

ESTUDIAR LA COMPETENCIA LÉXICA EN APRENDICES DE ELE: UNA NUEVA APROXIMACIÓN

GIULIA NALESSO
Università degli Studi di Padova

1. INTRODUCCIÓN

EL PROYECTO, enmarcado en el ámbito de la lingüística aplicada y de la didáctica de ELE, propone un instrumento de análisis empírico innovador para examinar la competencia léxica a través de la disponibilidad y de la riqueza léxica. Se trata de dos metodologías, combinadas por primera vez en un único estudio, capaces de determinar el nivel de aprendizaje y conocimiento del vocabulario español de aprendientes no nativos. La primera analiza en el plano paradigmático del léxico el alcance del dominio del vocabulario y su organización en el lexicón mental, mientras que la segunda observa en el plano sintagmático la variedad de palabras utilizadas en textos escritos u orales.

La pregunta «¿cómo se puede valorar la competencia léxica?» (Morante Vallejo, 2005: 50-53), a la que todavía no se ha encontrado una respuesta unívoca, ha impulsado el proyecto desarrollado en nuestra tesis doctoral (Nalesso, 2019) con el objetivo de arrojar luz sobre la cuestión por medio del análisis –cuantitativo, cualitativo y comparativo– de datos más significativos y completos.

El trabajo, además, se realiza desde los enfoques transversal y longitudinal, renovando el *modus operandi* de la investigación aplicada hasta ahora en este ámbito basado únicamente en estudios sincrónicos.

2. FUNDAMENTACIÓN DEL ESTUDIO

2.1. SOBRE LA COMPETENCIA LÉXICA

La competencia léxica es el «conocimiento del vocabulario de una lengua y la capacidad para utilizarlo» (MCER, 2002: 108), es decir, coincide tanto con el

conocimiento de la forma y del significado de las unidades léxicas como con el desarrollo de las destrezas que permiten la comunicación en la lengua objeto (estrategias de aprendizaje, de retención y de uso). Asimismo, el MCER establece que la competencia léxica permite el alcance de una habilidad comunicativa eficiente: no poder, o no saber, decodificar una palabra puede imposibilitar el éxito de la comunicación y, por tanto, decretar el fracaso de los procesos de enseñanza-aprendizaje. En esta óptica, se incluye dentro de la más amplia competencia lingüística (junto a la gramatical, la semántica, la fonológica, la ortográfica y la ortoépica), la cual, por su parte, es un componente de la competencia comunicativa.

De ahí que el conocimiento léxico sea imprescindible para poder comunicar y, por tanto, es necesario planificar el desarrollo de la competencia léxica del alumnado durante la adquisición lingüística. Claro está que la enseñanza del vocabulario de un idioma se reparte en una escala graduada según el nivel del grupo de discentes, a saber, según una gradación del conocimiento y de la capacidad de control que aumentan según avanza el dominio lingüístico. El aprendizaje no es estático, sino que es un proceso dinámico en el que las palabras se aprenden y se guardan de manera ordenada en un almacén que está en continua evolución, el lexicón mental (Baralo, 1999; 2001). En la etapa inicial del proceso, saber una palabra significa conocer su forma y significado y, sucesivamente, saber «realmente» una palabra abarca otros aspectos que comprenden la capacidad de usarla correcta y eficazmente en un educto, la producción de un acto de habla (Marconi, 1999; Nation, 2001; Jiménez Catalán, 2002; Rufat, 2018).

2.2. MEDIDAS DE EVALUACIÓN DE LA COMPETENCIA LÉXICA

La disponibilidad léxica es una metodología encaminada a descubrir qué y cuántas palabras sería capaz de usar un individuo en una situación determinada por el tema comunicativo a través de la recogida de su léxico disponible, aquel vocabulario «que se usa en una situación comunicativa específica, actualizándose sólo ante la necesidad de comunicar algo referente a esa situación concreta» (López Morales, 1999: 9). Es un método de evaluación de la competencia léxica que consiste en analizar listas de palabras que los informantes actualizan durante los test en función de estímulos semánticos denominados centros de interés (CI), pero ¿es suficiente para estudiar un fenómeno tan complejo como la competencia léxica?

Parece que no lo es, ya que «la prueba de obtención del léxico disponible nada nos dice sobre si los hablantes son capaces de usar un determinado término» (Hernández Muñoz, 2005: 13), pues «la disponibilidad léxica se perfila, en definitiva, como un indicador de gran utilidad para la evaluación del componente léxico, a pesar de que por sí sola no sea capaz de medir todos los aspectos que intervienen en el conocimiento de una unidad léxica» (Tomé, 2016: 96).

Sin embargo, en el panorama de los estudios sobre la lengua española –LM y LE– la disponibilidad ha tenido un gran éxito¹ y ha sido utilizada como metodología léxico-estadística cuyo producto principal es la recogida de listados de léxico disponible de hispanohablantes nativos o de estudiantes extranjeros en función de diferentes variables sociolingüísticas y entornos de residencia o de aprendizaje. En efecto, «el material léxico recogido mediante la metodología de la disponibilidad léxica se está empleando para profundizar en el conocimiento de la organización del lexicón mental, tanto en hablantes nativos como extranjeros» (Del Barrio, 2018: 174). Tales estudios tienen varios propósitos, en particular, los dedicados a no nativos se centran (i) en la medición del conocimiento léxico, su desarrollo, la organización del lexicón mental y (ii) en la eficacia de la propuesta didáctica a la que se somete el alumnado²:

Por lo que se refiere a la lengua extranjera, los estudios de disponibilidad léxica permiten conocer el vocabulario realmente disponible de un grupo de hablantes extranjeros en la lengua aprendida (L2). De esta manera, se convierte en un instrumento de evaluación de la competencia léxica en lengua extranjera y de la influencia de la metodología educativa en el desarrollo del conocimiento léxico. (López González, 2010)

Tales análisis permiten adquirir información valiosa ya que los listados de léxico disponible son «un auténtico diccionario de uso, no solo fiable por la metodología empleada, sino por su actualidad [...] con el valor añadido de que se trata de un trabajo que proporciona información dialectológica, sociolingüística y de frecuencia» (Gallos Camacho y Prado Aragonés, 2004: 372). Se componen de unidades léxicas encajadas en redes de relaciones formales y semánticas significativas, construidas con otras palabras, que conforman la competencia léxica de cada individuo.

