ÍNDICES LÉXICO-ESTADÍSTICOS PARA UN ESTUDIO DE LA DISPONIBILIDAD LÉXICA Y TERMINOLÓGICA MULTILINGÜE

GIULIA NALESSO
Università degli Studi di Padova

Introducción

Presentamos el proyecto DiLexTerM: Studio della Disponibilità Lessicale e Terminologica Multilingue in Studenti Universitari, que se enmarca en el ámbito temático del 32.º Congreso Internacional de ASELE, celebrado en la Universidad de Verona en 2022, "Humanidades digitales en el ámbito del español LE", ya que, aunque se trata de una investigación que en general se ocupará del aprendizaje del vocabulario de ELE —junto a otras LE— y de la medición de la competencia léxica de un grupo de estudiantes, tiene también la finalidad de crear y poner al alcance de los interesados[1] un recurso en línea, estructurado como una base de datos o repositorio dinámico e interactivo, que presente un diccionario de léxico disponible multilingüe. Para ello, se ha planteado la creación de un portal digital colaborativo, de acceso abierto, con los resultados de los análisis cuantitativo, cualitativo y comparativo que el grupo de investigación desarrollará. Los datos y materiales que se publicarán podrán ser útiles para la investigación y también para la didáctica de las lenguas, pues pueden ser aplicados como herramientas digitales de aprendizaje según las distintas posibilidades de explotación de la disponibilidad léxica en estos campos (Paredes García, 2015).

1. Objetivos

DiLexTerM, acrónimo de *disponibilidad léxica y terminológica multilingüe*, recientemente financiado por el Departamento de Estudios Lingüísticos y Literarios (DiSLL) de la Universidad de Padua, es un proyecto dedicado a la didáctica de las LE —español en particular— y dirigido al análisis cuantitativo y cualitativo de la competencia léxica multilingüe de estudiantes de lenguas extranjeras de esta misma institución.

Nuestro objetivo es la medición y la evaluación del vocabulario común y especializado en L1 (italiano) y en distintas LE estudiadas en este departamento (español, inglés, francés y portugués) a través de la disponibilidad léxica, una metodología encaminada a descubrir qué palabras sería capaz de usar un individuo en una situación comunicativa específica, a saber, su léxico disponible, que es el caudal de vocabulario que se actualiza ante la necesidad de comunicar algo referente a esa situación concreta (López Morales, 1999). En este campo apreciamos numerosos trabajos dedicados al español común LM y LE (para una reseña véase Nalesso, 2019), por tanto, este proyecto amplía el objeto de investigación en dos sentidos: hacia la competencia multilingüe y la terminología.

Con todo y con eso, el propósito último de DiLexTerM es el análisis de los índices de disponibilidad léxica general y terminológica, algo del que no tenemos constancia hasta ahora en estudios de campo realizados con aprendientes italianos de idiomas[2], de un grupo representativo de estudiantes matriculados en cursos de LE de la Universidad de Padua a través de la recopilación y el análisis de un corpus multilingüe de léxico disponible que coteje:

(1) por un lado, italiano, LM; español, portugués, francés, LE romances muy estudiadas en este ámbito educativo; inglés, LE más estudiada en general en Italia (y no solo en este país), esto es, la competencia léxica multilingüe;

[1] Al hablar de interesados nos referimos tanto a docentes y aprendientes de lenguas como a investigadores y profesionales dedicados a investigar y enseñar/aprender en nuestro caso específico el español como LE o, simplemente, los que tengan algún tipo de interés o curiosidad por estos temas.

[2] En este sentido, se planteó en el Centro de Enseñanza de Lenguas Extranjeras de la Universidad Nacional Autónoma de México un proyecto parecido dedicado a hispanohablantes nativos, *Estudios de Léxico Disponible* (Colín Rodea *et al*., 2010), finalizado a investigar el léxico disponible del español LM y de varias LE en proceso de adquisición.

(2) por otro lado, unidades léxicas comunes y lexías especializadas, es decir la competencia léxica general y la competencia léxica terminológica, desde una óptica interdisciplinar.

Se ha optado por aplicar la metodología de la disponibilidad léxica porque, en el ámbito de la didáctica de las LE, este tipo de estudio permite establecer cuantitativamente el nivel de competencia de los informantes y el grado de asociación cognitivo-semántica de los vocablos comunes y especializados, por lo que desde el eje de la psicolingüística es posible observar cualitativamente la organización del lexicón mental, el cual permite recuperar eficazmente los términos solicitados por los estímulos. En definitiva, pretendemos contestar a las siguientes preguntas:

(1) ¿Cuántos y cuáles vocablos de léxico común conocen los estudiantes?
(2) ¿Cuántos y cuáles vocablos de léxico especializado dominan?
(3) ¿Cómo se relacionan y asocian estos vocablos en su mente?

A este propósito, examinaremos los datos según índices estándar para diagnosticar el nivel de dominio del vocabulario y de asociación cognitiva de palabras comunes y voces técnicas en el lexicón mental cumpliendo los siguientes pasos:

(1) diseño y suministración de la prueba;
(2) edición del material;
(3) extracción informatizada de los índices;
(4) análisis cuantitativo, cualitativo, comparativo;
(5) presentación de los resultados.

