	183) <i>generate</i>	184) <i>explore</i>	185) <i>establish</i>	186) <i>interacting</i>
187) <i>represents</i>	188) <i>gives</i>	189) <i>analyzing</i>	190) <i>composed</i>	191) <i>label</i>	192) <i>spread</i>
193) <i>tags</i>	194) <i>increased</i>	195) <i>address</i>	196) <i>group</i>	197) <i>remains</i>	198) <i>deal</i>
199) <i>chosen</i>	200) <i>applying</i>	201) <i>want</i>	202) <i>assign</i>	203) <i>established</i>	204) <i>growing</i>
205) <i>understanding</i>	206) <i>benefit</i>	207) <i>investigate</i>	208) <i>attack</i>	209) <i>plays</i>	210) <i>quench</i>
211) <i>observe</i>	212) <i>involves</i>	213) <i>prevent</i>	214) <i>stated</i>	215) <i>showed</i>	216) <i>contribute</i>
217) <i>becomes</i>	218) <i>see</i>	219) <i>detected</i>	220) <i>needs</i>	221) <i>viewed</i>	222) <i>overcome</i>
223) <i>changing</i>	224) <i>check</i>	225) <i>start</i>	226) <i>display</i>	227) <i>weighting</i>	228) <i>derive</i>
229) <i>suggests</i>	230) <i>play</i>	231) <i>assess</i>	232) <i>achieve</i>	233) <i>freezing</i>	234) <i>affecting</i>
235) <i>know</i>	236) <i>measuring</i>	237) <i>takes</i>	238) <i>sizes</i>	239) <i>exist</i>	240) <i>binding</i>
241) <i>let</i>	242) <i>reduces</i>	243) <i>votes</i>	244) <i>access</i>	245) <i>designed</i>	246) <i>assume</i>
247) <i>connecting</i>	248) <i>describing</i>	249) <i>suffers</i>	250) <i>keep</i>	251) <i>focused</i>	252) <i>identifying</i>
253) <i>evolve</i>	254) <i>guarantee</i>	255) <i>transfer</i>	256) <i>vote</i>	257) <i>emerging</i>	258) <i>put</i>
259) <i>minimize</i>	260) <i>exchange</i>	261) <i>carried</i>	262) <i>front</i>	263) <i>estimated</i>	264) <i>estimate</i>
265) <i>providing</i>	266) <i>filter</i>	267) <i>add</i>	268) <i>match</i>	269) <i>seem</i>	270) <i>remain</i>
271) <i>producing</i>	272) <i>lack</i>	273) <i>evolves</i>	274) <i>exists</i>	275) <i>reproduce</i>	276) <i>implies</i>
277) <i>generates</i>	278) <i>manage</i>	279) <i>solve</i>	280) <i>extend</i>	281) <i>look</i>	282) <i>accept</i>
283) <i>tied</i>	284) <i>considers</i>	285) <i>uses</i>	286) <i>implement</i>	287) <i>decided</i>	288) <i>scheduling</i>
289) <i>monitoring</i>	290) <i>rates</i>	291) <i>looking</i>	292) <i>reach</i>	293) <i>devoted</i>	294) <i>reveals</i>
295) <i>calculate</i>	296) <i>gain</i>	297) <i>guide</i>	298) <i>notice</i>	299) <i>exhibit</i>	300) <i>becoming</i>
301) <i>making</i>	302) <i>contributing</i>	303) <i>matches</i>	304) <i>enhance</i>	305) <i>said</i>	306) <i>drive</i>
307) <i>suggest</i>	308) <i>forms</i>	309) <i>emerges</i>	310) <i>predict</i>	311) <i>decreases</i>	312) <i>jumps</i>
313) <i>studying</i>	314) <i>inverses</i>	315) <i>corresponds</i>	316) <i>branching</i>	317) <i>memorized</i>	318) <i>closed</i>
319) <i>bridge</i>	320) <i>studies</i>	321) <i>open</i>	322) <i>understood</i>	323) <i>mix</i>	324) <i>create</i>
325) <i>characterize</i>	326) <i>fit</i>	327) <i>presents</i>	328) <i>select</i>	329) <i>driving</i>	330) <i>count</i>
331) <i>rewiring</i>	332) <i>extract</i>	333) <i>tagging</i>	334) <i>turn</i>	335) <i>matching</i>	336) <i>working</i>
337) <i>giving</i>	338) <i>created</i>	339) <i>consist</i>	340) <i>peak</i>	341) <i>supervised</i>	342) <i>serve</i>
343) <i>connect</i>	344) <i>coincide</i>	345) <i>force</i>	346) <i>vary</i>	347) <i>emerge</i>	348) <i>connects</i>
349) <i>allowing</i>	350) <i>constitute</i>	351) <i>leads</i>	352) <i>published</i>	353) <i>synchronize</i>	354) <i>cooling</i>
355) <i>tail</i>	356) <i>transport</i>	357) <i>ending</i>	358) <i>places</i>	359) <i>reaches</i>	360) <i>recommend</i>
361) <i>decentralized</i>	362) <i>fire</i>	363) <i>disappear</i>	364) <i>obtaining</i>	365) <i>counted</i>	366) <i>suffering</i>

367) <i>explained</i>	368) <i>displays</i>	369) <i>dealing</i>	370) <i>reveal</i>	371) <i>assuming</i>	372) <i>introducing</i>
373) <i>stops</i>	374) <i>searching</i>	375) <i>comes</i>	376) <i>offer</i>	377) <i>representing</i>	378) <i>tackle</i>
379) <i>emit</i>	380) <i>spanning</i>	381) <i>ordered</i>	382) <i>head</i>	383) <i>extended</i>	384) <i>encode</i>
385) <i>survive</i>	386) <i>travel</i>	387) <i>simulate</i>	388) <i>wiring</i>	389) <i>maintain</i>	390) <i>overlap</i>
391) <i>incorporating</i>	392) <i>adapted</i>	393) <i>transform</i>	394) <i>involving</i>	395) <i>mentions</i>	396) <i>list</i>