Aun reconociendo sus logros, el léxico disponible tiene un carácter potencial que no es suficiente para medir en su totalidad dicha competencia, por lo que sería útil tener en cuenta también el vocabulario utilizado en una producción lingüística, el léxico activo. En efecto, pese a que una palabra disponible sea «una palabra

¹ El español es la lengua que más ha contribuido al desarrollo de la disponibilidad léxica. Para una reseña bibliográfica sobre el tema véanse Aabidi (2019), Hidalgo (2019), Nalesso (2019).

² Apreciamos las contribuciones de numerosos proyectos cuyo objetivo prioritario responde a finalidades educativo-pedagógicas relacionadas con el estudio de la competencia léxica en el ámbito de la didáctica de ELE, en lo que atañe a: (i) planificación lingüística; (ii) elaboración de materiales: eliminación del léxico superfluo de los contenidos didácticos y proposición de temas más apropiados; (iii) análisis de tipo cognitivo y definición de las etapas de aprendizaje; (iv) cotejo del vocabulario de los participantes con otros estudiantes, con nativos y con comunidades bilingües; (v) comparaciones culturales; (vi) evaluación de las propuestas didácticas; (vii) detección de los déficits léxicos y de los errores, las interferencias de la LM o de otras LE conocidas; (viii) ampliación de la cobertura de los estudios hacia la terminología, es decir, hacia el español para fines específicos.

que, sin ser particularmente frecuente, está lista para ser empleada y viene inmediata y naturalmente a la mente en el momento en que se tiene necesidad de ella» (Michéa, 1953: 310) y que, en este sentido, «un estudio de disponibilidad léxica implica centrarse en el plano de la producción lingüística» (Mena Osorio, 1986: 11) estaríamos analizando solo conocimientos que dependen del tipo de prueba a la que se someten los sujetos, la productividad léxica temática. Para ello, los análisis deberían complementarse con los de otra metodología que permitan examinar eductos reales y ofrezcan una visión más completa del dominio del léxico. La disponibilidad estudia exclusivamente el vocabulario activado ante un estímulo temático preciso y no aporta información sobre la capacidad de utilizar las palabras, esto es, la productividad léxica contextual. La unión crítica de diferentes análisis permitiría cotejar datos más completos:

[Aunque] el léxico está siendo centro de atención desde todos los puntos problemáticos del estudio de la lengua [sobre todo con] trabajos cualificados de vocabulario disponible (palabras aisladas) en escolares, recogidos por centros de interés, nos falta también un vocabulario recogido en textos producidos por los escolares [...], lo que daría nivel de uso en producción real de los escolares y representaría un conocimiento más exacto que el del vocabulario disponible. (Battaner, 1992 apud Cuba Vega, 2016: 81)

A este propósito, la riqueza léxica es una metodología que estudia la producción textual analizando cuantitativa y cualitativamente el vocabulario de un texto a través de índices estándar que evalúan la habilidad de manejar un repertorio de vocablos en un discurso oral o escrito: «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» (MCER, 2002: 86)³. Se basa en la longitud del texto, la diversidad del vocabulario y la alternancia de uso de palabras nocionales y funcionales del discurso (Read, 2000).

3. PROPUESTA DE ESTUDIO DE LA COMPETENCIA LÉXICA

Se plantea un cambio metodológico basado en la aplicación de las dos metodologías citadas cuyo acierto principal es la recogida de datos más significativos encaminados a la medición del nivel de aprendizaje y conocimiento del vocabu-

³ Las posibilidades de aplicación de la riqueza léxica son muchas en el ámbito de la investigación y didáctica de las lenguas extranjeras, entre las que contamos con: (i) averiguación de qué y cuánto vocabulario conocen los aprendientes; (ii) comprobación de la capacidad de uso del vocabulario; (iii) análisis y elaboración de materiales pedagógicos en función del vocabulario conocido; (iv) evaluación del desarrollo del léxico activo de cada alumno y del estudiantado en general.

lario de los informantes. La comparación del léxico disponible con los resultados de la riqueza léxica permite un estudio más preciso y coherente. Los análisis de disponibilidad léxica sirven para diagnosticar la cantidad y la calidad de lexemas que conocen los informantes, como el grado de asociación cognitiva de lo aprendido. Por su parte, los análisis de la riqueza permiten examinar eductos en la lengua objeto y averiguar si los mismos informantes saben una palabra y cómo la utilizan, a saber, si el conocimiento pasivo-receptivo inicial ha entrado a formar parte del conocimiento activo-productivo.

Asimismo, hasta la fecha, los estudios en este ámbito se limitaban a análisis transversales condicionados por el método de recogida de datos de la disponibilidad, sin tener en cuenta un enfoque longitudinal que proporcionase información para observar el desarrollo temporal de la competencia. Otra aportación de la presente propuesta es el hecho de considerar la evolución del conocimiento léxico temático y contextual tras un proceso de instrucción reglada del mismo grupo⁴.

3.1. MÉTODO DE LA INVESTIGACIÓN

3.1.1. *Recogida y edición de los datos*

La medición de la competencia léxica se realiza a través de una encuesta dividida en tres partes (Nalesso, 2019: 128-158). El primer apartado presenta un cuestionario sociológico encaminado a recoger información sobre el bagaje lingüístico de los participantes y datos personales que permiten la repartición según variables sociolingüísticas (que en nuestro caso fueron sexo, nivel de ELE, conocimiento de otras lenguas extranjeras).