2. Grupo de investigación

Los miembros del grupo de investigación de DiLexTerM son cinco, profesores e investigadores del DiSLL de la Universidad de Padua especializados en español, francés y portugués. En lo que se refiere al inglés, no hay nadie que en su trabajo se ocupe de la didáctica de este idioma, ya que los mencionados miembros tienen una competencia adecuada para los análisis previstos e, igualmente, no disponemos de docentes de italiano por la misma razón y porque contamos con tres hablantes nativos de italiano.

Además, estamos creando un grupo de investigación secundario formado por estudiantes de máster que están terminando la carrera y pueden contribuir al proyecto con su TFM a través de estudios piloto o investigaciones parecidas (por ejemplo, trabajando con alumnos de la enseñanza secundaria (Martini, 2023), hablantes nativos de otras lenguas o aprendientes de otros idiomas extranjeros).

3. Apartado teórico

Nos parece oportuno definir, brevemente, nuestro marco teórico a través de algunos conceptos necesarios para poder entender el proyecto que aquí se presenta —disponibilidad léxica, léxico disponible, competencia léxica, competencia terminológica— y establecer el estado de la cuestión de los trabajos desarrollados en estos ámbitos hasta la fecha.

3.1. La disponibilidad léxica

Ante todo, cabe precisar qué es la disponibilidad léxica siendo el punto de partida del proyecto DiLexTerM. Se trata de "[una] metodología nacida en Francia a mediados del siglo XX, se refiere al cómputo de palabras que un sujeto sabe y que, por tanto, es capaz de actualizar ante un estímulo temático dado, denominado *centro de interés* (CI)" (Nalesso, 2020, p. 609), cuyo propósito inicial era la selección de las unidades léxicas más rentables para el aprendizaje del francés como lengua segunda o extranjera (cf. *Français Fundamental*, 1964). Esto servía para superar los límites del criterio de frecuencia que se había revelado insuficiente puesto que lexías que presentaban índices de frecuencia bajos (que, por tanto, no se incluían en los diccionarios de frecuencia muy difundidos en la época como herramienta de aprendizaje) eran, por el contrario, comúnmente usadas por los nativos (Gougenheim *et al.*, 1964, pp. 31-35). El

mencionado *Français Fundamental* se publicó, por ende, como obra de consulta y de referencia para la enseñanza del léxico francés proporcionando el caudal y el tipo de vocabulario adecuado para los aprendientes no nativos (L2/LE).

Todo esto marcó un punto de inflexión en el campo de la léxico-estadística llevando a una renovación metodológica mediante el planteamiento de una nueva técnica de recogida de datos para la elaboración del léxico disponible de una comunidad (que se comentará a continuación ya que sigue vigente en este tipo de estudio, véase § 4.1). De ahí, la disponibilidad léxica se difundió hacia otras lenguas, en particular es en los proyectos dedicados al español, tanto LM como LE, donde más se aprovecha.

El punto de partida para el desarrollo de esta metodología fue la distinción entre *palabras temáticas* y *palabras atemáticas*: se definen *temáticas* aquellas unidades léxicas que se relacionan con el tema de la comunicación, que pueden ser más o menos frecuentes, y que se usan para desenvolverse en una situación determinada por este tema; mientras que son *atemáticas* las unidades léxico-gramaticales que aparecen en cualquier texto independientemente del tema y, en consecuencia, alcanzan un alto índice de frecuencia:

> En una situación determinada, las primeras palabras que vienen a la mente son aquellas que están relacionadas con tal situación y que la caracterizan […]. Una palabra disponible es una palabra que, sin ser particularmente frecuente, siempre está lista para ser utilizada, y acude inmediatamente a la mente cuando parece necesario. (Michéa, 1953, p. 342)

A este respecto, López Morales (1999, p. 11) afirma:

> Existe en el lexicón mental una serie de términos que no se actualizan a menos que sea necesario para comunicar una información muy específica. Se trata de un léxico "disponible", cuyo estudio no puede emprenderse manejando frecuencias, porque este factor es pertinente solo en el caso de las actualizaciones léxicas efectivas, no de las potenciales.

Entonces, la disponibilidad léxica se extendió hacia la lengua española a partir del planteamiento del Proyecto Panhispánico de Disponibilidad Léxica (López Morales, 1999) que está todavía desarrollándose con la finalidad de recopilar un diccionario de léxico disponible que abarque todas las áreas lingüísticas del mundo hispánico. En la época los análisis se centraban en el vocabulario de hispanohablantes nativos, pero rápidamente se originó una línea de investigación paralela dedicada a aprendientes de ELE que se reveló muy productiva a partir de los primeros estudios publicados a comienzos del siglo XXI de Carcedo González (2000) y Samper Hernández (2002).

Los resultados recabados de los estudios permiten colocarse en el plano paradigmático del léxico y analizar no solo de cuántas y cuáles palabras dispone un sujeto, sino cómo se organizan conceptualmente en su lexicón mental mediante la realización de una prueba asociativa basada en estímulos temáticos, los denominados centros de interés, que consienten la activación de vocablos en la mente. Esto facilita la recuperación de las unidades que los participantes pueden escribir en relación con el tema propuesto dentro de un límite temporal que suele ser de dos minutos.

