MINISTERIO DE CIENCIA E INNOVACIÓN

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN

MEMORIA CIENTÍFICO-TÉCNICA DE PROYECTOS INDIVIDUALES
Convocatoria 2021 - «Proyectos de Generación de Conocimiento»

AVISO IMPORTANTE - La memoria no podrá exceder de 20 páginas. Para rellenar correctamente esta memoria, lea detenidamente las instrucciones disponibles en la web de la convocatoria. Es obligatorio rellenarla en inglés si se solicita 100.000 € o más (en costes directos).

IMPORTANT – The research proposal cannot exceed 20 pages. Instructions to fill this document are available in the website. If the project cost is equal or greater than 100.000 €, this document must be filled in English.

1. DATOS DE LA PROPUESTA – *PROPOSAL DATA*

IP 1 (Nombre y apellidos)**:** Mercè LORENTE CASAFONT

IP 2 (Nombre y apellidos)**:**

TÍTULO DEL PROYECTO (ACRÓNIMO): *Procesos de lexicalización y complejidad de las comunidades de habla en el estudio diacrónico de la terminología y del discurso de especialidad en economía e informática* (EVOTERM)

TITLE OF THE PROJECT (ACRONYM): *Lexicalization processes and complexity of speech communities in the diachronic study of terminology and discourse in economics and informatics* (EVOTERM)

2. ANTECEDENTES, ESTADO ACTUAL Y JUSTIFICACIÓN DE LA PROPUESTA - *BACKGROUND, CURRENT STATUS AND JUSTIFICATION OF THE PROPOSAL*

El proyecto que planteamos se integra en una línea de investigación de la lingüística general, que es el estudio del cambio lingüístico, que tienen antecedentes muy claros y que aún está en pleno desarrollo para muchas lenguas, como puede observarse en la producción de libros y artículos de revistas. Pero hay que tener en cuenta que el estudio de cambio léxico, que se ha considerado un cambio menor desde la perspectiva historicista del lenguaje, es fundamental para el estudio del discurso de especialidad y su correlación con la evolución de las diversas disciplinas científicas y técnicas. Además se trata de un estudio innovador, ya que la terminología diacrónica justo empieza a desarrollarse en estos momentos.

Para poder justificar la propuesta, vamos a introducir a continuación algunos aspectos que guían nuestro proyecto:

Estudios sobre el cambio lingüístico

El cambio lingüístico es un objeto de estudio que tiene un largo recorrido: primero con las propuestas de los *junggrammatiker* de finales del siglo XIX, más tarde acompañando trabajos sobre tipología de lenguas o como foco temático central en la lingüística diacrónica de los años sesenta del siglo XX, y en la actualidad, ya que esta línea de investigación se revitaliza en los años noventa en el marco de la lingüística cognitiva y de la sociolingüística. En todas estas iniciativas, los estudios sobre cambio lingüístico se concentran básicamente en aspectos fonológicos, sintácticos, morfológicos y semánticos, y siempre en relación con la lengua común. (Hock 1991: 1-10)

Estos cambios suelen explicarse a partir de variables sociológicas (generacionales, económicas, de género, migratorias o de nivel sociocultural), de situaciones sociolingüísticas (contacto de lenguas, prestigio de variantes, entrada de préstamos, comunidad de hablantes) y de factores discursivos (la situación comunicativa, el canal, la tipología textual, el género discursivo, el propósito del emisor, el tipo de receptor, etc.). Todos estos elementos juegan un

papel concursal determinante para la puesta en marcha, la estabilización, la difusión y la implantación social de un cambio determinado.

Aunque desde la terminología y la historia de la ciencia ha habido aportaciones diversas sobre cambios que se han producido en el léxico de especialidad (Rodríguez-Rodilla 1998; Barona 2003; Carpi & Ramírez Luengo 2021, entre otros), no se ha estudiado, hasta ahora, el grado de correlación entre el cambio del léxico en la lengua común y el cambio de las unidades terminológicas. ¿Siguen las mismas tendencias? ¿Por las mismas causas? ¿Con la misma periodicidad? ¿En qué se asemejan y en qué difieren?

Autores, como Lehmann (2002), Hopper & Traugott (2003), Himmelmann (2004), Brinton & Traugott (2005), Croley & Bowern (2010) o Bybee (2010, 2015), coinciden en el establecimiento de los siguientes principios fundamentales del cambio lingüístico:

La naturaleza del cambio. El cambio lingüístico es un fenómeno natural de cualquier lengua y de cualquier variante de cada lengua. Los estudios de gramática histórica han mostrado que todas les lenguas cambian del mismo modo (Bybee 2015: 10) y comparten mecanismos, resultados y tendencias, lo que refuerza la idea del carácter universal del cambio lingüístico.

Las causas del cambio. Los factores que se han descrito para la explicación del cambio lingüístico se clasifican en dos bloques: factores internos, que apelan al propio sistema de la lengua y a los procesos cognitivos y pragmáticos que activan los hablantes en sus interacciones; y factores externos, que agrupan aspectos diversos referidos a los cambios socioculturales, a las estructuras sociológicas, a las condiciones sociolingüísticas y a los contextos comunicativos.

Los componentes lingüísticos afectados. De acuerdo con Bybee (2015), los cambios léxicos son seguramente los más obvios para los hablantes, pero no son ni los más sistemáticos ni los que tienen más impacto en la estructura de la lengua. Los cambios estructurales afectan los sistemas fonético y fonológico, la morfología, la sintaxis y, para una parte significativa de los autores consultados, la semántica.

La temporalidad del cambio. Los cambios lingüísticos son lentos y graduales, y los usuarios de una lengua no los podemos percibir en el momento en que se están produciendo. Esta idea ya la planteaba Bloomfield (1933: 347), cuando decía que « The process of lingüístic change has never been directly observed ».

La direccionalidad del cambio. Los estudios interlingüísticos del cambio han demostrado que algunos fenómenos siguen trayectorias bien establecidas, algunas de las cuales son unidireccionales y no reversibles. Los cambios fonéticos y fonológicos o los cambios sintácticos, los más relacionados con la estructura de la lengua, son en general los que trazan una direccionalidad más universal (Hopper & Traugott 1993: 99-139; Lehmann 2002: 17-23). Himmelmann (2004) defiende que gramaticalización y lexicalización no son procesos opuestos ni reversibles, y que ambos siguen en general trayectorias unidireccionales.

La variación como motor del cambio. Tanto la variación como el cambio son fenómenos naturales del lenguaje, pero no debemos confundir el uno con el otro. Muchos casos de variación impulsan la incoación de un cambio, pero no todos los cambios están relacionados con esquemas de variación. Hay consenso en considerar el análisis de la variación como una fuente de evidencia para la detección e, incluso, para la predicción del cambio lingüístico. Desde los postulados labovianos de los años setenta, el estudio empírico de la variación se convierte en una herramienta para el análisis del cambio, al lado de la metodología contrastiva propia de la gramática histórica.

El estudio del cambio en terminología, una novedad

Partimos de la idea de que el cambio terminológico se restringe al nivel del léxico, en un sentido amplio, dada su condición de cruce del resto de componentes: aspectos fónicos, de la morfología léxica, de la semántica, pero también cuestiones relacionadas con el comportamiento sintáctico de las unidades o con la combinatoria léxica. Esta delimitación del espacio del cambio no supone, en ningún caso, la reducción de los factores explicativos sobre el cambio que debemos ajustar al contexto del discurso de especialidad.

Los escenarios de análisis del cambio en terminología son los mismos que en el cambio léxico en general, con excepción de la gramaticalización:

- Desaparición en el uso de determinadas unidades terminologicas
- Neologismos y tipos de neologismos con recursos propios
- Incorporación de préstamos de otras lenguas
- Procesos de lexicalización progresiva
- Compleción de paradigmas léxicos
- Cambio semántico

Las hipótesis que planteamos para el análisis del cambio en terminología (Lorente 2019) son las que siguen:

(1) *Las comunidades de hablantes en ámbitos especializados*. La primera cuestión a plantearse es si en los dominios de especialidad tenemos una o varias comunidades de hablantes, y si estas comunidades son complejas u homogéneas. En el caso de ámbitos como la medicina o el derecho, estamos ante dos grandes configuraciones comunitarias: la comunidad de los expertos y la comunidad de los usuarios. La pertenencia a la primera es de carácter voluntario, a través de la formación y de la actividad profesional, mientras que la pertenencia a la segunda es circunstancial. Abogamos por la complejidad de las comunidades actuales de comunicación especializada en casi todos los ámbitos.