A continuación, se proporciona la prueba de disponibilidad léxica de tipo asociativo según el modelo tradicional. Se trata del clásico test de listas abiertas con los dieciséis centros de interés originales⁵ y con dos minutos para cada uno (Samper

⁴ Carcedo (1998) realiza un estudio sobre el desarrollo de la competencia léxica en ELE de estudiantes de lengua materna fina basándose en el cotejo de los resultados obtenidos en el test de disponibilidad léxica de cuatro grupos diferentes, dos de bachillerato y dos universitarios. Igualmente, Hidalgo (2019) propone un estudio diacrónico sobre el desarrollo del léxico disponible de aprendientes sinohablantes, pero compara los resultados de su grupo con los de Lin (2006).

⁵ CI01, «partes del cuerpo» [CUE]; CI02, «la ropa» [ROP]; CI03, «partes de la casa (sin muebles)» [CAS]; CI04, «los muebles de la casa» [MUE]; CI05, «alimentos y bebidas» [ALI]; CI06, «objetos colocados en la mesa para la comida» [MES]; CI07, «la cocina y sus utensilios» [COC]; CI08, «la escuela: muebles y materiales» [ESC]; CI09, «iluminación y calefacción» [ILU]; CI10, «la ciudad» [CIU]; CI11, «el campo» [CAM]; CI12, «medios de transporte» [TRA]; CI13, «trabajos del campo y del jardín» [TRC]; CI14, «los animales» [ANI]; CI15, «juegos y distracciones» [JUE]; CI16, «profesiones y oficios» [PRO].

Padilla, 1998). Para ello, nos basamos en los trabajos pioneros en ELE de Carcedo (2000) y Samper Hernández (2002) y en las aportaciones más recientes de Sánchez-Saus (2016) e Hidalgo (2019).

La prueba se compone de cuatro páginas formadas por cuatro columnas y un espacio sobre cada una, numerado de 1 a 16, en el que se escriben los nombres de los centros de interés que se presentan uno a uno a medida que se completan para que no se piense antes de tiempo en las palabras vinculadas con el estímulo siguiente. Los encuestados rellenan las fichas con las unidades léxicas que les vienen a la mente en relación con cada tema, siguiendo en el reverso del folio si la tabla no es suficiente para poder trabajar listas abiertas. Es importante exhortarles a escribir todas las palabras que les ocurren, aunque no estén seguros de la grafía o del significado.

La prueba de riqueza léxica consiste en un ejercicio de redacción de un texto narrativo simple, que se analiza según índices estándar (Read, 2000; Jarvis, 2013; Capsada y Torruella, 2017). Se propone un tema sencillo para maximizar la posibilidad de expresión de los participantes en función de su nivel lingüístico, un asunto complejo o especializado no aportaría los datos necesarios porque complicaría la tarea entorpeciendo la capacidad de la muestra⁶. Cada cuento debe presentar un máximo de cien palabras porque parece que en textos más largos los cálculos de la riqueza léxica se alteran: «[...] sabemos que los datos se desvirtúan al pasar un texto de las 100 palabras. [...] Necesitamos, pues, para tener metas claras, exámenes de textos escritos que se consideren ilustrativos, de los que se pueden tomar varias calas, pero nunca superiores a las 100 palabras» (López Morales, 2011: 24).

En el momento de la suministración es recomendable seguir el mismo procedimiento aplicando un protocolo estándar que garantice la repetitividad del experimento:

1. presentación del investigador y del estudio;
2. explicación de los contenidos y de las instrucciones;
3. aclaración de la anonimidad de la encuesta;
4. administración de las hojas;
5. realización de la prueba.

⁶ Laufer y Nation (1995: 308) destacan la importancia de medir la competencia de un sujeto según la familiaridad que tiene desarrollada sobre un cierto tema.

Se suministra la encuesta dos veces a los mismos informantes, una a comienzo y una a final del año académico, para capturar una imagen transversal y una longitudinal del dominio del vocabulario⁷.

Tras la recogida, se edita el corpus –las listas de léxico disponible por una parte y los cuentos por otra– someténdolo a una minuciosa revisión según las prácticas convencionales empleadas en la bibliografía sobre el estudio de la producción escrita en lengua extranjera.

En el análisis de la disponibilidad léxica, para que el corpus resulte homogéneo y consienta análisis fiables y cotejables, se aplican (i) criterios generales comúnmente utilizados en los proyectos en este ámbito (Samper Padilla, 1998)⁸; (ii) criterios específicos utilizados en trabajos dedicados a aprendientes no nativos de español (Samper Hernández, 2002; Sánchez-Saus, 2016)⁹; (iii) criterios particulares de la producción de los participantes por centro de interés, que cambian dependiendo de la muestra.

La edición de los textos para el estudio de la riqueza consiste en un análisis de errores, es decir, en la detección de la tipología de error para decidir si (i) tolerarlo y corregirlo, en el caso de faltas ortográficas, fonológicas o morfosintácticas que no perjudican el conocimiento del vocabulario o (ii) descartarlo en el caso de errores léxicos propiamente dichos (Barbasán Ortuño, 2016).

El tratamiento informático de los datos se realiza con los programas: (i) *Dispo-gen II* (Echeverría *et al.*, 2005), basado en la fórmula propuesta por López Chávez y Strassburger Frías (1987), que calcula los distintos indicadores de la disponibilidad

⁷ Laufer (1991) realiza un estudio sobre riqueza léxica donde examina diacrónicamente 47 aprendientes de inglés L2 en diferentes etapas de su carrera universitaria según el cual en intervalos de 14 y 28 semanas es posible detectar un desarrollo del vocabulario. Por tanto, la medición estadística y el cotejo de dos textos escritos por el mismo informante en dos etapas temporales permiten conseguir resultados útiles para medir la competencia léxica (Read, 2000: 197).

⁸ Se trata de: eliminación de los términos repetidos; corrección de la ortografía; unificación ortográfica; unificación de variantes meramente flexivas; unificación de derivados regulares; unificación de formas plenas y acortamientos; tratamiento de los sintagmas; tratamiento de las marcas comerciales.