397) <i>discovered</i>	398) <i>guild</i>	399) <i>concerning</i>	400) <i>center</i>	401) <i>examine</i>	402) <i>overlapping</i>
403) <i>auditing</i>	404) <i>combines</i>	405) <i>serves</i>	406) <i>revealed</i>	407) <i>led</i>	408) <i>enabling</i>
409) <i>starting</i>	410) <i>achieving</i>	411) <i>conclude</i>	412) <i>adapting</i>	413) <i>maps</i>	414) <i>practices</i>
415) <i>consuming</i>	416) <i>referred</i>	417) <i>closing</i>	418) <i>hidden</i>	419) <i>characterizes</i>	420) <i>denotes</i>
421) <i>arise</i>	422) <i>sent</i>	423) <i>demonstrate</i>	424) <i>refers</i>	425) <i>mitigate</i>	426) <i>repeated</i>
427) <i>counts</i>	428) <i>affects</i>	429) <i>seems</i>	430) <i>exchanges</i>	431) <i>consisting</i>	432) <i>keeping</i>
433) <i>lying</i>	434) <i>happened</i>	435) <i>exploit</i>	436) <i>mapping</i>	437) <i>induce</i>	438) <i>improves</i>
439) <i>recognize</i>	440) <i>accumulate</i>	441) <i>expect</i>	442) <i>expresses</i>	443) <i>retweets</i>	444) <i>depict</i>
445) <i>imposed</i>	446) <i>derived</i>	447) <i>enables</i>	448) <i>depicted</i>	449) <i>optimize</i>	450) <i>capture</i>
451) <i>generating</i>	452) <i>profit</i>	453) <i>lagged</i>	454) <i>agree</i>	455) <i>repeat</i>	456) <i>deployed</i>
457) <i>benefits</i>	458) <i>implements</i>	459) <i>belong</i>	460) <i>distributing</i>	461) <i>visualize</i>	462) <i>seed</i>
463) <i>compete</i>	464) <i>crossed</i>	465) <i>distinguish</i>	466) <i>explain</i>	467) <i>locate</i>	468) <i>report</i>
469) <i>produces</i>	470) <i>coordinate</i>	471) <i>fitted</i>	472) <i>trunk</i>	473) <i>reaching</i>	474) <i>cut</i>
475) <i>feed</i>	476) <i>clean</i>	477) <i>integrate</i>	478) <i>named</i>	479) <i>assessing</i>	480) <i>establishes</i>
481) <i>supports</i>	482) <i>varying</i>	483) <i>stay</i>	484) <i>quantify</i>	485) <i>reflects</i>	486) <i>sophisticated</i>
487) <i>evidences</i>	488) <i>involve</i>	489) <i>classifying</i>	490) <i>act</i>	491) <i>verify</i>	492) <i>encoding</i>
493) <i>care</i>	494) <i>exploring</i>	495) <i>granted</i>	496) <i>sort</i>	497) <i>covers</i>	498) <i>ground</i>
499) <i>rising</i>	500) <i>couple</i>	501) <i>stress</i>	502) <i>installed</i>	503) <i>indicate</i>	504) <i>peaks</i>

Linking- Palabras incluidas en la capa “palabras-enlace” ordenadas según el PageRank

1) <i>the</i>	2) <i>of</i>	3) <i>to</i>	4) <i>and</i>	5) <i>in</i>	6) <i>a</i>
7) <i>that</i>	8) <i>for</i>	9) <i>with</i>	10) <i>this</i>	11) <i>as</i>	12) <i>on</i>
13) <i>by</i>	14) <i>an</i>	15) <i>between</i>	16) <i>which</i>	17) <i>from</i>	18) <i>each</i>
19) <i>or</i>	20) <i>these</i>	21) <i>at</i>	22) <i>its</i>	23) <i>their</i>	24) <i>more</i>
25) <i>such</i>	26) <i>some</i>	27) <i>other</i>	28) <i>also</i>	29) <i>our</i>	30) <i>all</i>
31) <i>both</i>	32) <i>when</i>	33) <i>into</i>	34) <i>where</i>	35) <i>than</i>	36) <i>most</i>
37) <i>but</i>	38) <i>through</i>	39) <i>first</i>	40) <i>there</i>	41) <i>they</i>	42) <i>very</i>
43) <i>over</i>	44) <i>those</i>	45) <i>among</i>	46) <i>if</i>	47) <i>any</i>	48) <i>so</i>
49) <i>several</i>	50) <i>how</i>	51) <i>about</i>	52) <i>non</i>	53) <i>then</i>	54) <i>many</i>
55) <i>since</i>	56) <i>while</i>	57) <i>whole</i>	58) <i>within</i>	59) <i>however</i>	60) <i>because</i>
61) <i>without</i>	62) <i>therefore</i>	63) <i>out</i>	64) <i>above</i>	65) <i>usually</i>	66) <i>after</i>
67) <i>even</i>	68) <i>whose</i>	69) <i>thus</i>	70) <i>every</i>	71) <i>under</i>	72) <i>less</i>
73) <i>around</i>	74) <i>until</i>	75) <i>further</i>	76) <i>others</i>	77) <i>during</i>	78) <i>along</i>
79) <i>few</i>	80) <i>before</i>	81) <i>enough</i>	82) <i>who</i>	83) <i>towards</i>	84) <i>another</i>
85) <i>either</i>	86) <i>up</i>	87) <i>just</i>	88) <i>quite</i>	89) <i>still</i>	90) <i>per</i>
91) <i>finally</i>	92) <i>below</i>	93) <i>always</i>	94) <i>what</i>	95) <i>various</i>	96) <i>much</i>
97) <i>throughout</i>	98) <i>now</i>	99) <i>whether</i>	100) <i>rather</i>	101) <i>regarding</i>	102) <i>instead</i>
103) <i>although</i>	104) <i>previously</i>	105) <i>against</i>	106) <i>despite</i>	107) <i>together</i>	108) <i>moreover</i>
109) <i>off</i>	110) <i>already</i>	111) <i>again</i>	112) <i>though</i>	113) <i>whereas</i>	114) <i>furthermore</i>
115) <i>initially</i>	116) <i>nor</i>	117) <i>behind</i>	118) <i>almost</i>	119) <i>thereby</i>	120) <i>actually</i>
