

El desarrollo de la competencia léxica de estudiantes italianos universitarios de ELE

Giulia NALESSO
Università di Padova

Resumen

Este estudio se inscribe en el ámbito de la didáctica de ELE presentando un proyecto sobre aprendices italianos en contexto universitario, encajado en una más amplia investigación que examina diferentes aspectos del aprendizaje del vocabulario español por parte de estudiantes no nativos. Tras la exposición del marco teórico y de la metodología de la investigación, se comentan los resultados de un test de medición de la competencia léxica sometido a dos grupos de alumnos, correspondientes a los niveles B1 y B2 del MCER, lo cual permitirá conocer si existe una evolución del conocimiento de un nivel a otro y establecer cuál es el caudal léxico que se aprende. Efectivamente, para obtener una visión lo más completa posible se observará el desarrollo de los conocimientos mediante el análisis de la disponibilidad léxica (Carcedo, 2000; Samper Hernández, 2002) y la medición de la riqueza léxica (Guiraud, 1954 y 1960; López Morales, 1983 y 1984).

Palabras claves: ELE, léxico, disponibilidad léxica, riqueza léxica, competencia léxica.

Abstract

This paper presents a research in the field of Spanish as a Foreign Language (SFL) focused on Italian undergraduates, as a part of a wider survey that examines different aspects of the Spanish vocabulary learning by non-native students. After the introduction of the theoretical framework and the methodology, we will show the results of a lexical competence assessment test submitted to two groups of learners, corresponding to B1 and B2 levels of the CEFR. It will allow us to analyze if there is a development from one level to another and to establish the amount of learned lexis. Our main objective is to study the lexical proficiency through the analysis of lexical availability (Carcedo, 2000; Samper Hernández, 2002) and lexical richness (Guiraud, 1954 and 1960; López Morales, 1983 and 1984) to obtain the most complete view of the phenomenon.

Keywords: SFL, vocabulary, lexical availability, lexical richness, lexical competence.

1. INTRODUCCIÓN

En el panorama investigativo de las lenguas extranjeras es sabido que dominar el caudal léxico de otra lengua es un desafío (Alvar Ezquerro, 2003; Baralo, 2011; Sánchez López, 2013), por lo que la competencia léxica es una piedra angular del proceso de enseñanza/aprendizaje. Para ello, se plantea una encuesta transversal y longitudinal que analiza el desarrollo de la competencia léxica de dos grupos de aprendices italianos de nivel universitario para averiguar si existe un desarrollo del conocimiento entre el comienzo y el final del curso académico y si se detecta un crecimiento de un nivel a otro y, por último, qué factores sociolingüísticos inciden en la evolución de la competencia.

Por eso, se ha sometido a un test de medición de la competencia léxica a los alumnos –que componen la primera muestra de informantes del proyecto– con el fin de establecer cuál es el caudal de vocabulario que se aprende y corroborar las siguientes hipótesis:

- la disponibilidad y la riqueza léxica de los estudiantes de nivel B2 son mayores tanto en cantidad de palabras como de vocablos con respecto a las del grupo de B1;
- existe una relación asociativa entre el dominio de ELE y la cantidad de lexías aprendidas;
- se detecta una evolución de conocimientos de un nivel a otro.

2. MUESTRA

En esta fase inicial de la investigación se ha trabajado con la primera muestra de aprendices, procedentes del II y del III año de grado en lenguas extranjeras (precisamente *Lingue, letterature e culture moderne*, *Lingue, letterature e mediazione culturale* y *Mediazione linguistica e culturale*) de la Universidad de Padua (Italia), que acorde con el *Marco común europeo de referencia para las lenguas* (2002) pertenecen a los niveles B1 y B2, con lo cual se han recopilado el corpus B1 y el corpus B2¹.

En lo que atañe a las características de los estudiantes universitarios se sabe que son muy heterogéneas, por lo que se han analizado algunas variables, que coinciden con las utilizadas en estudios previos (Carcedo, 2000; Samper Hernández, 2002; Sánchez-Saus, 2011; Gallego Gallego, 2014; Rubio Sánchez, 2015; Hidalgo Gallardo, 2017), a fin de corroborar si influyen o no en el aprendizaje. De las muchas variables considerables² para el presente propósito investigativo solo se han aplicado tres³, como se indica en la tabla de abajo:

Variable	Variante y código	
Sexo	Mujer	1
	Hombre	2
Nivel de ELE	Nivel B1	1
	Nivel B2	2
LE conocidas	Conoce dos lenguas extranjeras	1
	Conoce más de dos lenguas extranjeras	2

Tabla 1. Codificación de las variables

3. MARCO TEÓRICO Y MÉTODO DE INVESTIGACIÓN

3.1 DISPONIBILIDAD LÉXICA

La disponibilidad léxica tiene como objetivo la recogida y el análisis del léxico disponible de una determinada comunidad de habla. Por tanto, el léxico disponible es el conjunto de palabras que los hablantes tienen en el lexicón mental y cuyo uso está condicionado por un argumento concreto. Mediante este tipo de análisis se pretende descubrir qué palabras sería capaz de usar un individuo en un cierto tema de comunicación, esto es, cuál es su léxico activo.

En este proyecto, el marco teórico es el de Carcedo (2000), cuya metodología práctica se apoya en técnicas de psicología asociativa que determinan el lexicón mental de un sujeto y establecen el nivel de su competencia léxica, la cual ha sido retomada por otros investigadores, entre los que citamos a Samper Hernández (2002) y Sánchez-Saus (2011).

¹ Conviene puntualizar que por el carácter introductorio del estudio se presentará el análisis de 20 pruebas elegidas aleatoriamente entre todas las entregadas.