⁹ Se registran todas las entradas que se entienden con facilidad, salvo que se trate de palabras españolas. En lo que atañe al tratamiento de los préstamos y de las interferencias de otras lenguas, se aceptan los neologismos universalizados presentes en el DRAE, tanto en la grafía adaptada fonológica y ortográficamente al español como en la grafía original. Asimismo, se admiten los extranjerismos utilizados con frecuencia por los hispanohablantes, aunque no se encuentren en el diccionario. Se descartan artículos, adjetivos y adverbios que acompañan los ítems activados, a excepción de los que forman unidades multipalabra según el criterio de la unificación en la presentación de las unidades. Por último, en función del criterio de la amplitud de las relaciones asociativas, se aceptan aquellas palabras activadas mediante relaciones secundarias, sobre todo en los campos léxicos considerados complejos para informantes extranjeros. Se eliminan solo las asociaciones inexplicables.

léxica; (ii) *AntConc* (Anthony, 2018) que permite el conteo de *types* y *tokens* necesario para poder computar los índices de variedad y densidad léxica de la riqueza; (iii) *Sketch Engine* (Kilgarriff et al., 2004) que combina estadísticamente diferentes criterios lingüísticos y permite recopilar listas de frecuencia y listas de palabras clave, entre otras funciones.

3.1.2. *Análisis de los datos*

En el estudio cuantitativo de la disponibilidad, los índices utilizados son (i) total de palabras y vocablos por CI, (ii) promedio de palabras y vocablos por CI, (iii) promedio de palabras y vocablos por informante. Asimismo, se examina el grado de concreción semántica de los centros –esto es, si el léxico del corpus es homogéneo– mediante (i) el índice de cohesión que determina la coincidencia de las respuestas¹⁰ y (ii) la densidad léxica que contabiliza la media de unidades repetidas¹¹.

A partir de los datos extraídos de los índices se desarrollan los siguientes análisis:

1. el análisis transversal examina los resultados generales y por variable;
2. el análisis longitudinal permite observar la evolución en el tiempo de la competencia léxica en los dos momentos de administración de la prueba y deducir en qué medida los procesos didácticos inciden en la adquisición del vocabulario;
3. el análisis comparativo coteja los resultados obtenidos con los de diferentes trabajos sobre disponibilidad realizados con distintos grupos de estudiantes de ELE de igual competencia, pero de distinta procedencia, lengua materna y contexto de aprendizaje;
4. el análisis descriptivo estadístico completa los datos examinando la simetría o la dispersión del comportamiento intragrupal e intergrupala de los participantes.

El estudio cualitativo transversal y longitudinal describe el grado de asociación de las respuestas en función del estímulo del centro de interés y las categorías gramaticales más difundidas en el corpus. Asimismo, se basa en los porcentajes de

¹⁰ Según la fórmula de Echeverría (1987) se obtiene dividiendo la media de palabras por informante por el número total de vocablos activados. Sus valores pueden oscilar entre 0 y 1: cuanto más se aproxima a 1, mayor cohesión se detecta porque las respuestas coinciden y, por tanto, el centro se define compacto. Al contrario, hay una menor coincidencia cuanto más se aproxima a 0, es decir una menor cohesión que supone un centro más difuso.

¹¹ Se computa dividiendo el número de palabras por el total de vocablos.

cardinalidad del conjunto (López Chávez, 1992) que revelan el nivel de compatibilidad del vocabulario activado en distintas agrupaciones de la muestra general.

Se propone también un cotejo con el *Corpus de Referencia del Español Actual* (CREA) de la RAE y el *Corpus de Aprendices de Español* (CAES) del Instituto Cervantes para averiguar si las unidades del léxico disponible de los informantes aparecen dentro de los corpus y, por tanto, si el vocabulario que dominan es léxico estándar de uso frecuente en el español actual y su correspondencia con el léxico conocido por otros aprendientes de ELE.

Siguiendo el mismo orden en la ejecución de los análisis, la riqueza léxica mide las diferencias cuantitativas entre (i) *tokens* y *types* (diversidad del vocabulario) y (ii) palabras funcionales o nocionales del discurso (densidad del vocabulario).

La variación léxica y la *Type/Token Ratio* (T^TR)¹² contabilizan el grado de diversidad léxica de cada cuento dividiendo el número de *types* por el total de *tokens*. Igualmente, el índice de hápax detecta las palabras empleadas una vez en un relato concurriendo a establecer la variedad del vocabulario a través de la división del número total de *types* por la suma de los hápax¹³.

Por otra parte, la densidad léxica calcula el porcentaje de unidades nocionales en un texto en proporción al número total de ítems que lo componen y el intervalo de aparición de palabras nocionales (IAT) determina el intervalo en el que aparecen las palabras de contenido semántico con respecto a las palabras funcionales, dividiendo el número total de *tokens* por el número de *tokens* léxicos¹⁴.

Si bien los datos procedentes de estos cálculos sean numéricos, aportan información cualitativa sobre la competencia léxica:

La llamada 'calidad de la escritura' está integrada por una serie de factores, entre los que destacan sin duda: la riqueza léxica, la madurez sintáctica, los esquemas de cohesión y la coherencia discursiva. La amplitud y variedad del vocabulario está muy apoyado en la disponibilidad léxica del hablante, la madurez sintáctica, en su grado de entrenamiento combinatorial de oraciones simples en el discurso, los esquemas de cohesión y la coheren-

¹² Este índice puede variar de 0 a 1. Es imposible llegar al 0 ya que siempre hay un *type* como mínimo en una muestra (al contrario, no habría ningún texto). De igual manera, es improbable llegar al 1 porque se obtendría en el caso de que no haya repeticiones, algo imposible en un texto que no sea una oración. En definitiva, el valor ideal sería el que más se acerca a 1.

¹³ Al tratarse de un indicador inversamente proporcional, la riqueza de un texto es mayor cuanto más pequeño es el índice.

¹⁴ Cuanto más se acerca a 1, mayor resulta el uso de palabras temáticas y, por consiguiente, la riqueza léxica (1 no es alcanzable porque significaría que todos los *tokens* de un texto fuesen palabras nocionales).

cia discursiva dependen esencialmente del 'orden' que quiera dársele conscientemente a los elementos constitutivos del discurso. (López Morales, 2011: 15)

Como colofón, se incluye un análisis cualitativo de los datos –poco común en los estudios de riqueza– que se suma a la propuesta metodológica para explorar aun más la competencia léxica (Nalesso, 2019: 342-355). Para ello, el punto de partida son los vocablos más frecuentes del conjunto textual que determinan morfológicamente las lexías más utilizadas, el área semántica a la que pertenecen y el grado de asociación con la trama del texto, cuyo tema debería poderse adivinar en función de las palabras clave detectadas.