(2) *Interlocutores y contextos*. La coexistencia dentro de un ámbito de una comunidad de especialistas y de una comunidad de usuarios, que pueden interactuar, nos lleva a reflexionar sobre cuáles son los verdaderos interlocutores de cada acto comunicativo. Así, mientras que no dudaremos en que los emisores y los receptores de un artículo científico son en ambos casos expertos en la materia, ejemplos como los del informe médico nos obligan a plantearnos que, en realidad, el emisor médico se dirige a múltiples perfiles de receptores (otros médicos, el paciente e incluso otros colectivos como los enfermeros o los fisioterapeutas). Si, por lo general, los cambios lingüísticos se originan en la lengua oral y pueden quedar reflejados en la lengua escrita, consideramos que, en el caso de los discursos de especialidad, a falta de disponer de fuentes orales, conviene distinguir entre géneros discursivos espontáneos, como el informe médico, y géneros discursivos controlados, como el artículo científico, por poner un ejemplo. Es necesario, por lo tanto, describir los perfiles comunicativos de expertos y otros agentes, los propósitos y los contextos comunicativos preeminentes.

(3) *Temporalidad del cambio*. El período de traspaso entre el siglo XVIII y el XIX supone una inflexión fundamental en la historia de las ideas, por un cambio profundo del saber con el desarrollo continuo de nuevos ámbitos científicos y técnicos. La necesidad de denominaciones terminológicas aumenta extraordinariamente en la medida en que las ciencias, técnicas e intercambios de información cambian y aceleran (Rey 1982: 9-12). A partir de ese momento hasta nuestros días, la evolución de las ciencias y de las técnicas, y de las formas de comunicación que las acompañan, es más que evidente. Tanto los expertos que la viven de primera mano, o la impulsan, como el resto de ciudadanos que la reciben somos capaces de observar cambios conceptuales y denominativos. Defendemos que en ámbitos especializados activos, científicos,

tecnológicos o comerciales, un periodo de entre 7 y 10 años son suficientes para detectar cambios en la terminología de un ámbito, debido a los avances y cambios de modelos.

(4) *Géneros textuales*. Los géneros más espontáneos o menos controlados se producen en circunstancias similares a las de la oralidad (inmediatez, rapidez), que refuerzan la máxima economía de los actos de habla. Por el contrario, los géneros controlados introducen una serie de factores que condicionan su resultado (la inscripción del autor en una determinada escuela de pensamiento o científica, la elección terminológica de escuela o de organismo vinculado, la búsqueda de la claridad, la revisión editorial, etc.). Para analizar el cambio terminológico es preciso seleccionar documentos de géneros textuales espontáneos y controlados.

Centramos pues nuestro proyecto en el análisis de las condiciones del cambio terminológico, que dentro del cambio lingüístico introduce algunas diferencias interesantes como la temporalidad reducida, la relevancia de géneros menos controlados o formales o el rol significativo de los interlocutores en las comunidades de habla o de comunicación especializada, que progresivamente han aumentado su complejidad. Concretamente, por las limitaciones propios de un proyecto de estas características, queremos enfocar nuestra atención en dos aspectos concretos relacionados con el cambio en terminología: un mecanismo lingüístico, la lexicalización, y una estructura sociolingüística, la complejidad de las comunidades de habla actuales de sectores de especialidad.

Respecto de la lexicalización

Con el término *lexicalización* se identifican dos fenómenos de orientación diferentes: el primero, sincrónico, consiste en la codificación léxica de estructuras gramaticales diversas; y el segundo, diacrónico, abarca todos los mecanismos que permiten incorporar nuevas unidades dentro del léxico de una lengua (Brinton & Traugott 2005: 18).

Dentro de la primera orientación, la lexicología restringe el uso del término *lexicalización* para los mecanismos de formación de palabras diferentes de la derivación o de la composición; es decir que lo reserva para denominar procedimientos de fijación de estructuras sintagmáticas o de cambio de categoría. La lexicalización de sintagmas generados en el componente sintáctico implica la fijación formal de la secuencia y que semánticamente apunte a un referente específico (vg. *cuello de botella, llave inglesa, aire acondicionado*). La conversión, procedimiento de formación de palabras que implica el cambio de categoría y de las posibilidades funcionales de una unidad léxica, también se ha denominado *lexicalización* o *derivación cero* (vg. *cantar* > *canto*; *presión* > *presionar*; Eng. *hammer* (tool) > to *hammer*; to *walk* > a *walk*). La conversión pertenece al conjunto de procesos morfológicos de libre dirección: de nombres a verbos, de verbos a nombres, de nombres a adjetivos, de adjetivos a nombres (Bybee 2015: 190). Un tercer procedimiento de lexicalización bastante productivo es la truncación de formas léxicas preexistentes, que incluye la reducción de unidades complejas (*hamburger* > *burger*), la acronimia (*radar*, *laser*), la siglación (*ADN, ONG*), la abreviación (*bici*, *metro*) o la creación onomatopéyica (*ziga-zaga*) (Cabré 1994: 95-96).

Los autores sobre cambio lingüístico reservan el nombre de *lexicalización* para la noción antitética a la de *gramaticalización*: +/- léxico vs. +/- gramatical o funcional. De hecho, esta aproximación diacrónica a la lexicalización abarca fenómenos de cambio que parten de los mecanismos de formación léxica anteriores y que muestran como las unidades lexicalizadas avanzan progresivamente hacia un mayor grado de fijación formal (*mesita de noche, coche bomba*), hacia una mayor capacidad de formación de derivados (*sida* > *sidoso*, *sídico*) y, en algunos casos, a la no composicionalidad (*ojo de buey*, *ojo de pez*). Este tipo de aspectos, más interesantes desde la perspectiva del cambio, no han sido prácticamente estudiados en terminología desde una perspectiva diacrónica.

Sobre la complejidad de las comunidades de hablantes en ámbitos especializados

Con el concepto de *comunidad de habla* (en inglés, *speech community*) los estudiosos de la etnografía de la comunicación denominan todo grupo social que comparte una misma variedad de una lengua y unos determinados modelos de uso. El concepto se origina en los trabajos de W. Labov y de J. Gumperz y D. Hymes sobre variación social del lenguaje y en sus estudios sobre grupos sociales que utilizan una determinada variedad del inglés.

Cada comunidad de habla genera las formas lingüísticas que configuran la variedad lingüística del grupo, que se acaba estableciendo como modelo y marca de identidad para sus componentes. La etnografía de la comunicación se interesa por estudiar comunidades reales y heterogéneas desde el punto de vista lingüístico, social y cultural. Las comunidades de habla se caracterizan por tendencias en su comportamiento lingüístico

Las comunidades de habla de ámbitos especializados han sido poco descritas. Habitualmente se ha discutido en terminología y discurso especializado sobre la delimitación del rol de emisor: para algunas orientaciones en discurso de especialidad (Ciapuscio 2003) los emisores de textos especializados pueden ser diversos, mientras que en terminología, y concretamente en la Teoría Comunicativa de la Terminología (Cabré 1999), se restringe el rol del emisor al experto que controla el conocimiento de su ámbito. En los últimos años, se ha desarrollado una línea de investigación que muestra, por un lado, la evolución y el carácter heterogeneo de las comunidades de expertos, con la integración de profesionales de perfiles ocupacionales y formativos diversos; y que ha puesto sobre la mesa, por otro lado, la incorporación en la comunicación especializada de otros agentes que, aunque no dominen el conocimiento experto, llegan a ser capaces de captar y reproducir algunos de sus componentes, como los mediadores lingüísticos (traductores e intérpretes, editores, correctores, redactores técnicos), los gestores de comunicación científica (periodistas científicos, managers de redes sociales), los aprendices de expertos (estudiantes universitarios y de formación profesional) o, incluso, en algunos sectores, los usuarios inmersos en procesos de empoderamiento social (comunidades de pacientes, de consumidores, de aficionados).

Se hace imprescindible empezar a describir con detalle las comunidades de habla de ámbitos especializados, sus integrantes y sus relaciones comunicativas.

Algunos ejemplos de la bibliografía

Atkinson, Mark; Gregory J Mills, Kenny Smith, Social Group Effects on the Emergence of Communicative Conventions and Language Complexity, *Journal of Language Evolution*, Volume 4, Issue 1, January 2019, Pages 1–18, https://doi.org/10.1093/jole/lzy010

Brinton, Laurel J.; Traugott, Elizabeth C. (2005) *Lexicalization and Language Change*, Cambridge, Cambridge University Press.

Bybee, Joan (2015) *Language Change*, Cambridge, Cambridge University Press.

Himmelmann, Nikolaus (2004) «Lexicalization and grammaticization: Opposite or orthogonal?», in Bisang, Himmelmann, Wiemer *What makes grammaticalization. A book from its fringes and its components*, Berlín, Mouton de Gruyter.