² En particular, no se ha considerado la lengua materna, una de las variables más utilizada en este ámbito, por suponer que los alumnos son italófonos de nacimiento o bien que, a pesar de ser extranjeros, manejan el italiano como lengua materna al estudiar a nivel universitario en Italia. Al contrario, otros estudios tratan grupos variados, donde las diferencias radican específicamente en la L1: Carcedo (2000) examina hablantes de finés y de sueco o bien Samper Hernández (2002), Sánchez-Saus (2011) y Rubio Sánchez (2015) cuentan con aprendices de diferentes L1 al mismo tiempo al trabajar en contextos de inmersión, con lo cual la distinción lingüística es imprescindible.

³ Las variables se han codificado mediante un código dividido en tres segmentos. Para poner un ejemplo, si tenemos un sujeto identificado con 12200 001 12 a través de los primeros tres dígitos podemos descifrar su género, su nivel de ELE y su conocimiento de otras lenguas (se ha añadido 00 porque el programa utilizado trabaja con cinco variables): se trata de una mujer con un nivel B2 de ELE, que conoce más de dos lenguas extranjeras. En cuanto al grupo de tres cifras que viene después, indica el número asignado al informante, mientras que las dos últimas representan el centro de interés analizado en el apartado de la disponibilidad léxica (en este caso, el número corresponde al centro de interés 12, los medios de transporte).

Ahora bien, cabe sacar a colación que los principales trabajos sobre léxico disponible en español están enfocados en su mayoría en la observación de hablantes nativos (véase el “Proyecto Panhispánico de disponibilidad léxica”, en adelante PPHDL, dirigido por Humberto López Morales) y que actualmente disponemos ya de una cantidad notable de estudios sobre aprendices de ELE en distintos países, si bien este no es el caso de Italia donde todavía este ámbito investigativo no ha tenido el seguimiento que merece, pese a las contribuciones de Caggiula (2013), Rubio Sánchez (2015) y Del Barrio (2016, 2017a, 2017b). De ahí, se ha incluido en este proyecto una prueba de disponibilidad léxica basada en los 16 centros de interés tradicionales. Este se compone de cuatro páginas con fichas formadas por cuatro columnas⁴ y un espacio sobre cada una numerado de 1 a 16 en el que los alumnos debían poner los campos semánticos según se les proporcionaban⁵. Sin embargo, a la hora de realizar la prueba se han presentado uno a uno a medida que se iban completando, para evitar que las respuestas fuesen el resultado de un proceso de reflexión y así asegurar mayor fiabilidad (Samper Hernández, 2002; Sánchez-Saus, 2011). Por ende, los informantes han rellenado las fichas en dos minutos para cada estímulo con todas las lexías que se les ocurrían.

Tras la recogida, los datos han sido tratados mediante el software *Dispogen II* (Echeverría *et al.*, 2005), después de una fase de edición del material léxico⁶ basada en criterios comunes de la investigación en disponibilidad léxica y criterios propios de la investigación en ELE (Samper Hernández, 2002):

- a) Normalización de las variantes flexivas, esto es, se ha registrado el infinitivo de los verbos y el singular masculino de los sustantivos y adjetivos, con excepción de los pluralia tantum.
- b) Eliminación de los términos repetidos y de los artículos.
- c) Corrección de la ortografía, con adaptación a la grafía propuesta por el DRAE (2017), ya que considerando el nivel general de competencia en ELE se han aceptado las palabras escritas con acentuación errata, incorrecta duplicación de consonantes y selección falsa de consonantes o vocales⁷:

**estanteria; *voleibol; *professor; *novidad; *lemón*

- d) Unificación ortográfica:

⁴ Cada columna permitía un máximo de 25 aportaciones, pero, con el fin de trabajar con listas abiertas, se hizo hincapié que si la ficha no era suficiente se podía seguir en el reverso del folio y, análogamente, se exhortó a escribir todas las unidades léxicas que les acudían a la mente, aunque no estuviesen seguros de la grafía correcta o del significado.

⁵ Se estudian los mismos centros de interés del PPHDL, que coinciden con las propuestas iniciales del *Français Fondamental* (1964): 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 trabajos [PRO]. Se ha optado por aplicar dichos campos semánticos tradicionales porque se pretende brindar una encuesta que resulte cotejable con otras (Callealta Barroso y Gallego Gallego, 2016; Jiménez Catalán, 2017; Paredes García, 2017), pese a que en diferentes trabajos se haya considerado oportuno reformular o sustituir algunos de ellos según necesidad (Mackey, 1971a; López Morales, 1999; Bombarelli, 2005; Sánchez-Saus, 2011; Paredes García, 2014; Tomé Cornejo, 2015). A este propósito, tómense de ejemplo las palabras de González Fernández (2014: 52) “[...] se pone de manifiesto como algunos centros de interés clásicos no son idóneos para los trabajos de disponibilidad léxica aplicada a la enseñanza del español como lengua extranjera. En las investigaciones de esta corriente se debería eliminar algunos centros de interés y modificar otros para poder obtener resultados más relevantes”. Al contrario, Bartol (2010) demuestra la correspondencia entre estos y los ámbitos de contenido de los que se trata en las indicaciones del MCER, por tanto aquí se intenta abarcar todos los intereses humanos y se han estudiado tal cual.

⁶ Al solucionar ciertas dificultades de comprensión de las respuestas se han tomado como referencia las siguientes obras de la Real Academia Española que tratan

- cuestiones léxicas: el *Diccionario de la lengua española* (2017) y el *Diccionario panhispánico de dudas* (2005);
- cuestiones gramaticales: la *Nueva gramática de la lengua española* (2009);
- cuestiones ortográficas: la *Ortografía de la lengua española* (2010);
- cuestiones de uso: el *Corpus de referencia del español actual* (2008, 2015).