121) <i>eventually</i>	122) <i>nevertheless</i>	123) <i>often</i>	124) <i>etc</i>	125) <i>entire</i>	126) <i>sometimes</i>
127) <i>except</i>	128) <i>hence</i>	129) <i>too</i>	130) <i>upon</i>	131) <i>hereafter</i>	132) <i>additionally</i>
133) <i>otherwise</i>	134) <i>besides</i>	135) <i>versus</i>	136) <i>across</i>	137) <i>whenever</i>	138) <i>yet</i>
139) <i>backwards</i>	140) <i>numerous</i>	141) <i>firstly</i>	142) <i>indeed</i>	143) <i>why</i>	144) <i>secondly</i>
145) <i>neither</i>	146) <i>down</i>	147) <i>whitout</i>	148) <i>meanwhile</i>	149) <i>fewer</i>	150) <i>regardless</i>
151) <i>amongst</i>	152) <i>henceforth</i>	153) <i>unless</i>	154) <i>perhaps</i>	155) <i>somewhat</i>	156) <i>thirdly</i>
157) <i>anyway</i>	158) <i>afterwards</i>	159) <i>whilst</i>	160) <i>till</i>	161) <i>nonetheless</i>	

Lista de las 500 primeras palabras incluidas en la capa *Remaining* (capa resto) ordenadas según su *PageRank*

1) <i>we</i>	2) <i>it</i>	3) <i>different</i>	4) <i>one</i>	5) <i>not</i>	6) <i>two</i>
7) <i>based</i>	8) <i>time</i>	9) <i>them</i>	10) <i>same</i>	11) <i>species</i>	12) <i>social</i>
13) <i>neural</i>	14) <i>obtained</i>	15) <i>considered</i>	16) <i>first</i>	17) <i>three</i>	18) <i>market</i>
19) <i>neurons</i>	20) <i>theory</i>	21) <i>us</i>	22) <i>users</i>	23) <i>possible</i>	24) <i>global</i>
25) <i>well</i>	26) <i>general</i>	27) <i>pipes</i>	28) <i>pipe</i>	29) <i>user</i>	30) <i>2</i>
31) <i>trade</i>	32) <i>i</i>	33) <i>fact</i>	34) <i>water</i>	35) <i>no</i>	36) <i>signal</i>
37) <i>markets</i>	38) <i>specific</i>	39) <i>authors</i>	40) <i>here</i>	41) <i>3</i>	42) <i>presented</i>
43) <i>terms</i>	44) <i>better</i>	45) <i>economic</i>	46) <i>distance</i>	47) <i>external</i>	48) <i>useful</i>
49) <i>differences</i>	50) <i>financial</i>	51) <i>necessary</i>	52) <i>ie</i>	53) <i>procedure</i>	54) <i>context</i>
55) <i>performed</i>	56) <i>countries</i>	57) <i>science</i>	58) <i>capital</i>	59) <i>others</i>	60) <i>previous</i>
61) <i>urban</i>	62) <i>stock</i>	63) <i>best</i>	64) <i>infection</i>	65) <i>neuron</i>	66) <i>bus</i>
67) <i>years</i>	68) <i>gas</i>	69) <i>last</i>	70) <i>cells</i>	71) <i>signals</i>	72) <i>4</i>
73) <i>second</i>	74) <i>strategy</i>	75) <i>applications</i>	76) <i>elections</i>	77) <i>5</i>	78) <i>event</i>
79) <i>techniques</i>	80) <i>technological</i>	81) <i>kind</i>	82) <i>least</i>	83) <i>able</i>	84) <i>friends</i>
85) <i>international</i>	86) <i>current</i>	87) <i>good</i>	88) <i>great</i>	89) <i>ones</i>	90) <i>resonance</i>
91) <i>electronic</i>	92) <i>computed</i>	93) <i>goal</i>	94) <i>description</i>	95) <i>items</i>	96) <i>madrid</i>
97) <i>expected</i>	98) <i>criterion</i>	99) <i>inside</i>	100) <i>epidemic</i>	101) <i>video</i>	102) <i>web</i>
103) <i>difficult</i>	104) <i>respect</i>	105) <i>10</i>	106) <i>survey</i>	107) <i>100</i>	108) <i>instance</i>
109) <i>itself</i>	110) <i>long</i>	111) <i>meter</i>	112) <i>song</i>	113) <i>semantic</i>	114) <i>world</i>
115) <i>country</i>	116) <i>fast</i>	117) <i>competition</i>	118) <i>scientific</i>	119) <i>6</i>	120) <i>7</i>
121) <i>tanzania</i>	122) <i>ontology</i>	123) <i>venezuelan</i>	124) <i>0</i>	125) <i>district</i>	126) <i>likely</i>
127) <i>end</i>	128) <i>messages</i>	129) <i>words</i>	130) <i>valuable</i>	131) <i>genetic</i>	132) <i>twitter</i>
133) <i>commodity</i>	134) <i>internet</i>	135) <i>successfully</i>	136) <i>ghana</i>	137) <i>students</i>	138) <i>voter</i>
139) <i>respectively</i>	140) <i>spanish</i>	141) <i>proteins</i>	142) <i>turbulence</i>	143) <i>zero</i>	144) <i>privacy</i>
145) <i>biological</i>	146) <i>software</i>	147) <i>her</i>	148) <i>companies</i>	149) <i>econometric</i>	150) <i>physical</i>
151) <i>ad hoc</i>	152) <i>interesting</i>	153) <i>districts</i>	154) <i>signatures</i>	155) <i>events</i>	156) <i>currently</i>
157) <i>computer</i>	158) <i>five</i>	159) <i>completely</i>	160) <i>automatically</i>	161) <i>technique</i>	162) <i>ten</i>
163) <i>equal</i>	164) <i>commonly</i>	165) <i>signature</i>	166) <i>pruning</i>	167) <i>prices</i>	168) <i>musical</i>

169) <i>wide</i>	170) <i>12</i>	171) <i>easily</i>	172) <i>tobago</i>	173) <i>available</i>	174) <i>cost</i>
175) <i>authorities</i>	176) <i>knowledge</i>	177) <i>consumer</i>	178) <i>easy</i>	179) <i>right</i>	180) <i>cold</i>