Kopaczyk, Joanna & Andreas H. Jucker (eds.), *Communities of practice in the history of English*. Amsterdam: John Benjamins, 2013. Pp. 291.

Kretzschmar, Jr, W. (2015). *Language and Complex Systems*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781316179017

Levin, B. and M. Rappaport Hovav (2019) «Lexicalization Patterns», in R. Truswell, ed., *Oxford Handbook of Event Structure*, Oxford University Press, Oxford, UK, 395-425.

Lipka, L., Handl, S., & Falkner, W. (2004). *Lexicalization & Institutionalization * The State of the Art in 2004*.

Mufwene, S., Coupé, C., & Pellegrino, F. (Eds.). (2017). *Complexity in Language: Developmental and Evolutionary Perspectives* (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press. doi:10.1017/9781107294264

Poplack, Shana; Sankoff, David; Miller, Christofer (1988) «The social correlates and linguistic processes of lexical borrowing and assimilation», *Linguistics*, 26, pp. 47-104.
Traugott, Elizabeth C.; Dasher, Richard B. (2004) *Regularity in Semantic Change*, Cambridge, Cambridge University Press.

Antecedentes y justificación de la propuesta

Este proyecto de investigación se insiere en la línea de investigación sobre terminología, y más concretamente en terminología diacrónica y evolución de la ciencia, del grupo IULATERM (Léxico y Tecnología), grupo consolidado del Mapa de Recerca de Catalunya, con la última referencia 2017 SGR 1530 (2018-2020), iniciada en el proyecto TERMMED. *Evolución del conocimiento científico en medicina: el cambio léxico y semántico* (FFI2017-88100-P) del período 2018-2020, cuyas investigadoras principales han sido Rosa Estopà y Mercè Lorente.

En este proyecto anterior se aplicaron las hipótesis sobre el cambio terminológico al análisis de un corpus en español sobre reproducción asistida dentro de la medicina, fragmentado en períodos de 7 años. Los escenarios del cambio analizados fueron: la neología de estructura sintagmática y su evolución; las siglas; la desaparición de términos; los mecanismos morfológicos y los patrones sintagmáticos relevantes; y la estabilización de términos mediante su incorporación en diccionarios.

Los resultados han mostrado que los factores externos (hitos históricos en reproducción asistida) condicionan progresivamente la ampliación de sintagmas terminológicos, la substitución de denominaciones y la desaparición de ciertas unidades. Entre los cambios explicables por factores internos destacamos la proliferación de siglas y la capacidad de formación de neologismos con recursos propios, con una baja presencia de préstamos del inglés. Otro resultado significativo ha sido la necesidad de ajuste de la temporalidad relevante para el análisis del cambio en terminología, que se había previsto para 7 años, y que observamos que puede ampliarse con resultados más ajustados a períodos de diez años.

En este nuevo proyecto queremos incidir en otros ámbitos de especialidad, la informatica y la economía, para poder generalizar algunas de las tendencias detectadas en el proyecto anterior. Por otro lado, ensayada ya la metodología de investigación, nos proponemos objetivos más ambiciosos en la aplicación (dos lenguas y dos ámbitos) y en la teoría (profundizar en el concepto de lexicalización en el cambio lingüístico y en el de comunidades de habla adaptadas a contextos comunicativos especializados).

3. OBJETIVOS, METODOLOGÍA Y PLAN DE TRABAJO - *OBJECTIVES, METHODOLOGY AND WORK PLAN*

Los **objetivos generales** del proyecto EVOTERM son los siguientes:

1. Establecer bases téoricas y metodológicas del estudio de la evolución del léxico de especialidad o terminología, en el marco del cambio lingüístico.
2. Analizar la variación y los procesos de cambio morfológico, sintáctico, semántico y pragmático relacionados con el concepto de lexicalización, con datos reales extraidos de corpus textuales de los ámbitos de la economía y de la informática en español y en catalán.
3. Caracterizar las comunidades de hablantes en estos sectores de especialidad (economía e informática), que han evolucionado desde comunidades homogéneas de expertos a comunidades complejas con agentes con distintos perfiles de formación, distintos perfiles de aplicación, distintos propósitos comunicativos y distintos modelos de referencia.

Estos objetivos generales se despliegan en los siguientes **objetivos específicos o paquetes de trabajo**:

1. Constitución de dos corpus de textos de economía actualizados y de dos corpus de textos de informática, ambos del período (2000-2020) y en español y catalán. Incorporación de estos textos actuales al Corpus Técnico del Institut de Lingüística Aplicada de la UPF. Puesta a disposición de estos corpus actualizados en e-repositorio de investigación.
2. Fragmentación en períodos de 10 años de los corpus completos actualizados.
3. Extracción automática de terminología de los corpus fragmentados.
4. Establecimiento del protocolo de análisis sobre procesos de lexicalización.
5. Realización de análisis comparativos a partir de los corpus fragmentados en cada lengua y en cada ámbito para cada proceso detectado.
6. Establecimiento del protocolo de análisis de las comunidades de cada ámbito especializado: detección de perfiles de formación, perfiles de aplicación y propósitos comunicativos a partir de la consulta de las fuentes de los corpus y de la consulta de expertos de cada materia (consultores del proyecto).
7. Desarrollo-ampliación de la herramienta de detección y desambiguación de siglas.
8. Desarrollo-ampliación de la herramienta de visualización de datos léxicos en evolución.
9. Análisis de la correlación entre procesos de lexicalización, contextos comunicativos y comunidades de hablantes establecidas.

Los **resultados aplicados** de este proyecto (recursos lingüísticos accesibles a toda la comunidad investigadora y a otros interesados) son:

1. Corpus Técnico del IULA-UPF actualizado en economía e informática hasta 2020 en http://bwananet.iula.upf.edu/
2. Corpus español de economía 2000-2020 en e-repositorio de datos de investigación
3. Corpus catalán de economía 2000-2020 en e-repositorio de datos de investigación
4. Corpus español de informática 2000-2020 en e-repositorio de datos de investigación
5. Corpus catalán de informática 2000-2020 en e-repositorio de datos de investigación
6. Listados de términos de cada corpus con índices de frecuencia y de terminologicidad en e-repositorio de datos de investigación
7. Listados de neologismos no incorporados en diccionarios con contextos de uso e información lexicográfica en e-repositorio de datos de investigación
8. Vocabulario catalán-español-inglés sobre economías alternativas, accesible por web
9. Herramienta para la detección y la desambiguación de siglas.
10. Herramienta de visualización de datos terminológicos diacrónicos.

La **metodología** que proponemos ya ha sido experimentada en el proyect oanterior:

a) El equipo investigador dentro del Institut de Lingüística Aplicada (IULA) de la UPF tiene mucha experiencia en la constitución, el etiquetado y la explotación de datos en **corpus textuales** especializados. El proyecto de constitución de un Corpus Técnico en español, catalán e inglés de textos especializados en genómica, medicina, derecho, economía, informática y medio ambiente del IULA se desarrolló de 1994 a 2004, con una financiación inicial de la Generalitat de Catalunya. En el marco de algunos proyectos del Plan Nacional, como TEXTERM-3 (HUM2006-09458), dirigido por Teresa Cabré Castellví, y RICOTERM-3 (HUM2007-65966-C02-01/FILO), dirigido por Mercè Lorente Casafont (IP de este proyecto), se ampliaron los corpus iniciales con nuevos textos más actualizados. Y, en el marco de proyecto METANET4U (2011-2013, UE270893): (1) el procesamiento del corpus se adaptó a las directrices del estándar LAF (Language resource management -- Linguistic annotation framework - ISO 24612:2012): formato XML y anotación "stand-off" y (2) se añadió el nivel de anotación sintàctica a más de 42.000 frases en castellano del corpus.
b) En el IULA se han desarrollado diversas herramientas de **extracción automática de terminología**, con estrategias distintas: YATE y WikiYATE son extractores ideados y

desarrollados por Jorge Vivaldi (miembro del equipo de trabajo de este proyecto) basados en la combinación de estrategias estadísticas y lingüísticas, entre las que se encuentran el acceso a ontologias léxicas o estructuras de clasificación documental. Terminus 2.0. es un sistema completo de gestión de corpus y terminología, ideado y ditigido por Teresa Cabré, que incluye un extractor automático de base estadística. Los tres extractores han sido adaptados a distintas lenguas y probados en distintos ámbitos de especialidad con resultados excelentes.