⁷ Los errores se indican con un “*”.

entre *hierba* y *yerba* se ha recopilado *hierba*

- e) Unificación de derivados regulares mediante la recogida de una sola forma que no supone alteración del significado:

mesita>*mesa*; *casita*>*casa*

- f) Unificación de formas plenas y acortamientos con paréntesis:

bicicleta>*bici(cleta)*; *bolígrafo*>*bolí(grafo)*

- g) Aceptación de “neologismos universalizados”, tal y como ha apuntado Samper Hernández (2002: 28), tanto adaptados como no gráfica y fonéticamente a la lengua española y de los extranjerismos que los hispanohablantes usan con frecuencia:

hámster, *cómic*, *PlayStation*

- h) Eliminación de los lexemas de la lengua madre o de otras lenguas conocidas:

**arbre*, **jeans*, **pizza*, **skateboard*, **t-shirt*

- i) Amplitud de las relaciones asociativas, es decir que se han aceptado algunas palabras fuera de contexto para demostrar la presencia de una unidad en el léxico activo del informante, como por ejemplo, se han aceptado *fruta*, *verdura* y *leche* en el CI13 (Trabajos del campo y del jardín [TRC]);⁸.

3.2 RIQUEZA LÉXICA

Ahora bien, la riqueza léxica contabiliza el vocabulario activo que un individuo emplea en un texto centrándose en la variedad del léxico utilizado. Según López Morales (2011: 15) esto permite estudiar la destreza comunicativa y el dominio léxico del sujeto ya que: “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.”

Actualmente, en Italia no ha despertado suficientemente el interés académico en este tipo de análisis en ELE, si excluimos una tesis de licenciatura inédita (Basso, 2017), a pesar de la atención de investigadores de otros países (García Rosas, 1996; Cuba Vega y Cuba, 2004; Berton, 2014; Lucha Cuadros y Díaz, 2016; Wang, 2016).

Por ende, se ofrece en ese apartado un estudio que remonta al marco teórico tradicional, aún vigente, de los pioneros Guiraud (1954, 1960) y López Morales (1983, 1984) que analizan la relación entre *variación* y *competencia léxica*, basándose en la *diversidad léxica* de Carroll (1938), quien *in primis* la define como “cantidad relativa de repetición de vocabulario en un texto, oral u escrito”. Análogamente, se comparte la teoría de Read (2000), el cual amplía la noción de riqueza léxica como un conjunto de aspectos: variación, sofisticación, densidad, número de errores, de los que se verán algunos en la fase analítica del presente estudio.

En lo que atañe a la recogida de datos, se encuentra en la última parte de la prueba mencionada el test de riqueza léxica, esto es, la redacción en 30 minutos de un texto sobre un viaje⁹. Es preciso señalar que, para obtener un corpus fiable, las composiciones no debían exceder las 150 palabras porque parece

⁸ Esta etapa de edición es llamativa ya que el investigador determina cuáles respuestas admitir y cuáles no: a partir de este dictamen se llega a resultados distintos. Efectivamente, representa un paso común de las investigaciones sobre disponibilidad léxica (Samper Padilla, 1998), pero al tratarse de aprendices no nativos procedentes de una cultura diferente y con un limitado conocimiento de la lengua española cobra una dimensión particular (Samper Hernández, 2002: 29). De ahí, se ha aceptado un gran número de asociaciones secundarias, en particular en los centros de interés más complejos.

⁹ La elección de este tema responde a la voluntad de indagar la capacidad real de uso del vocabulario: un argumento más complejo o especializado habría obstaculizado la producción de los informantes. Efectivamente, se pretendía maximizar la posibilidad de expresión, tal como plantean Laufer y Nation (1995: 308) al destacar la importancia de medir la competencia de un autor según la familiaridad desarrollada en torno a un tema.

que después de un cierto límite el cálculo se desvirtúa: “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). Sin embargo, los textos del corpus oscilan entre las 109 y 160 palabras y se han registrado como han sido entregados, sin alterar el contenido ya que durante las fases de edición no se han detectado errores tan graves que pudiesen alterar los resultados finales de la investigación¹⁰.

Como ya se ha dicho, el procesamiento analítico requiere una minuciosa edición de los datos, que se han tratado mediante *AntConc* (Anthony, 2018). Así que los textos han sido lematizados como las lexías del apartado anterior (infinitivo para los verbos y singular masculino para sustantivos y adjetivos) y se han aplicado las siguientes normas, conforme a las que se emplean convencionalmente en el análisis de la producción escrita en LE (Barbasán Ortuño, 2016) que han llevado a distinguir entre las faltas, que no perjudican el conocimiento global del vocabulario español, y los errores léxicos propiamente dichos, que descartamos:

a) Corrección de la ortografía:

**espectaculo; *incredulo; *trén; *passear; *árabos*

b) Eliminación de formas híbridas debidas a la transferencia negativa del italiano o de otras lenguas conocidas, palabras extranjeras, números en cifras:

**lunga; *shon; *skateboard*

c) Inclusión de morfología derivativa y de topónimos en español:

muchísimo; poquito; Nueva York; Londres; Inglaterra

d) Recopilación de las locuciones como una sola unidad para examinar el conocimiento de las combinaciones léxicas (Lewis, 1993, 1997, 2000; Higuera, 2006):

por la mañana; por la tarde; por la noche; echar de menos.

4. ANÁLISIS DE LA DISPONIBILIDAD LÉXICA

4.1 ANÁLISIS GENERAL

Con respecto a los resultados generales, los cómputos han revelado los siguientes valores, visibles en la Tabla 2 y en la Figura 1:

Informantes	20
Palabras totales	2.237
Palabras por informante	112
Vocablos totales	727
Índice de cohesión	3,61

Tabla 2. Resultados generales

¹⁰ Debido a la extensión reducida de los textos, no se aplican ni técnicas de normalización a las fórmulas ni medidas como las de Zipf (1935) o de Yule (1944).