3.2. APLICACIÓN DE LA PROPUESTA¹⁵

La investigación se basa en un experimento realizado a partir de la prueba de medición de la competencia léxica, diseñada según la propuesta presentada, que se sometió a cien aprendientes universitarios italianos de ELE de nivel intermedio, B1 y B2 según el MCER, matriculados en cursos de Lengua Española de grado.

Dicha prueba se proporcionó dos veces, a comienzo y a final del año académico, para analizar sincrónica y diacrónicamente la competencia léxica de los informantes. La primera administración se llevó a cabo con cincuenta alumnos de un nivel y cincuenta del otro, la segunda se suministró solo al grupo más avanzado.

Trabajamos con cuatro corpus: el «corpus general» consta de los datos de la primera prueba, del que se extrajeron los resultados de los B1 y los B2 para el análisis por variable («corpus B1» y «corpus B2_a») y para el estudio longitudinal («corpus B2_a» cotejado con el «corpus B2_b», que recoge los datos de la segunda prueba efectuada tras el curso).

Se organizó la investigación según los siguientes pasos:

1. selección de la muestra;
2. establecimiento de las técnicas de elicitación de datos, diseño de la prueba y de los procesos de recogida;
3. primera y segunda administración de la prueba;
4. fijación de los criterios de edición y de tratamiento informático del material;
5. despojo de los datos.

¹⁵ El estudio completo se halla en Nalesso (2019).

Antes de empezar, se fijaron algunas hipótesis que se pudieron corroborar o refutar al término de los análisis:

1. La competencia léxica aumenta según avance el nivel de español. Los estudiantes más avanzados presentarán una mayor productividad y riqueza léxicas, tanto en la activación de léxico disponible como en la producción escrita. Igualmente, el rendimiento será mejor a final del año académico.
2. Existe una relación asociativa entre el nivel de ELE y la cantidad de lexías aprendidas, a saber, activadas y utilizadas en las dos partes de la encuesta.
3. Considerado el nivel de ELE, los indicadores de riqueza léxica expresarán una apreciable variación de vocabulario gracias al desarrollo de una adecuada competencia léxica que se repercute positivamente en la competencia comunicativa de los estudiantes.
4. Discentes no nativos de distinta procedencia que comparten el dominio lingüístico poseen iguales capacidades expresivas y, a más y mejor, tienen desarrollada la misma competencia en ELE.
5. La correlación de los datos procedentes de los análisis de disponibilidad léxica, basados en parámetros de tipo paradigmático, y de riqueza léxica, basados en parámetros de tipo sintagmático, posibilitarán un pormenorizado estudio de la competencia léxica de los aprendientes de ELE.

3.2.1. Resultados parciales

Los análisis de la disponibilidad léxica revelaron un aumento del léxico disponible, en la mayoría de los centros de interés, de un nivel a otro (+7,51%) y entre las dos administraciones del test (+30,29%) indicando una mejora de la competencia léxica, aunque no sobresaliente como se esperaba debido a la afinidad entre LM y LE.

TABLA 1. Promedios de palabras activadas por CI

| CI | Corpus general | B1 | B2_a | B2_b |
|----------|----------------|-------|-------|-------|
| CI01.CUE | 13,41 | 11,88 | 14,94 | 15,70 |
| CI02.ROP | 7,91 | 7,28 | 8,54 | 11,88 |
| CI03.CAS | 6,10 | 6 | 6,20 | 8,42 |
| CI04.MUE | 6,60 | 6,32 | 6,88 | 8,88 |
| CI05.ALI | 14,63 | 13,38 | 15,88 | 17 |

| | | | | |
|----------|-------|-------|-------|-------|
| CI06.MES | 5,30 | 4,98 | 5,62 | 7,98 |
| CI07.COC | 3,65 | 3,22 | 4,08 | 5,76 |
| CI08.ESC | 8,03 | 7,84 | 8,22 | 13,78 |
| CI09.ILU | 3,42 | 3,46 | 3,38 | 6,76 |
| CI10.CIU | 11,71 | 11,92 | 11,50 | 14,20 |
| CI11.CAM | 6,85 | 6,98 | 6,72 | 8,04 |
| CI12.TRA | 8,08 | 7,92 | 8,24 | 10,44 |
| CI13.TRC | 2,52 | 2,14 | 2,90 | 4,16 |
| CI14.ANI | 10,80 | 10,36 | 11,24 | 14,58 |
| CI15.JUE | 7,84 | 8,34 | 7,34 | 9,68 |
| CI16.PRO | 8,94 | 9,22 | 8,66 | 12,56 |
| Promedio | 7,86 | 7,58 | 8,15 | 10,61 |

La riqueza se manifestó adecuada considerando el grado de dominio lingüístico de los informantes. Todos los textos alcanzaron un buen nivel de variedad léxica ya que las palabras diferentes cubren más del 50% de cada relato. Por su parte, la densidad léxica evidenció un rendimiento inferior. Estos datos comprueban que un texto con una buena diversidad de vocabulario no necesariamente arroja igual resultado en lo que se refiere al uso de las unidades léxicas nocionales.

Asimismo, no se produjo el mismo desarrollo de la competencia entre los dos niveles encuestados (+1,61% variedad, +2,16% densidad) y tampoco en el análisis longitudinal (-3,17% variedad, -1,08% densidad). En promedio, los índices no ascienden significativamente, o bien bajan, contradiciendo una de las hipótesis iniciales: parece que el avance de un nivel al otro ralentizó, hasta producir un bloqueo. Los alumnos B1 consiguieron un rendimiento mejor que los B2 y a principio del año académico.