c) El proyecto *TERMMED. Evolución del conocimiento científico en medicina: el cambio léxico y semántico* (FFI2017-88100-P), codirigido por Mercè Lorente y Rosa Estopà durante el período 2018-2020, supone un antecedente metodológico clave para este nuevo proyecto. Con él se abrió la línea de investigación en terminológia diacrónica dentro de nuestro grupo y se pusieron las bases de la metiodología para analizar la evolución del léxico de especialidad mediante corpus textuales como reflejo de la evolución de la ciència, en ese caso en el ámbito de la reproduccion asistida. Como hipótesis del proyecto TERMMED se estableció el período de 7 años como unidad idonea para la fragmentación de corpus de ámbitos científicos, y se pudo comprobar que las variables externas recomendarían ajustar esta medida a 10 años, que es lo que proponemos en este proyecto. Por otro lado, TERMMED suposo establecer otras hipótesis o escenarios concretos para el análisis del cambio léxico en el discurso de especialidad, que dieron muy buen resultado en el ámbito médico estudiado en ese proyecto, la reproducción asistida. Consideramos que este nuevo proyecto tendrá que validar estas hipótesis o escenarios del cambio, que exponemos a continuación:
    - Desaparición de un concepto, desuso de un objeto o abandono de una técnica, que se reflejan en el discurso por la desaparición de unidades terminológicas en períodos posteriores (terminología obsoleta)
    - Cambios conceptuales o modificaciones de técnicas, que suponen en el discurso la polisemia creciente de algunas unidades terminológicas (cambio semántico).
    - Mayor especialización y complejidad de conceptos, que condicionan en el discurso la proliferación creciente de sintagmas lexicalizados cada vez más largos y de sus variantes en forma de siglas (neología).
    - La difusión social de ciertas unidades terminológicas, su estabilizacion en el uso no experto y su entrada en diccionarios de lengua general (fijación lexicográfica).

La Universitat Pompeu Fabra (UPF) aporta, a través del Instituto de Lingüística Aplicada (IULA), la **infraestructura** necesaria para la ejecución del proyecto:

- Infraestructura informática: servidores, red informática, ordenadores portátiles, ordenadores de sobremesa, impresoras, escáner, grabadora de vídeo.
- Espacios: sala de trabajo del grupo de investigación IULATERM (10 espacios de trabajo) y sala de reuniones para ubicación de colaboradores externos.
- Gastos básicos aplicables al proyecto: teléfono, fungible (tóner de impresoras, papel, fotocopias), correo y mensajería.
- Aplicaciones informáticas (desarrolladas por el IULA-UPF): Bwananet (explorador del corpus); Terminus 2.0 (gestor de corpus y terminología); Yate y Wikiyate, dos versiones del extractor automático de terminología con estrategias estadísticas y lingüísticas; herramientas de procesamiento del lenguaje natural para el español y el catalán; Esten 2.0. (herramienta de análisis y visualizacion de datos léxicos discrónicos).

Además, los servicios centrales de la UPF aseguran el apoyo en tareas de edición de webs, a través de la Biblioetca-CRAI, en tareas administrativas, a cargo de la secretaría del Departamento de Traduccion y Ciencias del Lenguaje y del Servicio de Investigación de la UPF, y en tareas informáticas (copias de seguridad, mantenimiento de equipos y red, software, apoyo en el desarrollo de webs), a cargo del Servicio de Informática de la UPF del Campus de la Comunicación. No se prevé, por lo tanto, solicitar ningún equipamiento nuevo o singular para este proyecto.

La necesidad de dedicación intensiva a las tareas de preprocesamiento y de procesamiento estructural y lingüístico de los textos de los nuevos corpus justifica nuestra petición de incorporar la **contratación laboral de un investigador postdoc junior** para la primera anualidad del proyecto. Aunque los investigadores del proyecto durante esta primera anualidad se centrarán en las tareas de constitución de los cuatro corpus (economía e informática, español y catalán), los objetivos son ambiciosos y todos los investigadores tienen obligaciones docentes y de otra naturaleza, por lo que garantizar una continuidad diària a tiempo completo en las tareas de procesamiento de los corpus aseguraría la compleción de los objetivos parciales para la primera anualidad (objetivos específicos 1 y 2). Los requisitos de joven doctor nos permite preveer que los candidatos o las candidatas estén formados en tareas de procesamiento del lenguaje natural y constitución de corpus lingüísticos. Según hemos previsto en el cronograma, los tres primeros meses del proyecto se reservan para tareas de diseño, selección y recuperación de textos para los cuatro corpus. En este período se realizarían también las gestiones para la oferta de contratación y la selección de candidatos/-as.

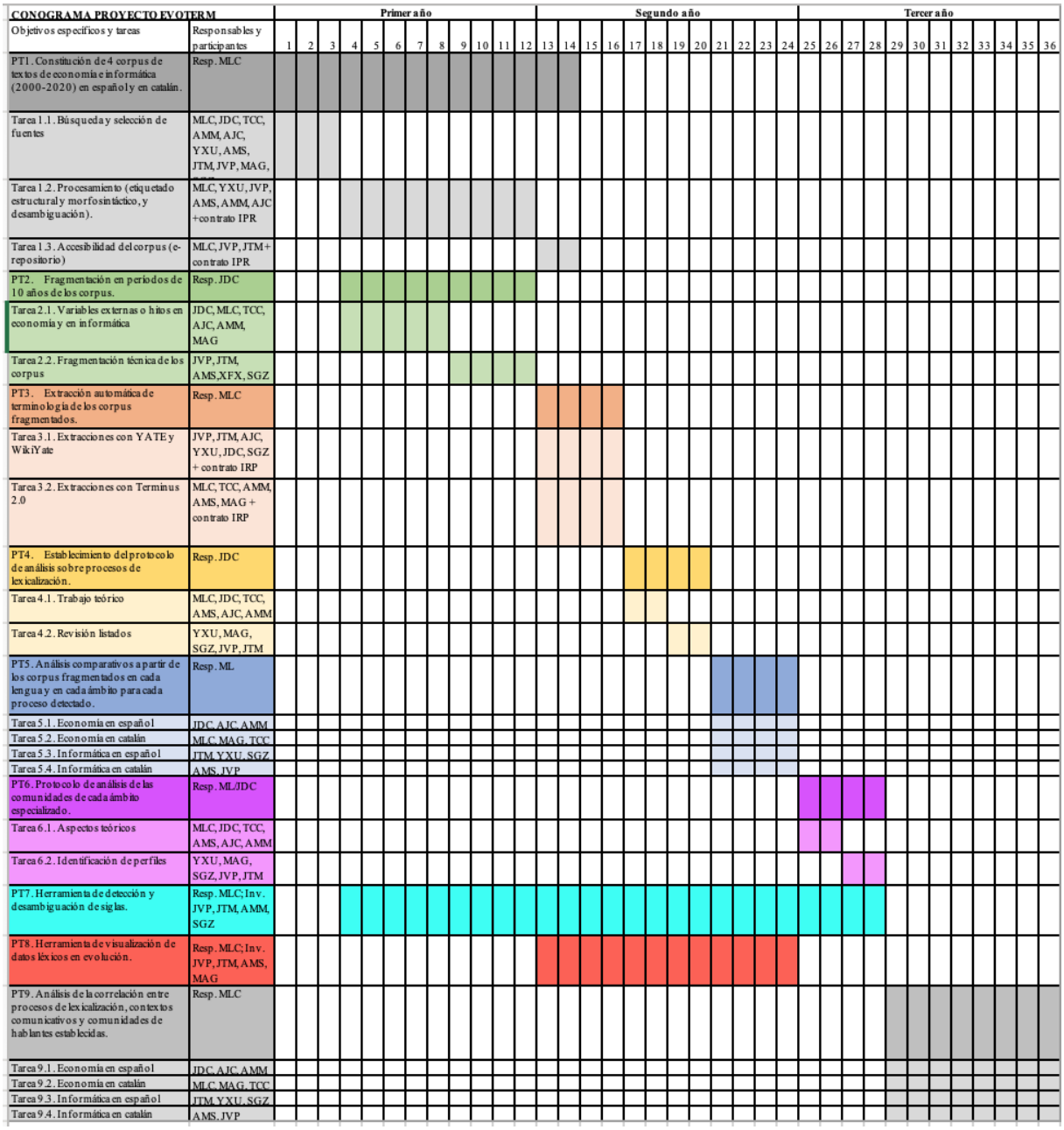
Entendemos que el proyecto se articula en una **orientación interdisciplinaria**, que combina aspectos lingüísticos o gramaticales (formacion de palabras, neología, procesos de lexicalización, adopción y adaptación de anglicismos y otros préstamos, truncaciones y siglas) con aspectos sociolingüísticos y de la etnografía de la comunicación, como la identificación y la descripción de la complejidad de las comunidades de habla propias de la economía y de la informatica actuales. Para conseguir la máxima integración interdisciplinaria se han previsto las siguientes acciones:

1. Seminarios de lecturas sobre complejidad, comunidades de habla y cambio lingüístico.
2. Seminarios invitados de expertos en etnografía de la comunicación, sociolingüística y complejidad para debatir los retos planteados en el proyecto.
3. Consultoría externa de algunos de los expertos (Lluís Payrató, M. Àngels Massip y Mar Cruz Piñol de la Universitat de Barcelona, Lucy Nausbam de la Universitat Autònoma de Barcelona).
4. Organización y celebración de una jornada sobre la interrelación entre comunidades de habla y variación lingüística en ámbitos especializados (prevista para la segunda anualidad).