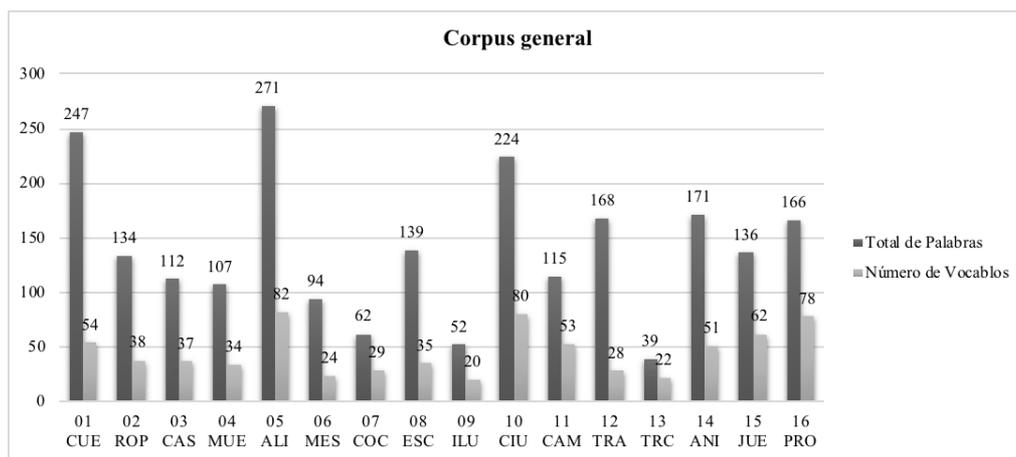


Figura 1. Resultados generales

Los centros de interés más productivos, que superan las 200 aportaciones son, en orden, CI14. Los animales, CI01. Partes del cuerpo y CI10. La ciudad; mientras que los que han aportado resultados más escasos, posiblemente al no ser temas tratados frecuentemente en aula, son CI13. Trabajos del campo y del jardín, CI09. Iluminación y calefacción.

Asimismo, como era esperable, examinando todo el conjunto de respuestas destaca la presencia masiva de sustantivos, en particular en los 20 lexemas más disponibles, pese a la presencia de adjetivos y verbos en CI09. Iluminación y calefacción, CI11. El campo, CI13. Trabajos del campo y del jardín, CI15. Juegos y distracciones: *frío, rural, cortar, regar, plantar, bailar, salir*.

5.2 ANÁLISIS POR VARIABLE

5.2.1 VARIABLE 'NIVEL DE ELE'

Pasando al análisis de la variable 'nivel de ELE', es interesante observar que el grupo B1 ha aportado un caudal mayor de palabras, y de vocablos, con respecto al B2.

El promedio de respuestas por cada informante del corpus B1 excede el del corpus B2, lo cual indica que el dominio de ELE en esta fase del aprendizaje no influye en la cantidad de palabras disponibles. Además, conviene resaltar que en las investigaciones sobre nativos el valor medio de palabras por sujeto es de 20 en cada centro de interés, pues bien, los informantes presentan un léxico disponible inferior (13,6) pero apreciable, con lo cual se espera que en la segunda aplicación de los test se alcancen resultados aún mejores.

Por último, el único cómputo que presenta valores parecidos es el índice de cohesión¹¹ que es elevado, por tanto, el aprendizaje resulta organizado y compacto.

	Corpus B1	Corpus B2
Informantes	10	10
Palabras totales	1168	1069
Palabras por informante	116,8	106,9
Vocablos totales	533	462
Índice de cohesión	3,61	3,81

Tabla 3. Resultados según la variable 'nivel de ELE'

¹¹ Su función es profundizar el escudriño de los datos midiendo el grado de coincidencia de las aportaciones (Echeverría *et al.*, 1987).

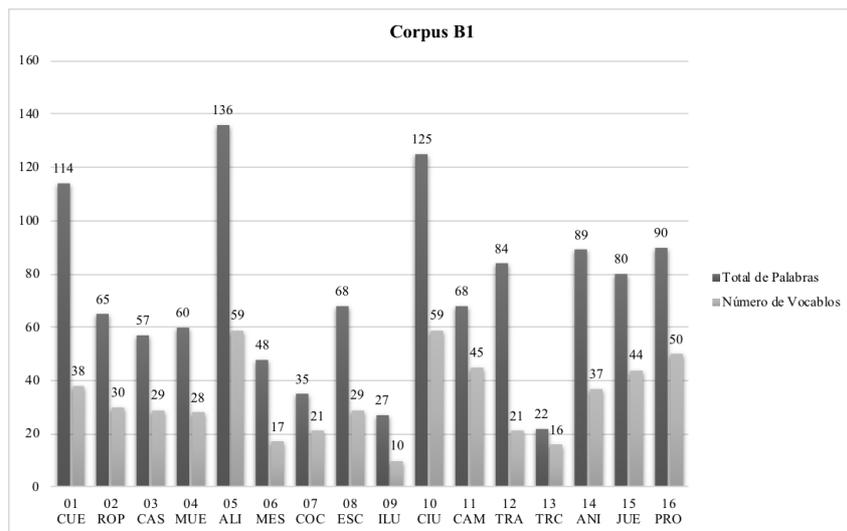


Figura 2. Resultados del Corpus B1

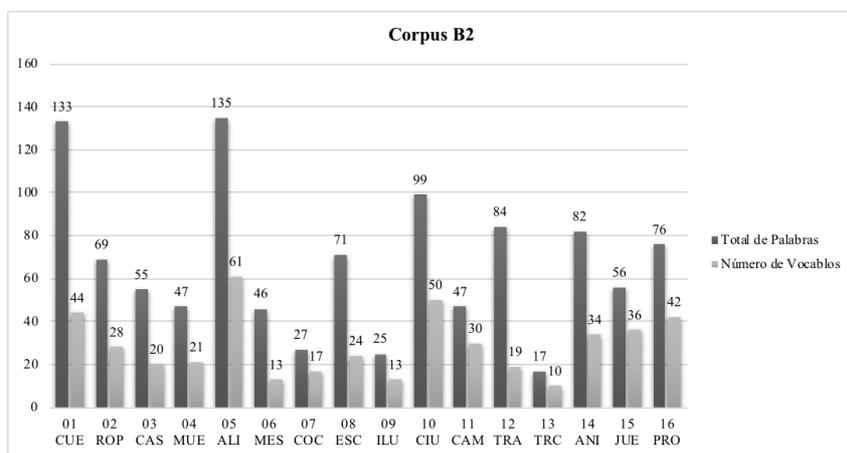


Figura 3. Resultados del Corpus B2

4.2.2 VARIABLE 'SEXO'

En lo que atañe al resultado de los cómputos que tenían en cuenta el factor 'sexo', cabe señalar una ligera supremacía de las mujeres sobre los varones que, al contrario, presentan un índice de cohesión un poco más alto y sistemático.