A estas alturas cabe señalar que las posibilidades de la disponibilidad en el campo de la didáctica de las LE abarcan una gran variedad de aplicaciones, teóricas y prácticas (Paredes García, 2015):

(1) analizar el léxico disponible de los aprendices en un determinado momento;
(2) definir las etapas de aprendizaje;
(3) estudiar las interferencias de la L1 y de otras LE conocidas;
(4) cotejar el vocabulario de los participantes con otros grupos de estudiantes y con los nativos;
(5) establecer cuáles son las unidades poco rentables o los déficits léxicos;
(6) proporcionar comparaciones culturales;
(7) examinar la adecuación de manuales y otros materiales pedagógicos;
(8) eliminar el léxico superfluo de los contenidos didácticos y seleccionar contenidos más apropiados;
(9) ampliar la cobertura de los análisis hacia la terminología (lenguas para fines específicos).

A este último propósito, han empezado a publicarse estudios que pasan de la disponibilidad léxica general a la disponibilidad léxica terminológica, esto es, que se centran en el léxico disponible específico de un

determinado sector científico o profesional desde el punto de vista de la terminología, entendida como disciplina y campo de trabajo, porque aumenta cada vez más el interés por la comunicación especializada.

Hasta hoy no conocemos muchas investigaciones realizadas con discentes extranjeros de español[3], pero disponemos de estudios dedicados a nativos que trabajan con el lenguaje especializado en los ámbitos de: comunicación y prensa (Gómez Sánchez y Guerra Salas, 2004; Guerra Salas y Gómez Sánchez, 2004); matemática (Urzúa Carmona, Sáez y Echeverría, 2006; Salcedo Lagos y Del Valle Leo, 2013); fisioterapia (Navarro Marrero, 2010); mecánica (Madrigal-Melchor, Rivera-Juárez, Enciso-Muñoz y López-Chávez, 2012); informática (Luján García y Bolaños Medina, 2014; Tomé Cornejo, 2016); viticultura (Toniolo y Zurita, 2019).

Antes de terminar este apartado, merece la pena destacar que la aplicación de la disponibilidad léxica no se limita a la lingüística y la didáctica de lenguas, sino que otras disciplinas pueden beneficiar de los datos que cada investigación aporta. La sociolingüística, por ejemplo, estudia el influjo de las variables sociales en el léxico disponible para examinar la variación de la lengua en función de distintos condicionantes extralingüísticos. En dialectología se utilizan estos datos para realizar comparaciones interdialectales y estudios de regionalismos. Ambas disciplinas, en efecto, concurren al desarrollo del ya citado Proyecto Panhispánico de Disponibilidad Léxica (López Morales, 1999).

Por su parte, la psicolingüística aborda análisis de corte cognitivo debido a la naturaleza de las pruebas, de tipo asociativo, siendo la producción del léxico disponible una tarea cognitiva compleja en la que intervienen diversos elementos psicológicos, más allá de los puramente lingüísticos. En este sentido, Hernández Muñoz (2005) describe esta metodología como "una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología". Se estudian en este marco: la organización del lexicón mental; el acceso y la selección del léxico en L1 y L2/LE; las estrategias de recuperación del léxico; las redes semánticas (Hernández Muñoz, 2005; Izura y Tomé Cornejo, 2014; Del Barrio de la Rosa, 2018; Gómez Devís, 2019).

Por último, la etnolingüística se apoya en la disponibilidad para estudiar la cultura y las costumbres de una comunidad que se reflejan en el vocabulario de los hablantes. En este ámbito resultan también de interés los proyectos realizados con aprendientes no nativos, ya que permiten observar la influencia de la cultura del país de origen y de la comunidad meta, a la que el alumnado se enfrenta (Herreros Marcilla, 2015).

3.2. La competencia léxica y terminológica

Es bien sabido que el fin último del aprendizaje de una lengua es el desarrollo de la competencia comunicativa: el estudiante es un agente social porque el proceso de adquisición lingüística está finalizado al alcance de todas las competencias que le permiten actuar como miembro de la sociedad, se trata de "la suma de conocimientos, destrezas y características individuales que permiten a una persona realizar acciones" (*MCER*, 2002, p. 20). Según la definición del *Diccionario de términos clave de ELE* del Instituto Cervantes la competencia comunicativa es:

> la capacidad para comportarse de manera eficaz y adecuada en una determinada comunidad de habla; ello implica respetar un conjunto de reglas que incluye tanto las de la gramática y los otros niveles de la descripción lingüística (léxico, fonética, semántica) como las reglas de uso de la lengua, relacionadas con el contexto sociohistórico y cultural en el que tiene lugar la comunicación.

Esta habilidad se compone de otras, entre las que contamos con la lingüística que, de nuevo según el *Diccionario de términos clave de ELE*, combina las siguientes competencias:

[3] Antes de plantear el proyecto DiLexTerM, realizamos un experimento preliminar (Nalesso, 2020) con un grupo restringido de alumnos de la enseñanza secundaria en el que se analizaron centros de interés para recoger el léxico disponible común (partes del cuerpo; la ropa; partes de la casa; los muebles de la casa; comidas y bebidas; objetos colocados en la mesa; la cocina y sus utensilios; la escuela; la ciudad; medios de transporte; los animales; juegos y diversión; profesiones y oficios) y especializado (*marketing*; mundo de la empresa; publicidad).

> […] la léxica, la gramatical, la semántica, la fonológica, la ortográfica y la ortoépica (capacidad de articular una pronunciación correcta partiendo de la forma escrita). En una detallada descripción de cada una de estas competencias y subcompetencias, el Marco Común Europeo ofrece unas escalas con seis niveles de dominio de cada una de ellas; así, describe en esos seis niveles la competencia lingüística general, la riqueza de vocabulario, el dominio del vocabulario, la corrección gramatical, el dominio de la pronunciación y el dominio de la ortografía.