Consideramos también que la propuesta tiene **orientación multidisciplinaria**, porque al trabajar con textos especializados, terminología y comunidades de habla en economía y en informática se hace necesario adentrarnos en las estructuraciones conceptuales que guían la organización de las actividades, las comunicaciones y la articulación teórica y metodologica de ambos sectores especializados. Para conocer mejor los entresijos de los ámbitos especializados se han previsto las siguientes acciones:

1. Consultoría de expertos en economía: Xavier Freixas de la Unversitat Pompeu Fabra, Eloi Serrano de la Cátedra de Economía Social y Eduard Arruga de la Societat Catalana d'Economia.
2. Consultoría de expertos en informática. Como hay dos miembros de equipo de trabajo doctores expertos en informática, se prevé completar el trio de expertos con la participación de Gerardo Sierra, director del Instituto de Ingeniería Lingüística de la Universidad Nacional Autónoma de México.
3. Organización y celebración de una jornada sobre comunicación, terminología y comunidades de habla en economía e informática (prevista para la tercera anualidad).
4. Publicación de algunos resultados del proyecto en revistas de estos ámbitos.

Cronograma



Responsables/Participantes	
Mercè Lorente Casafont	MLC
Janet De Cesaris Ward	JDC
Teresa Cabré Castellví	TCC
Albert Morales Moreno	AMM
Anna Joan Casademont	AJC
Jorge Vivaldi Palatresi	JVP
Juan Mnuel Torres Moreno	JMT
Àngels Martínez Salom	AMS
Yingfeng Xu	YXU
Mariona Arnau Garcia	MAG
Sangzhu Zhaxi	SGZ

4. IMPACTO CIENTÍFICO-TÉCNICO - *SCIENTIFIC-TECHNICAL IMPACT*

El impacto científico-técnico de los resultados de este proyecto se espera que tenga incidencia, en primer lugar, en las mismas disciplinas de la terminología y del análisis del discurso especializado, porque como se ha indicado la línea de estudios diacrónicos de épocas más recientes y con aplicación de la lingüística de corpus y de las tecnologías del lenguaje aun es muy incipiente. Consideramos que este proyecto y sus resultados puedan impulsar definitivamente esta perspectiva de estudio dentro de la terminología y el análisis del discurso.

En segundo lugar, la focalización del proyecto en los procesos de lexicalización y en el análisis de las comunidades de habla en economía e informática deberá recibir un impacto positivo en foros más generales de la lingüística teórica y de la lingüística aplicada, y en ámbitos más restringidos de los mismos, como el estudio del cambio lingüístico y la lexicografía.

En tercer lugar, el componente del proyecto dedicado al desarrollo de tecnologías del lenguaje (objetivos específicos / paquetes de trabajo 7 y 8) tendrá impacto asegurado en los foros de lingüística computacional (revistas y congresos internacionales).

Por otro lado, la estrategia de difusión de los resultados en publicaciones y congresos específicos de la economía y de la informática incide directamente en el impacto científico en los círculos interesados en la vertiente comunicativa y divulgadora de estas disciplinas. En esta línea, cabe destacar el compromiso de expertos externos en estos ámbitos como asesores para los contenidos del proyecto y para la difusión de los resultados.

Plan de difusión e internacionalización

El equipo de trabajo tiene la intención de concentrar la difusión de resultados en el ámbito internacional, mediante revistas de impacto y congresos internacionales de tradición consolidada. La previsión de mínimos para la difusión de los resultados del proyecto, que hemos indicado en la solicitud del proyecto, es la siguiente:

1. Organización de 2 eventos en la UPF, uno sobre lexicalización y otro sobre comunidades de habla en ámbitos especializados.
2. Publicación de 2 libros electrónicos (e impresión reducida a petición) con los resultados de los eventos anteriores y del proyecto: uno con financiación solicitada a cargo de este proyecto, otro con cofinanciación externa.
3. Capitulos de libro (previsión de 10) firmados por los investigadores del proyecto, en los libros anteriores y en otros proyectos editoriales monográficos internacionales.
4. Artículos en revistas de impacto (mínimo de 5), en revistas de lingüística, sociolingüística, tecnologías del lenguaje, terminología y lexicografía.
5. Otros artículos de divulgación en revistas de economía e informática (mínimo 2), que admitan temas de comunicación especializada.
6. Comunicaciones en congresos internacionales (previsión de 10) firmados por los investigadores del proyecto.
7. Publicaciones *open access* de datos de investigación (corpus, extracciones de términos) en el e-repositorio CORA. Repositori de Dades de Recerca.
8. Publicaciones *open access* en el e-repositorio de la UPF de los informes finales correspondientes a los resultados de los objetivos generales del proyecto: sobre el cambio léxico en terminología; sobre lexicalización; y sobre comunidades de habla en ámbitos especializados.

Planificación temporal de la difusión

Primera anualidad: Propuestas de publicaciones y participaciones en congresos sobre corpus fragmentados, extracción de terminología y análisis de la variación y de la evolución, y sobre la herramienta de visualización.

Seguna anualidad: Propuestas de publicaciones y participaciones en congresos sobre cambio terminológico, sobre procesos de lexicalización y sobre la herramienta de detección y desambiguación de siglas.
Tercera anualidad: Propuestas de publicaciones y participaciones en congresos sobre comunidades de habla en ámbitos especializados y sobre resultados generales del proyecto.

Previsión de algunas revistas de impacto (primeros objetivos)

Terminology
International Journal of Lexicography
Treballs de Sociolingüística Catalana
Linguistic Variation and Change
Computational Linguistics
Caplletra
Ibérica

Previsión de algunos congresos internacionales consolidados

Simposio Iberoamericano de Terminología (RITERM)
CINEO. Congreso Internacional de Neología de las Lenguas Románicas
Congreso Internacional de Lingüística de Corpus
EURALEX. Congreso Internacional de Lexicografía
LREC. Language Resources European Congress.

Resultados susceptibles de transferencia

- Diccionario multilingüe en economías alternativas (catalán, castellano e inglés)
- Corpus textual 2000-2020 en economía en español
- Corpus textual 2000-2020 en economía en catalán
- Corpus textual 2000-2020 en informática en español
- Corpus textual 2000-2020 en informática en catalán
- Extracciones terminológicas de los corpus con datos de frecuencia y de terminologicidad.

EPOs
Han manifestado el interés en los resultados de este proyecto las siguientes entidades EPO:

Instituto Cervantes
El Instituto Cervantes es la institución pública creada por España en 1991 para promover universalmente la enseñanza, el estudio y el uso del español y contribuir a la difusión de las culturas hispánicas en el exterior. En sus actividades, el Instituto Cervantes atiende fundamentalmente al patrimonio lingüístico y cultural que es común a los países y pueblos de la comunidad hispanohablante. Está presente en 88 ciudades de 45 países, a través de sus centros, aulas y extensiones, por los cinco continentes. Además, cuenta con dos sedes en España, la sede central de Madrid y la sede de Alcalá de Henares. Entre sus objetivos está publicar numerosos recursos digitales por medio del Centro Virtual Cervantes y poner a disposición del público interesado una biblioteca electrónica.

Sociedad Catalana de Economía
La Sociedad Catalana de Economía es una sociedad académica creada para fomentar el estudio y la investigación en el ámbito de la economía y procurar su difusión a lo largo del territorio de lengua y cultura catalanas. La Sociedad está adscrita a la Sección de Filosofía y Ciencias Sociales del Institut d'Estudis Catalans, y se financia a través de sus aportaciones, así como de las que hacen los socios y las socias. Entre sus objetivos están servir de foro académico donde las personas economistas catalanas, asociadas e invitadas puedan exponer y debatir todos los temas relativos a la ciencia económica que sean del interés de la Sociedad

y agrupar a las personas economistas que trabajan en el estudio y la investigación de la economía, en cualquiera de las ramas de la ciencia económica.

Plan de gestión de datos

Presentamos aquí el resumen del plan de gestión de datos de este proyecto, centrado en los corpus textuales que se pondrán a disposición de la comunidad investigadora en e-repositorio de datos de investigación para ser reutilizados. En el aplicativo, como documento anexo, pueden descargarse el plan completo.