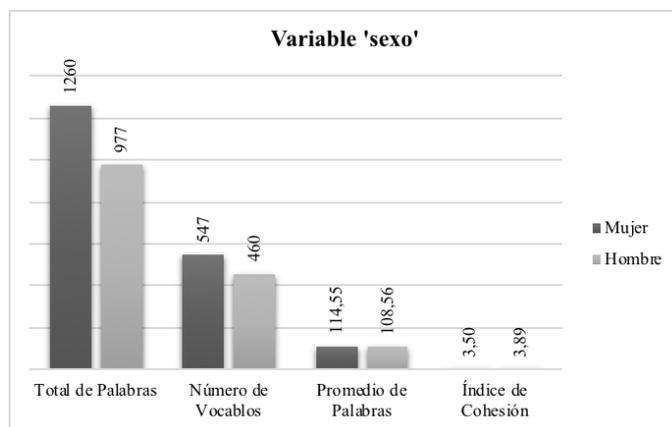


Figura 4. Resultados según la variable 'sexo'

4.2.3 VARIABLE 'CONOCIMIENTO DE OTRAS LENGUAS EXTRANJERAS'

Por fin, el 'conocimiento de otras lenguas extranjeras' revela una clara superioridad en todos los índices de los informantes que conocen más de dos lenguas, encima del español, lo cual confirma la *Hipótesis de la Interdependencia Lingüística* de Cummins (1979).

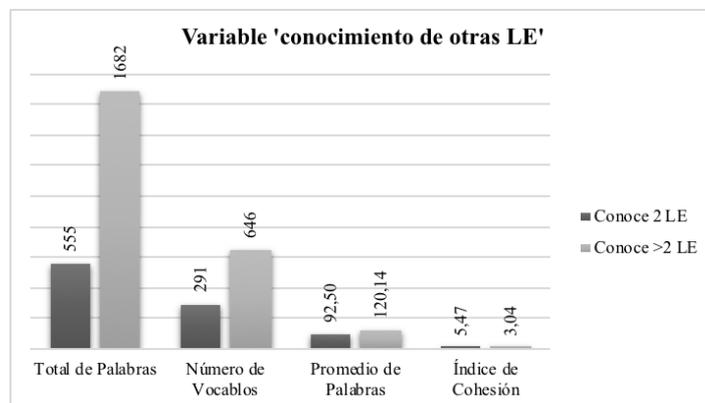


Figura 5. Resultados según la variable 'conocimiento de otras LE'

De lo expuesto hasta ahora, se puede inferir que, contrariamente a las hipótesis de partida, la disponibilidad de los estudiantes de nivel B2 no resulta mayor con respecto a la de los B1. Mientras que los factores que más inciden en la evolución del conocimiento es el género de los encuestados y, en particular, el hecho de conocer más lenguas extranjeras.

5. ANÁLISIS DE LA RIQUEZA LÉXICA

Pasando al análisis de la riqueza léxica, cabe mencionar que se comparte la teoría de Jarvis (2013) quien sostiene la aplicación de diferentes índices para obtener datos lo más completos posibles, por lo que se verá que la medición se ha llevado a cabo calculando la *Type/Token Ratio* (TTR), la variación léxica, la densidad léxica, el intervalo de aparición de palabras nocionales (IAT) y, por último, la originalidad léxica.

Así pues, el corpus general (2.566 *tokens*) se divide en:

- Corpus B1: 10 textos, 1319 *tokens*
- Corpus B2: 10 textos, 1247 *tokens*

5.1 ANÁLISIS DEL CORPUS B1

El corpus B1 se constituye por textos que van de 115 a 160 palabras, cuyo total de *tokens* es de 1319 (771 *tokens* léxicos) y de *types* es 765.

Informante	TTR	Variación léxica	Densidad léxica	IAT
001	0,66	66%	56%	1,77
002	0,50	50%	59%	1,70
003	0,67	67%	65%	1,53
004	0,62	62%	54%	1,86
005	0,58	58%	61%	1,64
006	0,52	52%	58%	1,73
007	0,57	57%	62%	1,61
008	0,58	58%	56%	1,80
009	0,57	57%	55%	1,83
010	0,56	56%	59%	1,70
<i>Promedio</i>	0,58	58%	58%	1,72

Tabla 4. Análisis del Corpus B1

Es sabido que el T^{*}T^rR y la variación léxica reflejan dos perspectivas del mismo resultado: la relación entre palabras repetidas y no repetidas, que llega a un valor de 0,58. Es decir que en la totalidad de texto el 58% está constituido por *types*, con lo cual la variedad del vocabulario es buena ya que cubre más de la mitad de una redacción. Como se ve en la Tabla 4, el informante 003 presenta el mejor nivel de riqueza con respecto a los demás, alcanzando resultados excelentes en todos los índices.

Luego, la media de densidad léxica es de 58%, por lo tanto, más de la mitad de las palabras de un texto es nocional según la taxonomía de Hallebeek (1986): el resultado es satisfactorio pese a que el promedio del IAT alcance el valor bastante alto de 1,72. Este indica la presencia de un ítem léxico casi cada dos gramaticales y no representa un alto nivel de riqueza ya que cuanto más el intervalo se acerca a 1, mejor es el índice: “un intervalo de voces de contenido léxico de 2,1 detecta menor riqueza léxica que si se consigue un intervalo entre lexías de 1,2” (Reyes Díaz, 2010: 148).