Centrándonos, ahora, en el ámbito que nos ocupa aquí —el aprendizaje del léxico— es oportuno establecer que la competencia léxica, dentro de la lingüística, es el "conocimiento del vocabulario de una lengua y la capacidad para utilizarlo" (*MCER*, 2002, p. 108) y por tanto en la didáctica de las LE "La riqueza, el alcance y el control del vocabulario son parámetros importantes en la adquisición de la lengua y, por ello, de la evaluación del dominio de la lengua que tiene el alumno, y de la planificación del aprendizaje y de la enseñanza de lenguas" (*ibid.*, p. 86).

Retomando el *Diccionario de términos clave de ELE*, el dominio del léxico está formado por dos componentes complementarios que remiten al concepto de léxico disponible:

(1) el vocabulario receptivo (pasivo o de recepción), dentro del lexicón mental, coincide con aquel conjunto de ítems que un hablante sabe y puede descifrar en la recepción de un aducto;
(2) el vocabulario productivo (activo o de producción) abarca el conjunto de unidades que están en el lexicón mental listas para ser utilizadas, de las que puede servirse un sujeto en su educto.

Concluyendo esta sección preliminar, en el marco didáctico del aprendizaje de las LE y, en particular, del léxico de estas lenguas, en los últimos años ha aumentado el interés por el conocimiento de la lengua para la comunicación profesional y, en consecuencia, de términos o vocablos propios de una determinada profesión, ciencia o materia, esto es, de disciplinas específicas en contextos especializados. Para el ya citado diccionario del Instituto Cervantes:

> La enseñanza de la lengua para fines específicos surge a finales de los años 60 del siglo XX con la aparición del enfoque comunicativo en el aprendizaje de lenguas extranjeras. La denominación EFE (español con fines específicos) se extiende a comienzos de la década de los 90 y desde entonces la didáctica de este tipo de comunicación ha pasado a constituir una de las ramas más relevantes de la enseñanza del español como lengua extranjera.

Por ello, como hemos mencionado previamente, se plantea aquí un estudio de disponibilidad léxica que incluye junto al léxico general el específico, es decir, aquel conjunto de vocabulario técnico que forma la competencia terminológica de un hablante. Sin embargo, para capacitar profesionalmente al estudiante, es necesario orientar los contenidos lingüísticos hacia el dominio de la terminología, siendo un instrumento indispensable en todo proceso de mediación lingüística.

Se trata, en otras palabras, de enseñar y aprender el repertorio terminológico que permite llevar a cabo la comunicación especializada, pues "alude a la terminología [y su respectivo conocimiento] propia de un ámbito de especialidad" (Cabré y Giraldo, 2005, p. 97, *apud* Navarro Marrero, 2010, p. 11). El mismo López Morales (1996, p. 245) en el marco de la disponibilidad léxica asevera que "el vocabulario de las lenguas de especialidad se incluye en el ámbito más amplio del léxico disponible, compuesto por términos que el hablante no actualiza a menos que se necesiten para comunicarse una información muy específica".

4. Apartado práctico

Pasamos ahora a describir la metodología de investigación que se ha planteado para el desarrollo de DiLexTerM reseñando en qué consiste la prueba, quiénes son los participantes, cuáles son los procesos de recogida de datos y los criterios de edición, cómo se realizarán los análisis y se discutirán los resultados. La metodología de trabajo prevé los siguientes pasos:

1. diseño de la prueba;
2. selección de la muestra;
3. establecimiento de las técnicas de elicitación de datos;
4. suministración y ejecución de la prueba;

5. fijación de los criterios de edición;
6. tratamiento informático-estadístico del material;
7. análisis de los datos;
8. creación del diccionario de léxico disponible;
9. discusión de los resultados;
10. publicación e implementación del portal.

4.1. La prueba

La parte empírica de la investigación se basa en un test de tipo asociativo creado *ad hoc* para DiLexTerM —a partir de las pruebas estandarizadas que se suelen utilizar en el ámbito de la disponibilidad léxica— con el propósito de recabar el léxico disponible multilingüe de la muestra a través de estímulos semánticos que tratan temas comunes/generales y especializados/sectoriales a fin de desarrollar un análisis de la competencia léxica que sea lo más completo posible en virtud de las carreras universitarias de los informantes (véase § 4.2.).

Las pruebas se realizarán de forma anónima y escrita, y tendrán dos partes: la primera encaminada a la obtención de datos sociológicos de los encuestados para la posterior caracterización de la muestra. Se proporciona un cuestionario que recoja datos personales e informaciones sobre el bagaje lingüístico que se utilizarán en el establecimiento de las variables lingüísticas y extralingüísticas. No se fija un límite temporal de manera que todos dispongan de tiempo suficiente para completar el informe antes de pasar a la ejecución de la prueba que ocupa la segunda parte de la encuesta, dedicada a la recogida del léxico disponible. El test se compone de distintas páginas formadas por columnas y un espacio sobre cada una donde se escriben los nombres de los CI propuestos. Los campos se presentan uno a uno a medida que se completan para evitar que se piense antes de tiempo en los vocablos relacionados con el tema sucesivo y que, por ende, las respuestas sean el resultado de una reflexión (Samper Hernández, 2002). Los informantes tienen que rellenar las fichas con las unidades léxicas que se les ocurren en función del tema presentado, empezando por la primera línea en alto de cada columna para que los ítems más disponibles sean los que aparecen en las primeras posiciones de las listas. En lo que al margen temporal se refiere, los participantes tienen dos minutos para completar cada CI y no pueden volver atrás una vez terminado el tiempo: iniciarán anotando sus respuestas en la primera columna y pasarán a las sucesivas cuando se les indique hasta agotar el tiempo concedido.