181) <i>trinidad</i>	182) <i>consumers</i>	183) <i>viruses</i>	184) <i>services</i>	185) <i>specialists</i>	186) <i>automatic</i>
187) <i>consumption</i>	188) <i>maximizers</i>	189) <i>economics</i>	190) <i>price</i>	191) <i>company</i>	192) <i>short</i>
193) <i>region</i>	194) <i>8</i>	195) <i>visualization</i>	196) <i>emergent</i>	197) <i>biology</i>	198) <i>fully</i>
199) <i>server</i>	200) <i>message</i>	201) <i>tlp</i>	202) <i>widely</i>	203) <i>face</i>	204) <i>wellknown</i>
205) <i>agreement</i>	206) <i>realtime</i>	207) <i>industrial</i>	208) <i>especially</i>	209) <i>overall</i>	210) <i>election</i>
211) <i>via</i>	212) <i>pollinators</i>	213) <i>government</i>	214) <i>metro</i>	215) <i>pixels</i>	216) <i>namely</i>
217) <i>gpr</i>	218) <i>papers</i>	219) <i>business</i>	220) <i>media</i>	221) <i>achieved</i>	222) <i>online</i>
223) <i>article</i>	224) <i>canada</i>	225) <i>year</i>	226) <i>valencia</i>	227) <i>human</i>	228) <i>spatiotemporal</i>
229) <i>domain</i>	230) <i>shell</i>	231) <i>existing</i>	232) <i>rentacar</i>	233) <i>admissible</i>	234) <i>mmca</i>
235) <i>drazin</i>	236) <i>desired</i>	237) <i>equipments</i>	238) <i>priori</i>	239) <i>equipment</i>	240) <i>clique</i>
241) <i>collaborations</i>	242) <i>videos</i>	243) <i>public</i>	244) <i>institution</i>	245) <i>50</i>	246) <i>importers</i>
247) <i>disease</i>	248) <i>four</i>	249) <i>stations</i>	250) <i>purpose</i>	251) <i>reported</i>	252) <i>educational</i>
253) <i>adult</i>	254) <i>songs</i>	255) <i>fluid</i>	256) <i>virtual</i>	257) <i>his</i>	258) <i>building</i>
259) <i>satisficers</i>	260) <i>comparing</i>	261) <i>food</i>	262) <i>40</i>	263) <i>universal</i>	264) <i>shortest</i>
265) <i>13</i>	266) <i>possibly</i>	267) <i>cfid</i>	268) <i>necessarily</i>	269) <i>mdax</i>	270) <i>contexts</i>
271) <i>own</i>	272) <i>seconds</i>	273) <i>dow</i>	274) <i>remarkable</i>	275) <i>flights</i>	276) <i>entries</i>
277) <i>followed</i>	278) <i>java</i>	279) <i>energy</i>	280) <i>empirical</i>	281) <i>harris</i>	282) <i>economies</i>
283) <i>crisis</i>	284) <i>hours</i>	285) <i>content</i>	286) <i>computation</i>	287) <i>little</i>	288) <i>steady</i>
289) <i>formalism</i>	290) <i>rca</i>	291) <i>actual</i>	292) <i>big</i>	293) <i>picture</i>	294) <i>verifiability</i>
295) <i>15</i>	296) <i>newcomer</i>	297) <i>satisficer</i>	298) <i>protein</i>	299) <i>physics</i>	300) <i>costs</i>
301) <i>skills</i>	302) <i>simultaneously</i>	303) <i>introduction</i>	304) <i>mobile</i>	305) <i>word</i>	306) <i>text</i>
307) <i>regulatory</i>	308) <i>segre</i>	309) <i>virus</i>	310) <i>specifically</i>	311) <i>steps</i>	312) <i>longterm</i>
313) <i>butterflies</i>	314) <i>attempt</i>	315) <i>thermal</i>	316) <i>politicians</i>	317) <i>ballots</i>	318) <i>mc</i>
319) <i>recently</i>	320) <i>greedy</i>	321) <i>something</i>	322) <i>once</i>	323) <i>sr</i>	324) <i>mentioned</i>
325) <i>complicated</i>	326) <i>interested</i>	327) <i>confirmed</i>	328) <i>fairly</i>	329) <i>geographical</i>	330) <i>third</i>
331) <i>table</i>	332) <i>meters</i>	333) <i>malware</i>	334) <i>products</i>	335) <i>alpine</i>	336) <i>technology</i>
337) <i>bachillerato</i>	338) <i>institutions</i>	339) <i>cell</i>	340) <i>bigger</i>	341) <i>welfare</i>	342) <i>reasonable</i>
343) <i>authority</i>	344) <i>straightforward</i>	345) <i>etc</i>	346) <i>parking</i>	347) <i>pages</i>	348) <i>score</i>
349) <i>banking</i>	350) <i>extinctions</i>	351) <i>treatment</i>	352) <i>correct</i>	353) <i>european</i>	354) <i>dimensional</i>
355) <i>apn</i>	356) <i>reference</i>	357) <i>perspectives</i>	358) <i>theories</i>	359) <i>reachable</i>	360) <i>lans</i>
361) <i>jones</i>	362) <i>70</i>	363) <i>improvement</i>	364) <i>challenging</i>	365) <i>implantation</i>	366) <i>fingerprints</i>

367) <i>smr</i>	368) <i>matter</i>	369) <i>20</i>	370) <i>exporters</i>	371) <i>worse</i>	372) <i>experiences</i>
373) <i>six</i>	374) <i>pilots</i>	375) <i>city</i>	376) <i>roi</i>	377) <i>euronet</i>	378) <i>plane</i>
379) <i>policies</i>	380) <i>brokers</i>	381) <i>estonian</i>	382) <i>henhouse</i>	383) <i>vaults</i>	384) <i>absence</i>
385) <i>color</i>	386) <i>feel</i>	387) <i>supported</i>	388) <i>beyond</i>	389) <i>norway</i>	390) <i>gene</i>
391) <i>dilation</i>	392) <i>programs</i>	393) <i>butterfly</i>	394) <i>texts</i>	395) <i>emergence</i>	396) <i>verification</i>