5.2 ANÁLISIS DEL CORPUS B2

El Corpus B2 consta de 10 textos cuya extensión oscila entre 109 y 146 palabras, esto es, 1247 *tokens* (697 *tokens* léxicos) y 729 *types*.

Informante		Variación léxica	Densidad léxica	IAT
011	0,62	62%	60%	1,66
012	0,50	50%	51%	1,95
013	0,60	60%	54%	1,86
014	0,58	58%	56%	1,79
015	0,56	56%	53%	1,90
016	0,58	58%	56%	1,79
017	0,53	53%	47%	2,11
018	0,57	57%	60%	1,67
019	0,66	66%	59%	1,68
020	0,64	64%	60%	1,66
<i>Promedio</i>	0,58	58%	56%	1,81

Tabla 5. Análisis del Corpus B2

Los cálculos revelan una media de T^{*}T^rR y de variación léxica idéntica con respecto al otro grupo (0,58 y 58%).

Mientras que en lo que se refiere a la densidad léxica es oportuno señalar un decrecimiento en el promedio de este corpus que alcanza el 56%, dos puntos porcentuales menos que el otro grupo, hecho que no sorprende ya que presenta globalmente valores más bajos. Igualmente, el análisis del IAT lleva a un promedio más alto que en B1, y negativo: 1,81.

Con todo esto, en términos cuantitativos resulta que el alumnado que se suponía más avanzado ha aportado resultados peores en relación al otro. Efectivamente, los *tokens* de los estudiantes B1 son mayores con respecto a los del B2: no aparece aún una relación asociativa entre el dominio de ELE y la cantidad de lexías aprendidas.

5.3 ORIGINALIDAD LÉXICA

Como colofón, se ofrece el cálculo del índice de originalidad léxica para comprobar si el vocabulario empleado es variado y cuánto. La Tabla 6 muestra los valores de ambos corpus, que oscilan de 9% a 12%, con lo que se desprende que la media es aproximadamente la misma en los grupos y no se detecta la

presencia de ningún texto con valor 0, o sea que no tiene lexías originales¹². En suma, se revela un índice de originalidad apreciable.

Corpus	Informante	Originalidad léxica
B1	001	9,55%
	002	11,22%
	003	8,72%
	004	8,87%
	005	9,10%
	006	11,52%
	007	10,77%
	008	9,40%
	009	8,72%
	010	12,13%
B2	011	11,71%
	012	9,38%
	013	9,38%
	014	8,74%
	015	9,30%
	016	11,07%
	017	9,14%
	018	11,55%
	019	10,26%
	020	9,46%

Tabla 6. Índice de Originalidad léxica

5.4 VARIABLE 'SEXO'

Para concluir, el análisis del factor 'sexo' no ha variado notablemente las pruebas de riqueza léxica, como se muestra muy resumidamente, mediante los promedios obtenidos de los cómputos, en las Tablas 7 y 8.

Nivel de ELE	Sexo	<i>Tokens</i>	<i>Types</i>	<i>Tokens léxicos</i>
B1	Mujeres	134,5	75,83	77,67
	Hombres	128	77,5	76,25
B2	Mujeres	125,8	72	70
	Hombres	123,6	73,8	69,4

Tabla 7. Promedios de las aportaciones según las variables 'nivel de ELE' y 'sexo'

Nivel de ELE	Sexo	TTR	Variación Léxica	Densidad Léxica	IAT
B1	Mujeres	0,57	57%	58%	1,74
	Hombres	0,61	61%	60%	1,68
B2	Mujeres	0,57	57%	55%	1,83
	Hombres	0,60	60%	56%	1,79

Tabla 8. Índices de riqueza léxica según las variables 'nivel de ELE' y 'sexo'

¹² Es preciso señalar que la gran mayoría de las palabras originales constituyen a menudo *hapax* ya que son utilizadas por un solo informante y, por tanto, resultan ser de baja frecuencia, esto es, ocurren en el corpus raramente cuando el mismo autor las reitera.

Manteniendo la supremacía relativa al análisis de la variable ‘nivel de ELE’, las mujeres han aportado un promedio más alto de *tokens* en ambos corpus, bien que la media de *types* favorece los varones como ocurre en todos los índices considerados, de lo cual se infiere que los hombres escriben menos, pero con resultados mejores y más variados.

Es necesario tener en cuenta que estos son los resultados de un estudio preliminar, por ende, se averiguará esta tendencia durante el análisis de todo el conjunto muestral, a fin de que tales conclusiones sean significativas.

6. CONCLUSIONES

Como ya se ha mencionado, este estudio presenta parte de un proyecto más amplio del que se ha expuesto un análisis exploratorio, cuyos resultados indican que contrariamente a las hipótesis de partida, la disponibilidad y la riqueza léxica de los estudiantes de nivel B2 no resulta mayor con respecto a la de los B1, por tanto, no parece existir una relación asociativa entre el dominio de ELE y la cantidad de lexías aprendidas, o bien que la mayor parte del esfuerzo docente dedicado a los alumnos de B2 se ha referido a otras áreas de la competencia, como podrían ser la gramatical o la metalingüística, lo que en todo caso implicaría que el léxico también se aprende en una acción didáctica dirigida. De todos modos, dado el nivel de ELE, los números se consideran en general positivos.

Análogamente, en los cómputos de la disponibilidad léxica, los factores más influyentes en el desarrollo son el sexo y, en particular, el conocimiento de más lenguas extranjeras. Además, en la medición de la riqueza léxica, el índice de originalidad es apreciable en ambos grupos.