4.2. La muestra

Como anticipado, la muestra que planeamos investigar abarca una cantidad que sea representativa de los estudiantes italohablantes nativos matriculados en cursos de grado y máster (*Lingue, letterature e mediazione culturale*, *Lingue e letterature europee e americane*, *Lingue moderne per la comunicazione e la cooperazione internazionale*) del DiSLL de la Universidad de Padua que estudian por lo menos dos de los idiomas contemplados de manera que puedan completar la prueba en mínimo tres lenguas (L1 + dos LE).

Cabe señalar, a estas alturas, que hemos podido plantear este proyecto multilingüe abarcando el análisis de la competencia terminológica porque los cursos de LE de estas carreras, además del aprendizaje de la lengua, a saber del desarrollo de la competencia lingüística y comunicativa, están encaminados a la adquisición de estrategias teóricas y prácticas de redacción y traducción tanto general como especializada de textos formales y técnicos enmarcados en sectores profesionales (comunicación, ciencias, comercio, publicidad, medicina, cocina, *marketing*, etc.) que requieren el conocimiento de una terminología adecuada. En suma, las clases de lenguas tratan también temas especializados cuyas actividades concurren al desarrollo de la competencia léxica terminológica del alumnado además de la competencia léxica propiamente dicha. Esto nos permite contestar a la interrogación de Urzúa Carmona, Sáez y Echeverría (2006, pp. 60-61), que se preguntan "¿qué pasa con [el léxico disponible de] los hablantes que se desenvuelven en ámbitos restringidos y que además reciben un entrenamiento lingüístico en este ámbito? No se trata de sujetos aislados, sino de sujetos que necesitan de un tipo definido de vocabulario en contextos determinados […]". En efecto, la lengua para fines específicos, o lengua de

especialidad, facilita el intercambio comunicativo en el mundo laboral y entre las profesiones, por lo que habría que concentrar los procesos de planificación didáctica en la adquisición de la competencia comunicativa de los discentes que deberían llegar a manejar una cantidad adecuada de léxico en virtud de los objetivos fijados y, en consecuencia, a un buen nivel de dominio terminológico.

4.3. Recogida de los datos

Para recabar el léxico disponible de la muestra, según los estándares de las investigaciones en disponibilidad léxica, como ya hemos explicado, se proporciona a los informantes una prueba asociativa en la que se pide que rellenen unos folios que contienen varias columnas, en función del número de CI que el investigador decide proponer, con todas las palabras que se les ocurren en relación con los estímulos propuestos.

A la hora de ejecutar la prueba, es esencial aclarar a los participantes que es importante escribir todas las palabras que sobrevienen, aunque no estén seguros de su grafía o significado, para maximizar la activación de asociaciones mentales y semánticas, ya que posteriormente se realizará una fase de edición y lematización del material.

Asimismo, en el momento de la suministración sería aconsejable seguir el mismo procedimiento, aplicando un protocolo estándar que garantice la repetitividad del experimento:

(1) presentación del investigador y del proyecto;
(2) explicación de las instrucciones;
(3) aclaración de la anonimidad;
(4) administración de las hojas;
(5) realización de la prueba.

4.4. Edición y tratamiento informático de los datos

Tras los procesos de recogida, el siguiente paso de la investigación consiste en la edición del material (1) antes del procesamiento informático, a saber, la aplicación de un *software* (2) que permite la extracción de los índices necesarios para los análisis (3).

(1) Ante todo, una vez recogidas las respuestas, se transcribirán y lematizarán de acuerdo con unos criterios de edición y homogeneización previamente establecidos. Aplicaremos los patrones que suelen emplearse en los estudios de disponibilidad léxica dedicados a estudiantes de ELE, basándonos en Samper Hernández (2002), ajustándolos a nuestras necesidades, ya que manejamos un corpus multilingüe.

El objetivo de esta etapa es homogenizar el material recogido, eliminando las palabras inexplicables, inexistentes o indescifrables para el cálculo informático, en general se tratará de: eliminación de términos repetidos y artículos; corrección de la ortografía; unificación ortográfica; unificación de variantes meramente flexivas; unificación de derivados regulares; unificación de formas plenas y acortamientos; eliminación de préstamos e interferencias de otras lenguas; eliminación de las marcas comerciales; amplitud de las relaciones asociativas.

(2) En la segunda fase se procesarán las respuestas editadas para obtener los índices de disponibilidad de los vocablos y otras medidas cuantitativas como el promedio de respuestas por informante y CI. El procesamiento informático se puede cumplir a través de distintas herramientas entre las que recordamos los pioneros Lexidisp (Moreno Fernández *et al.*, 1995) y DispoLex (López Morales *et al.*, s. f.), el más reciente Dispogen II (Echeverría *et al.*, 2005) y los novedosos DispoCen (Ávila Muñoz *et al.*, 2021) y LexPRO (Hernández Muñoz *et al.*, 2023), que emplearemos por ser los más innovadores y, posiblemente, tener funcionalidades inéditas para la extracción de los indicadores necesarios a nuestro proyecto.