397) <i>payoff</i>	398) <i>exporter</i>	399) <i>telecommunication</i>	400) <i>movements</i>	401) <i>car</i>	402) <i>holland</i>
403) <i>inherent</i>	404) <i>vision</i>	405) <i>university</i>	406) <i>economists</i>	407) <i>stephens</i>	408) <i>mass</i>
409) <i>status</i>	410) <i>ic</i>	411) <i>credit</i>	412) <i>machine</i>	413) <i>situations</i>	414) <i>panel</i>
415) <i>ecommerce</i>	416) <i>percent</i>	417) <i>larval</i>	418) <i>personal</i>	419) <i>physiological</i>	420) <i>florida</i>
421) <i>marketing</i>	422) <i>blocks</i>	423) <i>39</i>	424) <i>transactions</i>	425) <i>extremely</i>	426) <i>box</i>
427) <i>administrative</i>	428) <i>specialized</i>	429) <i>11</i>	430) <i>cancer</i>	431) <i>democracy</i>	432) <i>stored</i>
433) <i>exsoftware</i>	434) <i>cervical</i>	435) <i>governmental</i>	436) <i>atomization</i>	437) <i>vehicles</i>	438) <i>welldefined</i>
439) <i>colour</i>	440) <i>specially</i>	441) <i>surveys</i>	442) <i>adults</i>	443) <i>similarities</i>	444) <i>60</i>
445) <i>cities</i>	446) <i>political</i>	447) <i>correctly</i>	448) <i>extinct</i>	449) <i>desirable</i>	450) <i>implicitly</i>
451) <i>elearning</i>	452) <i>timespace</i>	453) <i>age</i>	454) <i>facultative</i>	455) <i>endoend</i>	456) <i>obligatory</i>
457) <i>liquid</i>	458) <i>details</i>	459) <i>specialist</i>	460) <i>pca</i>	461) <i>efforts</i>	462) <i>trophic</i>
463) <i>ballot</i>	464) <i>professional</i>	465) <i>air</i>	466) <i>celebrities</i>	467) <i>outcome</i>	468) <i>airports</i>
469) <i>overview</i>	470) <i>ultrasonic</i>	471) <i>really</i>	472) <i>posteriori</i>	473) <i>mexico</i>	474) <i>author</i>
475) <i>school</i>	476) <i>jordan</i>	477) <i>pilot</i>	478) <i>street</i>	479) <i>student</i>	480) <i>fisher</i>
481) <i>criterium</i>	482) <i>estonia</i>	483) <i>simply</i>	484) <i>regional</i>	485) <i>candidate</i>	486) <i>contagion</i>
487) <i>rp</i>	488) <i>fitness</i>	489) <i>fruit</i>	490) <i>spain</i>	491) <i>microeconomic</i>	492) <i>decades</i>
493) <i>ibex35</i>	494) <i>deep</i>	495) <i>leg</i>	496) <i>acceptance</i>	497) <i>nearest</i>	498) <i>mosttoleast</i>
499) <i>norwegian</i>	500) <i>opposite</i>	501) <i>socalled</i>	502) <i>aforementioned</i>	503) <i>education</i>	504) <i>false</i>

Lista de las 500 colocaciones más importantes según ordenadas según el *PageRank* (Asociaciones intra-capla lex-lex)

1) <i>complex-network</i>	2) <i>topological-properties</i>	3) <i>network-analysis</i>	4) <i>centrality-measures</i>
5) <i>network-structure</i>	6) <i>minumum-number</i>	7) <i>structural-controllability</i>	8) <i>multiplex-network</i>
9) <i>degree-distribution</i>	10) <i>explosive-percolation</i>	11) <i>mutualistic-networks</i>	12) <i>conditional-variance</i>
13) <i>mutualistic-networks</i>	14) <i>linear-system</i>	15) <i>network-topology</i>	16) <i>distribution-networks</i>
17) <i>concept-terms</i>	18) <i>complex-networks</i>	19) <i>community-detection</i>	20) <i>structural-controllability</i>
21) <i>personalization-vector</i>	22) <i>distribution-networks</i>	23) <i>simulated-networks</i>	24) <i>extinction-scenario</i>
25) <i>total-number</i>	26) <i>community-structure</i>	27) <i>pollination-networks</i>	28) <i>random-network</i>
29) <i>data-analysis</i>	30) <i>competitivy-graph</i>	31) <i>multiplex-networks</i>	32) <i>conditional-volatility</i>
33) <i>power-law</i>	34) <i>network-model</i>	35) <i>phase-transition</i>	36) <i>new-friends</i>
37) <i>connected-components</i>	38) <i>shape-classification</i>	39) <i>formal-languages</i>	40) <i>memory-phase</i>
41) <i>synthetic-networks</i>	42) <i>dynamical-system</i>	43) <i>large-networks</i>	44) <i>cayley-tree</i>
45) <i>phase-transitions</i>	46) <i>important-role</i>	47) <i>weighted-networks</i>	48) <i>relevance-measure</i>
49) <i>network-dimensions</i>	50) <i>network-oscillations</i>	51) <i>operational-capacity</i>	52) <i>structural-properties</i>
53) <i>clustering-coefficient</i>	54) <i>smallworld-networks</i>	55) <i>real-networks</i>	56) <i>graph-theory</i>
57) <i>similarity-matrix</i>	58) <i>highly-connected</i>	59) <i>positive-systems</i>	60) <i>positive-linear</i>
61) <i>variance-equation</i>	62) <i>collective-dynamics</i>	63) <i>sustainable-efficiency</i>	64) <i>ordinary-percolation</i>
65) <i>concept-terms</i>	66) <i>multiplex-graphs</i>	67) <i>degree-distributions</i>	68) <i>dynamical-processes</i>