Claro está que los escasos datos manejados llevan a unas conclusiones parciales, que se ampliarán, en primera instancia en la fase de estudio pormenorizado de todo el muestreo y, en segundo lugar, después de una segunda sumministrazione de la prueba. Efectivamente, el análisis detallado completará la investigación y ofrecerá una imagen del procesamiento del vocabulario con el reto de complementar la encuesta mediante el análisis tipológico de las listas de frecuencias para averiguar si las diferencias, o analogías, cuantitativas se corresponden con diferencias, o analogías, cualitativas.

BIBLIOGRAFÍA

- ALVAR EZQUERRA, Manuel (2003): *La enseñanza del léxico y el uso del diccionario*, Madrid: Arco/Libros, S. L.
- ANTHONY, Lawrence (2018): *AntConc* (Versión 3.5.7), Tokyo: Waseda University. Disponible en: <http://www.laurenceanthony.net/software>
- BARALO, Marta (2011 [1999]): *La adquisición del español como lengua extranjera*, Madrid: Arco/Libros, S. L.
- BARBASÁN ORTUÑO, Inmaculada P. (2016): *El error léxico en la interlengua de lenguas afines: italiano y español*, Valencia: Universitat Politècnica de València. (Tesis doctoral)
- BASSO, Valentina (2017): *La riqueza léxica en la producción escrita de aprendices itálfonos de E/LE*, Università degli Studi di Padova. (Tesis de licenciatura inédita)
- BERTON, Marco (2014): *La riqueza léxica en la producción escrita de estudiantes suecos de ELE*, Stockholms Universitet. (Memoria de máster Multiele)

- BOMBARELLI, Ángela (2005): *La disponibilidad léxica como herramienta didáctica: una propuesta de selección del vocabulario para un nivel umbral de ELE*, Salamanca: Universidad de Salamanca. (Memoria de máster Multiele)
- CAGGIULA, Silvia (2013): *El español como lengua extranjera: un estudio de disponibilidad léxica y su aplicación a la enseñanza*, Carolina del Norte: Lulu Press.
- CALLEALTA BARROSO, Francisco J.; GALLEGO GALLEGO, Diego (2016): “Medidas de disponibilidad léxica: comparabilidad y normalización”, *Boletín de filología*, 51(1), pp. 39-92.
- CARCEDO GONZÁLEZ, Alberto (2000): *Disponibilidad léxica en español lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*, Turku: Turun Yliopisto.
- CARROLL, John B. (1938): “Diversity of vocabulary and the harmonic series law of word frequency distribution”, *The Psychological Record*, 11, 3, pp. 79-86.
- CUBA VEGA, Lidia E.; CUBA, Yadira M. (2004): “Las pruebas de riqueza léxica y su aplicación en Español como Lengua Extranjera (ELE)”, en Díaz Mediavilla, Antonio (ed.): *Actas del VIII Congreso Sociedad Española de Didáctica de la Lengua y Literatura*, Alicante: Universidad de Alicante, pp. 815-830.
- CUMMINS, James (1979): “Linguistic Interdependence and the Educational Development of Bilingual Children”, *Review of Educational Research*, 49, pp. 221-251.
- CONSEJO DE EUROPA (2002): *Marco común europeo de referencia para las lenguas: Aprendizaje, Enseñanza, Evaluación*, traducción del Instituto Cervantes, Madrid: Ministerio de Educación, Cultura y Deporte. Disponible en: http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf
- DEL BARRIO DE LA ROSA, FLORENCIO (2016): “Algunas observaciones sobre la disponibilidad léxica en estudiantes itálicos de español”, en Sainz González E.; Solís García, I.; del Barrio de la Rosa, F.; Arroyo Hernández, I. (eds.): *Geométrica explosión. Estudios de lengua y literatura en homenaje a René Lenarduzzi*, Venezia: Edizioni Ca' Foscari, pp. 127-143.
- (2017a): “Los estudiantes itálicos de ELE y los estudios de disponibilidad léxica en español”, *Revista Nebrija de Lingüística Aplicada*, 22, pp. 52-57.
- (2017b) (ed.): *Palabras. Vocabulario. Léxico. La lexicología aplicada a la didáctica y a la diacronía*, Venezia: Edizioni Ca' Foscari.
- ECHEVERRÍA, Max et al. (2005): *Dispogen II. Programa computacional para el análisis de la disponibilidad léxica*, Concepción: Universidad de Concepción.
- GALLEGO GALLEGO, Diego J. (2014): *Léxico disponible de estudiantes de español como lengua extranjera en la Comunidad de Madrid*, Universidad de Alcalá. (Tesis doctoral inédita)
- GARCÍA ROSAS, M.^a Paloma (1996). *Riqueza léxica en textos de estudiantes de español como lengua extranjera*, Universidad de Alcalá. (Memoria de máster)
- GUIRAUD, Paul (1954): *Les Caractères Statistiques du Vocabulaire. Essai de méthodologie*, Paris: Presses Universitaires de France.
- GUIRAUD, Paul (1960): *Problemes et methodes de la statistique linguistique*, Paris: Presses Universitaires.
- GONZÁLEZ FERNÁNDEZ, Javier (2014): “Idoneidad de los centros de interés clásicos en los estudios de disponibilidad léxica aplicados al español como lengua extranjera”, *Revista Nebrija de Lingüística Aplicada*, 16, pp. 41-53.
- GOUGENHEIM, George; MICHEA René; RIVENC Paul; SAUVAGEOT, Aurélien (1964): *L'élaboration du français fundamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, Paris: Didier.
- HALLEBEEK, Jos (1986): “Las palabras funcionales del español”, *Boletín de la AEPE*, 34-35, pp. 205-216. Disponible en: https://cvc.cervantes.es/ensenanza/biblioteca_ele/aepe/pdf/boletin_3435_18_86/boletin_34-35_18_86_24.pdf