(3) Por último, como se verá en el siguiente apartado, la disponibilidad léxica se mide por medio de índices que permiten desarrollar análisis cuantitativos y cualitativos, cuyos resultados se corroborarán con comparaciones intra- e intermuestrales.

Asimismo, tras el cumplimiento de los pasos previos será posible pasar a la recopilación del diccionario de léxico disponible multilingüe que sirve para averiguar qué palabras y vocablos técnicos

podrían recuperar de su mente y utilizar los informantes a la hora de desenvolverse en una comunicación general o especializada en cada una de las lenguas conocidas.

4.5. Análisis de los datos

En primera instancia, en el plano cuantitativo se examinan los siguientes indicadores que utilizaremos para determinar la cantidad de léxico disponible de nuestros informantes:

(1) el total de palabras;
(2) el total de vocablos;
(3) los promedios por informante y por CI;

Igualmente, mediremos dos índices que informan sobre el grado de concreción semántica de las palabras en función de un estímulo:

(1) el índice de cohesión (Echeverría *et al.*, 1987), que estudia la coincidencia de las respuestas por campo semántico, esto es, si el léxico es homogéneo dentro del corpus dividiendo la media de palabras aportadas por informante por el total de vocablos[4];
(2) la densidad léxica, que contabiliza la media de repeticiones de las respuestas dividiendo el número de palabras por el número de vocablos.

A continuación, el objetivo de la centralidad léxica es analizar la disponibilidad tanto cuantitativa como cualitativamente, ya que se coloca entre los dos tipos de análisis dando cuenta de la *centralidad* de los términos en cada CI. Se trata de un índice cuantitativo que, a la vez, da cuenta de la calidad de las respuestas: la interpretación de los datos, en efecto, establece que un valor de 1, o muy cercano, respondería a la pertenencia al núcleo del vocabulario específico del estímulo, mientras que un valor próximo a 0 indicaría que se trata de un término poco accesible y, por tanto, descentralizado (Ávila Muñoz *et al.*, 2021).

Conjuntamente, según el análisis cualitativo, examinaremos el tipo de léxico activado observando cuáles son las temáticas y las categorías gramaticales más difundidas en cada campo mediante la extracción de los vocablos más disponibles, es decir los que alcanzan un índice de disponibilidad (ID) igual o mayor a 0,1. Para ello, compilaremos una lista que contiene los vocablos con ID $\geq$ 0,1 a partir de la cual calcularemos también el índice de la cardinalidad del conjunto, a saber, el porcentaje de compatibilidad del material recabado en las distintas agrupaciones de informantes (Ávila Muñoz y Sánchez Sáez, 2010 y 2011) con el fin de averiguar qué nivel de correspondencia existe entre las respuestas[5]. Todo esto llevará a analizar dos estrategias fundamentales de la competencia léxica: la creación de agrupaciones en el marco de la categoría propuesta como estímulo (*clusters*) y la realización de saltos entre las subcategorías evocadas (*switches*).

Para completar la investigación, el último paso del apartado analítico coincide con un cotejo entre los índices de disponibilidad en las lenguas examinadas y entre CI generales y especializados para obtener datos sobre la competencia léxica multilingüe de los informantes, los niveles de conocimiento del vocabulario común y técnico, y averiguar si un idioma alcanza resultados más satisfactorios que otros y de qué idioma se trata (incluso si es la LM o una de las LE). Esto valdrá también para valorar la efectividad de la acción didáctica propuesta en los cursos académicos, destacar eventuales déficits y proponer mejoras o ajustes a los programas curriculares.

[4] Los valores oscilan entre 0 y 1. Se detecta mayor cohesión cuanto más el índice se aproxima a 1 porque las respuestas coinciden en gran medida y, por tanto, el centro es más compacto (cerrado); al contrario, destaca menor coincidencia cuando el índice se aproxima a 0, con lo cual el centro es más difuso (abierto).

[5] Se calcula a partir del número de vocablos con ID $\geq$ 0,1 de cada grupo (cardinalidad) a través de la suma de las unidades convergentes en los grupos (intersección), de las que no están en todos, pero en varios (unión) y de las exclusivas de cada uno (complemento). La suma de la unión y de los complementos se denomina suma disyuntiva. La compatibilidad coincide con el peso de la intersección con respecto al total de vocablos (intersección + suma disyuntiva).

4.6. Difusión de los resultados

La difusión de los resultados se realizará a través de publicaciones, congresos y seminarios del grupo de investigación y, a más y mejor, se divulgarán en el portal en línea DiLexTerM que permitirá la consulta libre tanto de los índices computados como del diccionario multilingüe de léxico disponible general y técnico que podrán ser útiles para la investigación y también para la didáctica de las LE, pues pueden ser aplicados como herramientas de enseñanza y aprendizaje según las múltiples posibilidades de aplicación de la disponibilidad que hemos enumerado antes.

En conclusión, todos los datos se proporcionarán en un repositorio de tipo colaborativo para la consulta y la búsqueda de información en función de variables lingüísticas (como lengua, nivel, número de lenguas estudiadas) y también de factores sociales (como edad, curso, tipo de estímulo).

Conclusiones

Este apartado final, más que unas conclusiones —siendo DiLexTerM en su etapa inicial— presenta unas consideraciones finales y nuestras perspectivas futuras.