69) <i>strength-distributions</i>	70) <i>finite-size</i>	71) <i>generalist-core</i>	72) <i>phase-diagram</i>
73) <i>upper-bound</i>	74) <i>spatial-scales</i>	75) <i>vital-minimum</i>	76) <i>secondary-bifurcation</i>
77) <i>networks-theory</i>	78) <i>linear-continuous</i>	79) <i>linear-system</i>	80) <i>new-approach</i>
81) <i>new-nodes</i>	82) <i>real-sdh</i>	83) <i>numerical-simulations</i>	84) <i>network-growth</i>
85) <i>algebraic-topology</i>	86) <i>mutualistic-communities</i>	87) <i>leafwing-population</i>	88) <i>square-matrices</i>
89) <i>extinction-sequence</i>	90) <i>boundary-conditions</i>	91) <i>multiagent-systems</i>	92) <i>unconnected-node</i>
93) <i>synaptic-plasticity</i>	94) <i>propagation-algorithm</i>	95) <i>adjacent-nodes</i>	96) <i>random-networks</i>
97) <i>information-flows</i>	98) <i>information-processing</i>	99) <i>consensus-problem</i>	100) <i>percolation-model</i>
101) <i>percolation-models</i>	102) <i>attachment-model</i>	103) <i>reachability-controllability</i>	104) <i>functional-network</i>
105) <i>preferential-attachment</i>	106) <i>realworld-networks</i>	107) <i>routing-properties</i>	108) <i>downcomer-model</i>
109) <i>node-pairs</i>	110) <i>node-degree</i>	111) <i>node-extinction</i>	112) <i>population-growth</i>

113) <i>digital-environments</i>	114) <i>participation-coefficient</i>	115) <i>interest-area</i>	116) <i>dynamical-systems</i>
117) <i>dynamical-phase</i>	118) <i>large-number</i>	119) <i>mathematical-models</i>	120) <i>topological-attributes</i>
121) <i>edge-correlation</i>	122) <i>phase-space</i>	123) <i>local-centrality</i>	124) <i>adjacency-matrix</i>
125) <i>natural-languages</i>	126) <i>academic-performance</i>	127) <i>new-links</i>	128) <i>initial-conditions</i>
129) <i>algebraic-characterization</i>	130) <i>path-length</i>	131) <i>lag-conditional</i>	132) <i>transformation-dynamics</i>
133) <i>equisingularity-condition</i>	134) <i>closeness-centrality</i>	135) <i>pagerank-method</i>	136) <i>efficiency-measure</i>
137) <i>network-activity</i>	138) <i>network-dynamics</i>	139) <i>network-graphs</i>	140) <i>network-approach</i>
141) <i>network-performance</i>	142) <i>network-evolution</i>	143) <i>network-nodes</i>	144) <i>proposed-algorithm</i>
145) <i>proposed-system</i>	146) <i>antiattack-protocol</i>	147) <i>nodes-only</i>	148) <i>dimension-graphs</i>
149) <i>shortterm-synaptic</i>	150) <i>monte-carlo</i>	151) <i>sufficiently-large</i>	152) <i>friends-required</i>
153) <i>projected-graph</i>	154) <i>closely-related</i>	155) <i>site-percolation</i>	156) <i>considerably-slower</i>
157) <i>topologies-random</i>	158) <i>lower-level</i>	159) <i>connected-component</i>	160) <i>accurate-measurement</i>
161) <i>information-associated</i>	162) <i>information-retrieval</i>	163) <i>information-producers</i>	164) <i>information-flow</i>
165) <i>associated-network</i>	166) <i>scalefree-networks</i>	167) <i>collective-stochastic</i>	168) <i>percolation-transition</i>
169) <i>communications-network</i>	170) <i>matrix-associated</i>	171) <i>submitted-query</i>	172) <i>evolutionary-competitiveness</i>
173) <i>control-system</i>	174) <i>control-inputs</i>	175) <i>memory-mechanisms</i>	176) <i>squared-error</i>
177) <i>evaluation-methodology</i>	178) <i>fourier-spectrum</i>	179) <i>node-correlations</i>	180) <i>method-lrcm</i>
181) <i>huge-amount</i>	182) <i>intrinsic-dynamics</i>	183) <i>multiplex-pagerank</i>	184) <i>top-linked</i>
185) <i>simple-networks</i>	186) <i>large-amount</i>	187) <i>mathematical-representation</i>	188) <i>invariant-subspace</i>
189) <i>topological-features</i>	190) <i>edge-probabilities</i>	191) <i>lowest-pagerank</i>	192) <i>singular-system</i>
193) <i>individual-nodes</i>	194) <i>systems-present</i>	195) <i>critical-exponents</i>	196) <i>critical-role</i>
197) <i>local-nature</i>	198) <i>weighted-average</i>	199) <i>weighted-network</i>	200) <i>weighted-vector</i>
201) <i>weighted-graph</i>	202) <i>relevance-measures</i>	203) <i>continuous-phase</i>	204) <i>marked-class</i>
205) <i>marked-subspace</i>	206) <i>betweenness-centrality</i>	207) <i>bipartite-approach</i>	208) <i>bipartite-graph</i>
209) <i>dimensions-connectance</i>	210) <i>minimal-model</i>	211) <i>academic-level</i>	212) <i>linear-singular</i>