TABLA 2. Promedios de riqueza léxica

| Índice | Corpus general | B1 | B2_a | B2_b |
|------------------|----------------|------|------|------|
| TTR | 0,63 | 0,62 | 0,63 | 0,61 |
| Variación léxica | 63% | 62% | 63% | 61% |
| Índice de hápax | 1,39 | 1,39 | 1,39 | 1,39 |

| | | | | |
|-----------------|------|------|------|------|
| Densidad léxica | 55% | 55% | 54% | 55% |
| IAT | 1,84 | 1,82 | 1,85 | 1,83 |

El cotejo entre estudiantes de distinta procedencia¹⁶ puso de relieve que tanto las cantidades activadas de palabras como de vocablos de los italianos se revelaron escasas en todos los centros de interés, refutando la hipótesis según la cual su léxico disponible habría sido mayor gracias al parentesco lingüístico entre italiano y español, que habría favorecido la adquisición de la lengua objeto.

TABLA 3. Promedios de disponibilidad léxica en las investigaciones cotejadas

| Investigación | País | Promedio palabras | Promedio vocablos |
|--------------------------|-----------|-------------------|-------------------|
| Carcedo (2000) | Finlandia | 162,12 | 44,36 |
| Samper Hernández (2002) | España | 174,92 | 65,07 |
| Šifrar Kalan (2014) | Eslovenia | 174,24 | 28,22 |
| Sánchez-Saus (2016) | España | 285,66 | 47,17 |
| Del Barrio y Vann (2018) | Italia | 145,93 | 26,41 |
| Hidalgo (2019) | China | 284,99 | 42,60 |
| Nalesso (2019) | Italia | 125,79 | 14,90 |

Es posible que esto dependa de la atención que otorgan los programas curriculares al componente léxico, del método didáctico y del contexto de aprendizaje (país de origen *vs.* país hispanohablante). Destacan los resultados de los chinos que actualizan un número alto de palabras: puede que la didáctica insista en el vocabulario debido a la lejanía entre LM y LE que imposibilitaría la comprensión y la expresión de los discentes sin una base léxica sólida. En el caso opuesto, la afinidad lingüística de los italianos lleva a descuidar la enseñanza del léxico por la sensación de facilidad que causa la poca distancia tipológica.

¹⁶ El primero es un estudio de Carcedo (2000) dedicado a estudiantes finlandeses, hablantes de sueco y finés. Samper Hernández (2002) trabaja con aprendientes en inmersión en la Universidad de Salamanca, cuyas lenguas maternas eran inglés, italiano y japonés. Šifrar Kalan (2014) se centra en estudiantes eslovenos. Sánchez-Saus (2016) analiza el léxico disponible de informantes que estudiaban español en varias universidades andaluzas, nativos de alemán, finés, francés, inglés, italiano y polaco. Del Barrio y Vann (2018) se dedican a una muestra de estudiantes italianos e Hidalgo (2019) a sinohablantes.

En cuanto a la correlación de la capacidad productiva¹⁷, se validó que aprendientes no nativos de diferentes LM, que comparten el nivel de español, producen una riqueza semejante: los grados de variedad y densidad léxicas de sus textos son homogéneos.

TABLA 4. Promedios de riqueza léxica en las investigaciones cotejadas

| Investigación | TTR | IAT |
|-------------------------|------|-----------------|
| Cuba Vega y Cuba (2004) | 0,64 | 1,81 |
| Berton (2014) | 0,70 | — ¹⁸ |
| Wang (2016) | 0,60 | 1,80 |
| Basso (2017) | 0,60 | 2,35 |
| Nalesso (2019) | 0,63 | 1,84 |
| Promedio | 0,63 | 1,95 |

Los cuentos entregados presentaron el mismo buen nivel de diversidad, que supera el 50% en cada uno. De la misma manera, el cotejo de la densidad reveló datos similares, pero no igualmente satisfactorios porque suponen que las lexías con mayor presencia sean gramaticales con respecto a las nocionales. Extraña el escaso rendimiento de los sinohablantes de Wang (2016) si consideramos que en el análisis de la disponibilidad léxica los chinos (Hidalgo, 2019) alcanzaron datos sobresalientes.

Para terminar, destacan las conclusiones extraídas del contraste entre los resultados de las dos metodologías ya que proporcionaron datos inesperados y, por tanto, de interés. Mientras que con la instrucción aumenta el caudal léxico disponible, en el análisis léxico de la construcción textual (léxico activo) no sucede así en términos globales. Esto corrobora la utilidad de combinar los dos estudios cuya correlación crítica posibilita una detallada descripción de la competencia léxica.

¹⁷ Cuba Vega y Cuba (2004) examinan la producción de alumnos brasileños; Berton (2014) trabaja con aprendientes suecos; Wang (2016) estudia la riqueza en discentes chinos; Basso (2017) se focaliza en estudiantes itálfonos.

¹⁸ Berton (2014: 28) no contabiliza el IAT porque «la medida de la densidad léxica no parece útil para este estudio, ya que en las producciones de aprendices principiantes e intermedios como los informantes del presente estudio el uso de palabras funcionales, como por ejemplo los marcadores del discurso, es muy reducido».

4. CONCLUSIONES

El estudio empírico corroboró la utilidad de reunir bajo un mismo marco metodológico procedimientos experimentales distintos para lograr una mejor comprensión de un fenómeno complejo como es el de la competencia léxica, proporcionando un instrumento de análisis innovador para la comunidad científica.

El primer método, la disponibilidad léxica, se coloca en el plano paradigmático del léxico y permite analizar no solo de qué palabras dispone un sujeto, sino cómo se organizan conceptualmente en su lexicón mental en un momento dado. El segundo se denomina riqueza léxica y permite observar cuántas palabras distintas se utilizan en un texto, y cómo, ubicándose en el plano sintagmático.

Aunque ambas pertenecen a dos ámbitos bien consolidados e independientes, no se habían utilizado hasta ahora conjuntamente, de manera contrastiva y complementaria, para determinar la competencia léxica de los alumnos en un estudio transversal y longitudinal. En efecto, diferentes trabajos sobre disponibilidad léxica (Bartol Hernández, 2010; Tomé, 2015 y 2016; Paredes García, 2015; Sánchez Saus, 2016) subrayan la necesidad de comparar el léxico disponible con otras medidas evaluativas del vocabulario: el experimento perfecciona de alguna manera esta laguna y puede representar un avance en los estudios aplicados a la enseñanza de lenguas extranjeras o segundas.