- HIDALGO GALLARDO, Matías (2017): *La disponibilidad léxica como método de detección del vocabulario y de su elección en manuales: aplicación en una muestra de estudiantes sinobablantes de ELE*, Universidad de Jaén (Tesis doctoral inédita)
- HIGUERAS GARCÍA, Marta (2006): *Estudio de las colocaciones léxicas y su enseñanza en español como lengua extranjera*, Málaga: ASELE.
- JARVIS, Scott (2013): “Capturing diversity in lexical diversity”, *Language Learning*, 63, pp. 87-106.
- JIMÉNEZ CATALÁN, Rosa M. (2017): “Estudios de disponibilidad léxica en español y en inglés: revisión de sus fundamentos empíricos y metodológicos”, *Revista Nebrija de Lingüística Aplicada*, 22, pp. 16-31.
- LAUFER, Batia; NATION Paul (1995): “Vocabulary Size and Use: Lexical Richness in L2 Written Production”, *Applied Linguistics*, 16 (3), pp. 307-322. Disponible en: https://www.researchgate.net/profile/Batia_Laufer/publication/249237693_Vocabulary_Size_and_Use_Lexical_Richness_in_L2_Written_Production/links/544689af0cf22b3c14de3a94/Vocabulary-Size-and-Use-Lexical-Richness-in-L2-Written-Production.pdf
- LEWIS, Micheal (1993): *The Lexical Approach: The State of ELT and a Way Forward*, London: Language Teaching Publications.
- (1997): *Implementing the Lexical Approach: Putting Theory into Practice*, London: Language Teaching Publications.
- (2000): *Teaching collocations. Further developments in the Lexical Approach*, London: Language Teaching Publications.
- LÓPEZ MORALES, Humberto (1983): “Lingüística estadística”, en López Morales, H. (coord.): *Introducción a la lingüística actual*, Madrid: Playor, pp. 209-225.
- (1984): *La enseñanza de la lengua materna. Lingüística para maestros de español*, Madrid: Editorial Playor.
- (2011): “Los índices de ‘riqueza léxica’ y la enseñanza de lenguas”, en *Actas del XXI Congreso Internacional de la ASELE 2011 29 de septiembre-2 de octubre de 2010*, pp. 15-28. Salamanca: Universidad de Salamanca. Disponible en: http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/asele_xxi.htm
- LUCHA CUADROS, Rosa M.; DÍAZ RODRÍGUEZ, Lourdes (2016): “¿Progreso en mis textos escritos?: análisis de la riqueza y densidad léxica en producciones escritas mediante software informático”, *e-AESLA*, 2. Disponible en <https://cvc.cervantes.es/lengua/eaesla/pdf/02/01.pdf>
- MACKAY, William F. (1971): *Le vocabulaire disponible du Français*, 2 vols, Paris-Bruxelles-Montréal: Didier.
- MÜLLER, Charles (1968): *Estadística lingüística*, Madrid: Gredos.
- PAREDES GARCÍA, Florentino (2014): “A vueltas con la selección de ‘centros de interés’ en los estudios de disponibilidad léxica: para una propuesta renovadora a propósito de la disponibilidad léxica en ELE”, *Revista Nebrija de Lingüística Aplicada*, 16, pp. 54-59.
- (2017): “La comparabilidad de los trabajos de disponibilidad léxica”, en *Revista Nebrija de Lingüística Aplicada*, 22, pp. 71-77.
- READ, John A. S. (2000): *Assessing vocabulary*, Cambridge: Cambridge University Press.
- REAL ACADEMIA ESPAÑOLA: *Corpus de referencia del español actual*. Disponible en <http://corpus.rae.es/creanet.html>
- (2005): *Diccionario panhispánico de dudas*, Madrid: Santillana.
- (2009): *Nueva Gramática de la Lengua Española, vol. I: Morfología y Sintaxis*, Madrid: Espasa.
- (2010): *Ortografía de la lengua española*, Madrid: Espasa.
- (2017): *Diccionario de la Lengua Española*, 23.1ª edición. Disponible en: <http://dle.rae.es>

- REYES DÍAZ, María J. (2010): “Riqueza léxica de textos redactados por alumnos de bachillerato de Las Palmas de Gran Canaria”, *Anuario de Lingüística Hispánica*. XXIII-XXIV, pp. 147-163.
- RUBIO SÁNCHEZ, Roberto (2015): “Estudio de disponibilidad léxica en aprendices italianos de español: análisis cuantitativo”, en Gordejuela Senosiáin, A.; Izquierdo Alegría, D.; Jiménez Berrio, F.; de Lucas Vicente, A.; Casado Velarde, M. (eds.): *Lenguas, lenguaje y lingüística. Contribuciones desde la Lingüística General*, Pamplona: Servicio de Publicaciones de la Universidad de Navarra, pp. 430-441. Disponible en: <http://dadun.unav.edu/bitstream/10171/40043/1/35.Rubio%20Sánchez.pdf>
- SAMPER HERNÁNDEZ, Marta (2002): *Disponibilidad léxica en alumnos de español como lengua extranjera*, Málaga: ASELE.
- SAMPER PADILLA, José A. (1998): “Criterios de edición del léxico disponible: sugerencias”, *Lingüística*, 10, pp. 311-333.
- SÁNCHEZ LÓPEZ, Isabel (2013): *Léxico, diccionario y enseñanza/aprendizaje de ELE*, Málaga: ASELE.
- SÁNCHEZ-SAUS LASERNA, Marta (2011): *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas*, Universidad de Cádiz. (Tesis doctoral)
- TOMÉ CORNEJO, Carmela (2015): *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE*, Universidad de Salamanca. (Tesis doctoral inédita)
- WANG Xue Xue (2016): *Riqueza léxica en muestras de lengua escrita de estudiantes sinohablantes de pregrado y postgrado*, La Habana. (Memoria de máster)
- YULE, George U. (1944): *The statistical study of literary vocabulary*, Cambridge: Cambridge University Press.
- ZIPF, George K. (1935): *Psycho-biology of languages*, Cambridge, MA: MIT Press.