El proyecto tiene el objetivo empírico de recoger una cantidad representativa de datos sobre la disponibilidad léxica de aprendientes de LE, hablantes nativos de italiano, de la Universidad de Padua encaminada a la recopilación de un diccionario multilingüe en línea de léxico disponible, común y especializado, de utilidad para la creación (o la mejora) de programas didácticos adecuados al efectivo desarrollo de la competencia comunicativa de esta tipología de estudiante. Para ello, se plantea la creación de un portal digital, donde se publicarán los resultados en una base de datos interactiva que permita búsquedas cruzadas según variables lingüísticas y extralingüísticas a la vez, pues se proporcionarán las tablas con los índices computados y, por supuesto, el diccionario multilingüe de léxico disponible.

Asimismo, DiLexTerM tiene un objetivo teórico que coincide con la ampliación del campo de la investigación en disponibilidad léxica mediante el estudio y el cotejo del léxico disponible en la L1 y en distintas LE, por eso se trata de un proyecto multilingüe, y, a más y mejor, mediante el análisis y la comparación de la competencia léxica general y terminológica, esta última relacionada con aspectos clave del desarrollo del conocimiento en los campos especializados de ciencia, tecnología, economía, turismo, etc. puesto que la metodología actual de los estudios se limita a aprendientes de ELE de distinto origen y a temas generales de la vida cotidiana (Hidalgo Gallardo, 2019; Nalesso 2019; Nalesso, 2022).

En definitiva, los resultados de este proyecto permitirán sacar una fotografía de la competencia léxica en las LE estudiadas comparada con la LM y avanzar en el conocimiento de un fenómeno tan complejo como es el aprendizaje del vocabulario, tanto común como técnico.

Referencias bibliográficas

Ávila Muñoz, A. M. y Sánchez Sáez, J. M. (2010). La disponibilidad léxica. Antecedentes y fundamentos. En A. M. Ávila Muñoz y J. A. Villena Ponsoda (Eds.), *Variación social del léxico disponible en la ciudad de Málaga*. Málaga: Editorial Sarriá, pp. 35-81.

Ávila Muñoz, A. M. y Sánchez Sáez, J. M. (2011). La posición de los vocablos en el cálculo del índice de disponibilidad léxica: procesos de reentrada en las listas del léxico disponible de la ciudad de Málaga. *ELUA*, 25, pp. 45-74. https://rua.ua.es/dspace/bitstream/10045/21640/1/ELUA_25_02.pdf.

Ávila Muñoz, A. M., Sánchez Sáez, J. M., Odishelidze, N. (2021). Dispocen. Mucho más que un programa para el cálculo de la disponibilidad léxica. *ELUA*, 35, pp. 9-36. https://doi.org/10.14198/ELUA2021.35.1.

Carcedo González, A. (2000). *Disponibilidad léxica en español lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turku: Turun Yliopisto.

Colín Rodea, M. *et al.* (2010). El Corpus Multilingüe de Léxico Disponible de los alumnos de inglés, francés, español, portugués y ruso del CELE. UNAM. *Memorias del XI Encuentro Nacional de Estudios de Lenguas*.

Del Barrio de la Rosa, F. (2018). Pares léxicos en el léxico disponible de estudiantes italianos. *Lingue e linguaggi*, 26, pp. 173-196. https://iris.unive.it/retrieve/handle/10278/3702889/136443/2018_L%26L_Pares.pdf.

Echeverría, M. S. *et al.* (1987). Disponibilidad léxica en Educación Media. Resultados cuantitativos. *RLA. Revista de lingüística teórica y aplicada*, 25, pp. 55-115.

Echeverría, M. S. *et al.* (2005). *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Concepción de Chile: Universidad de Concepción.

Français Fundamental = Gougenheim, G. *et al.* (1964). *L'élaboration du français fundamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. París: Didier.

Gómez Devís, M. B. (2019). A propósito de las redes semánticas en el léxico disponible de escolares de primero de Educación Primaria. *Ogigia. Revista electrónica de estudios hispánicos*, 25, pp. 165-183. https://revistas.uva.es/index.php/ogigia/article/view/2917/2414.

Gómez Sánchez, M. E. y Guerra Salas, L. (2004). Disponibilidad y fines específicos: análisis del centro de interés prensa. En I. Sanz Sainz y Á. M. Felices Lago (Eds.), *Las nuevas tendencias de las lenguas de especialidad en un contexto internacional y multicultural. III Congreso Internacional de la Asociación Europea de Lenguas para Fines Específicos*. Granada: Universidad de Granada, pp. 695-703.

Guerra Salas, L. y Gómez Sánchez, M. E. (2004). Español de los medios de comunicación: aspectos de disponibilidad léxica. En H. Perdiguero y A. Álvarez (Eds.), *Medios de comunicación y enseñanza del español como lengua extranjera. Actas del XIV Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera*. Burgos: Servicio de publicaciones Universidad de Burgos, pp. 356-371. https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/14/14_0357.pdf.

Hernández Muñoz, N. (2005). La disponibilidad léxica: una herramienta fronteriza para el estudio del léxico en Lingüística y Psicología. En E. Díez Villoria, B. Zubiauz de Pedro y M. Á. Mayor Cinca (Eds.), *Estudio sobre la adquisición del lenguaje*. Salamanca: Ediciones Universidad de Salamanca, pp. 942-953.

Hernández Muñoz, N., Izura, C. y Tomé Cornejo, C. (2014). Cognitive Factors of Lexical Availability in a Second Language. En R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language*. Nueva York: Springer, pp. 169-188.