Los análisis llevaron a conclusiones inesperadas porque cuando se trató de evaluar la evolución diacrónica, los recursos léxicos disponibles habían aumentado con la educación, pero no ocurrió lo mismo en el análisis de la riqueza, en el que pareció que el desarrollo de la competencia había ralentizado.

Esta conclusión debería ser investigada más en profundidad, ya que es uno de los resultados que pueden marcar el camino hacia donde seguir investigando sobre los diferentes tipos de conocimiento léxico que desarrollan y manifiestan los hablantes de una lengua extranjera.

REFERENCIAS BIBLIOGRÁFICAS

- AABIDI, L. (2019). La disponibilidad léxica en español como lengua extranjera: dos décadas de investigación científica. *MarcoELE*, 28. https://marcoele.com/descargas/28/aabidi-disponibilidad_lexica.pdf
- ANTHONY, L. (2018) *AntConc*. (Versión 3.5.2.) [programa informático]. Waseda University. <http://www.laurenceanthony.net/software/antconcl>
- BASSO, V. (2017). *La riqueza léxica en la producción escrita de aprendices itálofonos de E/LE* (Tesis de máster inédita). Università degli Studi di Padova.
- BARALO, M. (1999). *La adquisición del español como lengua extranjera*. Arco Libros.

- BARALO, M. (2001). El léxico no nativo y las reglas de gramática. En S. PASTOR y V. SALAZAR (Eds.), *Estudios de Lingüística de la Universidad de Alicante. Tendencias y líneas de investigación en adquisición de segundas lenguas* (pp. 12-40). Universidad de Alicante.
- BARBASÁN ORTUÑO, I. P. (2016). *El error léxico en la interlengua de lenguas afines: italiano y español* [tesis doctoral, Universitat Politècnica de València]. Repositorio institucional. <https://riunet.upv.es/bitstream/handle/10251/63145/-Barbasán%20-%20El%20error%20léxico%20en%20la%20interlengua%20de%20lenguas%20afines%3A%20italiano%20y%20español.pdf?sequence=1>
- BARTOL HERNÁNDEZ, J. A. (2010). Disponibilidad léxica y selección del vocabulario. En R. CASTAÑER MARTÍN y V. LAGÜENS GRACIA (Eds.), *De moneda nunca usada. Estudios filológicos dedicados a José M^a. Engueta Utrilla* (pp. 85-107). Instituto Fernando el Católico. <http://ifc.dpz.es/recursos/publicaciones/29/95/09bartol.pdf>
- BERTON, M. (2014). *La riqueza léxica en la producción escrita de estudiantes suecos de ELE* [tesis de máster, Stockholms Universitet]. Repositorio institucional: <https://www.diva-portal.org/smash/get/diva2:729863/FULLTEXT01.pdf>
- CAPSADA BLANCH, R. y TORRUELLA CASAÑAS, J. (2017). Métodos para medir la riqueza léxica de los textos. Revisión y propuesta. *Verba*, 44, 347-408. <http://www.usc.es/revistas/index.php/verba/article/view/3155/4585>
- CARCEDO GONZÁLEZ, A. (1998). Desarrollo de la competencia léxica en español LE: análisis de cuatro fases de disponibilidad. *Pragmalingüística*, 5-6, 75-94. <https://revistas.uca.es/index.php/pragma/article/view/517/451>
- CARCEDO GONZÁLEZ, A. (2000). *Disponibilidad léxica en español lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Turun Yliopisto.
- CUBA VEGA, L. E. y CUBA, Y. M. (2004). Las pruebas de riqueza léxica y su aplicación en Español como Lengua Extranjera (ELE). En A. Díez MEDIAVILLA (Ed.), P. COUTO CANTERO, F. VIEITO LIÑARES y E. ARADAS CAROLLO (Coords.), *Actas VIII Congreso Internacional Sociedad Española de Didáctica de la Lengua y la Literatura, La Habana 5-9 de diciembre de 2004: Homenaje a María Zambrano y Alejo Carpentier* (pp. 815-830).
- CUBA VEGA, L. E. (2016). Sexo y vocabulario: relación en escolares habaneros. *Enunciación*, 21(1), 80-90. <https://revistas.udistrital.edu.co/index.php/enunc/article/view/10139/11753>
- DEL BARRIO DE LA ROSA, F. (2018). Pares léxicos en el léxico disponible de estudiantes italianos. *Lingue e Linguaggi*, 26, 173-196. <http://siba-ese.unisalento.it/index.php/linguelinguaggi/article/view/18194/16577>
- DEL BARRIO DE LA ROSA, F. y MAE VANN, M. (2018). *Disponibilidad léxica de los estudiantes de español en Italia. Estudio y diccionarios*. Aracne.
- ECHEVERRÍA, M. S., MORENO, P., OLIVA HERRERA, M.^a y PRADENAS, F. (1987). Disponibilidad léxica en Educación Media. Resultados cuantitativos. *RLA. Revista de lingüística teórica y aplicada*, 25, 55-115.
- ECHEVERRÍA, M. S., VARGAS, R., URZÚA, P. y FERREIRA, R. A. (2005). *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*. Universidad de Concepción.
- GALLOSO CAMACHO, M^a V. y PRADO ARAGONÉS, J. (2004). La estructura estadística del léxico disponible de informantes nativos e informantes de ELE. En M^a A. CASTILLO