213) <i>linear-model</i>	214) <i>periodic-systems</i>	215) <i>multilayered-network</i>	216) <i>corresponding-spatial</i>
217) <i>sdh-network</i>	218) <i>sdh-system</i>	219) <i>complex-system</i>	220) <i>real-nodes</i>
221) <i>real-data</i>	222) <i>modular-structure</i>	223) <i>undirected-graphs</i>	224) <i>undirected-network</i>
225) <i>theory-approach</i>	226) <i>unsupervised-classification</i>	227) <i>unsupervised-shape</i>	228) <i>density-function</i>
229) <i>evolutionary-rules</i>	230) <i>metzler-matrices</i>	231) <i>activity-state</i>	232) <i>error-statistics</i>
233) <i>solver-parameters</i>	234) <i>competitiveness-graphs</i>	235) <i>theoretical-result</i>	236) <i>internal-structure</i>
237) <i>kcore-decomposition</i>	238) <i>personalized-pagerank</i>	239) <i>mesh-refinement</i>	240) <i>reproduction-number</i>
241) <i>synchronized-hyperchaotic</i>	242) <i>synchronized-state</i>	243) <i>synchronization-properties</i>	244) <i>synchronization-behavior</i>

245) <i>system-parameters</i>	246) <i>system-division</i>	247) <i>system-structural</i>	248) <i>system-only</i>
249) <i>uniformly-distributed</i>	250) <i>workload-management</i>	251) <i>dependent-stimuli</i>	252) <i>transaction-systems</i>
253) <i>hierarchical-organization</i>	254) <i>mutation-probabilities</i>	255) <i>kuramoto-oscillators</i>	256) <i>implemented-solution</i>
257) <i>components-analysis</i>	258) <i>tokens-frequency</i>	259) <i>pagerank-algorithm</i>	260) <i>isolated-nodes</i>
261) <i>multi-agent</i>	262) <i>differing-numbers</i>	263) <i>connectivity-structure</i>	264) <i>modularitybased-algorithms</i>
265) <i>backup-mechanisms</i>	266) <i>flow-measured</i>	267) <i>ssn-model</i>	268) <i>network-data</i>
269) <i>network-formed</i>	270) <i>independent-experts</i>	271) <i>problem-networks</i>	272) <i>properties-nodes</i>
273) <i>attribute-value</i>	274) <i>structure-meaning</i>	275) <i>classification-methodology</i>	276) <i>meaningful-insights</i>
277) <i>links-removal</i>	278) <i>distribution-network</i>	279) <i>distribution-node</i>	280) <i>euclidean-space</i>
281) <i>inlet-boundary</i>	282) <i>densely-connected</i>	283) <i>proximity-graphs</i>	284) <i>nestedness-estimator</i>
285) <i>benchmark-graphs</i>	286) <i>benchmark-tests</i>	287) <i>directly-new</i>	288) <i>directly-linked</i>
289) <i>model-proposed</i>	290) <i>chaotic-system</i>	291) <i>multilayer-network</i>	292) <i>determinant-part</i>
293) <i>graph-information</i>	294) <i>graph-parameters</i>	295) <i>value-series</i>	296) <i>motivations-information</i>
297) <i>driver-agents</i>	298) <i>thorough-characterization</i>	299) <i>extinction-scenarios</i>	300) <i>extinction-patterns</i>
301) <i>precision-considerably</i>	302) <i>boundary-condition</i>	303) <i>centrality-measure</i>	304) <i>separable-problem</i>
305) <i>coupled-system</i>	306) <i>dynamics-related</i>	307) <i>signaturebased-information</i>	308) <i>similarity-measures</i>
309) <i>line-inflow</i>	310) <i>generations-node</i>	311) <i>detection-algorithms</i>	312) <i>evolution-rules</i>
313) <i>state-variables</i>	314) <i>state-space</i>	315) <i>state-feedback</i>	316) <i>input-units</i>
317) <i>small-number</i>	318) <i>small-fraction</i>	319) <i>small-amplitude</i>	320) <i>multiplexbased-attributes</i>
321) <i>availability-transaction</i>	322) <i>innovative-ways</i>	323) <i>hermite-matrix</i>	324) <i>applied-approach</i>
325) <i>nonequilibrium-dynamical</i>	326) <i>truncated-degree</i>	327) <i>utmost-importance</i>	328) <i>laplacian-matrix</i>
329) <i>networkbased-intersection</i>	330) <i>preference-terms</i>	331) <i>negative-examples</i>	332) <i>asymptotically-stable</i>
333) <i>controllable-equisingular</i>	334) <i>poisson-network</i>	335) <i>autonomous-systems</i>	336) <i>subway-networks</i>
337) <i>lognormal-gamma</i>	338) <i>lognormal-distributions</i>	339) <i>random-regular</i>	340) <i>connected-groups</i>
341) <i>connected-node</i>	342) <i>connected-adjacent</i>	343) <i>connected-network</i>	344) <i>similar-topological</i>
345) <i>support-systems</i>	346) <i>binary-images</i>	347) <i>infected-nodes</i>	348) <i>scenario-associated</i>
349) <i>scenario-produced</i>	350) <i>usual-pagerank</i>	351) <i>information-theory</i>	352) <i>information-diffusion</i>
353) <i>information-source</i>	354) <i>information-availability</i>	355) <i>information-required</i>	356) <i>hybrid-recommendation</i>