1.A State the purpose of the data collection/generation and its relation to the objectives of the project	Four textual corpora which consist of a number of specialized texts (Economics and Computer Science domains) available in Spanish and Catalan languages. These LSP corpora have been compiled with articles from open access specialized publications (PhD theses, Journals, Preprints, Reports, etc.) fron 2000 to 2020. Each corpus contains about 0,5 M words in 40 or 50 documents. This four corpora will result from specific objective 1 of the project.
1.B Specify the types and formats of data will generate/collect	These LSP textual corpora will be generated in textual format (txt ASCII).
1.C Specify if existing data is being re-used and how	These LSP corpora will be new data sources, because IULA Technical Corpus only contains LSP texts until 2000 in Economics and Computer Science. No data reuse is expected.
1.D Specify the origin of the data	As the data are collected, there are different sources from which they were extracted: data from the scientific literature and from partners, all of them in open access.
1.E State the expected size of the data	The expected size of the data is not currently known, but it is likely to be 4 GB.
1.F Outline the data utility: to whom will it be useful	The data will be suitable for use by other research groups working on the following topics from computer sciences: big data, terminology extraction, automatic translation, ontologies, NLP reusing.
2.2.F Specify where the data and associated metadata, documentation and code be deposited	The data generated under this project and information supporting information preservation and reuse will be deposited in CORA – Research Data Repository (RdR) (https://dataverse.csuc.cat/). It's a multidisciplinary data repository that allows Catalan universities to publish research datasets in FAIR mode and following the EOSC guidelines. This repository is designed to collect, disseminate, and provide persistent and reliable, long-term open access to the data. The CORA – Research Data Repository (RdR) assigns DOIs for persistent identification and identifiability of the dataset.
2.3.A Assess the interoperability of your data.	The data produced in the project will be interoperable as the datasets will adhere to standardised formats (ASCII, txt).
5.A Ethical or legal issues that may affect the collection and sharing of data	All the activities carried out under the <EVOTERM> project comply with ethical principles and relevant national, EU and international legislation, for example the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights. The tasks for <EVOTERM> only concern basic research activities and the project does not involve humans, animals or cells.

5. IMPACTO SOCIAL Y ECONÓMICO - *SOCIAL AND ECONOMIC IMPACT*

Beneficios, tangibles e intangibles

Los resultados aplicados del proyecto (corpus textuales, vocabulario multilingüe, extracciones de términos, herramienta de siglas, herramienta de visualización) serán beneficios tangibles para la comunidad investigadora, tanto para investigaciones y desarrollos en lingüística computacional, como en estudios de historia de la ciencia (economía e informática), comunicación y divulgación de la ciencia, lingüística descriptiva y teórica, especialmente en terminología, neología y discurso especializado.

Los resultados descriptivos y teóricos del proyecto, en materia de lexicalización en español y en catalán y sobre el tema de la configuración compleja de las actuales comunidades de habla o de comunicación especializada, aunque intangibles, serán beneficios claros en el ámbito de la terminología y el análisis del discurso especializado. Hay que recordar que la línea de terminología diacrónica es un tema nuevo y emergente, en el encuadre lingüístico de la terminología. Como hemos indicado, las acciones de difusión (publicaciones) tienen como objetivo transformar estos beneficios también en tangibles.

Creación de empleo

En este proyecto se solicita la incorporación de un FPI como nuevo doctorando vinculado al 100% al proyecto.

Además se prevé la contratación de un investigador doctor junior (IPR 1) a tiempo completo por un año de duración, que se iniciaría a partir del tercer mes de ejecución efectiva del proyecto. Esta iniciativa prevé ayudar a la incorporación en el mercado laboral de jovenes doctores en humanidades.

Dimensión de género del proyecto

El proyecto EVOTERM tiene en cuenta la dimensión de género en diversos elementos. El primero es el de la constitución del equipo (investigador y de trabajo) que se estructura en una proporción de 36,4% de hombres y 63,6% de mujeres, que refleja bastante fielmente la proporción de investigadores en el ámbito del léxico actualmente.

El segundo aspecto tiene en cuenta la consultoría de expertos externos al proyecto, que ofrecerán actividades de formación y de debate a los miembros del equipo y a otros públicos. En este sentido se ha controlado que la proporcionalidad de género de los consultores sea sostenible si lo ponemos en relación con la proporcionalidad de miembros del equipo. Así, las mujeres consultoras suponen un 37,5% y los hombres un 62,5%.

Y la tercera dimensión, la más importante, se refiere a la voluntad de impulsar la visibilidad de autorías femeninas en los textos seleccionados para la constitución de corpus textuales de ámbitos especializados. Precisamente en economía y en informática los datos y las percepciones sociales nos situan ante una imagen social de minorización para las mujeres, que se debería revertir en recursos lingüísticos puestos a disposición de la sociedad. Así, en el protocolo de selección de textos *open access* para la integración en los corpus textuales del proyecto, se prevé realizar acciones de equilibrio de género, siempre que sea posible.

Bienestar y mejora de formación de la población

La investigación en humanidades presenta carencias evidentes en la percepción positiva de su necesidad y su impacto social por parte de un sector de la ciudadanía e incluso de ciertos agentes de la gestión de la investigación. Así, en los últimos años, hemos visto como se asocia el bienestar de las personas y del planeta en general a temas como la salud, la tecnología o el medio ambiente. No obstante, las humanidades participan directamente en innovaciones y desarrollos que aportan bienestar a las personas en educación y cultura, y

también colaboran con otros expertos en proyectos interdisciplinarios, como los de alfabetización en la salud, mejoras de la comunicación entre médico y paciente, tecnologías del lenguaje aplicadas a la recuperación de información, robots asistentes (tecnologías del habla), traducción automática, gestión del multilingüismo, materiales formativos, etc.

Este proyecto se insiere en esta idea. La elaboración de recursos lingüísticos facilita el acceso a la información de los investigadores y del público en general, y puede revertiir en la mejora de productos tecnológicos de alto consumo. Por otro lado, la generación de conocimiento en cualquier ámbito tiene repercusiones directas en la formación de la población y por lo tanto del bienestar de la sociedad.

6. CAPACIDAD FORMATIVA - *TRAINING CAPACITY*

El grupo IULATERM (Léxico y Tecnología) ha mostrado un alta capacidad formativa a lo largo de su existencia. Sus investigadores imparten docencia en los grados de Traducción e Interpretación y de Lenguas Aplicadas y en los másters universitarios de Estudios del Discurso, de Lingüística Teórica y Aplicada y de Formación del profesorado de la Universitat Pompeu Fabra. Además organizan e imparten, desde 2004, el Máster online en Terminologia. En el marco del Instituto de Lingüística Aplicada, organizaron e participaron en el Programa de Doctorado en Lingüística Aplicada entre 1994 y 2007, año en que se integraron los programas de doctorado del IULA y del Departamento de Traducción y Ciencias del Lenguaje. El programa conjunto actual se denomina Programa de Doctorado en Traducción y Ciencias del Lenguaje.

Desde 1999 los investigadores del grupo IULATERM han dirigido un total de 75 tesis (62 defendidas en la Universidad Pompeu Fabra, 11 defendidas en otras universidades y 2 actualmente en depósito). En los últimos diez años, han dirigido 35 tesis defendidas, 2 en depósito y están dirigiendo 10 tesis en curso.

La investigadora principal del proyecto ha dirigido un total de 12 tesis doctorales (7 defendidas en los últimos diez años) y actualmente dirige 3 tesis en curso.

Tesis en curso dentro del grupo IULATERM

1. María del Carmen Vicens Català. Codirección: Ona Domènech y Yingfeng Xu
2. Francisco Javier Izquierdo Núñez. Dirección: Judit Freixa Aymerich.
3. Sangzhu Zhaxi. Codirección: Mercè Lorente y Yingfeng Xu
4. Mariona Arnau Garcia. Dirección: Mercè Lorente
5. Miguel Ángel Verón. Dirección: Joan Costa Carreras.
6. Ana María Tangarife. Codirección: Mercè Lorente y Núria Bel
7. Zebing Lu. Dirección: Rosa Estopà Bagot
8. Miguel Vázquez Sánchez. Dirección: Joan Costa Carreras.
9. Said Dani. Codirección: Ona Domènech y Carmen Piñeira (Université d'Artois)
10. Pedro Javier Bueno Ruiz. Dirección: Judit Freixa Aymerich.