- CARBALLO, O. CRUZ MOYA, J. M. GARCÍA, J. P. MORA GUTIÉRREZ y M.^a REGLA CORDERO RAFFO (Coords.), *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad*, *Actas del XV Congreso Internacional de ASELE* (pp. 370-375). Universidad de Sevilla-Secretariado de Publicaciones.
- HERNÁNDEZ MUÑOZ, N. (2005). *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos*. Ediciones Universidad de Salamanca.
- HIDALGO, M. (2019). *Factores y fuentes que inciden en el léxico disponible de estudiantes sinohablantes de ELE*. *Monografías ASELE*.
- INSTITUTO CERVANTES. [en línea]. *Corpus de aprendices del español (CAES)*. <http://galvan.usc.es/caes>
- JARVIS, S. (2013). Capturing diversity in lexical diversity. *Language Learning*, 63, 87-106.
- JIMÉNEZ CATALÁN, R. M.^a (2002). El concepto de competencia léxica en los estudios de aprendizaje y enseñanza de segundas lenguas. *Atlantis*, XXIV(2), 149-162. https://www.atlantisjournal.org/old/Papers/24_1/jimenez.pdf
- KILGARRIFF, A., RYCHLÝ, P., SMRZ, P. y TUGWELL, D. (2004). *The Sketch Engine*. [programa informático]. <http://www.sketchengine.eu>
- LAUFER, B. (1991). The development of L2 lexis in the expression of the advanced language learner. *The Modern Language Journal*, 75(4), 440-448.
- LAUFER, B. y NATION, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307-22.
- LIN, J. (2006). *El estudio de disponibilidad léxica de los estudiantes chinos de español como lengua extranjera* (Memoria de máster inédita) Universidad de Alcalá.
- LÓPEZ CHÁVEZ, J. (1992). Alcances panhispánicos del léxico disponible. *Lingüística*, 4, 26-124.
- LÓPEZ CHÁVEZ, J. y STRASSBURGER FRÍAS, C. (1987). Otro cálculo del índice de disponibilidad léxica. En *Presente y perspectiva de la lingüística computacional en México*. *Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada* (pp. 91-111). Universidad Nacional Autónoma de México.
- LÓPEZ GONZÁLEZ, A. M. (2010). La evaluación del desarrollo de la competencia léxica en L2 por medio de la disponibilidad léxica. *RedELE*, 18. <https://www.educacionyfp.gob.es/dam/jcr:4f5777a6-d513-4715-910a-01c047353995/2010-redele-18-01lopez-pdf.pdf>
- LÓPEZ MORALES, H. (1999). *Léxico disponible de Puerto Rico*. Arco Libros.
- LÓPEZ MORALES, H. (2011). Los índices de 'riqueza léxica' y la enseñanza de lenguas. En J. de SANTIAGO GUERVÓS, H. BONGAERTS, J. J. SÁNCHEZ IGLESIAS y M. SESEÑA (Eds.), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*. *Actas del XXI Congreso Internacional de la ASELE*, (vol. 1, pp. 15-28). Imprenta Kadmos. https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0015.pdf.
- MARCONI, D. (1999). *La competenza lessicale*. Editori Laterza.
- MENA OSORIO, M. (1986). *Disponibilidad léxica infantil en tres niveles de enseñanza básica* (Tesina inédita, Universidad de Concepción).
- MICHÉA, R. (1953). Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. *Les Langues Modernes*, 47, 338-344.

- MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE-INSTITUTO CERVANTES (2002). *Marco común europeo de referencia para las lenguas: Aprendizaje, Enseñanza, Evaluación*. Anaya. http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf
- MORANTE VALLEJO, R. (2005). *El desarrollo del conocimiento léxico en segundas lenguas*. Arco Libros.
- NALESSO, G. (2019): *Disponibilidad y riqueza léxica en un grupo de estudiantes universitarios italianos de ELE* [tesis doctoral, Università degli Studi di Padova]. Repositorio institucional. http://paduaresearch.cab.unipd.it/12776/1/Giulia_Nalesso_tesi.pdf
- NATION, P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- PAREDES GARCÍA, F. (2015). Disponibilidad Léxica y enseñanza de ELE: el léxico disponible como fuente curricular y como recurso en el aula. *Linred*, 13. http://www.linred.es/monograficos_pdf/LR_monografico13-articulo2.pdf
- READ, J. A. S. (2000). *Assessing vocabulary*. Cambridge University Press
- REAL ACADEMIA ESPAÑOLA. [en línea]. *Corpus de referencia del español actual (CREA)*. <http://corpus.rae.es/creanet.html>
- REAL ACADEMIA ESPAÑOLA Y ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA [en línea]. *Diccionario de la Lengua Española* <http://dle.rae.es>
- RUFAT, A. (2018). Enseñanza-aprendizaje del componente léxico-semántico. En M^a MARTÍNEZ-ATIENZA y A. ZAMORANO AGUILAR (Eds.), *Teoría y metodología para la enseñanza de ELE. Vol. II. Enseñanza-aprendizaje de los componentes lingüísticos* (pp. 73-108). En-Clave-ELE.
- SAMPER HERNÁNDEZ, M. (2002). *Disponibilidad léxica en alumnos de español como lengua extranjera*. ASELE.
- SAMPER PADILLA, J. A. (1998). Criterios de edición del léxico disponible: sugerencias. *Lingüística*, 10, 311-333.
- SÁNCHEZ-SAUS LASERNA, M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas*. Editorial Universidad de Sevilla.
- ŠIFRAR KALAN, M. (2014). Disponibilidad léxica en diferentes niveles de español/lengua extranjera. *Studia Romanica Posnaniensia*, 41(1), 63-85. <https://pressto.amu.edu.pl/index.php/srp/article/view/36/23>
- TOMÉ CORNEJO, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE* [tesis doctoral, Universidad de Salamanca]. Repositorio institucional. <http://hdl.handle.net/10366/128287>
- TOMÉ CORNEJO, C. (2016). Disponibilidad léxica y enseñanza de vocabulario. En N. DOMÍNGUEZ GARCÍA, C. FERNÁNDEZ JUNCAL, J. L. GARCÍA ALONSO (Eds.), *Innovación y Desarrollo En Español Como Lengua Extranjera* (pp. 89-105). Ediciones Universidad de Salamanca.
- WANG, X. (2016). *Riqueza léxica en muestras de lengua escrita de estudiantes sinohablantes de pregrado y postgrado* (Tesis de maestría inédita, Universidad de La Habana).