357) <i>hybrid-algorithm</i>	358) <i>frequency-analysis</i>	359) <i>destructive-methods</i>	360) <i>computing-recommendation</i>
361) <i>positive-continuous</i>	362) <i>positive-stable</i>	363) <i>positive-periodic</i>	364) <i>traditional-statistical</i>
365) <i>recurrence-relation</i>	366) <i>recommendation-algorithm</i>	367) <i>higher-number</i>	368) <i>higher-connectance</i>
369) <i>shape-images</i>	370) <i>accessorial-tool</i>	371) <i>scalefree-distribution</i>	372) <i>scalefree-strength</i>
373) <i>eigenvector-centrality</i>	374) <i>formal-framework</i>	375) <i>collective-response</i>	376) <i>consensus-algorithm</i>

377) <i>contact-process</i>	378) <i>percolation-threshold</i>	379) <i>invertible-part</i>	380) <i>identical-chaotic</i>
381) <i>matrix-formulation</i>	382) <i>pareto-positive</i>	383) <i>pareto-distribution</i>	384) <i>nilpotent-parts</i>
385) <i>compatibility-attachment</i>	386) <i>evolutive-rule</i>	387) <i>direct-correlation</i>	388) <i>size-data</i>
389) <i>control-systems</i>	390) <i>optimal-path</i>	391) <i>only-local</i>	392) <i>memory-network</i>
393) <i>uniform-distribution</i>	394) <i>routing-algorithm</i>	395) <i>routing-process</i>	396) <i>strong-modular</i>
397) <i>preexisting-network</i>	398) <i>garch-model</i>	399) <i>controlled-environment</i>	400) <i>controlled-state</i>
401) <i>metric-closely</i>	402) <i>sufficient-condition</i>	403) <i>sufficient-conditions</i>	404) <i>moment-matrix</i>
405) <i>node-removal</i>	406) <i>node-pair</i>	407) <i>method-proposed</i>	408) <i>huge-impact</i>
409) <i>identification-system</i>	410) <i>intrinsic-properties</i>	411) <i>intrinsic-growth</i>	412) <i>clear-difference</i>
413) <i>multiplex-nature</i>	414) <i>multiplex-relations</i>	415) <i>multiplex-information</i>	416) <i>people-organizations</i>
417) <i>simple-topological</i>	418) <i>simple-model</i>	419) <i>digital-networks</i>	420) <i>digital-network</i>
421) <i>digital-image</i>	422) <i>dynamical-properties</i>	423) <i>large-choice</i>	424) <i>mathematical-model</i>
425) <i>mathematical-definition</i>	426) <i>unique-opportunity</i>	427) <i>invariant-formulations</i>	428) <i>single-evolutive</i>
429) <i>single-line</i>	430) <i>testable-data</i>	431) <i>topological-measure</i>	432) <i>topological-invariants</i>
433) <i>topological-measures</i>	434) <i>computational-environments</i>	435) <i>computational-complexity</i>	436) <i>computational-approaches</i>
437) <i>classical-mathematical</i>	438) <i>classical-stagestructured</i>	439) <i>estimating-conditional</i>	440) <i>collaboration-networks</i>
441) <i>finite-partition</i>	442) <i>finite-family</i>	443) <i>edge-attributes</i>	444) <i>edge-regions</i>
445) <i>active-node</i>	446) <i>stochastic-process</i>	447) <i>nonlinear-terms</i>	448) <i>globalization-process</i>
449) <i>singular-linear</i>	450) <i>greatest-impact</i>	451) <i>descriptive-model</i>	452) <i>important-metric</i>
453) <i>important-actor</i>	454) <i>individual-level</i>	455) <i>individual-motivations</i>	456) <i>reliable-tool</i>
457) <i>lumped-model</i>	458) <i>statistical-analysis</i>	459) <i>spatial-restrictions</i>	460) <i>spatial-restriction</i>
461) <i>spatial-organization</i>	462) <i>query-concept</i>	463) <i>query-terms</i>	464) <i>standard-deviation</i>
465) <i>critical-modes</i>	466) <i>critical-behavior</i>	467) <i>critical-attributes</i>	468) <i>local-optima</i>
469) <i>local-information</i>	470) <i>local-field</i>	471) <i>local-measures</i>	472) <i>local-probabilities</i>
473) <i>novel-application</i>	474) <i>high-clustering</i>	475) <i>high-robustness</i>	476) <i>weighted-geometrical</i>
477) <i>analytical-expressions</i>	478) <i>average-pagerank</i>	479) <i>average-consensus</i>	480) <i>reduced-number</i>
481) <i>weak-sensory</i>	482) <i>operational-criteria</i>	483) <i>structural-relevance</i>	484) <i>directed-networks</i>
485) <i>directed-complex</i>	486) <i>directed-graph</i>	487) <i>bipartite-network</i>	488) <i>bipartite-networks</i>
489) <i>growths-slowly</i>	490) <i>natural-way</i>	491) <i>natural-frequency</i>	492) <i>existence-plausibility</i>
493) <i>certain-topological</i>	494) <i>underlying-network</i>	495) <i>academic-behavior</i>	496) <i>linear-network</i>
497) <i>linear-subsystems</i>	498) <i>circuit-strictly</i>	499) <i>new-state</i>	500) <i>new-mechanism</i>