Tesis defendidas y depositadas desde 2011 a 2021

Alejandra López Fuentes. Adecuación automática de términos biomédicos para personas no expertas: el caso de los informes médicos. Codirección: Rosa Estopà Bagot (UPF) i Julio Collado Vides (Centro de Ciencias Genómicas-UNAM). Universitat Pompeu. Tesis depositada en noviembre de 2021 (defensa prevista enero 2022)

Jorge Mario Porras Garzón. *El anglicismo terminológico. Consecuencias léxicas del uso de textos escritos en inglés en las clases de medicina en español.* Dirección: Rosa Estopà Bagot. Universitat Pompeu. Tesis depositada en noviembre de 2021 (defensa prevista enero 2022)

Nadia Belkacem. *Les verbes en Tamazight. Structure argumentale et sémantique.* Dirección: Mercè Lorente Casafont. Universitat Pompeu. Fecha de defensa: 27/09/2021.

Ivan Solivellas i Ugena. *Les partícules prefixals en català.* Dirección: Judit Freixa Aymerich. Universitat Pompeu. Fecha de defensa: 14/07/2021.

Laia Vidal Sabanés. *La terminologia en els textos mèdics per a pacients: el cas d'una comunitat virtual de dones amb càncer de mama.* Codirección: Rosa Estopà Bagot (UPF) i Cristian Ochoa Arnedo (Universitat de Barcelona). Universitat Pompeu. Fecha de defensa: 7/06/2021.

Maria Àngels Martínez Salom. *Combinacions lèxiques en llengua catalana: les col·locacions. Caracterització teòrica i empírica i proposta lexicogràfica.* Dirección: Mercè Lorente Casafont Universitat Pompeu. Fecha de defensa: 1/02/2021.

Yingfeng Xu. *Unidades léxicas especializadas para la didáctica del chino: Análisis de necesidades y propuesta.* Codirección: Mercè Lorente y Manuel Ollé. Universitat Pompeu Fabra. Fecha de defensa: 23/06/2020

Coralie Schneider. *Identification de facteurs d'intégration quantitatifs et qualitatifs des néologismes à la langue spécialisée médicale. Une analyse sur corpus diachronique entre 2007 et 2015*. Cotutela internacional: Université Paris-Diderot y Universitat Pompeu Fabra. Codirección: Rosa Estopà Bagot (UPF) i Nathalie Kluber (Paris-Diderot). París, Université Paris-Diderot. Data de defensa: 17-04-2020

Ségolène Demol. *Dinámica del léxico en las constituciones españolas.* Cotutela internacional: Universitat Pompeu Fabra y Université d'Artois. Codirección: Carmen Pineira (Université d'Artois) y Mercè Lorente (UPF). Universitat Pompeu Fabra. Fecha de defensa: 31-01-2020.

Òscar Pozuelo Ollé. *L'eventivitat en els textos i en els diccionaris especialitzats en biomedicina.* Codirección: Rosa Estopà Bagot y Mercè Lorente Casafont. Universitat Pompeu Fabra. Fecha de defensa: 13-11-2019.

Andrés Torres Rivera. *Detección y extracción de neologismos semánticos especializados Un acercamiento mediante clasificación automática de documentos y estrategias de aprendizaje profundo*. Codirección: Rosa Estopà Bagot (UPF) y Dr. Juan-Manuel Torres-Moreno (Université d'Avignon). Universitat Pompeu Fabra. Fecha de defensa: 31-10-2019.

Óscar Mendez Espinosa. *La terminología de Geografía en zapoteco: una nueva perspectiva para las escuelas de Oaxaca, México. La terminología zapoteca del siglo XXI.* Dirección: M. Teresa Cabré Castellví y Mercè Lorente Casafont. Universitat Pompeu Fabra. Fecha de defensa: 04-10-2019.

Suárez Gallo, Míriam. *La metáfora en el discurso especializado del genoma humano* Dirección: Mercè Lorente Casafont. Universitat Pompeu Fabra. Fecha de defensa: 28-06-2019.

Lima Moreira, Glauber. *Diccionario y enseñanza de ELE: propuesta de un modelo de artículo lexicográfico para estudiantes brasileños*. Dirección: Maria Teresa Cabre Castellvi y Maria Amor Montane March. Universitat Pompeu Fabra. Fecha de defensa: 03-12-2018.

Cao, ShuYuan. *Using annotated discourse information of a RST Spanish-Chinese treebank for translation and language learning tasks*. Dirección: Iria Da Cunha Fanego y Mikel Iruskieta. Universitat Pompeu Fabra. Fecha de defensa: 09-11-2018.

Fathi, Besharat. *Terminology planning evaluation: the case of Persian language*. Dirección: M. Teresa Cabré Castellví. Universitat Pompeu Fabra. Fecha de defensa: 15-09-2017.

Han, Jingyi. *Phrase Table Expansion for Statistical Machine Translation with reduced parallel corpora: the Chinese-Spanish case.* Dirección: Núria Bel Rafecas Universitat Pompeu Fabra. Fecha de defensa: 13-09-2017.

Fomicheva, Marina. *The role of human reference translation in machine translation Evaluation*. Dirección: Núria Bel Rafecas y Iria da Cunha. Universitat Pompeu Fabra. Fecha de defensa: 18-07-2017.

Llopart-Saumell, Elisabet. *La funció dels neologismes: revisió de la dicotomia neologisme denominatiu i neologisme estilístic*. Dirección: Judit Freixa Aymerich. Universitat Pompeu Fabra. Fecha de defensa: 04-11-2016.

Villena Araya, Belén. *Innovación léxica en mapudungún: genuinidad, productividad y planificación*. Dirección: M. Teresa Cabré Castellví. Universitat Pompeu Fabra. Fecha de defensa: 24-10-2016.

Cañete González, Paola Isabel. *Innovación léxica y género en textos periodísticos del español actual*. Dirección: Judit Freixa Aymerich. Universitat Pompeu Fabra, febrero de 2016.

Necsulescu, Silvia. *Automatic acquisition of lexical-semantic relations. Gathering information in a dense representation*. Dirección: Núria Bel Rafecas. Universitat Pompeu Fabra, febrero de 2016.

Vázquez Suárez, Silvia. *Pattern-Based Automatic Induction of Domain Adapted Resources for Social Media Analysis*. Dirección: Núria Bel Rafecas. Universitat Pompeu Fabra, enero de 2016.

Lauren Michele Romeo. *The structure of the lexicon in the task of the automatic acquisition of lexical information*. Dirección: Núria Bel Rafecas. Universitat Pompeu Fabra. Fecha de defensa: 16-10-2015.

Jorge Adrián Lázaro Hernández. *El ejemplo en terminología. Caracterización y extracción automática*. Codirección: M. Teresa Cabré Castellví (UPF) y Gerardo Sierra Martínez (UNAM). Universitat Pompeu Fabra. Fecha de defensa: 13-10-2015.

Lucimara Alves da Conceiçao Costa. *Reflexões sobre a variação terminológica na lexicografia corrente no brasil e a construção das bases teórico-metodológicas para o dicionário de lexicografia brasileira*. Tesis en cotutela académica. Codirección: M. Teresa Cabré Castellví (UPF) y Claudia Zavaglia (Universidade Estadual Paulista "Julio de Mesquita Filho"). Doble defensa. Fechas y lugares: 1-10-2015 en la Universitat Pompeu Fabra y 24-02-2015 en la Universidade Estadual Paulista "Julio de Mesquita Filho".

Albert Morales Moreno. *Estudi lexicomètric del vocabulari del procés d'aprovació de l'Estatut d'autonomia de Catalunya (2006)*. Codirección: M. Teresa Cabré Castellví (UPF) y Carmen Pineira (U. Artois). Universitat Pompeu Fabra, Fecha de defensa: 28-09-2015.

Hajar Benmakhlouf. *La comprensión lectora en los manuales de ELE en el sistema educativo marroquí: análisis y propuestas*. Dirección: M. Teresa Cabré Castellví (UPF) y Hassan Boutakka (Universitat Hassan II). Universitat Pompeu Fabra. Data de defensa: 28-07-2015.

Abolfazl Zarnikhi. *Towards a Model for Terminology Planning*. Dirección: M. Teresa Cabré Castellví. Universitat Pompeu Fabra. Fecha de defensa: 29-07-2014.

Diego Alberto Burgos Herrera. *Towards an Image-Term Co-occurrence Model for Multilingual Terminology Alignment and Cross-Language Image Indexing*. Dirección: Leo Wanner. Universitat Pompeu Fabra. Fecha de defensa: 19-05-2014.

Hèctor Martínez Alonso. *Annotation of regular polysemy: an empirical assessment of the underspecified sense*. Tesis en cotutela académica: Universitat Pompeu Fabra y Universitat de Copenhaguen. Codirección: Núria Bel Rafecas (UPF) y Bollette Sandford Pedersen (U. Copenhaguen). University of Copenhaguen. Fecha de defensa: 19-12-2013.