Hernández Muñoz, N., Tomé Cornejo, C., López García, M. y Bartol Hernández, J. A. (2023). *Manual de LexPRO*. https://dispogram.usal.es.

Herreros Marcilla, M. (2015). Los centros de carácter cultural en los estudios de disponibilidad léxica: análisis y nueva propuesta. En A. Gordejuela Senosiáin, D. Izquierdo Alegría, F. Jiménez Berrio, A. De Lucas Vicente y M. Casado Velarde (Eds.), *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General*. Pamplona: Servicio de Publicaciones de la Universidad de Navarra, pp. 279-290.

Hidalgo Gallardo, M. (2019). *Factores y fuentes que inciden en el léxico disponible de estudiantes sinohablantes de ELE*. Colección Monografías, 23, ASELE.

Instituto Cervantes (s. f.). *Diccionario de términos clave de ELE*. [Diccionario en línea]. https://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/default.htm.

López Morales, H. (1996). Los estudios de disponibilidad léxica: pasado y presente. *Boletín de Filología de la Universidad de Chile*, 35, pp. 245-359.

López Morales, H. (1999). *Léxico disponible de Puerto Rico*. Madrid: Arco Libros.

López Morales, H. *et al.* (s. f.). *DispoLex*. [Programa informático]. Universidad de Salamanca. http://www.dispolex.com.

Luján García, C. I. y Bolaños Medina, A. (2014). Disponibilidad léxica y anglicismos informáticos en los centros de interés: *internet, software y hardware*. *Odisea*, 15, pp. 101-126. http://repositorio.ual.es/bitstream/handle/10835/3948/CARMEN%20ISABL%20LUJAN.pdf?sequence=1.

Madrigal-Melchor, J., Rivera-Juárez, J. M., Enciso-Muñoz, A., López-Chávez, J. (2012). Disponibilidad léxica para medir el crecimiento conceptual de electricidad. *Latin-American Journal of Physics Education*, 6 (4), pp. 648-651. http://www.lajpe.org/dec2012/21_LAJPE_706_Jesus_Madrigal_preprint_corr_f.pdf.

Martini, E. (2023). *Disponibilidad léxica en el español de negocios en cuatro grupos de aprendices italianos*. [Trabajo de fin de máster]. Padua: Universidad de Padua.

MCER = Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*.

Madrid: MECD/Anaya. http://cvc.cervantes.es/Ensenanza/Biblioteca_Ele/marco/cvc_mer.pdf.

Michéa, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les Langues Modernes*, 47, pp. 338-344.

Moreno Fernández, F., Moreno Fernández, J. E. y García de las Heras, A. J. (1995). Cálculo de disponibilidad léxica. El programa LexiDisp. *Lingüística*, 7, pp. 243-250.

Nalesso, G. (2019). *Disponibilidad y riqueza léxica en un grupo de estudiantes universitarios italianos de ELE*. [Tesis doctoral]. Padua: Universidad de Padua. https://www.research.unipd.it/retrieve/handle/11577/3424862/568695/Giulia_Nalesso_tesi.pdf.

Nalesso, G. (2020). Disponibilidad léxica terminológica en ELE: una propuesta de análisis. *Orillas. Rivista d'Ispanistica*, 9, pp. 609-631. https://www.orillas.net/orillas/index.php/orillas/article/view/90.

Nalesso, G. (2022). *La competencia léxica de estudiantes italianos de ELE: un estudio sobre la disponibilidad y la riqueza léxica*. Padua: Padova University Press. https://www.padovauniversitypress.it/system/files/download-count/attachments/2023-09/9788869383311.pdf.

Navarro Marrero, Y. (2010). Terminología especializada en el área de fisioterapia: acercamiento desde la metodología de la disponibilidad léxica específica. *Interlingüística*, 20, pp. 1-11. https://clt.uab.cat/pagines_clt/xxivajl/Interlinguistica/Encuentro%20XXIV/Navarro_Marrero_REVF.pdf.

Paredes García, F. (2015). Disponibilidad Léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula. *Linred*, 13, pp. 1-32. https://linred.web.uah.es/numero13_monografico_Art2.html.

Salcedo Lagos, P. y Del Valle Leo, M. (2013). Disponibilidad léxica matemática en estudiantes de enseñanza media de Concepción, Chile. *Atenas*, 1 (21), pp. 1-16. https://www.redalyc.org/pdf/4780/478048957001.pdf.

Samper Hernández, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. Málaga: ASELE.

Tomé Cornejo, C. (2016). Vocabulario de la informática y las nuevas tecnologías. Caracterización desde la disponibilidad léxica. *Caracteres: estudios culturales y críticos de la esfera digital*, 5 (1), pp. 112-139.

Toniolo, M.ª T. y Zurita, M.ª E. (2019). Léxico vitícola y vitivinícola de uso en Córdoba, Argentina. Estudio léxico de especialidad. En M.ª L. Perassi y M. Tapia Kwiecien (Eds.), *Palabras como puentes: Estudios lexicológicos, lexicográficos y terminológicos desde el Cono Sur*. Córdoba: Buena Vista Editores, pp. 85-100.

Urzúa Carmona, P., Sáez, K. y Echeverría, M. S. (2006). Disponibilidad léxica matemática. Análisis cuantitativo y cualitativo. *Revista Lingüística Teórica y Aplicada*, 44 (2), pp. 59-76.