Alba Coll Pérez. *La norma lingüística en els diccionaris: contrast del discurs lexicogràfic de llengua catalana i de llengua castellana*. Codirección: M. Teresa Cabré Castellví y Rosa Estopà Bagot. Universitat Pompeu Fabra. Fecha de defensa: 06-09-2013.

M. Amor Montané March. *Terminologia i implantació: anàlisi d'alguns factors que influencien l'ús dels termes normalitzats de la informàtica i les TIC en llengua catalana*. Dirección: M. Teresa Cabré Castellví. Universitat Pompeu Fabra. Fecha de defensa: 11-12-2012.

Natalia Andrea Seghezzi. *Variación terminológica y canal de comunicación. Estudio contrastivo de textos especializados escritos y orales sobre lingüística*. Dirección: Judit Freixa Aymerich. Universitat Pompeu Fabra. Fecha de defensa: 27-10-2011.

Hada Rosabel Salazar Burgos. *Descripción y representación de los adjetivos deverbales del participio en el discurso especializado*. Dirección: Rosa Estopà Bagot. Universitat Pompeu Fabra. Fecha de defensa: 05-09-2011.

Gabriela Resnik. *Los nombres eventivos no deverbales en español*. Codirección: Núria Bel Rafecas (IULA) y Alexandre Alsina Keith (DTCL). Universitat Pompeu Fabra. Fecha de defensa: 09-03-2011

Sabela Fernández Silva. *Variación terminológica y cognición. Factores cognitivos en la denominación del concepto especializado*. Codirección: M. Teresa Cabré Castellví y Judit Freixa Aymerich. Universitat Pompeu Fabra. Fecha de defensa: 14-02-2011

El 75% de los egresados entre 2011 y 2021 disfrutaron de becas predoctorales o ayudas para su contratación como personal investigador en formación. Las fuentes de financiación fueron AGAUR-Generalitat de Catalunya, Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i (1 FPI); CAPES de Brasil; CONACYT de México, CONICYT de Chile, Fundación Ford, y las propias ayudas del Institut Universitari de Lingüística Aplicada o del Departament de Traducció i Ciències del Llenguatge de la UPF, todas en convocatorias competitivas.

Desarrollo científico o profesional de los doctores egresados

Entre los 37 doctores egresados en los últimos diez años, un 51,35% se han incorporado como profesores e investigadores en otras universidades (5 en universidades españolas y 14 en universidades extranjeras); un 10,81% han seguido trayectorias profesionales en empresas, todas del ámbito de las tecnologías; un 8,10% se dedican a la enseñanza de lenguas; y un 24,32% son profesionales en diversas actividades (2 en lexicografía; 1 en servicios lingüísticos; 1 en edición de revistas científicas; 1 en un centro de investigación en biomedicina; 2 en traducción; 1 en enseñanza, y 1 en gestión de másteres universitarios). El restante 5,42% (2 personas) aun tienen pendiente su adscripción postdoctoral.

Algunas de las universidades receptoras de los doctorandos de IULATERM son las de Ginebra (Suiza), Sheffield y Durham (Reino Unido), General Sarmiento (Argentina), Pontificia de Valparaiso y Concepción (Chile), Sao Paulo y Federal do Delta do Parnaíba (Brasil), Sonora y Nacional Autónoma de México (México), Tongji (China), Barcelona, Alicante y Pompeu Fabra

(España). Conviene destacar también en lexicografía los destinos de Oxford University Press o de la Academia de la Lengua Persa.

Contexto científico-tecnico y formativo del equipo

El grupo IULATERM dispone de espacios propios para la investigación dentro del Institut de Lingüística Aplicada de la Universitat Pompeu Fabra, en el Campus de la Comunicación. En este espacio hay 10 espacios de trabajo individuales dotados con ordenadores personales conectados a la red. La sala está dotada además de impresora y escáner y alberga un anexo con la biblioteca Espai Teresa Cabré, que contiene los fondos cedidos por Teresa Cabré en materia de terminología y otros ámbitos de la lingüística.

Los doctorandos del grupo IULATERM que disfrutan de beca o de contrato tienen asegurado su espacio de trabajo individual en los espacios del grupo. El resto de doctorandos pueden acceder a un espacio complementario dotado con ordenadores también. Todos los doctorandos se integran en las actividades del grupo: seminarios quincenales, organización de congresos y jornadas, proyectos de investigación y de transferencia. La política del grupo es apoyar a los doctorandos y planificar con ellos publicaciones y participaciones en congresos relacionadas con las respectivas tesis doctorales. El grupo concede ayudas para inscripciones, viajes y dietas de asistencia a congresos, a cargo de proyectos o del fondo de ayuda COFRE concedido anualmente por la universidad. Además, los doctorandos tienen acceso también a ayudas del Departament de Traducció i Interpretació para congresos y estancias predoctorales. Otros apoyos del grupo a sus doctorandos son la adquisicion de bibliografía, el establecimiento de convenios de movilidad con universidades extranjeras y el asesoramiento técnico en el desarrollo, la aplicación de recursos lingüísticos y herramientas de procesamiento y extracción, y la transmisión regular de información a través de un boletín semanal.

Los doctorandos actuales, si no han realizado formación máster en la Universitat Pomepu Fabra, reciben complementos de formación durante el primer año del doctorado en Traducción y Ciencias del Lenguaje. Además los doctorandos de IULATERM reciben formación complementaria en estadística, tecnologías del lenguaje, terminología y neología.

Programa de formación previsto en el contexto del proyecto

En este proyecto se solicita la incorporación de una beca predoctoral FPI para poder desarrollar una tesis con vinculación directa al tema del proyecto. Una posibilidad (según preferencias del candidato) sería profundizar en el tema de los procesos de lexicalización en el marco del cambio lingüístico; otra sería abordar el tema de las comunidades de habla especializadas. En este sentido el proyecto sería una buena ayuda para introducirse en estos temas y formarse específicamente en metodología de investigación de orientación diacrónica.

En caso de ser concedida esta beca FPI, la idea es hacer una llamada pública para que alguna persona interesada y con buen currículum pudiera inscribirse en el programa de doctorado en Traducción y Ciencias del Lenguaje del curso 2022-2023, y realizar sus estudios de doctorado durante el trienio 2022-2025, período del proyecto.

La planificación de la formacion de este doctorando becado vinculado al proyecto sería la siguiente:

Primer año: Planificación del alcance de la tesis (objeto de estudio, hipótesis, objetivos, ámbitos de especialidad complementarios a los del proyecto). Integración en el proyecto. Formación complementaria. Presentación del plan de tesis ante una comisión evaluadora.

Segundo año: Avances de la tesis. Reuniones quincenales con la dirección. Estancia predoctoral de investigación en universidad extranjera, en un grupo adecuado para el tema elegido. Consultorías en tecnologías y en estadística, si fuera necesario. Primeros resultados provisionales en congresos internacionales.

Tercer año: Redacción y revisiones de la tesis. Publicaciones asociadas en revistas de impacto.

Otras dos tesis en curso están vinculadas al proyecto parcialmente:

- Mariona Arnau Garcia, que ha iniciado su doctorado en el curso 2021-22, investiga sobre las ideologías en diccionarios, con un planteamiento innovador que incluye temas hasta ahora no analizados en este ámbito. El tema que vincula su tesis al proyecto es la economía y la representación de términos de la economia en diccionarios generales y especializados.
- Sangzhu Zhaxi, que inició su doctorado en el curso 2020-21, investiga sobre la variación terminológica y la complejidad en el comercio internacional de flores cortadas. Sus aportaciones sobre variación y necesidades comerciales en un ámbito tambien vinculado a la botánica y la variación geolectal serán fundamentales en el anállisis de la complejidad de comunidades de hablantes en economía e informática y su correlación con la variación terminológica.

Medios de difusión para la oferta del contrato FPI

La oferta de contrato predoctoral se difundirá a través de los siguientes medios:

- Oferta pública a través del Servicio de investigación y de la secretaría del Departament de Traducció i Ciències del Llenguatge de la UPF.
- Listas de difusión: Linguist List, AESLA (Asociación Española de Lingüística Aplicada), AETER (Asociación española de Terminología).
- Boletines de terminología: TermCoord_UE; Asociación europea de terminología (EAFT/AET); Red internacional de terminología y neología de lenguas románicas (REALITER); Red Iberoamericana de Terminología (RITERM).
- Webs: Grupo IULATERM e Institut de Lingüística Aplicada (IULA).

7. CONDICIONES ESPECÍFICAS PARA LA EJECUCIÓN DE DETERMINADOS PROYECTOS – *SPECIFIC CONDITIONS FOR THE EXECUTION OF CERTAIN PROJECTS*